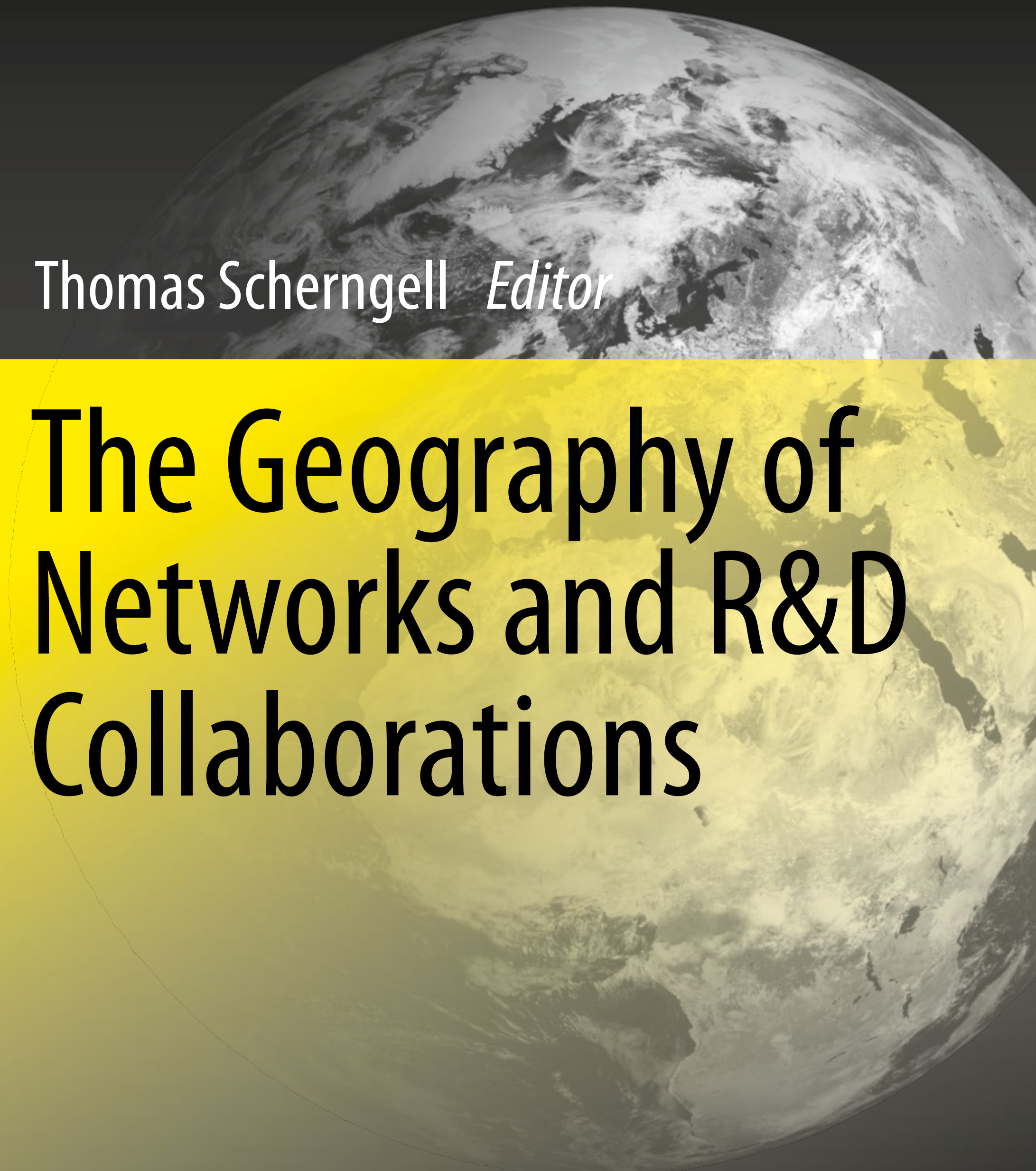
Advances in Spatial Science

Thomas Scherngell *Editor*

The Geography of Networks and R&D Collaborations



Springer

Advances in Spatial Science

For further volumes:
http://www.springer.com/series/3302

Thomas Scherngell
Editor

The Geography of Networks and R&D Collaborations

Springer

Editor
Thomas Scherngell
Innovation Systems Department
AIT Austrian Institute of Technology
Wien, Austria

ISSN 1430-9602
ISBN 978-3-319-02698-5 ISBN 978-3-319-02699-2 (eBook)
DOI 10.1007/978-3-319-02699-2
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013958071

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The spatial dynamics of the web of interactions between organisations conducting joint Research and Development (R&D) activities – referred to as R&D networks – has recently evolved to one of the 'hot topics' in modern research of the *Geography of Innovation* literature. After the era of mainly focusing on direct, dyadic relations between actors performing joint R&D, emphasis is nowadays increasingly shifted to a network perspective. The latter extends the focus on dyads to the structure of indirect relations in a network of actors and its systemic implications. Recognising the importance of indirect ties and their potential role as channels for knowledge and information flows, the structure of these indirect ties is of major interest to understand and describe knowledge diffusion processes. Special interest is devoted to the interplay between spatial effects and structural effects at the network level in explaining the development of collaborative R&D and knowledge production activities.

In this context, network analytic methods and tools have increasingly come into play for the investigation of the spatial dimension of R&D interactions. By this, the field has become much more interdisciplinary, particularly in methodological terms. The more traditional spatial analysis techniques, spatial econometric approaches and spatial interaction models – which are without doubt still essential to investigate the spatial character of R&D networks – are increasingly augmented, sometimes merged with network analytic approaches, mainly comprising a set of tools stemming from graph theory. The realms of Complex Network Analysis (CNA) and Social Network Analysis (SNA) are essential to meet the aspiration of taking into account network structural effects that influence the spatial structure of R&D collaborations. In recent spatial studies of R&D networks, such network analytic methods are often combined with most recent advances in spatial analysis and spatial econometric modelling, for instance, by relating network structural effects – as captured by network analytic indicators – to spatial effects within a spatial econometric modelling framework.

In essence, the present volume explicitly reflects this recent development in spatial studies of R&D collaborations and networks. It constitutes a joint product of scholars analysing the geography of R&D networks from different angles, from distinct disciplinary backgrounds, using a diverse set of methodologies and producing a range of policy conclusions in diverse spatial and sectoral environments. By this, it represents – on the one hand – a quite unique collection of articles presenting methodological advancements for the analysis of R&D networks from different disciplines and – on the other hand – a distinguished anthology of novel empirical contributions on the relationship between geography and network structures as well as the impact of such networks on knowledge creation and innovative performance of firms, regions or countries.

The initial stimulus for the preparation of this volume was given at the congress of the European Regional Science Association (ERSA) in Bratislava in 2012. The emphasis on the geography of networks and R&D collaborations has been highlighted in various presentations and sessions of the congress. This volume is mainly the outgrowth of works that have been presented there, extended by an exclusive selection of invited works. The contributors come from all over the world and from a range of different disciplines, including economists, physicists, geographers and sociologists. They provide fresh ideas on the analysis of the geography of networks and R&D collaborations, both from a theoretical and a methodological perspective.

At this point, I would like to thank Folke Snickars and Manfred M Fischer for suggesting to propose such a volume to the *Advances in Spatial Science* series of Springer. Further, my warmest gratitude goes to all contributors of the volume, not only for their fine contributions in their chapters, but also for their motivating encouragement, stimulating discussions and smooth collaboration. My thanks also go to Barbara Fess, senior editor for economics and political science at Springer, for her ongoing support during the production process, and to Ramya Prakash, project manager at SPi Content Solutions – SPi Global, for her fine editing and production work.

Vienna, Austria
December 2013

Thomas Scherngell

Contents

Contributors

Corinne Autant-Bernard Université de Lyon, Lyon, France; Université Jean Monnet, Saint-Etienne, France; CNRS, GATE Lyon Saint-Etienne, Ecully, France

Pierre-Alexandre Balland University of Utrecht, Eindhoven Centre for Innovation Studies, Utrecht, Netherlands

Michael J. Barber Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria

Bernhard M. Dachs Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria

Thomas Brenner Economic Geography and Location Research, Philipps University Marburg, Marburg, Germany

Stefano Breschi Department of Management and Technology & CRIOS, Università L. Bocconi, Milan, Italy

Tom Broekel Institute of Economic and Cultural Geography, Leibniz University of Hanover, Hanover, Germany

Uwe Cantner Friedrich Schiller University Jena, Germany and University of Southern Denmark, Jena, Germany

Simone Caschili UCL QASER Lab and Centre for Advanced Spatial Analysis, University College London, London, UK

Alessandro Chessa IMT Lucca Institute for Advanced Studies, Lucca, Italy; Linkalab, Complex Systems Computational Laboratory, Cagliari, Italy

Yongwan Chun University of Texas, Dallas, TX, USA

Joan Crespo Political Science Institute, University of Toulouse, Toulouse, France

Manfred M. Fischer Institute for Economic Geography and GIScience, Vienna University of Economics and Business (WU Wien), Vienna, Austria

Koen Frenken School of Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands

Holger Graf Friedrich Schiller University Jena, Germany and TU Bergakademie Freiberg, Jena, Germany

Matté Hartog Section of Economic Geography, Faculty of Geosciences, Utrecht University, Utrecht, The Netherlands

Çilem Selin Hazir Université de Lyon, Lyon, France; Université Jean Monnet, Saint-Etienne, France; CNRS, GATE Lyon Saint-Etienne, Ecully, France

Barbara Heller-Schuh Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria

Slavomir Hidas Institute for Economic Geography and GIScience, Vienna University of Economics and Business (WU Wien), Vienna, Austria; Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria

Susanne Hinzmann Friedrich Schiller University Jena, Jena, Germany

Jarno Hoekman Department of Pharmaceutical Sciences & Department of Innovation Sciences, Utrecht University, Utrecht, The Netherlands

Manuela Korber Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria

Rafael Lata Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria

Sandra Leitner Vienna Institute for International Economic Studies, Vienna, Austria

Camilla Lenzi Department of Architecture, Built Environment and Construction Engineering, Politecnico di Milano, Milan, Italy

Andrea De Montis Dipartimento di Agraria, Università degli Studi di Sassari, Sassari, Italy

Manfred Paier Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria

Florian Reinold Institute for Economic Geography and GIScience, Vienna University of Economics and Business (WU Wien), Vienna, Austria; Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria

Thomas Scherngell Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria

Tamás Sebestyén MTA-PTE Innovation and Economic Growth Research Group and Department of Economics and Regional Studies, Faculty of Business and Economics, University of Pécs, Pécs, Hungary

Robert Stehrer Vienna Institute for International Economic Studies, Vienna, Austria

Raphaël Suire University of Rennes 1, CREM-CNRS, Rennes, France

Attila Varga MTA-PTE Innovation and Economic Growth Research Group and Department of Economics and Regional Studies, University of Pécs, Pécs, Hungary

Jérôme Vicente University of Toulouse, Toulouse Business School, and Political Science Institute, Toulouse, France

Iris Wanzenböck Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria

Martyna Wolska Institute for Economic Geography and GIScience, Vienna University of Economics and Business (WU Wien), Vienna, Austria; Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria

About the Editor

Dr. Thomas Scherngell is senior scientist in the Innovation Systems Department of the AIT Austrian Institute of Technology and lecturer at the Institute for Economic Geography and GIScience of the Vienna University of Economics and Business. He holds a venia docendi (habilitation) in Economic Geography and Regional Science, received from the Vienna University of Economics and Business in 2012. He received his M.A. (2003) (Mag. phil.) and his Ph.D. in Economic Geography (2006) from the same university (Dr. rer. nat.). Prior to working at the AIT, Thomas Scherngell was a full-time university assistant in the Department of Economic Geography and GIScience at the Vienna University of Economics and Business (2002–2007). Further, he held a postdoc fellowship at the University of Macao, China (2012–2013). He is an expert in the economics of innovation and technological change as well as in regional science and spatial analysis. Over the past five years, Thomas' research focus was on the structure and dynamics of R&D networks from a spatial perspective. He has been working extensively on spatial analysis methods, with a special emphasis on spatial econometric methods and spatial interaction modelling, and on network analysis methods. His works on R&D networks in Europe and China have been published and cited in major international journals. He is member of the Regional Science Association International (RSAI) and reviewer for leading international journals including *Science*, *Papers in Regional Science*, *Journal of Regional Science*, *Regional Studies* and *The Annals of Regional Science*.

Part I
Editorial Introduction

Chapter 1
The Networked Nature of R&D in a Spatial Context

Thomas Scherngell

1.1 Rising Interest in the Geography of R&D Networks

Starting with the seminal works of Feldman (1994) and Audretsch and Feldman (1996), the *Geography of Innovation* has – without doubt – evolved to one of the main research fields in Economic Geography and Regional Science. A great deal of theoretical and empirical literature has been followed in this area, drawing on significant methodological advancements in spatial analysis, spatial statistics and spatial econometrics as well as on the availability of novel, systematic information sources on the innovative activity of firms, regions and countries. The Geography of Innovation literature describes the role of proximity and location for innovative activity. It is emphasised that spatial studies of innovation provide pivotal anchor points for understanding and explaining the space-economy (see Feldman and Kogler 2010).

Over the past decade, we have observed an increasing research interest within the Geography of Innovation literature on the spatial dimension of networks and collaborations between actors conducting joint Research & Development (R&D) activities. This subfield has meanwhile become an essential and fascinating domain for advanced research on the spatial and temporal evolution of innovation systems at different spatial scales. Special emphasis is placed on interactions between organisations performing joint R&D, for instance in the form of collaborative research projects, joint conferences and workshops, or shared R&D resources in the form of labour and capital. Such interactions have attracted a burst of attention in the last decade, both in the scientific and in the policy sector (see, for instance, Autant-Bernard et al. 2007). With the focus on networks and R&D collaborations, the Geography of Innovation literature clearly has become more interdisciplinary – in particular in methodological terms – involving a multiplicity of scientific fields

T. Scherngell (✉)
Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria
e-mail: thomas.scherngell@ait.ac.at

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_1,

such as economics, geography, social sciences, physics and complex systems research (see Reggiani and Nijkamp 2009).

The research focus on the geography of R&D networks has been triggered by various considerations in theoretical and empirical literature in Economic Geography and Regional Science in the 1980s and 1990s (see, e.g. Clark et al. 2000). When we recapitulate the development of this literature stream, two arguments for the focus on networks are central:

First, innovation, knowledge creation and the diffusion of new knowledge are the key vehicles for sustained economic growth of firms, industries or regions, and, thus, are essential for achieving sustained competitive advantage in the economy (see, e.g., Romer 1990; Lucas 1988; Grossman and Helpman 1991). The theory of endogenous growth and the geography-growth synthesis both consider that economic growth and spatial concentration of economic activities emanate from localised knowledge diffusion processes (Autant-Bernard et al. 2007). The fundamental neoclassical assumption of constant or decreasing returns to scale is contested, assuming that knowledge may be subject to increasing returns because of the externalities inherent in its production and use. In this respect, the value of the geographically localised knowledge base increases due to network effects and the characteristics of knowledge. Network effects come into play, since a diversified set of local actors may gain access to new knowledge. The properties of knowledge crucial for this argument are non-excludability – knowledge is accessible to actors that invest in the search for it – and non-rivalry – knowledge can be exploited by different innovating actors simultaneously (see Feldman and Kogler 2010).

Second, interactions, research collaborations and networks of actors have become an essential element for successful innovation (see, for instance, Fischer 2001). Long viewed as a temporary, inherently unstable organisational arrangement, R&D networks have become the norm rather than the exception in modern innovation processes (Powell and Grodal 2005). Organisations must collaborate more actively and more purposefully with each other in order to cope with increasing market pressures in a globalizing world, new technologies and changing patterns of demand. In particular, firms have expanded their knowledge bases into a wider range of technologies (Granstand 1998), which increases the need for different types of knowledge, so firms must learn how to integrate new knowledge into existing products or production processes (Cowan 2004). It may be difficult to develop this knowledge alone or acquire it via the market. Thus, firms form different kinds of co-operative arrangements with other firms, universities or research organisations that already have this knowledge to access it faster.

The fundamental importance of networks for generating innovations is also reflected in the various systems of innovation concepts (see Lundvall 1992 among many others). In this conception, the sources of innovation are often established between firms, universities, suppliers and customers. Network arrangements create incentives for interactive organisational learning, leading to faster knowledge diffusion within the innovation system and stimulating the creation of new knowledge or the combination of pieces of existing knowledge in a new way. Participation in innovation networks reduces the high degree of uncertainty present in innovation processes, providing fast access to different kinds of knowledge, in particular tacit knowledge (see, for example, Kogut 1988).

Science, Technology and Innovation (STI) policies have recently followed this trend, shifting emphasis to the support of networks and collaborative arrangements between innovating actors, in particular between universities and firms. At the European level, the Framework Programmes (FPs) for Research and Technological Development (RTD) are the prime examples of policy programmes to support collaborative knowledge production across Europe. This has led to the establishment of a pan-European network of actors performing joint R&D (see, e.g., Scherngell and Barber 2009). From this background, not only the scientific domain, but also the policy sector shows increasing interest in network structures and network dynamics driven by public funds. In a European policy setting, particular interest is devoted to the geography of such networks, bearing in mind the overall policy goal of an integrated European Research Area (ERA).

The focus of this volume is on the geographical dimension of interactions in networks and R&D collaborations. While early contributions to the Geography of Innovation literature highlight the localised character of knowledge production and diffusion, one of the most fundamental questions of current research is how the structure of formal and informal networks modifies and influences the spatial and temporal diffusion of knowledge (see Autant-Bernard et al. 2007). As highlighted by Reggiani and Nijkamp (2009), the foundation for an interpretation of the economy as an interdependent complex set of economic relationships has long been underpinned by the "first law of geography" (Tobler 1970), stipulating that everything in space is related to everything else, but nearby things are more related than distant things. However, advances in network theory may challenge or – at least – extend this statement, assuming that in certain network typologies distant things may be more related than near things.

In the Geography of Innovation literature, such considerations are referred to as the local buzz vs. global pipelines nature of knowledge creation. This concept describes the interplay between the interaction behaviour of localised innovating actors, mainly driven by spatial proximity, and the access and transfer of more distant knowledge, mainly distributed via alternative channels, often in more formalised form as, for instance, by networks of joint R&D projects between firms providing complementary, highly specialised knowledge (Bathelt et al. 2004). Assuming that the relative importance of such geographically dispersed and more distant knowledge sources – transferred over network channels – increases, certain network structures may be considered as essential determinants of how knowledge diffuses in geographical space, and why some actors, regions or countries benefit more than others due to certain network positions.

However, these theoretical considerations rest on a small base of empirical evidence (see Feldman and Kogler 2010), which may be related to methodological limitations as well as to a lack of data and insufficient information on different types of R&D networks and collaboration patterns. In methodological terms, we need to combine existing spatial analytic tools with methods coming from sociology, in particular Social Network Analysis (SNA) (see Ter Wal and Boschma 2009), or from physics and complex systems research (see, e.g., Reggiani and Nijkamp 2009). However, until now it remains in many aspects unclear in which way and how these different methodological streams can complement each other in a meaningful way.

1.2 Motivation, Objective and Structure of the Book

From this perspective, the motivation of this book is to bridge the research gap discussed above. There are two objectives: *First*, the volume aims to advance the theoretical basis and the methodological toolbox for the investigation of the geography of networks and R&D collaborations. *Second*, it aims to provide novel empirical evidence on spatial network structures and the impact of R&D networks on knowledge creation and diffusion which is particularly to be interpreted in respect to current European STI policies. In this sense, the books brings together a selection of articles providing novel theoretical and empirical insights into the geographical dynamics of networks and R&D collaborations, using new, systematic data sources, and employing cutting-edge spatial analysis, spatial econometric and network analysis techniques. It simultaneously provides a collection of high-level recent research on the spatial dimension of R&D collaboration networks, and contributes to the recent debate in Economic Geography and Regional Science on how the structure of formal and informal networks modifies and influences the spatial and temporal diffusion of knowledge.

Given the focus of the book on the geography of networks and R&D collaborations, with the aim to methodologically advance analytic approaches for the analysis of such networks in a spatial context, and to provide novel empirical evidence on structure and impact of R&D networks, the volume comprises three major parts. Initially, Part II shifts attention to methodological advancements from an interdisciplinary perspective, while Parts III and IV are two thematic sections focusing on structure and impact of R&D networks in a STI policy context.

Part II, entitled *Analytic advances and methodology*, comprises a selection of articles providing insight into novel and advanced methodologies for the analysis of R&D networks – formally defined as a set of nodes, most often representing organisations, inter-linked by a set of edges, most often representing joint R&D activities – in a spatial context. One essential element of this section is to bring together methodological approaches from different disciplines, ranging from advanced spatial analysis tools to network analysis approaches coming from statistical physics, sociology and complex systems research. Part II highlights different modelling approaches for investigating the spatial structure of R&D networks and how it changes over time. From this perspective, the section significantly addresses a research issue raised by many economic geographers and regional scientists in the recent past, inspiring a look at alternative methodological and analytical approaches coming from related disciplines for the spatial analysis of networks, such as, for instance, Social Network Analysis (SNA) techniques (see, e.g., Bergman 2009; Ter Wal and Boschma 2009).

Part III, entitled *Structure and spatial characteristics of R&D networks*, shifts emphasis to the empirical analysis of real world R&D networks from a geographical perspective, employing advanced methods of spatial analysis, spatial econometrics and network analysis, some of them introduced in Part II in an abstract manner. By this, the articles gathered in Part III provide new insight into the research questions raised above, as, for instance, on the effects of different forms of proximity on the constitution of R&D networks at different spatial scales and in different economic sectors of activity. Another common focus of the articles in this section is that they use novel, systematic data and information sources on different kinds of R&D networks, such as, for instance, project-based R&D networks constituted under the heading of the European Framework Programmes (FPs).

Part IV, entitled *Impact of R&D networks and policy implications*, puts emphasis on the crucial question on how structure and dynamics of R&D networks affects knowledge creation and inventive behaviours of innovating actors. Since modern STI policies have shifted their focus on supporting such networks, this section provides important implications in a STI policy context, particularly at the European level. This is of crucial importance, since the realisation of an integrated ERA is one of the major goals of the STI policy strategy of the European Commission (see, e.g., Hoekman et al. 2013). Networks of actors performing joint R&D should span the territory of the EU – stimulating the circulation of knowledge and researchers in a Europe-wide system of innovation – and, thus, the analysis of the spatial dimension of European R&D networks shows direct European policy relevance. In this sense, the articles gathered in Part IV address the essential points: how to interpret results from empirical investigations of spatial R&D networks in a STI policy context, and how potential policy implications and measures may be derived.

1.3 Overview of the Chapters

As mentioned in the previous section, Part II of the volume focuses on analytic and methodological advances – from an interdisciplinary perspective – for the investigation of R&D networks and R&D collaborations in a spatial context. After this introductory chapter, Part II begins with a contribution by Autant-Bernard and Hazir (Chap. 2) focusing on different modelling approaches and underlying conceptions for network formation in a geographical context. The article provides a review – as a reasonable starting point for Part II – on recent works that investigate network formation in space and time but reveal a high variation in terms of methodological and analytical approaches. In doing so, the authors discuss the different aspects of the relationship between geography and networks, and discuss in some detail the distinct methodological approaches and their capability to investigate this relationship. Chapter 3 authored by De Montis, Caschili and Chessa shifts attention to a complex systems research perspective for investigating spatio-temporal network dynamics, in particular for spatial systems with a very large number of nodes and vertices. The authors present a state-of-the art summary in the field of complex network analysis, laying special emphasis on the issue of community detection in networks which is of crucial interest when describing R&D network structures (see also Chap. 9 of this volume by Barber and Scherngell). Communities, defined as homogenous, densely connected sub-networks, are a key element for understanding the network structure as a whole. The authors demonstrate this by means of a case study employing a network community detection approach to study the problem of regionalisation on the island of Sardinia (Italy).

Part II continues with two contributions introducing two distinct analytical approaches for the investigation of spatial network structures that have initially been applied mainly in an a-spatial context. Initially, Broekel and Hartog (Chap. 4) focus on exponential random graph models (ERGM) to analyse the determinants of cross-region R&D collaboration networks. The authors lay special emphasis on advantages and disadvantages of this approach in comparison to a spatial interaction modelling perspective that is often used to disentangle the influence of different types of proximities on R&D network structures (see, e.g., Scherngell and Barber 2009). The solidity of the ERGM approach is demonstrated by means of an illustrative example focusing on the structure of cross-region R&D networks of the German chemical industry. After that, Sebestyén and Varga (Chap. 5) develop a novel index, labelled Ego Network Quality (ENQ), for measuring the quality of network position and node characteristics in spatial R&D networks. The authors demonstrate that the ENQ is an integrated measure for the network position of a specific node in a spatial context, very much resembling to the solution applied in the well-established index of eigenvector centrality in an a-spatial context. Robustness and weighting schemes of the index are tested via simulation and econometric techniques.

Chapter 6, authored by Chun, discusses the notion of network autocorrelation, referring to a situation when network links from a particular origin may be spatially autocorrelated with other flows that have the same origin, and, similarly, network links into a particular destination may be correlated with other flows that have the same destination. The author argues that this invalidates the independence assumption of network flows, raising the need for a proper modelling method which can account for network autocorrelation. The eigenvector spatial filtering method is presented as an effective way to incorporate network autocorrelation in linear regression and generalised linear regression models. Chun illustrates these methods with applications to interregional commodity flows and interstate migration flows in the U.S.

Part II closes with a contribution by Crespo, Suire and Vicente (Chap. 7) on the assortativity and hierarchy in localised R&D collaboration networks. By this, the authors focus on two important structural properties and present a combination of two SNA measures, degree distribution and degree correlation, to study whether such localised networks are allowed to avoid technological lock-in.

The contributions gathered in Parts II and III comprise a selection of articles providing novel empirical evidence on real world R&D networks from a spatial perspective. Initially Part III shifts attention to the investigation of spatial network structures and dynamics. The section opens with a contribution by Lata, Scherngell and Brenner (Chap. 8) that puts emphasis on observing integration processes in European R&D from a network perspective. The authors investigate co-patent and project based R&D networks, and estimate the evolution of separation effects over the time period 1999–2006 that influence the probability of cross-region collaborations in these distinct networks. They use Poisson spatial interaction models accounting for spatial autocorrelation among network links. Chapter 9, authored by Barber and Scherngell, employs community detection (see Chap. 3 of this volume) to characterise the structure of the European R&D network using data on R&D projects funded by the fifth European FP (FP5). Communities are subnetworks whose members are more tightly linked to one another than to other members of the network. The identified communities are analysed with respect to their spatial distribution and by means of spatial interaction models.

Chapter 10, authored by Leitner, Stehrer and Dachs, focus on the global R&D network, proxied by R&D investment flows between countries. The authors analyse internationalisation patterns of business R&D for OECD countries and identify specific home- and host-country characteristics that are conducive or obstructive to cross-border R&D expenditure of foreign affiliates.

Chapters 11, 12 and 13 investigate spatial aspects of different networks constituted under the heading of the FPs at an organisational and R&D project specific level. Initially, Reinold, Paier and Fischer (Chap. 11) explore determinants of inter-organisational knowledge generation – proxied by joint publications or patents resulting from joint FP projects – by means of a binary response model using novel data from a survey among FP5 participants. Chapter 12 by Hazir presents an empirical investigation on the formation of multilateral FP collaboration networks in the Biotechnology field employing exponential random graph models (ERGM). The author focuses on the question how geography and heterogeneity in institution types affect the way organisations form R&D networks. Chapter 13, authored by Vicente, Balland and Suire, completes Part IV adopting a SNA perspective to analyse collaborative projects funded in FP5 and FP6. They study the properties both of the network of organisations and the network of collaborative projects, focusing on the particular case of Global Navigation Satellite Systems (GNSS) in Europe.

Part IV turns to the impact of R&D networks on knowledge creation and inventive behaviours of organisations, and its consequences for STI policy. As a starting point, the contribution of Hoekman and Frenken (Chap. 14) frames the geography of scientific research networks laying special emphasis on empirical studies that evaluate policy efforts to support the creation of ERA. The authors introduce a logic of proximity, intended to provide researchers with a way to coordinate their networks, and a logic of stratification, intending to provide pathways for researchers to get involved in networking. The chapter presents an overview of recent empirical findings to illustrate the interplay between proximity and stratification of European R&D networks, and discusses potential implications for future ERA policies. Chapter 15 by Wanzenböck and Heller-Schuh connects very well to this discussion, as it stresses the importance of specific network positions to gaining access to knowledge located further away in geographical space. They analyse the position of regions in the European network of R&D collaboration within the FPs in the time period 1998–2006. By means of a panel version of the Spatial Durbin Model (SDM), the authors identify determinants that push a region in a specific, favourable network position to gain access to region-external knowledge.

Chapters 16 and 17 are among the first contributions that aim to establish a direct link between network structures and network impact in terms of knowledge creation and inventive behaviours of innovating organisations. Chapter 16 by Breschi and Lenzi analyses R&D networks among 331 US cities using patent data for the period 1990–2004. The authors investigate the impact of network participation in driving the spatial diffusion of scientific and technological knowledge. They propose new indicators that are intended to capture US cities' propensity to engage not only in local, but also global, knowledge exchanges, and relate these propensities to cities' inventive and economic performance. The contribution of Hidas, Wolska, Fischer and Scherngell (Chap. 17) is in a similar spirit in that it aims to explain inventive performance by means of network participation. The authors identify and measure effects of research collaboration networks on knowledge production at the level of European regions, using a panel data SDM relationship for empirical testing.

Chapters 18 and 19 focus on different types of policy induced R&D networks, and the impact of policy initiatives on network formation and innovative outcome. Cantner, Graf and Hinzmann (Chap. 18) analyse the impact of governmental funding on cooperation networks in Germany under the heading of the so-called Leading-Edge Cluster Competition. The authors identify the extent of policy influence for selected clusters on the network of the most important cooperation partners, its geographic reach, and network dynamics. Chapter 19 by Korber and Paier provides an alternative approach to investigate the relationship between STI policy funding schemes, R&D collaborations and innovative performance. The contribution presents an agent-based simulation model to explore the relationship between a specific type of policy-induced networking, so called competence centres, and innovative outcome in the Viennese Life Sciences innovation system.

The volume closes with Chap. 20, which provides a synthesis of the main empirical results, methodological advancements and policy implications. Furthermore, ideas for a future research agenda are presented, emphasising the need for further crossing of disciplinary boundaries for the future investigation of the spatial dimension of R&D networks and R&D collaborations.

References

Audretsch D, Feldman M (1996) R&D spillovers and the geography of innovation and production. Am Econ Rev 86:630–640

Autant-Bernard C, Mairesse J, Massard N (2007) Spatial knowledge diffusion through collaborative networks. Pap Reg Sci 86:341–350

Bathelt H, Malmberg A, Maskell P (2004) Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. Prog Hum Geography 28:31–56

Bergman E (2009) Embedding network analysis in spatial studies of innovation. Ann Reg Sci 43:559–565

Clark GL, Feldman MP, Gertler M (2000) Handbook of economic geography. Oxford University Press, Oxford

Cowan R (2004) Network models of innovation and knowledge diffusion, Unu-MERIT working paper series, 2004–016

Feldman M (1994) The geography of innovation. Kluwer, Boston

Feldman M, Kogler DF (2010) Stylized facts in the geography of innovation. In: Hall BH, Rosenberg N (eds) Handbook of the economics of innovation, vol 1. North Holland, Amsterdam, pp 381–410

Fischer MM (2001) Innovation, knowledge creation and systems of innovation. Ann Reg Sci 35:199–216

Granstand O (1998) Towards a theory of the technology-based firm. Res Policy 27:465–489

Grossman GM, Helpman E (1991) Quality leaders in the theory of growth. Rev Econ Stud 58:43–61

Hoekman J, Scherngell T, Frenken K, Tijssen R (2013) Acquisition of European research funds and its effect on international scientific collaboration. J Econ Geogr 13:23–52

Kogut B (1988) Joint ventures: theoretical and empirical perspectives. Strateg Manage J 9:319–332

Lucas RE (1988) On the mechanics of economic development. J Monet Econ 22:3–42

Lundvall BA (1992) National innovation systems. Pinter, London

Powell WW, Grodal S (2005) Networks of innovators. In: Fagerberg J, Mowery DC, Nelson RR (eds) The Oxford handbook of innovation. Oxford University Press, Oxford, pp 56–85

Reggiani A, Nijkamp P (2009) Simplicity in complex spatial systems. In: Reggiani A, Nijkamp P (eds) Complexity and spatial networks. Advances in spatial science series. Springer, Berlin, pp 1–7

Romer PM (1990) Endogenous technological change. J Politi Econ 98:71–102

Scherngell T, Barber M (2009) Spatial interaction modelling of cross-region R&D collaborations. Empirical evidence from the 5th EU Framework Programme. Pap Reg Sci 88:531–546

Ter Wal ALJ, Boschma R (2009) Applying social network analysis in economic geography: framing some key analytical issues. Ann Reg Sci 43:739–756

Tobler W (1970) A computer movie simulating urban growth in the Detroit region. Econ Geography 46:234–240

Part II
Analytic Advances and Methodology

Chapter 2
Network Formation and Geography: Modelling Approaches, Underlying Conceptions, Recent and Promising Extensions

Corinne Autant-Bernard and Çilem Selin Hazir

Abstract Due to the strong polarisation of economic activities in space and rise in collaborative behaviour, increasing attention has recently been devoted to the relationship between geography and network formation. The studies conducted on this topic reveal a high variation in terms of methodologies. Putting special emphasis on R&D networks, the aim of this chapter is to review the different methods and assess their ability to address the issues raised by the relationship between network and space. We first discuss the different facets of the relationship between geography and networks. Then, we detail the methodological approaches and their capability to test each effect of geography on network formation. We argue that the effect of distance on dyads have received the major attention so far, but the development of block modelling and top-down approaches opens new research perspectives on how distance or location might affect formation of more complex structures. Moreover, recent improvement in temporal models also offers opportunities to better separate spatial effects from that of influence over time.

2.1 Introduction

In the field of economics, the relationship between geography and network formation attracts attention in order to understand how knowledge flows in a space of social interactions relate to regional growth and innovation. So far a number of studies have been conducted to elucidate this relationship. Even a glimpse on these studies reveals a high variation in terms of methodologies.

C. Autant-Bernard (✉) • Ç.S. Hazir
Université de Lyon, Lyon F-69007, France

Université Jean Monnet, Saint-Etienne F-42000, France

CNRS, GATE Lyon Saint-Etienne, Ecully F-69130, France
e-mail: corinne.autant@univ-st-etienne.fr; cilem.selin.hazir@univ-st-etienne.fr

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_2,

On the one hand, this variety stems from the fact that the term "*geography*" contains a number of meanings in it. Sometimes geography is associated with physical separation, sometimes it refers to locations as a material and relational context for economic action, and sometimes geographical units themselves are considered as nodes in a network. The way it is conceived, in turn affects the way it is related to network formation and constrains model choices as some models are not capable of testing all kinds of effects. On the other hand, the variety in methodologies results from addressing the same phenomenon; i.e. formation of a network, through different analytical perspectives.

From a "learning perspective" these differences enclose invaluable information on the evolution of the way that the research community has conceived and addressed the geographical dimension of network formation, and on possible future directions. In this regard, this chapter will try to disclose this information by elaborating how different meanings associated to geography can yield different conceptualizations of geography-network relationship. Hence in Sect. 2.2 we will address alternative ways of relating geography to network formation. In Sect. 2.3, we will try to identify main distinctions between different methodologies and compare models that are widely used in the study of spatial dimension of R&D networks. Our aim here is not to provide a full-fledged list and a hierarchy of network formation models but rather to highlight main differences in analytical approaches putting emphasize on their ability to address the issues raised by the relationship between network and space. Finally, we will review some recent methodological advances that loom large regarding their potential future contributions to understand knowledge flows in space.

2.2 Relating Geography to Network Formation

2.2.1 A Tie Covariate: Physical Distance

One of the meanings associated to geography is the physical distance, which is the relative position or physical separation of two entities. Under this definition, space is perceived to be homogenous and exogenous to the network formation process due to the fact that regardless of the configuration of the network, the physical distance among nodes remains unchanged. Then, the role of geography is conceptualized as the effect of an attribute of a possible tie; i.e. the length of a tie.

High levels of this attribute is hypothesized to have a negative effect on the utility out of being connected[1] due to the fact that there exists a *tacit* component of knowledge (Polanyi 1966) and some interaction is necessary for its transmission. Therein, physical proximity is considered to be a facilitator of face-to-face

[1] As shall be seen in the succeeding section, this utility either refers to a utility obtained out of a tie (see binary choice models), or to the utility out of the overall network (see ERGM).

interactions, which in turn eases the transmission of tacit knowledge (Feldman and Florida 1994) and hence increases the utility of being connected.[2] Also, physical proximity is assumed to increase this utility via enabling cross-fertilization of ideas (Feldman and Florida 1994) and timely inflows of information (Feldman 1993) and by decreasing the cost of collaboration (Hoekman et al. 2009).

However, the fact that physical distance is just one of the many dimensions of separation (Boschma 2005) and in particular the embeddedness of economic relations between firms and individuals in social relations (Granovetter 1985) has modified this hypothesis. Thus, it has become a matter of interest to know whether physical distance still plays a role on the utility of being connected when the effects of other dimensions of separation are controlled for.

2.2.2 A Node Covariate: Local Context

Another meaning that is associated to geography is the physical context that economic agents are embedded in. Once the context that embraces networking agents is taken into account, then the network becomes embedded in a physical space. One way to relate this embeddedness to network formation is to consider the physical space as an exogenous setting, which affects the attractiveness of the organizations as potential partners or their capacity to establish connections. In that case, the role of geography is conceptualized as the effect of a node attribute on network formation. In the literature, this effect is formulated in a number of ways such as the effect of agglomeration economies, knowledge externalities, system of innovations, or innovative "milieu".

Although, considering geography as an exogenous node attribute simplifies the analytical processes to study network formation; obviously the local processes and network processes are not mutually exclusive. On the one hand, the black-box of advantages that a location provides might also include the outcomes or impacts of network activity of its constituents. On the other hand, some local processes might not only work through increasing node attractiveness or capacity but also through creating tie dependence as will be discussed in the sequel.

[2] However, if proximity is often associated with the tacit dimension of knowledge, we must avoid an overly simplistic view (Massard and Mehier 2009). There are probably complementarities between tacit and codified knowledge, any two being transmitted both locally and remotely. The link between proximity and knowledge can then lie in the way of combining the tacit and codified nature of knowledge.

2.2.3 A Factor Affecting Tie Dependence: Physical Distance and Local Context

Pattison and Robins (2002) argue that each network tie could be associated with a "*social locale*", which refers to "*a complex relational entity that links the geographical, social, cultural and psychological aspects of the context for social action*". They argue further that these social locales overlap with each other due to the fact that "*the outcome of processes in one locale may have some impact on processes within another locale*". Therein, a local context might be considered as a joint social locale for ties created within it, as they all share a number of intermingled local processes such as social, economic, political, historical processes. The outcomes of these processes might be heterogeneous across space and they may create, enhance or even dampen dependencies among ties. Similarly, being spatially proximate could be associated with overlaps in social locales as being spatially proximate might mean sharing similar local features.

Hence, in this case the role of geography can be conceptualized as the effect of tie dependence on network formation. Unlike considering the role of geography as the effect of a tie attribute, in this conceptualization the specific role played by distance is not disentangled from the role of other types of proximities or processes that co-exist or interact with geographical proximity.

2.2.4 Regions as Nodes Themselves

As a matter of fact, geographical units may themselves constitute the nodes in a network. In the case of networks representing economic relations, regions as nodes symbolize the aggregate behaviour of individuals. Hence, all three types of roles discussed above might be relevant to study the inter-regional networks. The role played by the distance between two regions or existence of a common border might again be considered as an exogenous tie property. Regional properties that might affect the aggregate performance of individuals can be considered as exogenous node attributes under the assumption that network processes and these properties are mutually exclusive. Finally, contiguity or co-location in a wide geographic area can be conceptualized as a factor affecting overlaps in social locales.

2.3 Approaches to Model Network Formation

Networks attract attention from a wide range of fields like medicine, biology, computer science, sociology, political science, economics, etc. Accordingly, a number of different analytical approaches have been suggested to model their formation. A major distinction among these approaches stems from considering the network as an outcome of “choice” or “chance” (Jackson and Wolinsky 1996). In the first view, formation of a network is explained on the basis of individual incentives (costs and benefits) (Jackson and Wolinsky 1996). A number of strategic and game theoretic models have been developed along this view. On the other hand, graph-theory has bestowed various random graph models in line with the second view, where the observed network is considered as just one realization among all possible network configurations. Beside random graphs, complex network analysis has been developed along the same line. Finally, the usual econometric models and spatial econometrics have also been applied to study the network, where both views are in play.

These approaches may also be classified into two as static approaches and dynamic approaches. The former works on a snapshot of the network; whereas the latter considers the evolution of the network in time. Among those, some models allow creation of new nodes in time as in the case of preferential attachment model (Barabási and Albert 1999). Some others allow studying the dynamics stemming from creation and dissolution of ties among a fixed set of nodes in time as in the case of stochastic actor-based models (Snijders et al. 2010).

As a third classification, these approaches can be considered in two groups as top-down approaches and bottom-up approaches. Top down approaches focus on the topology of the network as a whole and try to identify global features rather than modelling the network on the basis of individuals. Complex network analysis or block modelling (Nowicki and Snijders 2001), where the aim is to identify groups, members of which are equivalent in terms of their connection patterns, may illustrate this approach. On the other, hand bottom-up approaches focus on processes taking place in components of the network. Therein, a further distinction can be made among bottom-up approaches with respect to the types of components that they focus. In some approaches the network configuration is explained by focusing on the behaviour of actors, ex: stochastic actor-based models (Snijders et al. 2010). Whereas in some others the focus is either on formation of a single tie or a local pattern (a subset of ties).

Another distinction among these approaches could be made with respect to underlying assumptions on tie dependence. Some models base on the assumption that the stochastic processes behind formation of ties work independently. Some others assume that the outcomes of these stochastic processes are correlated. Finally a third group assumes that some ties are realized jointly through the same stochastic process.

In the sequel, we will focus mainly on the empirical studies that investigate the role of geography in R&D networks. We will discuss them under three headings: network as the equilibrium of choices; network as an outcome of choice and random effects, and network as an outcome of a random process. We will try to highlight the differences in the analytical process among these models in terms of the above-mentioned criteria and their capacity to handle alternative ways of relating geography to network formation

2.3.1 Network as the Equilibrium of Choices

As mentioned earlier the game theoretic approach considers the network as the outcome of individual choices. Among these models the seminal work by Jackson and Wolinsky (1996) has considerable influence on both theoretical and empirical work on the geographical dimension of R&D networks. Their model, known as the connections model, explains the formation of a network on the basis of individual incentives (costs and benefits) and bases on the idea that agents do not only benefit from those they are linked directly; but also from those they are linked indirectly. The benefit they can obtain from others decreases with distance; but direct links are costly implying a trade-off between the benefits and costs of a direct link.

The spatial extensions of this model is provided by Johnson and Gilles (2000) and Carayol and Roux (2007). In these extensions the role of geography is investigated in a static network, where the number of nodes is fixed. Geography is considered as the geographical distance and its role is hypothesized as an exogenous factor affecting the cost of maintaining a link. Based on this conception on the geography-network formation relationship, these theoretical models suggest that for a wide range of intermediary values of decay in transmission of knowledge, a particular stable network structure called “small world” emerges. Carayol and Roux (2007) also provide some empirical evidence by fitting the model to actual co-inventions that took place during 1977–2003 with at least one inventor located in France.

2.3.2 Network as an Outcome of Choice and Random Effects

While in the game theoretic models the network is considered as the equilibrium of individual choices, in some statistical models used to study connections among nodes we see an expression of the utility that an individual can obtain out of its choice and some notion of randomness in making that choice. In the sequel, these models will be explained briefly and their capacity to integrate the geographical dimension will be discussed.

Binary Choice Models. The use of Binary Choice Models illustrates the application of usual econometric tools to study network formation (Geuna 1998; Powell et al. 2005; Mairesse and Turner 2005; Autant-Bernard et al. 2007; Paier and Scherngell 2008). These models aim at explaining the factors that affect realization of a single tie; hence they analyse formation of a network by focusing on its smallest unit. Factors that are symmetric for a pair of nodes, i.e. tie attributes, are the easiest ones to test with these models. Some practical problems arise in studying the effect of node attributes since the explanatory variables have to be symmetric and hence insensitive to the changes in the order of indexation. Finally, these models allow studying the effect of the observed network configuration on tie formation but under the assumption that it is an exogenous factor. This stems from the fact that in these models realization of a tie is supposed to be a Bernoulli process, meaning that ties are realized independently of each other.

Therein, the capacity of Binary Choice Models to investigate the role of geography mainly lies in the ability to study how physical distance affects the probability that a tie is created given the effect of other factors. This ability complies with the research interest to demarcate the role of geographical proximity from that of other proximity dimensions. In these models geography may also be included as a node attribute as long as they are defined symmetrically for the pair of nodes. Finally, due to the tie independence assumption, with these models it is not possible to study the role of geography in terms of tie dependence.

Poisson Regression Models and Gravity Models. The analytical process and the assumptions in Poisson Regression Models are the same as those in Binary Choice Models except for the fact that the objective is to explain the intensity of interaction among a pair of nodes rather than its existence (Powell et al. 1996). Hence, they allow studying the role of geography on the intensity of interactions, where this role could be introduced as a tie or node property (Mairesse and Turner 2005; Frachisse 2010). Once distance is accounted for, Poisson models can be interpreted as gravity models. As in the case of Poisson Regression Models, the objective in Gravity Models is to explain the strength of interaction among two spatial units. Hence, the approach undertaken to explain for the network builds upon ties among pairs. This type of models can be applied to individual choices or aggregated behaviour. It is worth noticing however that much attention has been devoted so far to study inter-regional networks, hence focusing on aggregated data.

The use of these models illustrates an application of spatial analysis techniques to study network formation. The earlier studies using Gravity Models assume that the stochastic process behind tie formation works identically and independently; i.e. any pair of ties, among the same pair of nodes or not, are independent (Ponds et al. 2007; Maggioni et al. 2007; Scherngell and Barber 2009; Hoekman et al. 2010). More recent applications (Scherngell and Lata 2011) take the spatial autocorrelation among flow residuals into account and corrects for this by using eigenvector filtering. Hence, the extension with spatial filtering rests upon weaker assumptions on tie dependence since it handles the correlation among ties sharing the same node.

As Gravity Models include two mass terms and a separation function; they allow studying the role of geography as a node itself with some attributes and as a tie attribute. The extensions dealing with spatial autocorrelation might allow controlling for correlations among intensity of interactions resulting from the topology of regions. Hence, the specific role played by the physical distance might be identified better as suggested in Chap. 11 of this book.

Stochastic Actor-Based Models. Stochastic Actor-based Models are statistical models to study tie dynamics in networks of fixed size (Snijders et al. 2010). As the name implies they focus on the behaviour of actors and model the formation of the network by means of changes that actors make in their outgoing ties. These changes are explained by means of two functions. The former is the rate function showing the frequency at which a change occurs. Whereas, the latter refers to the objective function, which shows the probabilities of alternative courses of action given the opportunity to make a change. This function is expressed in terms of "effects", which are tendencies (like reciprocity, closure, multi-connectivity etc.) taking place locally.[3] Both functions may depend on network position of actor and some actor attributes.

These models assume that actors act independently; hence the changes they make are not coordinated yet sequential. However, as the outcomes of their decisions change each other's environment, in time their actions depend on each other. Thus, unlike Binary Choice Models, where for each pair of agents the rest of the network is considered exogenous simultaneously; the sequential nature of Stochastic Actor-based Models allow handling dynamism in choices and dependencies on the environment.

Geographical dimension might be introduced in these models through both the rate and the objective function. A rate function differentiated with respect to location of actors might enable spatial heterogeneity in frequency of tie changes. On the other hand, the objective function might be modified either by integrating the distance as a dyadic covariate (Ter Wal 2013), or location as a node attribute (Balland 2012) which in turn might be used to study the effect of co-location and some network effects arising from being co-located.

2.3.3 Network as an Outcome of a Random Process

As mentioned earlier the graph theoretic approaches consider the observed network as an outcome of a random process. Hence, these approaches do not base on utility functions of micro agents but on the distribution of probabilities. Nevertheless, it should also be noted that although a utility function is not specified in these models, the distribution of probabilities can be constrained using a theoretical basis on preferences of agents.[4] Below, graph theoretic approaches used to study geographical dimension of R&D networks are discussed.

Exponential Random Graph Models (ERGM or p*). ERGMs are (Frank and Strauss 1986; Wasserman and Pattison 1996) more recent types of random graph

[3] These effects are similar to the "local configurations" in Exponential Random Graph Models that will be discussed in the sequel.

[4] As shown by Park and Newman (2004) random graph models can be expressed as a constrained maximum entropy problem; which maximizes the entropy in the probability distribution of observing a particular network configuration. In the earlier random graph models (Erdös and Renyi 1959) the problem is constrained only by the number of edges in the network and a probability distribution which assigns the same probability to all networks with the same number of edges is obtained. However, in more recent models as shall be seen in subsection on Exponential Random Graph Models, the preferences of actors for homophily, central agents, closure, etc. can be used as additional constraints by defining local configurations accordingly.

models. They allow studying networks with a fixed set of nodes. However, as will be discussed in Sect. 2.4, temporal extensions that allow dynamism in terms of tie creation and dissolution have recently become available. The idea behind these models is that the observed network is just one realization of all possible configurations of connections among a given set of nodes. Hence, as stated by Cranmer and Desmarais (2011), there is a "conceptual leap" from the Binary Choice and similar models to ERGM. While in the former the vector of interest is a series of values drawn from a univariate distribution; in ERGM it is considered as a single draw from a multivariate distribution. This leap allows relaxation of the tie-independence assumption and provides ERGMs capacity to tackle with even complex dependence structures among ties (see realization dependence assumptions by Pattison and Robbins 2002).

An ERGM explains the formation of a network by means of local configurations, which are some small and regular patterns. Among all possible network configurations it gives a higher probability to those that are similar to the observed network in terms of these small structures. In defining these local configurations, ERGM is capable of differentiating ties and nodes with attributes (Robins et al. 2007).

Therein, the capacity of ERGMs to investigate the role of geography is three-folds. First, it may be studied as the role of physical distance by means of a distance interaction function (Daraganova et al. 2012). Second, geography may be included as a node attribute. Third, geography can be considered as a spatial setting that imposes limits on tie dependence, hence on local configurations (Pattison and Robbins 2002). The studies by Broekel and Hartog in Chap. 4 and Hazir in Chap. 13 illustrate the applications of these models on R&D networks.

Preferential Attachment Model. Preferential Attachment Model (Barabási and Albert 1999) is a graph theoretical model explaining dynamic networks with growing number of nodes. The model in its original form explains the formation of a network as a process where the degrees of existing nodes increase proportional to their magnitude and result in a scale-free degree distribution. Hence, it considers a single factor; i.e. degree affinity of agents, to explain for the network via explaining one of its macro properties; i.e. its degree distribution. The extension by Vinciguerra et al. (2010) integrates the effect of geographical distance and co-location in the same country to the probability that a node receives connections as the network grows.

Complex Network Analysis (CNA). While the models reviewed so far aim at explaining the formation of a network by means of micro processes, Complex Network Analysis focuses on the overall topological structure of complex networks. Hence CNA aims at identifying and explaining key global features like degree distribution, diameter, clustering, and communities.

A number of studies revealed that R&D networks display a scale-free degree distribution, "small-world" property in terms of diameter and high "clustering" (Goyal et al. 2006; Newman 2001; Gay and Dousset 2005). On the one hand, theoretical models (Johnson and Gilles 2000; Carayol and Roux 2007), the above mentioned spatially extended preferential attachment model, and possibly ERGMs illustrate how a model explaining the effect of geography on micro processes can also explain for these global properties. On the other hand, CNA adopts a macro perspective to study the spatial dimension of such properties. The study by Barber and Scherngell in Chap. 10 illustrates studying the heterogeneity in the spatial configuration of communities in an R&D network. Whereas, the study by De Montis et al. in Chap. 3 illustrates the use of CNA to investigate whether similar geographical contexts give rise to similar global network properties or not.

2.4 Conclusions and Future Directions

In this chapter, we considered the relationship between geography and network formation but our focus was on how to investigate this relationship. Hence, we reviewed different meanings of geography and different conceptualizations of this relationship. Then we provided an overview on different approaches through which network formation is explained. Our aim was neither to provide a complete list or a hierarchy of network formation models nor to identify best models. Rather we were interested in two aspects. First, leaving all the practical issues and formal definitions of models aside, we aimed at identifying the grand avenues that a researcher can follow in studying network formation. We identified that whether to consider it as an outcome of choice or chance; whether to consider it as a dynamic or a static process; whether to explain it from bottom-up or top-down; whether to study its complex interdependencies or simplify it are the major decisions to be made by the researcher in making a model choice. Second, all these choices suggest a different capacity to study the role of geography. Hence, we reviewed applied studies with a particular interest on those on R&D networks to highlight these analytical differences, the evolution of analytical frameworks (if any) and to identify future directions.

One of the main conclusions that could be derived from this review is that so far the research community made use of mainly bottom-up approaches to study the role of geography in formation of R&D networks. In other words, the emphasis is given to explain how geography affects the formation processes at the micro level. Although global topological features of these networks have attracted attention, spatial heterogeneities in these global features or heterogeneities in spatial patterns of components of networks have received less attention. Apart from those techniques used by Barber and Scherngell in Chap. 10 and by De Montis et al. in Chap. 3; block modelling might also be used to study the relationship between geography and network components, members of which are equivalent in terms of their connection patterns.

Another conclusion could be derived on the evolution in the analytical processes that are adopted to study the role of geography in formation of R&D networks. While there is not a clear cut distinction, it is noteworthy that the community has recently shown interest in models that can allow dependence among ties. This enables demarcating the effect of dependence from other factors of interest; hence improves estimates for the role of geography. Furthermore, the ability of these models to handle tie dependence results in models which can explain the global topology of the network (such as clustering, degree distribution, etc.) as well as the local process in focus. However, the relationship between geography and tie dependence is far from being exploited. So far the effect of distance on dyads have received the major attention, leaving how distance or location might affect formation of more complex structures than dyads aside.

A third conclusion stems from the temporal dimension of networks. As a matter of fact most applied studies consider an R&D network as a static object, where neither new nodes are added nor ties created or dissolved. Applications of stochastic actor-based models relaxed this assumption and considered the tie dynamics among a fixed set of nodes. These models indeed possess a capacity to analyse not only the determinants of tie formation but also tie dissolution by means of an endowment function (Snijders et al. 2010). Hence, these models may well be used to study the role of geography on tie dissolution. Although not applied to study the geographical dimension of R&D networks some recent temporal extensions of ERGM also suggest similar possibilities. Among these Hanneke et al. (2010) provides Temporal Exponential Random Graph Model (TERGM), which allows studying the evolution of a network of fixed size. Whereas, Krivitsky and Handcock (2010) enables separating tie formation and dissolution processes in a TERGM.

Apart from these some other model extensions suggests additional explanatory capacity for the field. Among these, extension of ERGMs for valued networks (Krivitsky 2012) stands as another tool to study the effect of geography on the intensity of connections, which has been studied so far by means of Poisson regression models and gravity models. The ability of this tool to handle tie dependence might be useful for better treatment of network effects and demarcate the role of geography more properly. In addition to that, Steglich et al. (2010) extended Stochastic Actor-based Models to distinguish partner selection from social influence in a dynamic network. This extension basis on the idea that two actors showing the same behaviour might be collaborating due to similarity in their behaviour, or one gets similar to the other as a result of being connected. The ability to separate those two processes might be valuable in better demarcation of spatial effects from that of influence over time.

In addition, by improving our understanding of network formation and evolution, all these developing techniques may also contribute to a better comprehension of the mechanisms that generate network outcomes. A growing literature tries to understand how some of the particular topological network properties (such as density, clustering, connectivity of the network, degree distribution of nodes or degree assortativity) influence economic performances at the regional level (Breschi and Lenzi 2011; Crespo et al. 2013). However, as argued by Ahuja, Soda and Zaheer (2012), “without a comprehension of the logic that drives network creation, scholarly understanding of their outcomes remains incomplete” (p. 34). In particular, as it is difficult to identify whether the network structure implies the outcome or the reverse, we have to consider both aspects together. To this respect, the contributions of spatial econometrics to the field of network analysis may extend beyond gravity models as suggested by (Autant-Bernard 2012). Spatial tools can indeed provide valid instruments allowing endogenous effects to be separated from exogenous ones (see for instance Bramoullé and Fortin 2009). In the same line, the temporal extensions of the above reviewed network approaches are also very promising in order to cope with this causality problem.

Finally, it is a matter of fact that model choices are strongly constrained by the nature of data and data availability. Assumptions of a model might be severe or reasonable depending on the nature of the data and on the properties of the economic process through which it is generated. Hence, there is no one-for-all answer on how to study the spatial dimension of network formation.

Acknowledgements This research has received funding from the European Community's Seventh Framework Programme under grant agreement no. 266834.

References

Ahuja G, Soda G, Zaheer A (2012) The genesis and dynamics of organizational networks. Organ Sci 23:434–448, http://didattica.unibocconi.it/mypage/upload/

Autant-Bernard C (2012) Spatial econometrics of innovation: recent contributions and research perspectives. Spat Econ Anal 7(4):405–421

Autant-Bernard C et al (2007) Social distance versus spatial distance in R&D cooperation: empirical evidence from European collaboration choices in micro and nanotechnologies. Pap Reg Sci 86(3):495–519

Balland PA (2012) Proximity and the evolution of collaboration networks: evidence from research and development projects within the Global Navigation Satellite System (GNSS) industry. Reg Stud 46(6):741–756

Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286 (5439):509–512

Boschma RA (2005) Proximity and innovation: a critical assessment. Reg Stud 39(1):61–74

Bramoullé Y, Fortin B (2009) The econometrics of social networks, Centre Interuniversitaire sur le Risque, les Politiques Économiques et l'Emploi working papers 09-13. http://www.cirpee.org/fileadmin/documents/Cahiers_2009/CIRPEE09-13.pdf

Breschi S, Lenzi C (2011) Net and the city. Co-invention networks and the inventive productivity of US cities, Mimeo. http://cep.lse.ac.uk/seminarpapers/09-12-11-SB.pdf

Carayol N, Roux P (2007) The strategic formation of inter-individual collaboration networks: emprical evidence from co-invention patterns. Annales d'Economie et de Statistiques 87–88:75–301

Cranmer SJ, Desmarais BA (2011) Inferential network analysis with exponential random graph models. Politi Anal 19(1):66–86

Crespo J, Suire R, Vicente J (2013) Lock-in or lock-out? How structural properties of knowledge networks affect regional resilience. J Econ Geogr, (forthcoming). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2034901

Daraganova G et al (2012) Networks and geography: modelling community network structures as the outcome of both spatial and network processes. Soc Netw 34(1):6–17

Erdös P, Renyi A (1959) On random graphs. I. Publicationes Mathematicae Debrecen 6:290–297

Feldman MP (1993) An examination of the geography of innovation. Ind Corp Change 2(3):451–470

Feldman MP, Florida R (1994) The geographic sources of innovation: technological infrastructure and product innovation in the United States. Ann Assoc Am Geogr 84(2):210–229
Frachisse D (2010) L'analyse des réseaux sociaux appliquée aux PCRDT: Enjeux méthodologiques et outils pour l'évaluation. Dissertation, CREUSET, Université de Saint-Etienne
Frank O, Strauss D (1986) Markov graphs. J Am Stat Assoc 81(395):832–842
Gay B, Dousset B (2005) Innovation and network structural dynamics: study of the alliance network of a major sector of the biotechnology industry. Res Policy 34:1457–1475
Geuna A (1998) Determinants of university participation in EU-funded R&D cooperative projects. Res Policy 26:677–687
Goyal S, van der Leij MJ, Moraga-González JL (2006) Economics: an emerging small world. J Politi Econ 114(2):403–412
Granovetter M (1985) Economic action and social structure: the problem of embeddedness. Am J Sociol 91(3):481–510
Hanneke S, Fu W, Xing EP (2010) Discrete temporal models of social network. Electron J Stat 4:585–605
Hoekman J, Frenken K, van Oort F (2009) The geography of collaborative knowledge production in Europe. Ann Reg Sci 43(3):721–738
Hoekman J, Frenken K, Tijsen RJW (2010) Research collaboration at a distance: changing spatial patterns of scientific collaboration within Europe. Res Policy 39:662–673
Jackson MO, Wolinsky A (1996) A strategic model of social and economic networks. J Econ Theory 71:44–74
Johnson C, Gilles R (2000) Spatial social networks. Rev Econ Des 5:273–299
Krivitsky PN (2012) Exponential-family random graph models for valued networks. Electron J Stat 6:1100–1128
Krivitsky PN, Handcock MS (2010) A separable model for dynamic networks. arXiv:1011.1937v1
Maggioni MA, Nosvelli M, Uberti TE (2007) Space vs. Networks in the geography of innovation: a European analysis. Pap Reg Sci 86:471–493
Mairesse J, Turner L (2005) Measurement and explanation of the intensity of co-publication in scientific research: an analysis at the laboratory level. NBER working paper 11172
Massard N, Mehier C (2009) Proximity and innovation through an "Accessibility to Knowledge" Lens. Reg Stud 43(1):77–88
Newman MEJ (2001) Clustering and preferential attachment in growing networks. Phys Rev E 64(2):025102
Nowicki K, Snijders T (2001) Estimation and prediction for stochastic blockstructures. J Am Stat Assoc 96:1077–1087
Paier M, Scherngell T (2008) Determinants of collaboration in European R&D networks: empirical evidence from a binary choice model perspective. NEMO working paper #10
Park J, Newman MEJ (2004) The statistical mechanics of networks. Phys Rev E 70(6):066117
Pattison P, Robins G (2002) Neighbourhood-based models for social networks. Sociol Methodol 32:301–337
Polanyi M (1966) The tacit dimension. Doubleday, Garden City/New York
Ponds R, van Oort F, Frenken K (2007) The geographical and institutional proximity of research collaboration. Pap Reg Sci 86(3):423–443
Powell WW, Koput KW, Smith-Doerr L (1996) Inter-organizational collaboration and the locus of innovation: networks of learning in biotechnology. Adm Sci Q 41(1):116–145
Powell WW et al (2005) Network dynamics and field evolution: the growth of inter-organizational collaboration in the life sciences. Am J Sociol 110(4):1132–1205
Robins G et al (2007) An introduction to exponential random graph (p*) models for social networks. Soc Netw 29:173–191
Scherngell T, Barber M (2009) Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU Framework Programme. Pap Reg Sci 88(3):531–546

Scherngell T, Lata R (2011) Towards an integrated European Research Area? Findings from Eigenvector spatially filtered spatial interaction models using European Framework Programme data, Papers in regional science. doi:10.1111/j.1435-5957.2012.00419.x
Snijders T et al (2010) Introduction to stochastic actor-based models for network dynamics. Soc Netw 3:44–60
Steglich C et al (2010) Dynamic networks and behaviour: separating selection from influence. Sociol Methodol 40(1):329–393
Ter Wal ALJ (2013) The dynamics of the inventor network in German biotechnology: geographical proximity versus triadic closure. J Econ Geogr. doi:10.1093/jeg/lbs063
Vinciguerra S, Frenken K, Valente M (2010) The geography of internet infrastructure: an evolutionary simulation approach based on preferential attachment. Papers in evolutionary economic geography 10.06, Utrecht University Urban and Regional Research Centre
Wasserman S, Pattison P (1996) Logit models and logistic regression for social networks: I. An introduction to Markov graphs and p*. Psychometrika 61(3):401–425

Chapter 3
Recent Developments of Complex Network Analysis in Spatial Planning

Andrea De Montis, Simone Caschili, and Alessandro Chessa

Abstract In the last years, we acknowledge a great scientific interest on complex network analysis, a method able to characterise systems with very large numbers of entities (the nodes or vertices) interlaced by a series of connections/relationships (the links or edges). The objects of analyses as such are biological (predator-pray); information (internet); social (actor-in the same movie); transportation (railway and road networks) systems. While in general a network is an abstract (topo) logical object, spatial networks belong to an important class of systems that includes nodes and edges with a clear reference to space. Recently the interest of scientists has focused on methods able to define and investigate on communities emerging from the structure of a network. In this respect the spatial factor can emerge both as the result of the topological community structure that maps back onto geography in the form of sensible spatial regions, or just as spatial clusterisation of nodes in principle embedded in space. In this essay, the authors aim at presenting a state of the art summary of the last advances in the field of network community detection methodologies with a detailed view to the case of spatial networks. Secondly, the paper will report on a case study concerning a major issue for policy makers and planners: the delimitation of sub-regional domains showing a sufficient level of homogeneity with respect to some specific territorial features. We compare some intermediate body partitions of the island of Sardinia (Italy) with the patterns of the communities

A. De Montis (✉)
Dipartimento di Agraria, Università degli Studi di Sassari, Viale Italia, 39, 07100 Sassari, Italy
e-mail: andreadm@uniss.it

S. Caschili
UCL QASER Lab and Centre for Advanced Spatial Analysis, University College London, Gower Street, London WC1E 6BT, UK
e-mail: s.caschili@ucl.ac.uk

A. Chessa
IMT Lucca Institute for Advanced Studies, Lucca 55100, Italy

Linkalab, Complex Systems Computational Laboratory, 09129 Cagliari, Italy

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_3,

of workers and students, by applying grouping methodologies based on the characterisation of the Sardinian commuters' system as a complex weighted network.

3.1 Introduction

Over the last 15 years there has been a great scientific interest in complex network analysis, a method able to characterize systems with a very large number of entities (the nodes or vertices) interlinked by a series of connections/relationships (the links or edges). The objects of this analysis have regarded biological systems (predator-pray); information systems (Internet); sociological systems; and transportation systems (railway and road networks). While, in general, a network is an abstract topological object, spatial networks belong to an important class of systems that includes nodes and edges with a clear reference to space. Recently the interest of scientists has focussed on methods able to define and investigate communities emerging from the structure of a network. In this respect the spatial factor can emerge both as the result of the topological community structure that maps back onto geography in the form of sensible spatial regions, or just as spatial clustering of nodes in principle embedded in space.

In this essay, the authors aim to present a state of the art summary of the last advances in the field of network analysis and network community detection methodologies focusing on spatial networks. We will review a case study concerning a major issue for policy makers and planners: the delimitation of sub-regional domains showing a sufficient level of homogeneity with respect to some specific territorial features. We compare some intermediate administrative bodies of the island of Sardinia (Italy) with the patterns of the communities of workers and students, by applying grouping methodologies based on the characterization of the Sardinian commuter system as a complex weighted network.

This essay unfolds as follows. In the next section, we develop a brief state of the art summary on social networks with a focus on Research and Development (R&D) networks. At the end of this section, we introduce the reader to the main concept of the essay, i.e. spatial networks displaying a clear geographical reference. In the third section, we review the recent advancements in the field of complex network analysis as well as its adoption in geography, spatial and regional planning. In the fourth section, we report on the latest advances regarding community detection methodologies able to cluster nodes into homogeneous groups. The fifth section presents a case study about the application of a network community detection approach to study the problem of regionalisation. Commuter basins in the island of Sardinia (Italy) are used to scrutinise the relevance of administrative subdivisions at the provincial level.

3.2 Complex, Social and R&D Networks

Complex network analysis (CNA) consists of a set of methods and tools, grounded in graph theory, that enable scientists to model systems as networks. This approach emphasizes the role of agents (i.e. the nodes or vertices) and the relations between them (i.e. the edges or links). More and more, CNA has become the go to tool for a number of scholars in many disciplines, ranging from macro infrastructures such as road systems and gas pipeline frameworks, to the Internet, the world wide web and micro ensembles, such as: genomic protein-amino acid and DNA chains (for a review, see Albert and Barabàsi 2002; Newman 2003).

Social science has similarly applied graph theory to the study of many issues in sociology, business administration, industrial management, anthropology and psychology. In these works, nodes signify individuals and edges signify patterns of acquaintances between them. One of the cornerstone findings in social science, the 'six degrees of separation' experiment (i.e. small world phenomena) was carried out by the psychologist Stanley Milgram (1967) using graph theory. After Milgram's work, many authors further investigated collaborative social systems (for a review of methods and applications, see Wasserman and Faust 1994). In these studies scientists, engineers, or inventors are modelled as vertices and the links are collaborative ties between them. Inter alia, Bloch (2005) scrutinised individual behaviours of agents and collective dynamics of wide organizations.

In the remainder of this section, we discuss five research articles that have demonstrated that individual behaviour in productive domains is clearly affected by relational roles played by each agent both directly, on their local neighbour, or indirectly, on the global network.

Hanaki et al. (2010) studied spillover effects arising from R&D collaborations in the U.S. Information and Technology industry. Starting from the analysis of patents granted from 1985 to 1995, they investigated the dynamics of the inter-firm ensembles through a topological network representation. Firms are modelled as nodes and edges represent collaboration ties (two firms were connected if they had at least one inventor in common). Hanaki et al. (2010) demonstrated that the U.S. IT R&D network belongs to the class of "small world" networks (Watts and Strogatz 1998). In the U.S. IT R&D network, the number of collaborations has increased over the past few years generating a denser and more interconnected system. Nodes display behaviours similar to the preferential attachment rule (Barabàsi and Albert 1999). In the case of R&D networks, the more connected nodes have patterns of collaboration choices that are affected by closure and preferential attachment (Barabàsi and Albert 1999).

Jin et al. (2011) referred to R&D networks of scientific collaborations. They scrutinised research on bio-, and nano-technology from the R&D national data of South Korea. Jin et al. detected and characterized nine communities of scientists applying a divisive method introduced by Newman and Girvan (2004) for network grouping and generalized for weighted networks. CNA showed that this R&D network exhibits properties typical of scale free networks, similar to the network of citations between scientific papers (Derek de Solla Price 1965). The authors presented an interesting description of the relationships between scientists working in different fields and between their clusters. CNA approach allowed them to indicate the most prominent members in each cluster and the most promising sectors of R&D, which are worth interest and funding.

König et al. (2012) studied networks of firms focussing on costly R&D collaborations. In this study the authors described the inventive activities of firms belonging to technology intensive industries. They demonstrated that stability and efficiency is clearly dependent upon the cost of R&D collaborations and the topology of the network. The authors also argued that "the complete graph is stable in small industries and for low collaboration costs, while the class of size-homogeneous disconnected cliques and the star are stable in large industries" (König et al. 2012, p. 707).

Smith-Doerr et al. (2004) scrutinised the social network of project managers belonging to the R&D laboratory of a Fortune 500 company and leading six projects. They developed a CNA to study the centrality of each manager under four points of view: instrumental, expressive, technical advice, and organizational advice. In particular, the authors calculated the in-degree centrality (Freeman 1979) of all the 42 members of the laboratory and discovered that project leaders' average centrality is by far higher than the corresponding figure of all lab members. A relevant result of this work is that network centrality matters. The project leader, who has a high centrality in almost all networks, is the only manager able "to look at the big picture and generally reflect on how to think about R&D project success or failure" (Smith-Doerr et al. 2004, p. 74).

We conclude this section focusing on spatial networks which are at the core analysis of this manuscript. In the field of R&D networks, scholars have often ignored the contribution of space and geography to this topic. Oerlemans and Meeus (2005) investigated inter-organizational networks and the effects generated by spatial proximity on firm performance. They argued that innovation agreements with intra-regional firms matters but in a specific way. Firms that use intra- and interregional agreements tend to outperform other firms in the same sector. But firms that only depend on intra-regional or on interregional innovation ties do not perform better than other firms in their sector. A combination of intra- and interregional innovative ties are essential for the commercial success of a firm (Sternberg and Arndt 2001; Oerlemans and Meeus 2005). The influence of space on Research & Development collaboration has been studied by Chessa et al. (2013). They took into consideration the evolution of geographical collaboration networks under the European Research Area (ERA) framework. They scrutinized the network generated by patent and scientific publication data by applying network community detection methods (we will discuss this methodology in Sect. 3.4). Results show that since 2003 the level of collaboration within and outside European countries is stable which has resulted in poor research collaboration among European countries.

Under this context, we are interested in inspecting the influence of space on networks. In the next section we review the latest advances in the field of CNA with a focus on geography and regional planning.

3.3 Complex Networks Modelling and Spatial Planning

Two hundred years ago only 5 % of the world's population lived in cities, today more than 50 % lives in urban areas. This trend is likely to increase as scientists have forecasted that more than 80 % of the population will live in cities by the end of the twenty-first century.[1] Apart from the modification of our economic activities (200 years ago people lived in rural areas and their main activities were generation of food), living in compact settings has intertwined the daily activities of people with "soft" and "hard" infrastructures. Cities are composed of interlinked systems that can be conceptualised under the lenses of networks modelling. People move for work or leisure using transport systems (trains, buses, roads, flights etc.), communicate and exchange information through land lines and digital networks (i.e. the Internet); our lives are powered by electricity, sustained by utility systems and kept safer through CTV camera networks. Financial systems, education systems, health care systems, systems of government, as well as emergency services (i.e. "soft" infrastructures) all contribute to maintain economic, health, cultural and social activities in a territory. Because of their interactive nature, all these systems can be seen as networks that are part of our daily life. Thus, we are surrounded and immerged in networks that have intrinsic spatial features (Barthélemy 2011). Within the field of spatial planning, several authors have studied spatial networks with different aims such as scrutinising the network centrality of streets in a city and the correlation with economic activities (Porta et al. 2010); the disease contagion through human mobility networks (Bajardi et al. 2011); impacts, accessibility and network patterns generated by movements of commuters among regional units (De Montis et al. 2007, 2011; Caschili and De Montis 2013); urban transport networks – i.e. bus, subway (von Ferber et al. 2009; Kurant and Thiran 2006; Latora and Marchiori 2001); the structure and vulnerability of power grids (Crucitti et al. 2004; Albert et al. 2004) and water distribution networks (Yazdani and Jeffrey 2010).

The popularity of network modelling and analysis results from three factors: (i) availability of large real-world data sets (also geographically referred), (ii) accessibility of cheap high computational resources and (iii) opportunity, also for non-computer scientists and mathematicians, to scrutinise large non-linear systems. Within the field of spatial and regional planning, scientists and practitioners have applied complex network analysis with two approaches: the first derives from the statistical mechanics field and aims to explain observed

[1] Source: http://web.unfpa.org/swp/2007/english/introduction.html

hierarchical structure of settlements and to advance urban models using stochastic approaches (Andersson et al. 2006; Porta et al. 2006a, b). A second branch of studies adds to urban planning from a social network perspective. A few scholars have also studied the interdependencies and influences among diverse actors in planning processes (Booher and Innes 2002; Innes and Booher 2010).

From a modelling view point, classic planning approaches have long revolved around the understanding of (i) patterns and rationale of economic activities distributed in urban regions, (ii) forces that influence spatial configuration, and (iii) urban structures and functions. Monocentric (Muth 1969; Mills 1972), polycentric (Heikkila et al. 1989; Wang, 2000) and dispersed models (Lang 2003) have been used to describe the hierarchical organisation of urban settings. The relationship between land use, economic activities and mobility are conceptualised through rank-size rules (Zipf 1949), gravitational models (Putman 1983; Anderstig and Mattsson 1991; Martinez 1996), spatial interaction models (Wilson 2000) and discrete choice models (McFadden 1974).

In the remainder of this section we discuss the contribution of the two classes of studies under the research framework of complex network analysis that have been used in spatial and urban planning. Urban morphology has been at the core of these studies. Seminal works that scrutinised urban morphology with a network approach, date back to the 1960s with the work of Nystuen and Dacey (1961). They used networks of commuters, goods and communications to quantify the degree of association between cities. Kansky (1963) proposed a number of measures based on graph theory to characterise transportation networks. A decade later, Space Syntax methodology (Hillier et al. 1976; Hillier and Hanson 1984; Hillier 1996) introduced a pioneering approach to measure the relation between different components of urban structure using planar graphs. The novelty of this approach consisted in measuring the cognitive complexity of a spatial graph through non-local network measures. Space Syntax has multiple applications in a variety of fields such as architecture, planning, transport and interior design (Hillier 1996). Advancing the research framework of Space Syntax, Porta et al. (2006a, b, 2010) introduced the concept and methodology of Multiple Centrality Assessment (MCA). The aim of this new methodology was to include metric measures to understand street networks and to use the spatial geographical representation of a network instead of dual representations. In fact, Space Syntax scrutinises dual graphs of real networks: axes are turned into nodes and intersections into links, thus losing the geographic content of a network (Porta et al. 2010). It is interesting to note that while urban growth has been investigated with a number of methods, such as agent-based (Benenson 1998), spatial statistic modelling (Lopez et al. 2001; Wu and Yeh 1997), neural networks (Pijanowskia et al. 2002) and fractal based modelling (Batty and Longley 1994; Makse et al. 1998), the contribution of network analysis to this topic is still scant. Focusing on microscopic mechanisms of urban growth that generate macroscopic structures, Barthélemy and Flammini (2008) proposed a network model which combines an optimisation process with pattern formation. The main assumption of this model is that road networks evolve converging into mass centres in an efficient and economic way. Andersson et al. (2006) used complex networks combined with a cellular automata and verified that their model is consistent with large-scale regularities such as power laws and fractality. With this approach they have been able to verify that hierarchical urban structures can be explained as ‘a stationary property of a stochastic evolutionary process rather than as equilibrium points in a dynamic process’ (Andersson et al. 2006). Finally network analysis has also been used to address the so-called ‘problem of regionalisation’, i.e. to group local administrative units in upper level clusters. In Sect. 3.5 we present a case study (De Montis et al. 2013) which has applied network community detection as a tool to investigate the problem of regionalisation.

The activities of urban planners are not only focused on modelling land uses, population growth, mobility, and social and economic activities, but also around regulations, stakeholder engagement and more in general public engagement. In this direction, the contribution of social network analysts have involved collaborative activities across public authorities (Scholz et al. 2008), public participation in planning processes (Holman 2008; Davies 2002) and social community detection (Wellman 2001). According to Dempwolf and Lyles (2011) three ‘broad planning issues’ can be addressed with the tool of social network analysis. First, it is important for a planner to understand the dimension and composition of a community which, after all, is the beneficiary of planning activities. Thus planners should take into consideration not only the various categories (young, adult, elderly, employed, unemployed etc.) that compose a population but also the links between them which generate the complex phenomena that we observe in a territory. A second issue regards public participation. The use of network maps enhances capital interaction among the ‘actors’ of a planning process and allows planners to pay attention to their position in the network. Finally, a third issue regards the creation of spatial and social dimension which generates innovation. Eraydin et al. (2008) show that social networks among governmental and nongovernmental actors instil a positive economic effect in a territory.

We conclude this section with some final remarks. Despite its importance as a suitable tool for analysis in planning, complex network analysis seems scarcely applied to territorial planning and processes. Much has still to be done for this technique to be fully integrated in the tools used by planners.

3.4 Community Detection in Networks

In network analysis, starting from the network topological structure, it is possible to extract various types of information. Beyond the well known centrality measures, a way to characterize the internal network organization is to look at the cluster formation among the node components, i.e. group of nodes that are well connected among themselves with few outgoing links toward the other groups. Under the Complex Networks Theory field, the task of finding these clusters goes under the name of Community Detection (see Fortunato 2010, for a review). For example, in the World Wide Web communities correspond to websites pertaining to related subjects (Flake et al. 2002); in social networks as clusters of individuals connected by similar activities (Girvan and Newman 2002; Lusseau and Newman 2004), while in metabolic networks communities behave as functional modules (Guimerà and Amaral 2005; Palla et al. 2005), and compartments in food webs (Pimm 1979; Krause et al. 2003).

Generally speaking, we can define three main categories: local, global, and based on vertex similarities. In local definitions, the local connectivity of nodes is inspected, disregarding the rest of the graph. In global definitions, the graph is analyzed as a whole and the communities are regarded as structural units of the graph. Definitions based on node similarity select communities' membership whenever nodes are similar each other, according to a quantitative/qualitative criterion. In general, community detection aims to identify communities through an analysis of the topology of a graph. New advances also propose to extend the detection of communities in weighted networks, where not only the topology shapes the cluster structure but also the weight of each link.

Indeed, community detection may become a complex activity, if we consider systems with a large number of nodes and links. Communities tend to overlap each other showing some nodes in common throughout the network (Palla et al. 2005; Fortunato 2010). Another case is that of large networks for which nodes have various levels of organization. Communities can have hidden internal cluster organization, i.e. a community may include recursively other smaller communities. In this case the community structure is characterized by a hierarchical structure (Sales-Pardo et al. 2007).

In the literature, we can find three main classes of methods: divisive algorithms, optimization methods, and spectral methods. Alternative approaches that do not fit in the above classification are the following: clique percolation, random walk, maximum likelihood, Q-state Potts model, Markov cluster algorithm, and L-shell method (see Fortunato 2010).

In the study of the regionalization processes we envisage an interest of analysts and planners for network based community detection methods. These tools are able to detect patterns starting from the analysis of similarities among the basic elements under investigation (the nodes) intertwined in a known topology. This goal can be achieved, because network community detection methods cover additional information, in comparison to traditional clustering methods adopted for identifying sub-regions.

There are various community detection methods and algorithms; one of the most important, adopted in many applications, is the modularity optimization introduced by Newman and Girvan (Newman and Girvan 2004). This method has been widely adopted, because it has a very straightforward implementation. However, it is generally extremely difficult to find the best network partition. It has been found that the optimization process is an NP-complete problem (Brandes et al. 2006). In this case, it is probably impossible to find the solution in a time growing polynomially with the size of the graph. In this respect the best approach is to use a heuristic procedure able to approximate the solution. Moreover, the methods based on modularity optimization have a drawback related to the existence of a resolution limit (Fortunato and Barthelemy 2006), which prevents it from detecting smaller modules. The modularity function is defined as follows:

$$Q_w = \frac{1}{2W} \cdot \sum_{ij} \left(w_{ij} - \frac{s_i s_j}{2W} \right) \cdot \delta\left(c_i, c_j\right)$$

where w_{ij} is the weight associated to the edge connecting the node i and the node j, $s_i = \sum w_{ij}$ (node strength) is the sum of the weights of the edges attached to the node i, $W = \frac{1}{2} \sum w_{ij}$ is the sum of all the edge weights, and $\delta(c_i, c_j)$ is a function equal to one, when vertices i and j belong to the same community, and to zero otherwise.

The modularity function quantifies the goodness of a network subdivision among all possible ones, by computing, for a particular subdivision, how many edges are inside the communities, with respect to the random case. The maximum value attainable is 1 (an ideal case for which the clusters are perfectly isolated) and can take also negative values. The 0 value corresponds to a single partition that will coincide with the whole graph. A negative value means that the communities will typically have few internal edges and many edges lying between them and so there is no community structure whatsoever.

Once the optimization function has been defined, we need an efficient method to maximize it. One of the most successful algorithms is the so called 'Louvain algorithm' as proposed by Blondel et al. (2008).

The Louvain algorithm is quite interesting, since it allows one to successfully approach two critical issues of optimization methods: detecting communities in large networks in a short time and taking into account hierarchical community structure. The number of communities at each hierarchical level emerges naturally from the algorithm and has not to be imposed at the beginning, as in other clustering approaches. Moreover, this bottom up approach can possibly help in preventing the resolution limit problem found by Fortunato and Barthelemy (2006). This algorithm may be used for both weighted and un-weighted networks.

The modularity is extremely useful in regional studies since, as we will see in the following sections, it is able to reconstruct territorial clusters starting just from the topological features of the network. Even if the nodes are not explicitly embedded in space, when it comes to exploiting the aggregation features of the network, the space emerges in the shape of sensible spatial domains. There are cases for which it could be of interest to take explicitly into account the presence of space, and cancel it in order to discover hidden interactions beyond the spatial correlations. To this end, new modularity definitions have been recently introduced that include the spatial factor (Expert et al. 2011; Cerina et al. 2012).

In the general case, valid for an un-weighted network, one usually chooses $P_{ij} = k_i k_j / 2m$, which allows one to take as a null model a random network with the same degree sequence as the original network. In order to introduce spatial features, the idea is to change the null model defined by P_{ij} and to compare the actual network with this null model. Recently, such a proposal was made by Expert et al. (2011), who obtained the quantity P_{ij} directly from the data describing the network. More precisely, Expert et al. adopted the following form:

$$P_{ij}^{Data} = N_i N_j f(d_{ij})$$

where N_i is related to the importance of the node i (such as the population, for example) and f is the probability to have two nodes i and j connected at the distance d_{ij}. This form recalls the gravitational model for traffic flows, where flows are directly proportional to the product of populations and, inversely, to the distance. In this specific case, extracting the node spatial dependencies from the real link distribution present in the network data is the most effective way to subtract the spatial component. Otherwise if there are any correlations between space and node attributes, the data contains in an unknown proportion information on both space and attribute and the method needs to be reformulated. One possible way to overcome this problem is to explicitly determine a spatial dependency of the link distribution and to put it as an independent factor in the optimization function definition. In order to be able to deal with the correlated case and to remove spatial effect only, one can introduce the following explicit function of space for P_{ij}

$$P_{ij}^{Spatial} = \frac{1}{Z} k_i k_j g(d_{ij})$$

where Z is the normalization constant, k_i the degree of the node i, d_{ij} the Euclidean distance between node i and node j. The function $g(d)$ decreases with distance and its role is to remove the spatial effect. A simple form of $g(d)$ is chosen as follows

$$g(d) = e^{-d/<l>}$$

where $<l>$ is the average Euclidean distance between nodes in the network. Of course, $<l>$ is a rough approximation of the typical community size, but it is enough to capture the essence of the spatial signature of the network. In the next section we present a case study for the application of network community detection methodology in the field of regional planning.

3.5 Community Detection in Spatial Social Systems: Regionalisation and Commuter Networks

Planning urban settlements is considered a complex process because it concerns several intertwined issues (Hinloopen et al. 1983). The complexity of urban and territorial phenomena makes this task even harder. Planners and scholars now have access to new tools derived from complexity science. Among various techniques and tools, Complex Networks paradigm, and network community detection methods provide valuable instruments to solve classical problems of regionalization, i.e. the assessment of appropriate territorial units (Duque et al. 2007).

In this paragraph we review the case study of delimitation of provinces in the Region of Sardinia as proposed by De Montis et al. (2013) using network community detection techniques applied to commuter networks. Although we show that community detection methods are useful tools to plan homogenous territorial units, it is worth noting that planners need to check and combine models' results with political goals (Palermo 1980).

3.5.1 Devolution and Regionalisation in the Italian Provinces

Modern states are constantly in search of optimal internal administrative configurations in order to optimise resources and provide more efficient services to their citizens. Devolution[2] is the concept at the base of this process which has also been achieved through in-between administrative sub units. In the European Union there are four levels of in-between units: regional (NUTS 2 units) and local bodies (NUTS 3, LAU 1 and LAU 2 units).[3] Historically in Europe those in-between districts have different names and carry out different tasks: in France "Le departement" dates back to the Napoleonic age, counties are part of the Anglo-Saxon tradition, "Regierungsbezirk" in Germany, "Provincia" in Italy, "Disputaciones" in Spain. Those sub divisions are identified according to both normative and analytical regulations. In fact sub administrative units have similar spatial and demographic features; for instance in Italy a province is identified as an administrative body with a population from 100 to 500,000 citizens which live at a like distance around a big town. In this paper we focus on the Italian administrative hierarchical organization. The case study that we discuss is based on the application of a network community detection method for the recognition of productive, social and administrative territorial units in Sardinia (De Montis et al. 2013). As of 2012, Italy is divided into 20 regions (Regione in Italian) that are further divided into 110 provinces (Provincia) and 8,100 municipalities (Comune).

Nevertheless while regions and municipalities kept a strong configuration since they have been founded, provinces with changing fortunes and cyclical successes assumed different roles and strategies in the Italian territorial organization. During the 60s, 70s and 80s, the institution of new in-between bodies similar to provinces but smaller such as "comprensori", "comunità montane" (mountain community), "unità sanitarie" (health districts), "distretti scolastici" (school districts), raised the discussion about which body could better represent and meet the demands of local

[2] Devolution is the process used by a central state to grant power to sub national administrative levels such as regions, counties, provinces etc.

[3] NUTS and LAU are two classifications introduced by EUROSTAT for dividing up EU territory.

communities. “Provincia” is an administrative body with functions of (i) economic-financial planning, (ii) territorial planning, and (iii) promotion and coordination of projects among groups of municipalities. A recent adjustment of the legal institute of Italian provinces was published in 2000 (“Testo Unico degli Enti locali”). With this last act, the central government assigned the following duties to provinces:

- Equality;
- Autonomy;
- Relevance of constitutional governmental power;
- Subsidiarity;
- Sustainability;
- Self-sufficiency.

Thus Provincial bodies have been granted power for administrative functions that regard spatial regulations. Those functions pertain to: land protection, land enhancement, prevention of natural disasters, management of water resources, energy development, enhancement of cultural heritage, transport infrastructures, protection of fauna and flora, hunting and fishing, waste management and school building. This shift in power from the central government gave to local communities, which are more knowledgeable about citizens’ needs, more independence. The Italian central government has met a growing request for devolution which has also resulted in creating new provinces. Since their institution the number of provinces has always increased: in the last 20 years, seven new provinces have been established. This administrative reorganisation has generated a new map of the Italian national administrations. Nevertheless, the administrative devolution has introduced new problems, such as redundant duties carried by different bodies at different levels (regional, provincial and local administrations) and the waste of public money. Under this background, in 2001 the Region of Sardinia decided to double the number of Provinces to eight units. In this manuscript we discuss a method to verify the goodness of the new regional administrative configuration of Sardinia. The results that we present are based on a case study by De Montis et al. (2013).

3.5.2 Regionalisation and Commuting Networks in Sardinia, Italy

Sardinia is the second largest Mediterranean island with an area of approximately 24,000 km^2 and 1,600,000 inhabitants. Its geographical location and morphological features have fostered an important history of commercial and cultural relations with international communities. As of 2012 the island is partitioned into eight provinces and 377 municipalities. The Sardinian economy is progressively losing competitiveness compared to other Italian regions and other European countries. Sardinia is nevertheless in a slightly better position than average southern Italian regions (the poorest part of Italy). The Sardinian economy is primarily based on the tertiary sector (67.8 % of employment), with commerce, services, information technology, and especially tourism, which represents the main industry of the island with 2,721 firms (Banca d'Italia 2012).

According to recent studies (De Montis et al. 2007, 2011), the inter-municipal commuting system of Sardinia, (i.e. the daily movements of workers and students) can be conceived as a network. The Sardinian inter-municipal commuting network (SMCN) is composed of a set of vertices corresponding to towns, and a set of edges representing the flow of commuters between towns. The SMCN is undirected and weighted. The weight of each link represents the number of commuters that generally move between two municipalities. The SMCN is built using information from the national census (ISTAT). SMCN's characteristics can be summarized as follows:

- The SMCN is similar to a small-world random graph in terms of topology (i.e. connectivity). However the level of local interconnection between nodes diverges from usual random networks and is like typical technological networks. These networks show a hierarchy of nodes. In the case of SMCN small municipalities are locally densely interconnected. Moreover the SMCN may be defined as a disassortatively mixed network (Newman 2000), where hub nodes preferentially connect to nodes with a low connectivity and centrality ranking.
- The SMCN behaves as a scale-free network when it is conceived as a weighted network. The analysis of probability distributions of weights and strengths (the sum of weights' links attached to each node) fit a power-law. The traffic is thus gathered on a few links. This signals the presence of hub-behaviour over the busiest travelled nodes.

In Fig. 3.1, we report a geographical representation of the SMCN in the years 1981, 1991 and 2001. The networks were pruned of the less important links (connections with a few commuters compared to the average values). It is worth mentioning that the system has strongly improved its topological structure becoming more complex as time passes. This can be explained by looking at some improvements in the Sardinian economy, for example an increase in the number of cars owned, upgrades in infrastructure (especially for the road system, less for the rail system) and the per capita income.

The research idea that we discuss in this essay is based on the concept that network communities can be seen as productive basins of mobile agents. Commuter movements generate diversified scenarios depending on socio-economic peculiarities of the territory involved. Censis (2008) showed a positive correlation between the level of commuting in a territory and GDP. Richer regions have higher percentage of commuters. On the contrary, a negative correlation is detected between number of commuters and unemployment rate. The structure of commuter basins provides a functional redefinition of the administrative and cultural divisions based on the idea that the strongest interrelationships link administrative units that belong to the same cluster (i.e. municipalities are clustered into provinces). In order to understand the composition and significance of network commuting communities,

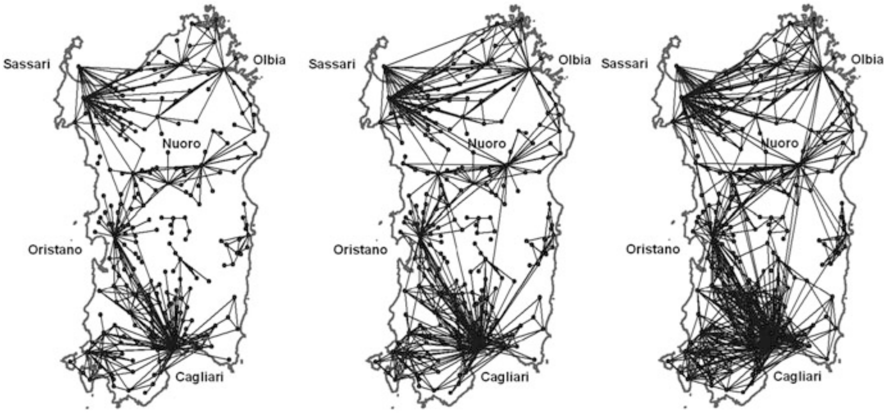


Fig. 3.1 Geo-referred representation of the SMCN in 1981 (*left*), 1991 (*centre*) and 2001 (*right*) (Source: De Montis et al. (2013))

we confront them with administrative and cultural divisions, such as provinces, "Piano di Rinascita" units, "Comprensori", "Profili d'Area" and Historical regions. Fifteen territorial units were proposed with "Piano di Rinascita" during the 1960s. In the 70s, 25 Comprensori were conceived and designed (although they were never made official): Comprensori represented a new intermediate administrative level between Regione and Comuni (municipalities). The historical regions are homogeneous geographical areas that group together Sardinian municipalities with a similar history, language, and cultural identity.

De Montis et al. (2013) applied the network community detection method proposed by Blondel et al. (2008) to the SMCN and correlate the results with the above mentioned administrative units. Figure 3.2 visually overlays the limits of provinces before the 2001 reform (old provinces) and after (new provinces) with the eight communities detected by the Blondel method for the SMCN in 2001.[4]

The Adjusted Rand Index (Hubert and Arabie 1985) was used in order to quantitatively assess similarities between the Lauvain and the administrative partitions. Results show that the highest similarities of Lauvain partition are with the new configuration of provinces and Profili d'Area. The highest similarities are detected in the SMCN's partitions of years 1991 and 2001. In brief, we can assert that the recent institution of the four new provinces better suits the actual socio-economic dynamics of the Sardinian territory. With this case study, we have shown that community detection methods are helpful tools in spatial and urban planning. They provide guidance for analysts, planners and stakeholders to read, understand and depict territorial dynamics. Such models and methods applied to planning

[4] See De Montis et al. (2013) for further visualisations and results on Comprensori, Piano di Rinascita, Profili d'Area, historical regions and communities detected over the SMCN.

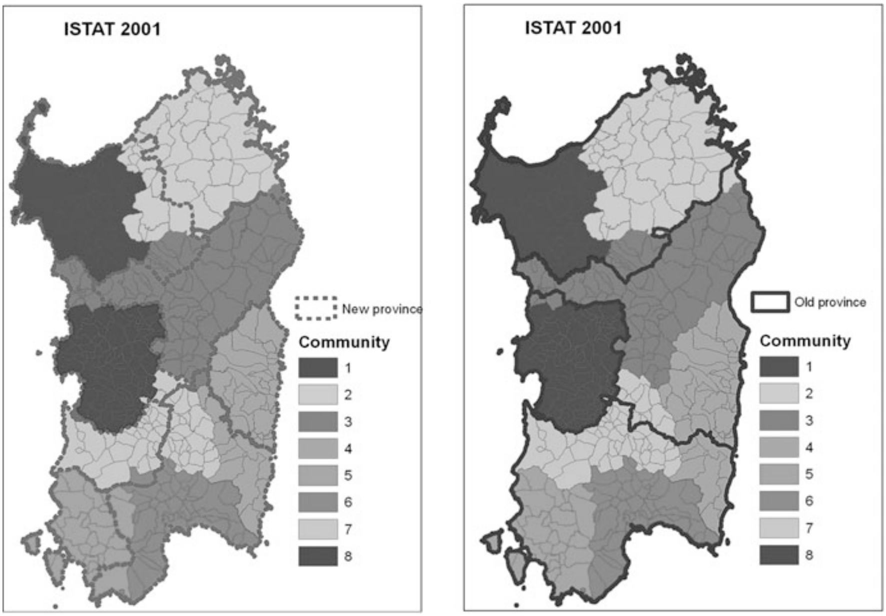


Fig. 3.2 Overlay of SMCN communities with old and new provinces (Source: De Montis et al. (2013))

cannot replace the work of analysts but can facilitate to unearth and implement solutions for complex spatial problems in regional planning.

3.6 Conclusion

In this chapter, we have presented and illustrated evidence of a case study regarding the detection of commuters' communities in Sardinia starting from the characterization of a social network sui generis, where nodes stand for origin and destination towns and edges correspond to commuting flows between them. We have verified that SMCN exhibits properties typical of other social networks. The presentation of the case study is provided with a review of (i) applications of CNA to social and R&D networks that show a clear reference to space, (ii) recent integration between spatial and network analysis fields and (iii) the last acquisitions in the field of community detection methodologies adopted to partition, in particular, spatial networks.

The development of the case study application demonstrates that the applied community detection methodology is able to profile homogeneous and contiguous clusters of municipalities that are commuters' basins which generally mirror provincial bodies. In a backward looking vein, this method has allowed us to spot critical situations arising from spatial discrepancies between commuter's basins and other relevant partitions historically adopted by planners. In a forward looking perspective, community detection proves to be a tool able to support planners in shaping ideal spatial units or subdivision, a very important issue for regionalization and regional planning.

In supporting the foregoing remarks for the unforeseeable future, Complexity and Complex Network theory have to be more integrated into urban and regional planning methods. This can be achieved through an extensive application of these concepts to model urban phenomena. This practice might change the perceptions of urban phenomena and the manner that urban planning is practiced. The case study presented in this manuscript is one of the first attempts for a fruitful integration between the complex network paradigm and regional science. This case study encourages us to extend the analysis by including an economic framework into the analysis. We would like to verify whether the detected clusters are also economically sustainable. Furthermore, the method needs to be validated in other realms, both in other Italian regions and in international settings.

References

Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47–97

Albert R, Albert I, Nakarado GL (2004) Structural vulnerability of the North American power grid. Phys Rev E 69:025103

Andersson C, Frenken K, Hellervik A (2006) A complex network approach to urban growth. Environ Plann A 38(10):1941–1964

Anderstig C, Mattsson LG (1991) An integrated model of residential and employment location in a metropolitan region. Pap Reg Sci J RSAI 70(2):167–184

Bajardi P, Poletto C, Ramasco JJ, Tizzoni M, Colizza V et al (2011) Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. Plos One 6(1):e16591

Banca d'Italia (2012) Economie regionali: L'economia della Sardegna. Banca d'Italia Eurosistema 22:1–102

Barabàsi AL, Albert R (1999) Emergence of scaling in random networks. Science 286:509–512

Barthélemy M (2011) Spatial networks. Phys Rep 499(1–3):1–101

Barthélemy M, Flamini A (2008) Modeling urban street patterns. Phys Rev Lett 100(138702):1–4

Batty M, Longley P (1994) Fractal cities: a geometry of form and function. Academic, San Diego

Benenson I (1998) Multi-agent simulations of residential dynamics in the city. Comput Environ Urban Syst 22(1):25–42

Bloch F (2005) Group and network formation in industrial organization: a survey. In: Demange G, Wooders M (eds) Group formation in economics: networks, clubs and coalitions. Cambridge University Press, New York, pp 335–353

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 10:10008–10020

Booher DE, Innes JE (2002) Network power in collaborative planning. J Plann Edu Res 21:221–236

Brandes U, Delling D, Gaertler M, Görke R, Hoefer M, Nikoloski Z, Wagner D (2006) On modularity-np-completeness and beyond. Univ., Fak. für Informatik, Bibliothek, Tech. rep., Faculty of Informatics, Universitat Karlsruhe, Germany

Caschili S, De Montis A (2013) Accessibility and Complex Network Analysis of the U.S. commuting system. Cities 30:4–17

Censis (2008) Pendolari d'Italia: scenari e strategie. FrancoAngeli, Milano
Cerina F, De Leo V, Bartélemy M, Chessa A (2012) Spatial correlations in attribute communities. Plos One 7(5):e37507. doi:10.1371/journal.pone.0037507
Chessa A, Morescalchi A, Pammolli F, Penner O, Petersen AM, Riccaboni M (2013) Is Europe evolving toward an integrated research area? Science 8(339):650–651
Crucitti P, Latora V, Marchiori M (2004) A topological analysis of the Italian electric power grid. Phys A 338:92–97
Davies A (2002) Power, politics and networks: shaping partnerships for sustainable communities. Area 34(2):190–203
De Montis A, Barthélemy M, Chessa A, Vespignani A (2007) The structure of interurban traffic: a weighted network analysis. Environ Plann B Plann Des 34(5):905–924
De Montis A, Caschili S, Chessa A (2011) Time evolution of complex networks: commuting systems in insular Italy. J Geogr Syst 13(1):49–65
De Montis A, Caschili S, Chessa A (2013) Commuter networks and community detection: a method for planning sub regional areas. Eur Phys J Spec Top 215:75–91
de Solla Price DJ (1965) Networks of scientific papers. Science 149(3683):510–515
Dempwolf CS, Lyles LW (2011) The uses of social network analysis in planning: a review of the literature. J Plann Lit 27(1):3–21
Duque JC, Ramos R, Suriñach J (2007) Supervised regionalization methods: a survey. Int Reg Sci Rev 30(3):195–220
Eraydin A, Koroglu A, Ozturk H, Yasar S (2008) Network governance for competitiveness: the role of policy networks in the economic performance of settlements in the Izmir region. Urban Stud 45:2291–2321
Expert P, Evans TS, Blondel VD, Lambiotte R (2011) Uncovering space-independent communities in spatial networks. Proc Natl Acad Sci U S A 108:7663–7668
Flake GW, Lawrence S, Lee GC, Coetzee FM (2002) Self-organization and identification of web communities. IEEE Comput 35(3):66–71
Fortunato S (2010) Community detection in graphs. Phys Rep 486:75–174
Fortunato S, Barthelemy M (2006) Resolution limit in community detection. Physics 104 (8):0607100
Freeman LC (1979) Centrality in social networks: conceptual clarification. Soc Netw 1:215–239
Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci U S A 99:7821–7826
Guimerà R, Amaral LA (2005) Functional cartography of complex metabolic networks. Nature 433:895–900
Hanaki N, Nakajima R, Ogura Y (2010) The dynamics of R&D network in the IT industry. Res Policy 39:386–399
Heikkila E, Gordon P, Kim J, Peiser R, Richardson H (1989) What happened to the CBD-distance gradient?: land values in a polycentric city. Environ Plann A 21:221–232
Hillier B (1996) Space is the machine: a configurational theory of architecture. Cambridge University Press, Cambridge, UK
Hillier B, Hanson J (1984) The social logic of space. Cambridge University Press, Cambridge, UK
Hillier B, Leaman A, Stansall P, Bedford M (1976) Space syntax. Environ Plann B 3:147–185
Hinloopen E, Nijkamp P, Rietveld P (1983) Qualitative discrete multiple criteria choice models in regional planning. Reg Sci Urban Econ 13(1):77–102
Holman N (2008) Community participation: using social network analysis to improve developmental benefits. Environ Plann C Government policy 26(3):525–543
Hubert L, Arabie P (1985) Comparing partitions. J Classif 2:193–218
Innes JE, Booher DE (2010) Planning with complexity: an introduction to collaborative rationality for public policy. Routledge, London
Jin JH, Park SC, Pyon CU (2011) Finding research trend of convergence technology based on Korean R&D network. Expert Syst Appl 38:15159–15171

Kansky KJ (1963) Structure of transportation networks: relationships between network geometry and regional characteristics. Research Paper 84, Department of Geography, University of Chicago, Chicago
König M, Battiston S, Napoletano M, Schweitzer F (2012) The efficiency and stability of R&D networks. Games Econ Behav 75:694–713
Krause A, Frank K, Mason D, Ulanowicz R, Taylor W (2003) Compartments revealed in food-web structure. Nature 426:282–285
Kurant M, Thiran P (2006) Extraction and analysis of traffic and topologies of transportation networks. Phys Rev E 74:036114–10
Lang R (2003) Edgeless cities: Exploring the elusive metropolis. Brookings Institution, Washington, DC
Latora V, Marchiori M (2001) Efficient behavior of small-world Networks. Phys Rev Lett 87:198701–198704
Lopez E, Bocco G, Mendoza M, Duhau E (2001) Predicting land cover and land use change in the urban fringe: a case in Morelia city Mexico. Landsc Urban Plann 55(4):271–285
Lusseau D, Newman MEJ (2004) Identifying the role that animals play in their social networks. Proc R Soc Lond B 271:477–481
Makse HA, Andrade JS, Batty M (1998) Modelling urban growth patterns with correlated percolation. Phys Rev E 58(6):7054–7062
Martínez FJ (1996) MUSSA: a land use model for Santiago city. Transp Res Rec 1552:126–134
McFadden D (1974) Conditional logit analysis of qualitative choice behaviour. In: Zarembka P (ed) Frontiers in econometrics. Academic, New York, pp 105–142
Milgram S (1967) The small world problem. Psychol Today 2:60–67
Mills S (1972) Studies in the Structure of the urban economy. Johns Hopkins Press, Baltimore
Muth R (1969) Cities and housing. University of Chicago Press, Chicago
Newman MEJ (2000) Assortative mixing in networks. Phys Rev Lett 89:208701
Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45:167–256
Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69(2):026113
Nystuen JD, Dacey MF (1961) A graph theory interpretation of nodal regions. Pap Proc Reg Sci Assoc 7:29–42
Oerlemans L, Meeus M (2005) Do organizational and spatial proximity impact on firm performance? Reg Stud 39(1):89–104
Palermo P (1980) La delimitazione territoriale dei comprensori, in Comprensori ristrutturazione istituzionale e territorio. Franco Angeli, Milano.
Palla G, Derenyi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435:814–818
Pijanowskia BC, Brown DG, Shellitoc BA, Manikd GA (2002) Using neural networks and GIS to forecast land use changes: a land transformation model. Comput Environ Urban Syst 26 (6):553–575
Pimm SL (1979) The structure of food webs. Theor Popul Biol 16:144–158
Porta S, Crucitti P, Latora V (2006a) The network analysis of urban streets: a dual approach. Phys A Stat Mech Appl 369:853–866
Porta S, Crucitti P, Latora V (2006b) The network analysis of urban streets: a primal approach. Environ Plann B 33:705–725
Porta S, Latora V, Strano E (2010) Networks in urban design six years of research in multiple centrality assessment. In: Network science complexity in nature and technology. Springer, London, pp 107–130
Putman SH (1983) Integrated urban models. Pion, London
Sales-Pardo M, Guimera R, Moreira AA, Amaral LAN (2007) Extracting the hierarchical organization of complex systems. Proc Natl Acad Sci U S A 104(39):15224–15229
Scholz JT, Berardo R, Kile B (2008) Do networks solve collective action problems? Credibility, search, and collaboration. J Polit 70(2):393–406

Smith-Doerr L, Manev IM, Rizova P (2004) The meaning of success: network position and the social construction of project outcomes in an R&D lab. J Eng Technol Manage 21:51–81
Sternberg R, Arndt O (2001) The firm or the region: what determines the innovation behavior of European firms? Econ Geogr 77:364–382
von Ferber C, Holovatch T, Holovach Y, Palchykov V (2009) Public transport networks: empirical analysis and modelling. Eur Phys J B 68:261–275
Wang F (2000) Modeling commuting patter in Chicago in a GIS environment: a job accessibility perspective. Prof Geographer 52:120–133
Wasserman S, Faust K (1994) Social networks analysis: methods and applications. Cambridge University Press, Cambridge, UK
Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393:440–442
Wellman B (2001) The persistence and transformation of community: from neighbourhood groups to social networks. Report to the Law Commission of Canada
Wilson AG (2000) Complex spatial systems: the modelling foundations of urban and regional analysis. Prentice Hall, Harlow
Wu F, Yeh AGO (1997) Changing spatial distribution and determinants of land development in Chinese cities in the transition from a centrally planned economy to a socialist market economy: a case study of Guangzhou. Urban Stud 34(11):1851–1879
Yazdani A, Jeffrey P (2010) A complex network approach to robustness and vulnerability of spatially organized water distribution networks. arXiv 1008.1770
Zipf GK (1949) Human behaviour and the principle of least effort. Addison-Wesley Press, Cambridge, MA

Chapter 4
Determinants of Cross-Regional R&D Collaboration Networks: An Application of Exponential Random Graph Models

Tom Broekel and Matté Hartog

Abstract This study investigates the usefulness of exponential random graph models (ERGM) to analyze the determinants of cross-regional R&D collaboration networks. Using spatial interaction models, most research on R&D collaboration between regions is constrained to focus on determinants at the node level (e.g. R&D activity of a region) and dyad level (e.g. geographical distance between regions). ERGMs represent a new set of network analysis techniques that has been developed in recent years in mathematical sociology. In contrast to spatial interaction models, ERGMs additionally allow considering determinants at the structural network level while still only requiring cross-sectional network data.

The usefulness of ERGMs is illustrated by an empirical study on the structure of the cross-regional R&D collaboration network of the German chemical industry. The empirical results confirm the importance of determinants at all three levels. It is shown that in addition to determinants at the node and dyad level, the structural network level determinant "triadic closure" helps in explaining the structure of the network. That is, regions that are indirectly linked to each other are more likely to be directly linked as well.

T. Broekel (✉)
Institute of Economic and Cultural Geography, Leibniz University of Hanover, Hanover, Germany
e-mail: broekel@wigeo.uni-hannover.de

M. Hartog
Section of Economic Geography, Faculty of Geosciences, Utrecht University, P.O. Box 80115, 3508 TC Utrecht, The Netherlands
e-mail: m.hartog@uu.nl

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_4,

4.1 Introduction

There is growing scientific interest in the creation of knowledge and its diffusion among organizations. In the new growth theory, new knowledge is regarded as being pivotal to economic growth by generating increasing returns (Romer 1990). In evolutionary economics, the re-combination of existing knowledge from different sources is argued to be crucial for new innovations to occur (Nelson and Winter 1982). These theories and the according empirical evidence also impacted the policy level. For instances, one of the most well known policy instruments to stimulate knowledge diffusion and innovation are the Framework Programmes of the European Union. These programs have been in existence since 1984 and are used to fund thousands of collaborative research projects between organizations in the EU.

Such R&D collaboration networks, which are induced by policy, alter the spatial diffusion of knowledge. This put the investigation of their spatial structures on the agenda of regional economists and economic geographers (Autant-Bernard et al. 2007). The geographical structures of inter-organizational collaboration networks are frequently analyzed from an organizational perspective (cf. Giuliani and Bell 2005) and a regional perspective, the latter focusing on cross-regional R&D collaboration networks (cf. Scherngell and Barber 2009, 2011; Hoekman et al. 2010). In order to investigate factors explaining the structure of cross-regional networks, most commonly used are spatial interaction models, which allow for considering factors at the node and dyad level. An example of a factor at the node level is the size of a region that matters as regions with more organizations are also more likely to have links to regions elsewhere. At the dyad level, most attention has been paid to the effect of geographical distance, which has been found to have a negative impact on the chance of research collaboration (cf. Ponds et al. 2007; Scherngell and Barber 2009; Hoekman et al. 2009, 2010).

In addition to the node and dyad level, factors at the structural network level may also be important, though. That is, the creation of new links might not only depend on attributes of regions or region pairs, but may also be influenced by the existing structure of the cross-regional network. For instance, a key hypothesis in organizational network science is the tendency towards triadic closure (or transitivity), which implies in this context that regions, which are indirectly linked, are more likely to link themselves as well. However, factors at the structural network level cannot be included in spatial interaction models.

This chapter presents exponential random graph models (ERGM) as an alternative empirical tool to investigate this. These models have been developed in mathematical sociology in recent years (Snijders et al. 2006; Robins et al. 2006, 2007; Wang et al. 2012) and are increasingly used across scientific disciplines, for example in bioscience (Saul and Filkov 2007), political science (Desmarais and Cranmer 2012) and organization science (Uddin et al. 2012). The advantage of these models is that they allow for simultaneously estimating the effect of factors at the node, dyad, and structural network level for networks that are observed at one point in time. We illustrate the usefulness of ERGMs by exemplarily investigating the structure and its determinants of the cross-regional R&D collaboration network in the German chemical industry between 2005 and 2010.

The study is structured as follows. The second section gives an overview of the literature on spatial structures of R&D collaboration networks and their determinants. This includes a brief discussion of factors at the node, dyad, and structural network level that may impact network structures. The third section elaborates on the exponential random graph model that we subsequently use to investigate the structure of the cross-regional network. We present the empirical data in the fourth section. It is followed by the discussion of the results in the fifth section and some concluding remarks in the sixth section.

4.2 Determinants of Cross-Regional R&D Collaboration

The structural determinants of cross-regional R&D collaboration networks can be distinguished at three different levels. These are the node level, the dyad level, and the structural network level. In the following, we elaborate on the factors effective at these three different levels.

Node level factors are properties of network entities themselves. With respect to regional R&D collaboration networks, regions' size and research intensity are particularly important. Regions with more organizations can be expected to have more ties because they have more collaboration opportunities. Such a size effect also applies at the firm level, as large organizations are likely to have more ties than small organizations because their position in the industry is more prominent and have more resources at their disposal to create and maintain ties. For instance, Boschma and Ter Wal (2007) find that larger organizations are more central in the knowledge network of footwear producers in Barletta. Secondly, the research intensity of organizations in a region matters. At the firm level, Giuliani and Bell (2005) show that organizations with a more advanced knowledge base are more frequently approached by other organizations to exchange knowledge because they are perceived to be 'technological leaders'. A similar argument can be applied to the regional level: the research intensity of a region is generally characterized by a large number of R&D employees, many organizations being engaged in R&D-intensive activities, and by the presence of universities or other research institutes. These characteristics are likely to increase the number of research collaboration links organizations have with other organizations in the same region (regional collaboration) as well as with organizations located elsewhere (cross-regional collaboration), with the latter representing a region's (degree) centrality in the cross-regional collaboration network. Accordingly, it can be expected that the absolute numbers of regional and cross-regional links are strongly correlated.

Factors at the dyad level are characteristics of relationships between two entities (nodes) in a network. In the context of the paper it refers to the relation between two regions. A key idea in sociology is that entities are more likely to link when they have similar attributes, known as homophily effect (McPherson et al. 2001). For instance, regions with organizations that operate with similar routines and under comparable incentive mechanisms are more likely to be linked in R&D collaboration. Another example are universities, which are subject to different incentive frameworks than firms when it comes to knowledge creation and diffusion as they aim to publish new knowledge, whereas firms have an incentive to keep new knowledge secret. Hence, because of their institutional proximity (Metcalfe 1995), universities are more likely to collaborate with others and especially with other universities (cf. Broekel and Boschma 2012; Broekel and Hartog 2013). This is likely to translate to the regional level as regions rarely house more than one university. Accordingly, university regions have a higher likelihood of being linked to each other.

In addition to institutional proximity, other forms of proximity may also be relevant, namely: geographical proximity, technological proximity, and social proximity. Many studies confirm that cross-regional R&D collaboration is more likely when regions are located close to one another in space (e.g. Maggioni et al. 2007; Scherngell and Barber 2009; Hoekman et al. 2009, 2010). This may be due to a variety of reasons, for instance geographical proximity facilitates face-to-face contact, which stimulates the diffusion of information about potential collaboration partners. The likelihood of cross-regional R&D collaboration is shown to increase when regions have similar technological profiles and specializations (Fischer et al. 2006; LeSage et al. 2007; Scherngell and Barber 2009). A potential explanation is that organizations are more prone to collaborate with organizations with related knowledge assets. Similar technological profiles (technological proximity) ensure that two organizations can easily communicate and learn from each other (Cohen and Levinthal 1990; Nooteboom 2000). Social proximity may also increase the likelihood of R&D collaboration (cf. Autant-Bernard et al. 2007). People already knowing each other find it easier to develop trust-based relations, which in turn facilitate knowledge exchange and ease interactions across regional boundaries (Maskell and Malmberg 1999; Sobrero and Roberts 2001; Breschi and Lissoni 2009).

In addition to these factors at the node and dyad level, factors at the structural network level may also matter for the structure of cross-regional R&D collaboration. Such factors relate to properties of the entire network. Three factors are commonly put forward in this context: triadic closure (transitivity), multi-connectivity, and preferential attachment (cf. Ter Wal and Boschma 2009; Glückler 2010). Triadic closure predicts that partners of organizations are likely to become partners themselves as well. As a result, a network will consist of many triangles, i.e. dense cliques of strongly interconnected organizations (Ter Wal 2011). Such cliques can be regarded as a sign of social capital (Coleman 1988) that may enhance trust and willingness among actors to invest in mutual goals, such as research collaboration. In contrast, multi-connectivity suggests that organizations will connect to others in multiple ways to decrease the dependency on a single link. It implies that multiple paths are formed amongst organizations leading to multiple reachability. Evidence for this is found in the creation of inter-firm alliances between US biotech firms (Powell et al. 2005). Preferential attachment means that organizations with many links are more likely to create or attract new links in the future. If a network is shaped by this factor, its degree distribution follows a power law (Barabasi and Albert 1999). Gulati (1999) shows that in the case of multinational firms, the likelihood of creating new alliances increases the better organizations are connected in the network. Hence, the network of alliances among multinational firms is subject to preferential attachment processes.

In contrast to most of the discussed factors at the node and dyad level, these factors are not regional in nature. Concepts like transitivity, preferential attachment or reciprocity do not apply to the regional level. However, in most empirically observed cross-regional networks, links are constructed from regionally aggregated inter-organizational relations. To the extent that these inter-organizational relations involve organizations being located in different regions such effects will naturally be translated to the cross-regional network. Accordingly, they need to be taken into account when analyzing the network structure as multi-connectivity, preferential attachment, and triadic closure also shape the empirically observed cross-regional networks.

To estimate the relative impact of the above factors on the structure of a network, they need to be simultaneously incorporated in the empirical model. This is not possible with the models most frequently used to investigate cross-regional collaboration: spatial interaction models in general and gravity models in particular (cf. Scherngell and Barber 2009). These models can account for factors at the node and dyad level. However, they cannot be used to evaluate factors at the structural network level. In light of the theoretical relevance of factors at the structural network level, we therefore argue that network analysis modeling techniques represent a powerful alternative because they are able to simultaneously incorporate factors at all three levels.

When longitudinal data is available, a stochastic actor-based network approach can be used. It models the change of a network from one point in time to another as part of an iterative Markov chain process (see for technical details: Snijders et al. 2010). When it comes to the analysis of research collaboration networks of regions, however, such an approach is less useful. By aggregating collaboration data to the regional level and creating cross-regional networks, researchers generally are interested in approximating the relational interaction structures of regions and investigate their structures and determinants. Such networks are unlikely to drastically change within short time periods, though, as they are results of long-term social, regional, and industrial evolution processes. Hence, even when longitudinal data on these cross-regional networks structures are available, it is unlikely to cover a sufficiently long time period. It may include multiple time periods (years) and thereby principally allow for employing longitudinal network analysis to study changes in the underlying cross-regional interaction structures.[1] However, the

[1] The relational data derived from the 5th, 6th, and 7th EU-Framework Programmes are (currently) a good example in this respect. While they represent longitudinal data, it covers only a limited

results generated with stochastic actor-based network approaches are unlikely to yield meaningful insights because the empirically observed changes in the network structures are dominated by short-term fluctuations that are of little interest to the researcher. We therefore argue that exponential random graph models are the preferred option when investigating the structure of cross-regional interaction on the basis of data with a cross-sectional nature and factors at the structural network are to be considered. We elaborate on these models in the next section.

4.3 Exponential Random Graph Models

Exponential random graph models are stochastic models that approach link creation as a time-continuous process. They regard a network observed at one point in time as one particular realization out of a set of multiple hypothetical networks with similar properties. This allows applying these models to purely cross-sectional network data.

The aim of exponential random graph models is to identify factors that maximize the probability of the emergence of a network with similar properties as the structure of the observed network. The general form of exponential random graph models is defined as follows (Robins et al. 2007):

$$\Pr(Y = y) = \left(\frac{1}{\kappa}\right)\exp\left\{\sum\nolimits_{\mathrm{A}}\eta_{\mathrm{A}}\, g_{\mathrm{A}}(y)\ \right\} \tag{4.1}$$

where $\Pr(Y = y)$ is the probability that the network (Y) generated by an exponential random graph is identical to the observed network (y), κ is a normalizing constant to ensure that the equation is a proper probability distribution (summing up to 1), η_{A} is the parameter corresponding to network configuration A, and $g_{\mathrm{A}}(y)$ represents the network statistic. Network configurations can be factors at the node level, dyad level and structural network level.

Estimation of the parameters is done with maximum pseudo likelihood or a Markov Chain Monte Carlo Maximum Likelihood Estimation procedure. The latter has been developed most recently and is regarded as the preferred procedure as it is often most accurate (Snijders 2002; Van Duin et al. 2009). It is based on the generation of a distribution of random graphs by stochastic simulation from a starting set of parameter values, and subsequent refinement of those parameter values by comparing the obtained random graphs against the observed graph. This process is repeated until the parameter estimates stabilize. If they do not, the model might fail to converge and hence becomes unstable (see for technical details, Handcock 2003; Hunter et al. 2008).

time-period (1998–2013). Of course, this may change when data on future programs will become available.

Checking whether the parameters predict the observed network well, i.e. evaluating a model's goodness of fit, is done by comparing the structure of the simulated networks to the structure of the observed network. According to Hunter et al. (2008), the comparison consists of the degree distribution, the distribution of edgewise shared partners (the number of links in which two organizations have exactly k partners in common, for each value of k), and the geodesic distribution (the number of pairs for which the shortest path between them is of length k, for each value of k). The more the distributions of the simulated networks are in line with those of the observed network, the more accurate are the parameters of the exponential random graph model. In the next section, we construct an exponential random graph model to investigate the structure of the network of subsidized R&D collaboration in the German chemical industry.

4.4 Determinants of Cross-Regional R&D Collaboration in the German Chemical Industry

4.4.1 Data

We analyze R&D collaboration that has been funded by the German federal government. As in most other advanced countries, the government actively supports public and private R&D activities with subsidies. While the Federal Ministry of Education and Research (BMBF) is the prime source of subsidies, the Federal Ministry of Economics and Technology (BMWi) and the Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) contribute as well. The federal ministries publish comprehensive information about subsidized projects in the so-called "Förderkatalog" (subsidies catalog). This catalog contains detailed information on more than 150,000 individual subsidies that have been granted between 1960 and 2012. The catalog also includes information on the cooperative nature of projects. It specifically indicates if projects are joint projects realized by consortia of organizations. Participants in joint projects agree to a number of regulations that guarantee significant knowledge exchange between the partners. Accordingly, two organizations are defined to cooperate if they participate in the same joint project. Hence, the original network is a two-mode network (project-organizations links), which we transform into a one-mode projection of the network (organization-organization links). All organizations can be assigned to labor market regions allowing for regionalizing the network (see for more details on the data Broekel and Graf 2012). The data is comparable to the EU Framework Programmes (EU-FP) data by and large, which is extensively used to model research collaboration networks (cf. Scherngell and Barber 2009). In contrast to the EU-FP data, the data at hand exclusively covers collaboration between German organizations.

To construct the network of subsidized R&D collaboration in the German chemical industry, we first identify all firms in the data that are classified as being involved in the 2-digit NACE code C20 'Manufacture of chemicals and chemical products'. Subsequently, all cooperative projects are extracted in which at least one of these firms participates. On the basis of the joint appearance in a project, we construct the inter-organizational network among all chemical firms participating in these projects. We only consider links among firms: links to universities, research organizations, associations, and to firms belonging to other industries are excluded. We believe that this approach provides the most conservative picture of the (subsidized) R&D collaboration network in the chemical industry. Alternatively one may consider all organizations active in joint projects in which at least one firm of the chemical industry is participating. However, such seems to be a very broad definition of an industry-specific network, which makes the definition of appropriate empirical variables more difficult. We acknowledge that the links to organizations in other industries are also likely to shape the intra-industry network, but as our main focus is on the impact of the factors at the three different levels (node, dyad, structural) rather than on knowledge exchange as such, we leave this for future research.

The corresponding inter-organizational undirected network is subsequently aggregated to the regional level using information on organizations' location in the 270 German labor market regions. The 270 labor market regions are defined by the German Institute for Labor and Employment (e.g. Greif and Schmiedl 2002). We construct the network that existed between 1 January 2005 and 31 December 2010. In this period, 775 projects were subsidized in which at least one firm of the chemical industry was involved. These projects are split into 975 individual funds allocated to 557 German firms belonging to the chemical industry.[2] 133 of the 775 projects are joint projects, which involve on average 2.8 firms. The resulting cross-regional R&D collaboration network is shown in Fig. 4.1.

The network is dichotomized, as we are only interested in whether or not a link exists between regions. The figure shows that the large agglomerations of the Ruhr Area, Frankfurt am Main, and Munich are important nodes in the network. In addition, a number of central regions are located along the Rhine River in the west. The region of Dresden is a central node in East Germany. All these regions are well-known centers of the chemical industry in Germany. Some additional descriptive statistics of the network are presented in Table 4.2 in the Appendix.

[2] This figure is based on the number of executing organizations ("*Ausführende Stelle*") as given in the data. Many of these organizations are part of larger organizations. This has however little relevance for the results as all data are aggregated to the regional level.

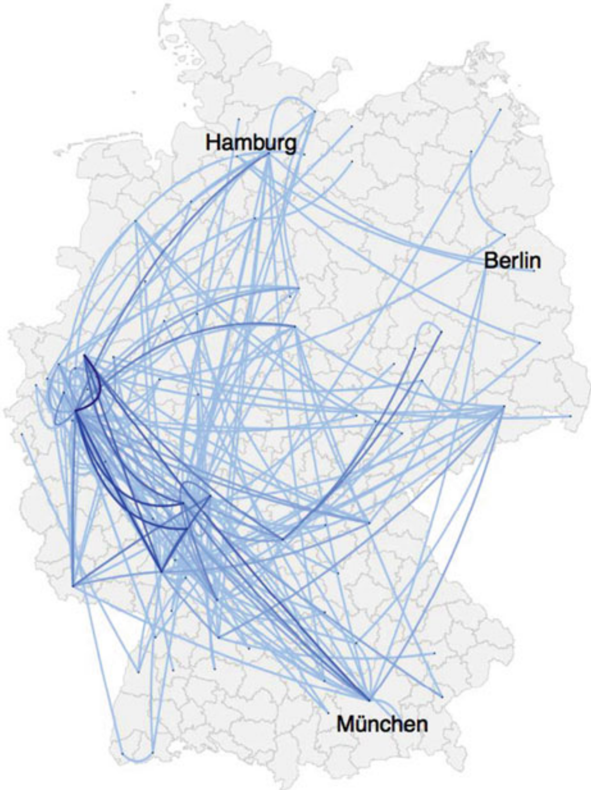


Fig. 4.1 Network of subsidized R&D collaboration among firms in the German chemical industry (2005–2010)

4.4.2 Construction of Empirical Variables

Node Level Variables

The most important node-level factors likely are the intensity of regional R&D and innovation activities in the field of chemistry. Foremost, this is because undertaking R&D activities is necessary to receive R&D subsidies. Regions with large R&D activities are likely to host more organizations that are involved in R&D collaboration. Moreover, such regions may also be the location of the most successful innovators, which are preferred collaboration partners. We therefore consider the number of applied patents in chemistry by regional organizations as proxy for the intensity of regional R&D activities in this field. The regionalized data on patent applications are published in Greif and Schmiedl (2002) and Greif et al. (2006), which include applications to the German as well as to the European Patent Office, with a correction for double counts. The patents are assigned to labor market regions according to the inventor principle. The patent data is organized according to IPC-classes, which is matched to the 2-digit NACE industry using the concordance of Broekel (2007). Lacking the data for the years 2005–2010, we construct the first node-level variable as the summed number of patents of regional firms in the field of chemistry in the years 2001–2005.[3] The variable is denoted as PATS.

We take into account the effect of urbanization by including population density (POP) and the gross-domestic product (GDP) of a region in the year 2005. The corresponding data are obtained from the German Federal Institute for Research on Building.

Firms located in regions with strong public research infrastructure may also be more likely to link across regions. For instance, being co-located with public research institutes may induce knowledge spillovers and give better access to highly qualified personnel (e.g. Fritsch and Slavtchev 2007). Accordingly, firms in these regions may be more prone to conduct R&D, engage in R&D collaboration, and be more successful in terms of innovation. In order to approximate this, we measure regions' public R&D infrastructure with three variables. The presence of universities in a region is modeled by counting their numbers of graduates in natural sciences in 2005 (UNI). Similarly, the analysis includes the number of employees working in regional research institutes of the Max Planck Society (MPG) and the Fraunhofer Society (FHG). More precise, only the numbers of employees working in the institutes' technological or natural science institutes in the year 2005 enter the analysis.[4]

Dyad Level Variables

We construct three variables at the dyad level. We measure geographical proximity with the physical distance between two regions' geographic centers. The variable is denoted as (GEO_DIST). The chance of two regions being linked is expected to decrease with geographical distance. Geographical proximity frequently correlates with social proximity (Boschma 2005), which needs to be considered in the interpretation.

We also include the variable SAME_REG that has a value of 1 if both regions are located in the same federal state (i.e. NUTS 1 region), and 0 if not. SAME_REG not only accounts for geographical proximity. It is likely to represent institutional proximity as well, as regions in the same federal state are probably similar in their R&D-related institutional framework. The reason for this is the significant role the federal level is playing in the German R&D landscape. For instance, each federal state ("Bundesland") is responsible for its own resource endowment of universities and has its own R&D policies.

We also take into account that two regions with universities may be more likely to be linked. Firms in such regions are probably structurally more similar than two

[3] The latest version of the "*Patentatlas*" was published in 2006 and includes the patent data up to 2005. We use the aggregated numbers for 2001–2005 to minimize annual fluctuation.

[4] The employment numbers are relatively stable over time. Using data for a single year is therefore considered appropriate.

firms of which one is not located in a university region. It can be expected that firms in university regions are more R&D intensive and technologically more advanced as are more probable to benefit from knowledge spillovers (cf. Jaffe 1989). To take this into account, we include the variable UNI_1, which has a value of one if both regions have a university and zero otherwise.

Notably, we do not construct a measure of technological similarity, which has been shown to make regions more likely to be linked (Scherngell and Barber 2009). This is primarily motivated by data constraints. We analyze a network among firms of the same industry aggregated at the regional level. Hence, for the construction of a meaningful technological similarity measure we need information about the technological profiles of all regional firms in the chemical industry. Unfortunately, we miss such information and have to leave this issue to future research.

Structural Network Level Variables

We include four variables at the structural network level. Triadic closure (or transitivity) is captured by the geometrically weighted edgewise-shared partner statistic (GWESP-statistic: Snijders et al. 2006; Hunter et al. 2008). It measures the number of triangles in the network whilst taking into account the number of links that are involved in multiple triangles (multimodality) (see for details: Hunter et al. 2008). It thereby captures how frequently two nodes are connected by a direct link as well as by an indirect connection of length 2 (i.e. "two-path") through another node (e.g. Hunter 2007). If a positive coefficient is found for this statistic, there is a tendency towards triadic closure in the network.

We consider the geometrically weighted dyad shared partner statistic (GWDSP), which is an advanced version of the alternating k-two-path statistic put forward by Snijders et al. (2006). It measures the extent to which a network shows a tendency of nodes not directly linked to each other being at least indirectly linked. In other words, the statistic approximates whether multiple paths exist between such nodes. Accordingly, it captures multi-connectivity for nodes that are not tied directly.

Another variable at the network level is EDGES. It equals the number of links in the network and should always be included in exponential random graph models. Moreover, EDGES represents a so-called k-star(1) parameter. K-stars are essential configurations in networks. For instance, a k-star(2), or 2-star, corresponds to three nodes of which one is linked to each of the other two. Accordingly, a k-star(3) shows as four nodes with one node being linked to the other three. A triangle, i.e. three mutually connected nodes, logically includes three k-stars(2). This means that these configurations are hierarchically related (cf. Snijders et al. 2006; Hunter 2007). While the EDGES parameter corresponds to a type of intercept parameter in the model, it is especially useful when considering the GWDEGREE statistic as well.

GWDEGREE is the geometrically weighted degree statistic, which helps modeling the observed network's degree distribution. Notably, the statistic can also be seen as an equivalent to the more traditional k-star statistic (Hunter 2007). When being considered alongside the EDGES statistic, GWDEGREE (broadly) allows modeling preferential-attachment processes. More precise, if this statistic obtains a positive coefficient it signals the presence of preferential-attachment and a negative coefficient indicates anti-preferential attachment (Hunter 2007).

For all three statistics, GWESP, GWDSP, and GWD, decay parameters have to be specified. Because few attempts have been made to systematically identify the best fitting parameter combinations (cf. Wright 2010), researchers commonly rely on a manual iterative trail-and-error process of estimating varying model specifications. These specifications differ in terms of included variables as well as decay parameters of the GWDSP, GWESP and GWDEGREE statistics. This process ends when the best fitting model is identified. The best fitting model is a model that is stable and converges (when the Markov Chain Monte Carlo approach is used, the parameter traces should be horizontal) and provides the most appropriate goodness-of-fit statistics (matching degree, edgewise shared partners, and geodesic distributions) given the empirical data (observed network). In other words, the best fitting model most accurately predicts the structure of the observed network.

Once this model is identified the final goodness-of-fit statistics and MCMC trace plots are generated exclude all variables that are not significant at the 0.05 level in the original model. These variables are excluded because they represent noise that may distort the model and thereby bias the according statistics (cf. Wright 2010). This "refined" model is used to generate all goodness-of-fit related statistics. We present the best fitting ERG-model for the cross-regional R&D collaboration network in the next section.

4.5 Results

Table 4.1 presents the results of the final, i.e. best fitting, model and those of its refined variant. Included are factors at the node, dyad, and structural network level. The model is stable and converges. Moreover, it is characterized by appropriate goodness-of-fit statistics (matching degree, edgewise shared partners, and geodesic distributions (Fig. 4.2 in the Appendix) and horizontal parameter traces (Figs. 4.3, 4.4, 4.5, and 4.6 in the Appendix).

Before we discuss the variables with significant coefficients, it is worthwhile to take a brief look at the insignificant ones. The insignificance of GDP implies that the economic prosperity of regions does not impact the structure of the cross-regional R&D collaboration network in the German chemical industry. The measure of the absolute physical distance (GEO_DIST) between regions better captures the effect of geographic distance than when considering whether two regions are part of the same federal state (SAME_REG), as the latter's coefficient is insignificant while that of the first is not. The finding moreover questions the role of institutional proximity, which we argued to be reflected by SAME_REG.

The measure of the network's degree distribution (GWDEGREE) does not help in explaining the structure of the network. This means that we do not find evidence for preferential attachment processes, i.e. well-connected regions are not more

Table 4.1 Results of exponential random graph model with dyad level, node level and structural network level variables

	Main model				Refined model		
Variable	Estimate	Std. error	p-Value	Sign.	Estimate	Std. error	Sign.
Node level							
PATS	0.00056	0.00013	< 1e-04	***	0.00028	0.00008	***
UNI	−0.00069	0.00017	< 1e-04	***	−0.00119	0.00015	***
POP_DEN	0.00009	0.00004	0.022735	*	0.00022	0.00001	***
GDP	−0.00113	0.00159	0.478296				
MPG	0.00037	0.00011	0.000882	***	0.00071	0.00009	***
FHG	0.00064	0.00026	0.013101	*	0.00135	0.00016	***
Dyad level							
GEO_DIST	−0.00164	0.00021	< 1e-04	***	−0.00072	0.00018	***
SAME_REG	0.07019	0.10950	0.521505				
Nodematch. UNI_1	0.30200	0.07094	< 1e-04	***	0.14760	0.07873	*
Structural network level							
EDGES	−4.36800	0.17230	< 1e-04	***	−7.24000	0.20440	***
GWESP, 0.69, fix	1.04400	0.06772	< 1e-04	***	2.02	0.00902	***
GWDEGREE	−2.86600	14.81000	0.846554				
GWDSP, 0.15, fix	0.02133	0.02736	0.435589				
Null deviance:	50343.3 on 36,315 degrees of freedom				50343.3 on 36,315 degrees of freedom		
Residual deviance:	1753.3 on 36,302 degrees of freedom				1619.3 on 36,305 degrees of freedom		
Deviance:	48589.o on 13 degrees of freedom				48724.0 on 9 degrees of freedom		
AIC:	1779.3				1639.3		
BIC:	1889.8				1724.3		

*Significant at 95 %; ***Significant at 99 %

prone to gain additional links than sparsely connected regions. The same applies to the GWDSP-statistic suggesting that two regions without a direct link are unlikely to be indirectly connected. Accordingly, we observe insignificant coefficients for variables at all three levels (node, dyad, and structural network level).

Now, we turn towards the significant variables reported in Table 4.1. As expected, regions with R&D intensive firms (PATS) tend to have more links. The same applies to urban regions (POP_DEN) and regions in which institutes of the Max-Planck (MPG) and Fraunhofer (FHG) societies are located. The according coefficients of PATS, POP_DEN, MPG, and FHG are all positive and significant. UNI obtains a negative significant coefficient suggesting that university regions tend to have fewer links. While this contradicts our expectations, it is essential to also consider the positive significant coefficient of the dyad-level variable UNI_1 in the explanation. Accordingly, university regions generally have less links but they are more likely to link to other university regions. The latter is in line with our expectations and signals the presence of a homophily effect.

The dyad-level variable GEO_DIST is characterized by a negative significant coefficient. Hence, geographical distance hampers link creation, which confirms existing empirical studies (cf. Maggioni et al. 2007; Ponds et al. 2007; Scherngell and Barber 2009; Hoekman et al. 2009, 2010; Broekel and Boschma 2012).

We argued above that the main advantage of exponential random graph models is their ability to take into account factors at the structural network level in addition to factors at the node and dyad level. The significant coefficients of two variables at the structural network level empirically confirm this level's relevance. The coefficient of EDGES is negative and significant. By being similar to an intercept variable, EDGES represents the overall density of the network when all other effects are excluded. Its negative coefficient is a common feature of networks established by social processes indicating that such networks tend to be less dense than exponential random networks (cf. Varas 2007).

In addition, we find a positive and significant coefficient of the GWESP-statistic. It means that triangles are a common feature of the network, which corresponding to the visual inspection of the network (see Fig. 4.1). In other words, regions that are directly linked are also more likely to link through indirect connections. Hence, the result suggests that triadic closure is a driving force in the network formation processes. There might however be an alternative explanation. When constructing the empirical network, we transformed a bipartite network into a one-mode type. Such transformation more or less automatically increases the likelihood of triplets in the final one-mode network. Accordingly, the positive GEWSP-statistic might pick up this effect and act as a kind of control parameter for the one-mode projection procedure. However, we pointed out in Sect. 4.1 that on average less than three firms (2.8) are jointly participating in a cooperative project. Hence, it is most likely a combination of both effects that explains the statistic's significance. In any case, this structural network factors significantly helps in modeling the structure of the network.

In sum, we find that the structure of the network is best explained by factors at the node level, dyad level, and structural network level. Moreover, the coefficients (which can be translated into odd-ratios by taking the exponential) underline that in comparison to factors at the dyad, factors at the structural network level have greater explanatory power. It shows the crucial importance of these factors for the structure of the cross-regional R&D collaboration network in the German chemical industry. This result thereby also highlights the usefulness of exponential random graph models as a tool for analyzing the structure of such types of networks.

4.6 Conclusion

The aim of this study was to discuss exponential random graph models (ERGM) as promising tools for the investigation of cross-regional collaboration networks. We pointed out that most existing studies focus on the evaluation of factors at the node and dyad level. However, network science suggests that factors at the structural network level may also be relevant in this respect. Such factors cannot be considered in methods commonly applied in this context. For instance, spatial interaction models allow only for factors at the node and dyad level. We argued that ERG-models represent a powerful alternative as they take into account factors at all three levels and require only cross-sectional network data.

We illustrated the application of ERGMs by analyzing the structure of the cross-regional R&D collaboration network in the German chemical industry between 2005 and 2010. By using an exponential random graph model, we considered factors at all three levels that might influence the network's structure. At the node level, it was shown that urban regions (reflected by population density) and regions with high research intensities are more likely to be linked to other regions. At the dyad level, we found regions to be more likely being linked when they have a university. Moreover, our results confirmed the negative impact of geographical distance on the likelihood of research collaboration. Finally, at the structural network level, evidence was provided for the existence of a triadic closure (transitivity) effect: regions that are indirectly linked to each other are likely to be directly linked as well.

A challenge for future research is the projection of networks among individuals and organizations to the regional level. This particularly concerns the question about what factors impact link formation at the level of the individual (e.g. trust, reciprocity), at the level of the organization (e.g. reputation, absorptive capacity), and at the spatial (regional) level (e.g. image, collective identity). In the present paper, and in most of the corresponding literature, these factors are all translated to the same level, i.e. that of the chosen unit of analysis. However, this ignores their relevance at different observational levels. For instance, a general finding is that regions with high research intensity are more likely to be linked to each other, but in theory it could be that the actual linkages between those regions are created by organizations that in contrast to all other organizations in their respective regions, show little or no research intensity (although this is unlikely). The same applies to the factors at the structural network level. For instance, if three organizations in three different regions link with each other a triangle will be observed in the network that might suggest the presence of a triadic closure effects. However, if two of the three organizations are located in the same region, the cross-region network shows a single link instead, which does not supports this interpretation. In this sense, the question of what is the most appropriate unit of analysis (and level of aggregation) becomes evident. This clearly lays the path for future research focusing on changing network structure when moving from one level of node aggregation to another. Researchers will have to adjust the level of node aggregation in correspondence to the objective of their investigation until reliable insights on this matter are available.

Clearly, the study is only a first step towards understanding the role factors at the structural network level play for the formation of cross-regional collaboration networks. It nevertheless underlines the usefulness of exponential random graph models for future research endeavors on this subject.

Appendix

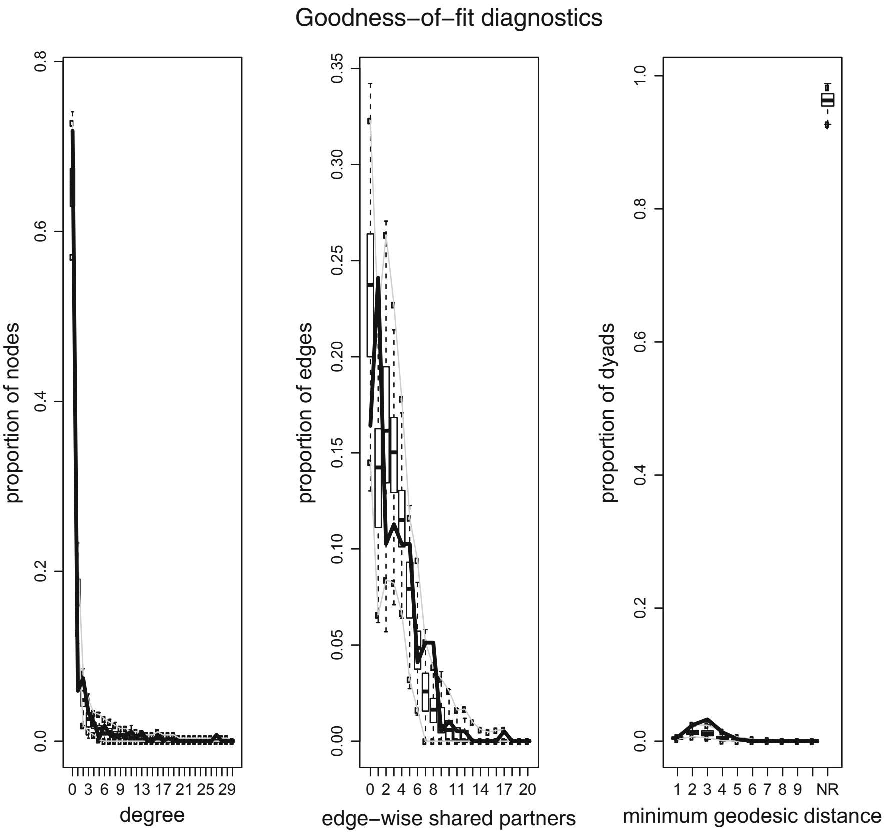


Fig. 4.2 Goodness of fit of exponential random graph model with dyad level, node level + structural network level variables

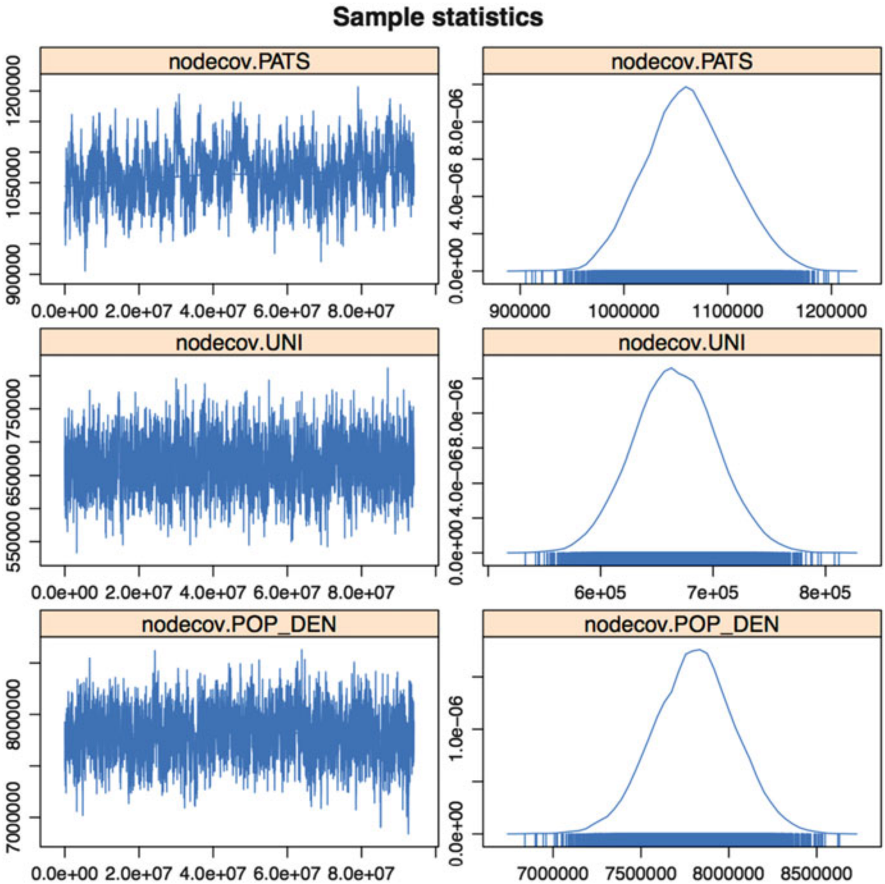


Fig. 4.3 MCMC-Statistics of exponential random graph model with dyad level, node level and structural network level variables

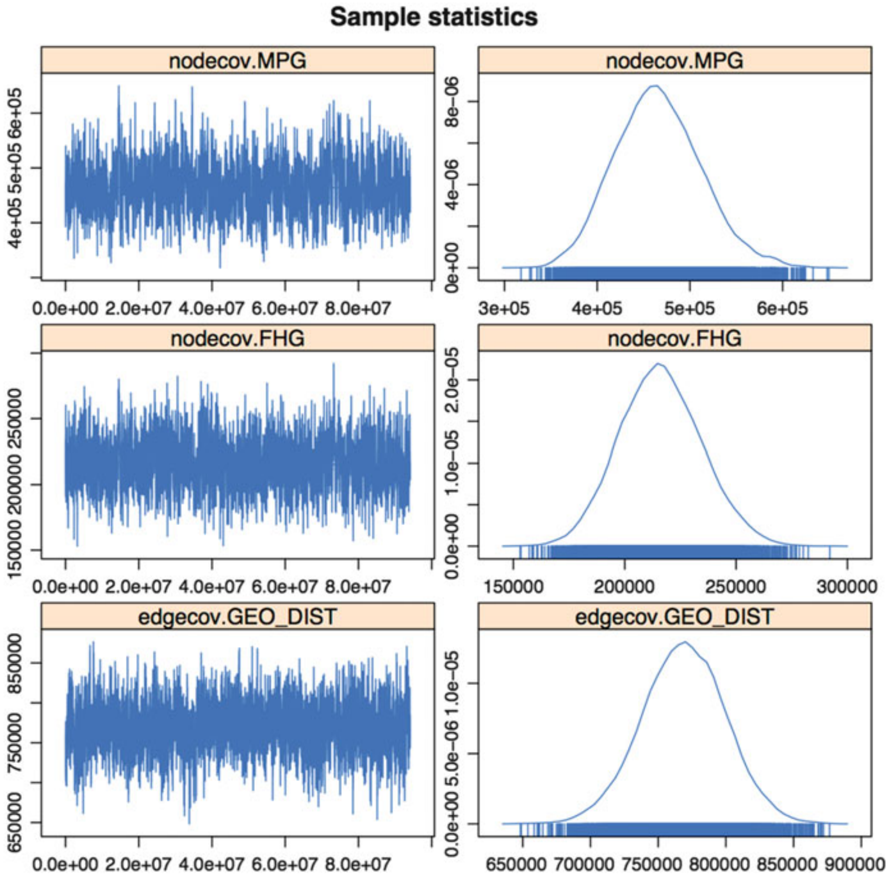


Fig. 4.4 MCMC-Statistics of exponential random graph model with dyad level, node level and structural network level variables

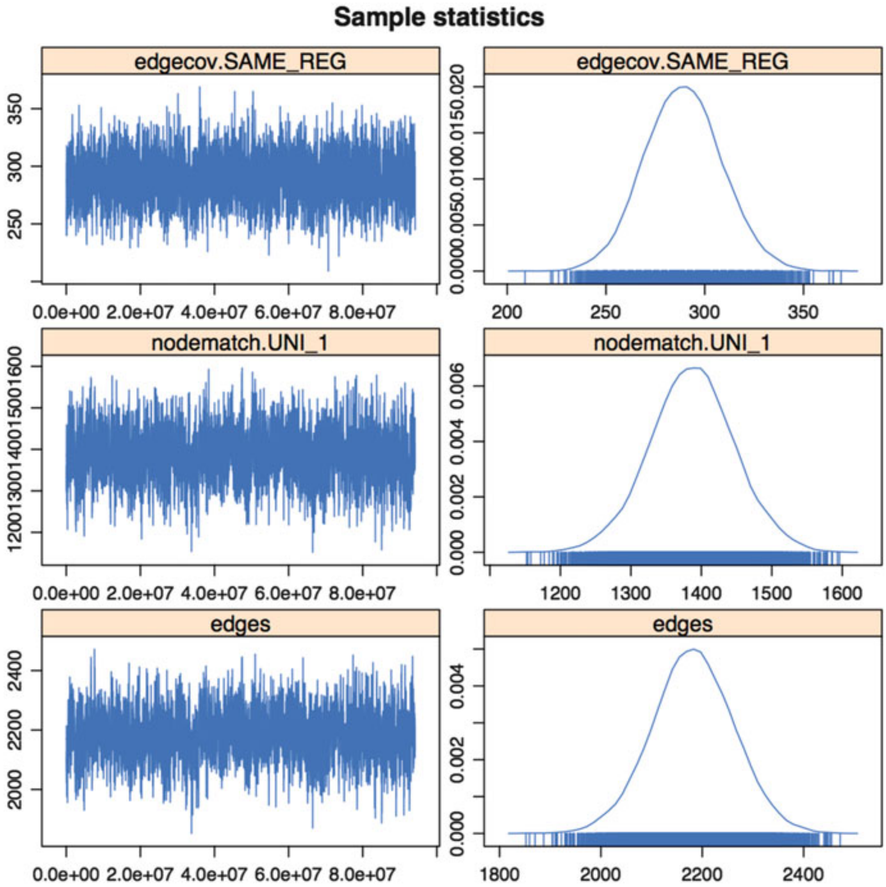


Fig. 4.5 MCMC-Statistics of exponential random graph model with dyad level, node level and structural network level variables

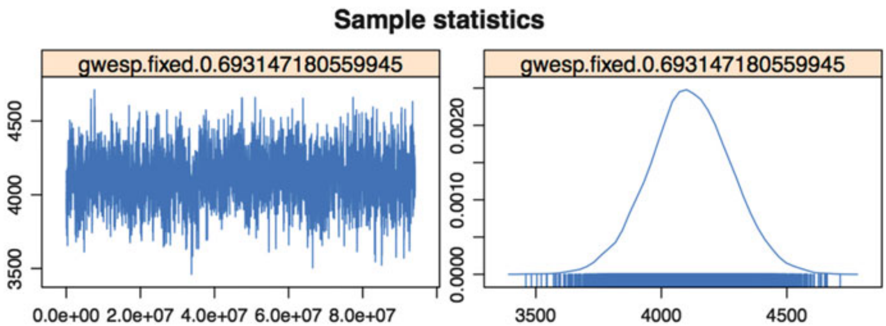


Fig. 4.6 MCMC-Statistics of exponential random graph model with dyad level, node level and structural network level variables

Table 4.2 Descriptives of empirical variables

Variables	n	Mean	St. deviation	Median	Min	Max	Skew	Kurtosis
PATS	270	69.55	199.12	12.48	0	1,691.31	5.34	32.55
POP_DEN	270	825.35	1,265.19	244.5	40	8,523	3.06	11.44
GDP	270	40.46	33.58	26.75	14.1	296.9	3.66	19.83
UNI	270	101.51	244.73	0	0	1,812	3.46	15.55
MPG	270	49.12	248.08	0	0	3,438	10.20	128.50
FHG	270	30.81	123.52	0	0	978	5.22	29.24
GEO_DIST	72,900	379.81	186.03	368.54	0	977.45	0.29	−0.52
SAME_REG	72,900	0.11	0.31	0	0	1	2.49	4.22
UNI_1	72,900	0.62	0.49	1	0	1	−0.47	−1.77

References

Autant-Bernard C, Billand P, Frachisse D, Massard N (2007) Social distance versus spatial distance in R&D co-operation: empirical evidence from European collaboration choices in micro and nanotechnologies. Pap Reg Sci 86:495–519

Barabasi AL, Albert R (1999) Emergence of scaling in random networks. Science 15:509–512

Boschma R (2005) Proximity and innovation: a critical assessment. Reg Stud 39(1):61–74

Boschma RA, Ter Wal ALJ (2007) Knowledge networks and innovative performance in an industrial district: the case of a footwear district in the south of Italy. Ind Innov 14:177–199

Breschi S, Lissoni F (2009) Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. J Econ Geogr 9:439–468

Broekel T (2007) A concordance between industries and technologies – matching the technological fields of the Patentatlas to the German industry classification. Jenaer economic research papers, 2007-013

Broekel T, Boschma R (2012) Knowledge networks in the Dutch aviation industry: the proximity paradox. J Econ Geogr 12:409–433

Broekel T, Graf H (2012) Public research intensity and the structure of German R&D networks: a comparison of 10 technologies. Econ Innov New Technol 21(4):345–372

Broekel T, Hartog M (2013) Explaining the structure of inter-organizational networks using exponential random graph models: does proximity matter? Industry Innov 20(3):277–295

Cohen WM, Levinthal DA (1990) Absorptive capacity: a new perspective on learning and innovation. Adm Sci Q 35:128–152

Coleman JS (1988) Social capital in the creation of human capital. Am J Sociol 94(Suppl): S95–S120

Desmarais BA, Cranmer SJ (2012) Micro-level interpretation of exponential random graph models with an application to estuary networks. Policy Stud J 40. doi: 10.1111/j.1541-0072.2012.00459.x

Fischer MM, Scherngell T, Jansenberger E (2006) The geography of knowledge spillovers between high-technology firms in Europe. Evidence from a spatial interaction modeling perspective. Geogr Anal 38:288–309

Fritsch M, Slavtchev V (2007) Universities and innovation in space. Ind Innov 14:201–218

Giuliani E, Bell M (2005) The micro-determinants of meso-learning and innovation: evidence from a Chilean wine cluster. Res Policy 34:47–68

Glückler J (2010) The evolution of a strategic alliance network: exploring the case of stock photography. In: Boschma R, Martin R (eds) Handbook of evolutionary economic geography. Edward Elgar, Cheltenham, pp 298–315

Greif S, Schmiedl D (2002) Patentatlas 2002 Dynamik und Strukturen der Erfindungstätigkeit. Deutsches Patent- und Markenamt, München
Greif S, Schmiedl D, Niedermeyer G (2006) Patentatlas 2006. Regionaldaten der Erfindungstätigkeit. Deutsches Patent- und Markenamt, München
Gulati R (1999) Network location and learning: the influence of network resources and firm capabilities on alliance formation. Strateg Manag J 20:397–420
Handcock MS (2003) Statistical models for social networks: degeneracy and inference. In: Breiger RL, Carley KM, Pattison P (eds) Dynamic social network modeling and analysis: workshop summary and papers. National Academies Press, Washington, DC, pp 229–240
Hoekman J, Frenken K, Van Oort F (2009) The geography of collaborative knowledge production in Europe. Ann Reg Sci 43:721–738
Hoekman J, Frenken K, Tijssen RJW (2010) Research collaboration at a distance: changing spatial patterns of scientific collaboration within Europe. Res Policy 39:662–673
Hunter DR (2007) Curved exponential family models for social networks. Soc Netw 29 (2):216–230
Hunter DR, Goodreau SM, Handcock MS (2008) Goodness of fit for social network models. J Am Stat Assoc 103:248–258
Jaffe A (1989) Real effects of academic research. Am Econ Rev 79(5):957–970
LeSage J, Fischer MM, Scherngell T (2007) Knowledge spillovers across Europe. Evidence from a Poisson spatial interaction model with spatial effects. Pap Reg Sci 86:393–421
Maggioni MA, Nosvelli M, Uberti TE (2007) Space versus networks in the geography of innovation: a European analysis. Pap Reg Sci 86:471–493
Maskell P, Malmberg A (1999) Localized learning and industrial competitiveness. Camb J Econ 23:167–186
McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Ann Rev Sociol 27:415–444
Metcalfe S (1995) The economic foundations of technology policy: equilibrium and evolutionary perspectives. In: Stoneman P (ed) Handbook of the economics of innovation and technological change. Basil Blackwell, Oxford, pp 409–512
Nelson R, Winter S (1982) An evolutionary theory of economic change. Belknap, Cambridge, MA
Nooteboom B (2000) Learning and innovation in organizations and economies. Oxford University Press, Oxford
Ponds R, Van Oort F, And Frenken K (2007) The geographical and institutional proximity of research collaboration. Pap Reg Sci 86:423–443
Powell WW, White DR, Koput KW, Owen-Smith J (2005) Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences. Am J Sociol 110:1132–1206
Robins R, Snijders T, Wang P, Handcock M, Pattison P (2006) Recent developments in exponential random graph (p*) models for social networks. Soc Netw 29:192–215
Robins G, Pattison P, Kalish Y, Lusher D (2007) An introduction to exponential random graph (p*) models for social networks. Soc Netw 29:173–191
Romer PM (1990) Endogenous technological change. J Polit Econ 98:71–102
Saul ZM, Filkov V (2007) Exploring biological network structure using exponential random graph models. Bioinformatics 23:2604–2611
Scherngell T, Barber MJ (2009) Spatial interaction modeling of cross-region R&D collaboration: empirical evidence from the 5th EU framework programme. Pap Reg Sci 88:531–546
Scherngell T, Barber MJ (2011) Distinct spatial characteristics of industrial and public research collaborations: evidence from the fifth EU framework programme. Ann Reg Sci 46:247–266
Snijders TAB (2002) Markov Chain Monte Carlo estimation of exponential random graph models. J Soc Struct 3. Web journal available from: http://www2.heinz.cmu.edu/project/INSNA/joss/index1.html
Snijders TAB, Pattison PE, Robins G, Handcock MS (2006) New specifications for exponential random graph models. Sociol Methodol 36:99–153

Snijders TAB, van de Bunt GG, Steglich CEG (2010) Introduction to stochastic actor-based models for network dynamics. Soc Netw 32:44–60

Sobrero M, Roberts EB (2001) The trade-off between efficiency and learning in interorganizational relationships for product development. Manag Sci 47:493–511

Ter Wal ALJ (2011) The dynamics of inventor networks in German biotechnology: geographical proximity versus triadic closure. Pap Evol Econ Geogr 11:02

Ter Wal ALJ, Boschma RA (2009) Applying social network analysis in economic geography: framing some key analytic issues. Ann Reg Sci 43:739–756

Uddin S, Hamra J, Hossain L (2012) Exploring communication networks to understand organizational crisis using exponential random graph models. Comput Math Organ Theor. doi:10.1007/s10588-011-9104-8

Van Duin MAJ, Gille KJ, Handcock MS (2009) A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. Soc Netw 31:52–62

Varas MLL (2007) Essays in social space: applications to Chilean communities on inter-sector social linkages, social capital, and social justice. Dissertation at University of Illinois, Urbana-Champaign

Wang P, Pattison P, Robins G (2012) Exponential random graph model specifications for bipartite networks – a dependence hierarchy. Soc Netw. http://dx.doi.org/10.1016/j.socnet.2011.12.004

Wright D (2010) Repression and network science: tools in fight against terrorism. Dissertation at University of Michigan: http://deepblue.lib.umich.edu/bitstream/handle/2027.42/77898/dewright_1.pdf?sequence=1

Chapter 5
A Novel Comprehensive Index of Network Position and Node Characteristics in Knowledge Networks: Ego Network Quality

Tamás Sebestyén and Attila Varga

Abstract While developing the comprehensive index of Ego Network Quality (ENQ) Sebestyén and Varga (Ann Reg Sci, doi:10.1007/s00168-012-0545-x, 2013) integrates techniques mainly applied in a-spatial studies with solutions implemented in spatial analyses. Following the theory of innovation they applied a systematic scheme for weighting R&D in partner regions with network features frequently appearing in several (mostly non-spatial) studies. The resulting ENQ index thus reflects both network position and node characteristics in knowledge networks. Applying the ENQ index in an empirical knowledge production function analysis Sebestyén and Varga (Ann Reg Sci, doi:10.1007/s00168-012-0545-x, 2013) demonstrate the validity of ENQ in measuring interregional knowledge flow impacts on regional knowledge generation. The aim of this chapter is twofold. First we show that ENQ is an integrated measure of network position and node characteristics very much resembling to the solution applied in the well-established index of eigenvector centrality. Second, we test the robustness of the weighting schemes in ENQ via simulation and empirical regression analyses.

5.1 Introduction

Network analytic tools have been increasingly employed in studying the flows of knowledge in two, more or less separately developed scientific literatures. 'A-spatial' approaches mostly appearing in the science and technology literature

T. Sebestyén
MTA-PTE Innovation and Economic Growth Research Group and Department of Economics and Regional Studies, Faculty of Business and Economics, University of Pécs, 80. Rákóczi str, Pécs, 7632, Hungary

A. Varga (✉)
Department of Economics and Regional Studies and MTA-PTE Innovation and Economic Growth Research Group, University of Pécs, 80. Rákóczi str, Pécs, 7632, Hungary
e-mail: vargaa@ktk.pte.hu

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_5,

study the impacts of different characteristics such as size, centrality, density, tie strength or knowledge diversity of collaboration networks among firms, research institutions and alike. The influence of different network characteristics are examined individually and the selection of particular network features studied is usually related to actual research questions and do not seem to follow an underlying theoretical agreement in the field.

On the other hand, several of the 'spatial' studies appearing in the regional economics and economic geography literature show a different origin and a somewhat different interest. The focus of this literature is not much on the architecture of knowledge networks but more on the characteristics of network partners. Knowledge level of network partners is considered their main feature determining the flow of knowledge through networks. This empirical approach is significantly influenced by spatial econometric techniques applied in some of the first papers studying the geography of knowledge flows (e.g., Anselin et al. 1997). Many of the spatial network studies apply the same intuition by replacing spatial weights matrices with matrices representing interregional collaborations. With this technique it became possible to study the impact of R&D carried out by partners in different spatial units on knowledge produced in the region (Maggioni and Uberti 2011; Varga et al. 2013; Ponds et al. 2010).

While developing the comprehensive index of Ego Network Quality (ENQ) Sebestyén and Varga (2013) integrate techniques mainly applied in a-spatial studies with solutions implemented in spatial analyses. Following the theory of innovation they applied a systematic scheme for weighting R&D in partner regions with network features frequently appearing in several (mostly non-spatial) studies (such as tie strength, number of edges, density of interactions, network distance, knowledge diversity).

The resulting ENQ index thus reflects both network position and node characteristics in knowledge networks. This index is a measure of knowledge accessible by the agents from their interregional network. Thus the interest behind ENQ is the same as in the spatial studies (i.e., the impact of R&D in partner regions). The difference is in the broader set of network features that we take into account in the analysis. Applying the ENQ index in an empirical knowledge production function analysis Sebestyén and Varga (2013) demonstrate the validity of ENQ in measuring interregional knowledge flow impacts on regional knowledge generation.

The aim of this chapter is twofold. First we show that ENQ is an integrated measure of network position and node characteristics very much resembling to the solution applied in the well-established index of eigenvector centrality. Second, we test the robustness of the weighting schemes in ENQ via simulation analyses and empirical regressions.

The second section introduces the concept of ENQ. The third section positions ENQ in traditional network centrality measurement. Results of simulation analyses with respect to the robustness of the weighting schemes applied in the formula of ENQ are reported in the fourth section. The fifth section presents some empirical underpinning of the simulation results, while summary concludes the chapter.

5.2 The Ego Network Quality Index

The theory of innovation emphasizes the role of interactions among different actors in innovation. These interactions follow a system and the characteristics of the system determine the efficiency of new knowledge production to a large extent (Lundvall 1992; Nelson 1993). An extensive survey-based empirical literature evidences that innovation is indeed a collective process where the knowledge and expertise of partners as well as the intensity of collaborations among them largely determine the production of new, economically useful knowledge (e.g., Diez 2002; Fischer and Varga 2002). Representing actors as nodes and their connections as ties, interactions of collaborating agents can be mapped as a network. On the basis of this representation the application of network analysis extends the frontiers of the study of knowledge interactions well beyond the possibilities of traditional innovation surveys.

Behind the concept of ENQ there are three intuitions directly influenced by the theory of innovation. The first intuition is that the level of knowledge in an agent's network is in a positive relationship with the agents' productivity in new knowledge generation. The second intuition is that the structure of collaboration among partners in the agent's network is the source of further growth of knowledge available from the network. Following the third intuition we assume that partners in the ego network not only increase the amount of knowledge accessible, but also contribute to its diversity through building connections to different further groups not linked directly to the ego network.

Therefore we structure ENQ around basically two dimensions, which are then augmented with a related third aspect. The two dimensions are: Knowledge Potential and Local Connectivity. Knowledge Potential (KP) measures knowledge accumulated in the direct neighbourhood and it is related to the number of partners and the knowledge of individual partners. Local Connectivity[1] (LC) is associated with the strength of ties and the intensity of interactions among partners. The third aspect is called Global Embeddedness (GE) as it intends to capture the quality of distant parts of the network (beyond immediate partners). However, this aspect is implemented by applying the concepts of KP and LC for consecutive neighbourhoods of indirect partners in the network.[2]

[1] Note that connectivity is used here in a broader sense than in graph theory. In graph theory connectivity refers to the number of vertices the removal of which disconnects the graph. In our case, this term refers to a similar concept but with a less strict definition. By connectivity we simply mean the extent of ties connecting a given group of vertices.

[2] Before moving forward, though, we have to make a terminological clarification. Generally the term 'neighbourhood' of a node refers to the group of nodes connected *directly* to a specific node. In our study by neighbourhood we mean not only the directly but also the indirectly connected nodes. As this definition in itself would mean that the term neighbourhood refers to the totality of nodes in the graph, we refine the definition and use the more specific term 'neighbourhood at distance d' which refers to the nodes exactly at distance d from a specific node.

The notation in the proceeding formulation is as follows. As usual, we represent the network under question by the adjacency matrix $\mathbf{A} = [a_{ij}]$, where the general element a_{ij} describes the connection between nodes i and j. Generally, the elements of the matrix are considered as weights, normalized to the interval between 0 and 1. A special case of this formulation is thus the binary network, where the elements of the adjacency matrix can be either zero or one. We have to note here, that only undirected networks are dealt with in this chapter, i.e. the adjacency matrix is assumed to be symmetric. The adjacency matrix defines the matrix of geodesic distances (lengths of shortest paths) between all pairs of nodes, which we denote by $\mathbf{R} = [r_{ij}]$.[3] In order to account for knowledge levels, we use $\mathbf{k} = [k_i]$ as the vector of knowledge at each specific node of the network.

Given the conceptual model presented above, we can formalize ENQ as follows:

$$ENQ^i = \sum_{d=1}^{M-1} W_d LC_d^i KP_d^i \tag{5.1}$$

In this formula superscript i refers to the node for which ENQ is calculated and subscript d stands for distances measured in the network (geodesic distance). M is the size of the network, W_d is a weighting factor used for discounting values at different d distances from node i,[4] whereas KP_d^i and LC_d^i are the respective Knowledge Potential and Local Connectivity values evaluated for the neighbourhood at distance d from node i.

As a consequence of the formulation in Eq. 5.1, we emphasize that the proposed formula for ENQ is a distance-weighted sum of Local Connectivity-weighted Knowledge Potentials evaluated for neighbourhoods at different distances in the network. By directly differentiating between immediate and indirect partners in the network, we can reformulate ENQ as follows:

$$ENQ^i = W_1 LC_1^i KP_1^i + \sum_{d=2}^{M-1} W_d LC_d^i KP_d^i = LC_1^i KP_1^i + GE^i \tag{5.2}$$

where $W_1 = 1$ is the (assumed) weighting factor for the immediate neighbourhood. Everything beyond the immediate neighbourhood can be labelled as Global Embeddedness (GE). In what follows, the two basic concepts, Knowledge Potential and Local Connectivity are introduced in more detail.

[3] In this chapter we use a non-weighted algorithm for the calculation of geodesic distances, i.e. the distance of two nodes is regarded as the number of ties connecting them, irrespective of the weights associated with these ties.

[4] The weighting factor is defined to be unity for $d = 1$ and descending towards zero as d increases. There is no unique best choice with regards the decay function. We present some illustrative simulations related to the choice of the decay function later in this chapter.

5.2.1 Knowledge Potential

The concept of KP relates to the amount of knowledge an agent's partners possess. Using the notation presented before, the concept of KP can be formulated in the following way:

$$KP_d^i = \sum\nolimits_{j:r_{ij}=d} k_j \tag{5.3}$$

The Knowledge Potential, as perceived by node i, can thus be calculated for the neighbourhoods at different d distances from node i, and for all these distances it is the sum of knowledge possessed by nodes at these distances.

5.2.2 Local Connectivity

As mentioned before, Knowledge Potential defined by Eq. 5.3 is going to be weighted by the Local Connectivity of direct and indirect neighbours. It is assumed that not only the knowledge levels of partners are of positive value to the node under question but also the cooperation between neighbours. More specifically we assume that each crosscutting tie has a positive value, depending on the weight the tie has. Local Connectivity is therefore the sum of the tie weights present in a given neighbourhood, normalized by the size of this neighborhood. The concept can be formulated as follows:

$$LC_d^i = \frac{1}{N_d^i}\left(\sum\nolimits_{j:r_{ij}=d-1}\sum\nolimits_{l:r_{il}=d} a_{jl} + \frac{\sum\nolimits_{j:r_{ij}=d}\sum\nolimits_{l:r_{il}=d} a_{jl}}{2}\right) \tag{5.4}$$

where N_d^i is the number of nodes laying exactly at distance d from node i. The expression in the parenthesis is made up of two parts. The first term counts the (weighted) ties between nodes at distance $d - 1$ and d.[5] This reflects the intensity at which two adjacent neighbourhoods are linked together. The second term counts the (weighted) number of ties among nodes at distance d.[6] As a result, Local Connectivity captures the intensity with which the (possibly indirect) neighbours at distance d are linked together and linked to other neighbourhoods.[7]

---

[5] Distances are always measured from node i.

[6] Division by two is required because matrix **A** is symmetric, and thus we can avoid duplications in the counting. This division is not required in the first term because the definition there counts only links from distance $d - 1$ to distance d and not vice versa.

[7] It is worth devoting a word to the inclusion of distance-crossing ties (the first term in the expression). Our intuition behind the concept of Local Connectivity is that collaboration among partners enhances knowledge sharing and this leads to a better environment for knowledge

To better capture the specific meaning of the expression in Eq. 5.4 recall that we employ LC as a weighting factor to KP. Assume for example that node i has N_1^i direct partners and the links connecting it to these partners are of strength 1 ($a_{ij} = 1$ for all j in the direct neighborhood). If these partners have no connections among each other, then the second term in the parenthesis is zero, and the first term is N_1^i (because all connections have weight 1). Thus, LC is unity, which reflects the intuition that the knowledge levels of partners are fully absorbed. If the connections linking node i to its partners were less strong ($a_{ij} \leq 1$ for all j in the direct neighborhood) then LC would be lower than one, contributing to a lower weighting factor. This reflects the fact that in this case partners' knowledge is not fully accessible. Assume now, that the partners establish some links among each other. In this case the second term in the parenthesis starts to increase, and the weighting factor (LC) increases too. This reflects our previous concept, namely that a higher level of collaboration among partners contributes to the knowledge attainable from a network.

In the case of indirect neighbourhoods, i.e. when $d > 1$, this normalization bears a different meaning. Nodes at distance d must be connected with nodes at distance $d - 1$ with at least as many links as many nodes there are at distance d, i.e. N_d^i. Therefore, if all these links connecting nodes at distance $d - 1$ and d are of unit strength, the first term in the parenthesis of Eq. 5.4 will be at least unity. However, it still holds that the weighting factor LC is unity in the special case if nodes at distance d are linked to nodes at distance $d - 1$ through connections of unit strength *and* with the minimum number of connections required. It is also still valid that interconnections in the neighbourhood at distance d increase the weighting factor and weaker connections between the different neighborhoods decrease the weighting factor. The only difference is that in these cases there is an extensive margin: the number of connections between the neighbourhoods can also increase, and this increases the value of LC resulting in a higher weighting factor.

To sum up, LC is a weighting factor, which describes how well-connected a node is to its neighbours and how well these neighbours are connected to each other. However, the weighting is done according to a reference point: the weight is taken to be unity for the special case if links to the partners are of unit strength and the network around the node in question is a tree, with only one link attaching each node to the previous (in the sense of distance) neighbourhood and no cooperation among partners.

---

creation. In the case of the direct neighborhood, the links connecting the node in question and its neighbors are clearly relevant in the general case of weighted ties: the amount of knowledge learnt from the immediate partners depends on the intensity of interactions with those partners. On the other hand we argue that in our concept the question is how dense the tissue of the network around the node is. We are going to attach less weight to this connectivity the farther away it is from the node, but the main point is that better connectivity among nodes is of higher value, and this connectivity is not necessarily restricted to connectivity among nodes at a specific distance.

5.2.3 Summing Up: Ego Network Quality

We can define the quality of the neighbourhood at distance d (denoted by Q_d^i), which is the Knowledge Potential of the nodes at distance d from node i weighted by the Local Connectivity of the neighborhood at distance d. If we are looking only at the direct neighbourhood of node i, we can write:

$$Q_1^i = KP_1^i LC_1^i \tag{5.5}$$

This expression reflects the knowledge level of direct neighbours, weighted by the interaction among these neighbours.[8] However, as noted earlier, the level of knowledge attained from direct neighbours is enhanced by the level of knowledge these neighbours attain from their individual networks. Therefore, we augment our quality measure with the same connectivity-weighted knowledge levels of further indirect neighbourhoods, using the distance weights as introduced before. According to this, the index of Ego Network Quality is defined as follows, which comes back to our starting definition in Eq. 5.1:

$$ENQ^i = \sum\nolimits_d W_d Q_d^i = \sum\nolimits_d W_d KP_d^i LC_d^i \tag{5.6}$$

5.3 Structural and Node Characteristics in ENQ

Consider the special case when distance weighting is applied in the ENQ formula but differences in node characteristics (i.e., knowledge levels) are disregarded, which is the assumption behind traditional network position measures. In this section we show that in this special case ENQ measures network position in a way very much similar to the intuition behind eigenvector centrality (Bonacich 1972, 2007).

For the sake of simplicity we normalize knowledge levels to unity, so we have $k_i = 1$ for all i, therefore $KP_d^i = \sum\nolimits_{j:r_{ij}=d} k_j = N_d^i$.[9] Now we have the following formula for ENQ:

$$ENQ^i = \sum\nolimits_{d=1}^{M-1} W_d \left(\sum\nolimits_{j:r_{ij}=d-1} \sum\nolimits_{k:r_{ik}=d} a_{jk} + \frac{\sum\nolimits_{j:r_{ij}=d} \sum\nolimits_{k:r_{ik}=d} a_{jk}}{2} \right) \tag{5.7}$$

The first term in the parenthesis counts the number of links connecting nodes at distance d with nodes at distance $d - 1$, whereas the second term gives the number

[8] Note, that the weight for $d = 1$ is unity by definition.

[9] It is easy to see that using (identical) knowledge levels different from unity would change the results by a multiplicative constant compared to the situation with the normalized levels.

of links between nodes at distance d. In Eq. 5.8 we add the expression $\sum_{j,r_{ij}=d}\sum_{k,r_{ik}=d+1}a_{jk}$ to those in the parenthesis, and multiply the second term by 2. We will denote this new expression by $DIST^i$:

$$DIST^i = \left(\sum_{d=1}^{M-1} W_d \sum_{j:r_{ij}=d-1}\sum_{k:r_{ik}=d}a_{jk} + \sum_{j,r_{ij}=d}\sum_{k,r_{ik}=d+1}a_{jk} + \sum_{j:r_{ij}=d}\sum_{k:r_{ik}=d}a_{jk}\right) \tag{5.8}$$

This expression simply counts the weighted number of links between nodes at distance $d - 1$ and d (first expression in the parenthesis), the weighted number of links between nodes at distance d and $d + 1$ (second expression in the parentheses) and as the result of the multiplication double-counts the links among nodes at distance d (third expression in the parentheses). In other words, after this modification the number in the parenthesis gives the (weighted) sum of links which nodes at distance d have (double counting the links within the neighbourhood at distance d as these links belong to two nodes in this neighbourhood), which is simply the sum of (weighted) degrees of nodes at distance d. Using all this, we can write

$$DIST^i = \sum_{d=1}^{M-1} W_d \left(\sum_{j:r_{ij}=d}\sum_{k}a_{jk}\right) = \sum_{d=1}^{M-1} W_d \sum_{j:r_{ij}=d} DEG_j \tag{5.9}$$

where DEG_j is the degree of node j. This expansion, however, results in no substantial change in the ENQ measure. The original measure counts every link once, whereas the modified version counts every link twice. If the adjacency matrix is symmetric, which we assume, this modification then means a simple multiplication by 2 on the level of the overall index. To sum up, the expression in Eq. 5.9 is the distance-weighted sum of degrees in the network. On the other hand, the previous reasoning clearly shows that this measure is twice the ENQ measure in this special case with no knowledge weights at the nodes:

$$DIST^i = 2ENQ^i \tag{5.10}$$

On the other hand, the expression in Eq. 5.9 has some similarity with eigenvector centrality (Bonacich 1972, 2007), which also reflects a distance-weighted sum of degrees in a network, although it uses a recursive definition with implicit exponential weights leading to an eigenvector problem.[10] This means that our ENQ index,

[10] Eigenvector centrality is defined by the following recursive concept. Let x_i denote the centrality of node i and let this centrality be determined by the cetralities of adjacent nodes: $x_i = 1/\lambda \sum_j a_{ij}x_j$.

Written for all nodes we end up with the matrix equation $\mathbf{x} = 1/\lambda\mathbf{Ax}$, which is an eigenvector problem. The eigenvector corresponding to the largest eigenvalue (which rules out x_i s of opposite signs) gives the required centrality measures. It is easy to see that this recursive definition discounts the centrality value of distant nodes exponentially (given that $\lambda > 1$). In addition, if we consider the partners' centrality indices identical, the centrality index of node i is proportional to its degree, whereas relaxing the assumption of identical centrality measures in the direct

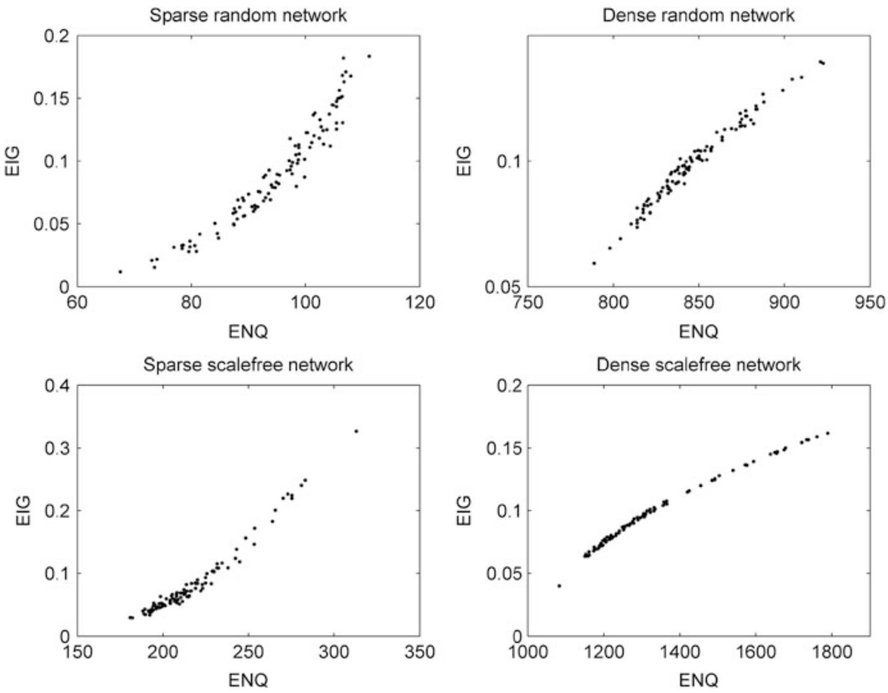


Fig. 5.1 Correlations of ENQ (*horizontal axes*) and eigenvector centrality (*vertical axes*) in different networks, under the assumption of unit knowledge weights for all nodes

when knowledge levels are homogenous, reflects similar properties to eigenvector centrality, which is a comprehensive measure of network position taking into account the whole structure around a given node from its immediate neighbourhood to farther parts of the network.

Figure 5.1 illustrates this point on a random graph (based on the Erdős-Rényi (1959) algorithm) and on a scale free graph (based on the preferential attachment mechanism proposed by Barabási and Albert (1999)), both of size 100. Both graphs are represented for a sparse and a dense case.[11] As it is clear from the figure, there is a tight positive correlation between ENQ and eigenvector centrality.[12]

neighborhood but retaining it in the consecutive ones, the index for node *i* turns out to be the sum of degrees of direct partners, and so on. This is not to prove that the expression in Eq. 5.9 and eigenvector centrality are the same, but the underlying concepts have common characteristics.

[11] The sparse network is simulated at 5 % density and the dense network at 30 % density. These two values were picked as follows. Density 5 % is the threshold approximately at which random networks of size 100 become connected, so that the whole network is likely to be connected at 5 % density. The 30 % density value corresponds to the density of interregional co-patenting networks as presented in Sebestyén and Varga (2013).

[12] Note that these illustrations are created for the case when ENQ is calculated with linear distance weights and homogenous knowledge levels across the nodes. Linear distance weights are chosen

The correlation is stronger for the denser networks and for the scalefree structures (with the dense scalefree case being almost deterministic).

The arguments and results presented above show that our measure captures the position of the nodes in the network, which results from the structure of the network around specific nodes, taking into account both the direct and indirect neighbourhoods. However, the ENQ index is also capable of taking into account node-specific characteristics, captured by knowledge potential in our context, which is not part of the traditional measures of network position (especially of eigenvector centrality analysed here).

5.4 ENQ with Different Weighting Methods: A Systematic Comparative Analysis

ENQ involves three weighting dimensions. The first is weighting at the node level and is captured by the knowledge of each connected node. That is, we do not consider nodes as homogenous with respect to their inherent characteristics apart from their position in the network, but take their heterogeneity implicitly into account. The two other dimensions of weighting correspond to the structural properties of the network. The second dimension is weighting by distance and the third one is weighting by the local structure captured by the Local Connectivity element in the ENQ measure. Though the first dimension (i.e., knowledge level) is taken as exogenous the other two dimensions are tightly related to the structural properties of the network around the specific nodes. In this section we study the impacts of different approaches measuring these structural properties on ENQ.

5.4.1 Analytical Framework: The Modified Preferential Attachment Model

The analyses to be presented are placed in the framework of a modified version of the well-known preferential attachment model, originally proposed by Barabási and Albert (1999). The reason for using this framework is twofold. First, we can simulate the behaviour of the ENQ index under different network structures and second, it is possible to define those structures and the corresponding characteristics of the ENQ index which seem to be relevant in the context of knowledge networks.

The model used offers the opportunity to build networks which range from the random topology (according to the Erdős-Rényi algorithm) to centralized topologies (with a connected core and peripheral actors tied only to the core with sparse or

because in this case the weight of GE in the ENQ index is the highest (see later), thus the differences in the GE element are captured the best in this case.

no connection within the periphery). Between the two extremes the model reproduces the properties of the preferential attachment model.[13] The detailed description of the model can be found in Sebestyén (2011), and a brief outline is presented in the Appendix.

Figure 5.2 illustrates the change in the average clustering coefficient and the average path length (distance) as we move from random to centralized graphs along the modified preferential attachment model.[14] On the horizontal axis we move from random to centralized structures, the meaning of the scale moving from 0 to 1 is described in the Appendix. On the vertical axis, the two measures are normalized, with the values corresponding to the random structure being unity. Average path length measures the average distance of the nodes from each other in the network and the average clustering coefficient captures that on average how well connected the direct partners of the nodes are.[15] As the figure shows, there is a monotonous decrease in average path length and a monotonous increase in the average clustering coefficient.

The decrease in path lengths stems from the centralized nature of the networks on the right wing of the figure: path lengths are the shortest possible in this topology, and any departure from this, including the random structure, results in an increase in the path lengths.[16] However, the change in average distances is not so marked due to the fact that random networks are already characterized by short average path lengths, so there is relatively less room for a further decrease. The considerable increase in the clustering coefficient comes from two facts. First, the clustering coefficient in a random network is typically low as the randomness of ties

[13] It is important to highlight that the proposed model is not capable of capturing all characteristics of the empirical knowledge networks one encounters in practice. For example, the networks generated are characterized by one core group and multiple cores are not accounted for. Also, hierarchical structures often found in real networks are not present in the simulated structures. The goal, however, is not to provide a network model which generates topologies that precisely reflects empirical ones, but to establish a relatively simple method to span a reasonably wide range of network structures and to test the behaviour of the ENQ index under these structures. On the other hand, the choice of the underlying network model seems reasonable as it comes up with topologies reflecting those characteristics often found in reality. First, it accounts for preferential attachment in its intermediate range (which is found to be a robust driving force behind real world networks) and second, it also accounts for centralized structures with connected cores and marked periphery which is a typical pattern in knowledge networks. Additionally, although less relevant from an empirical point of view, but as an extreme case the random topology is accounted for.

[14] The network size is 100 and the density is 28 % in this specific illustration (corresponding to the empirical network analyzed by Sebestyén and Varga (2013)) but further simulations show that the tendency visible in the figure is robust across different network sizes and densities.

[15] Average path length is the average of the shortest paths measured between every pair of nodes in the network. The clustering coefficient measures the density of the direct neighborhood of a node and the average clustering coefficient is simply the mean of these local coefficients (see Wasserman and Faust (1994) for details).

[16] Take the star network as an extreme example. In this topology average path length is somewhat smaller than two as the majority of the nodes are at distance two from all other nodes except from the central one and the central node is at distance one from all other nodes.

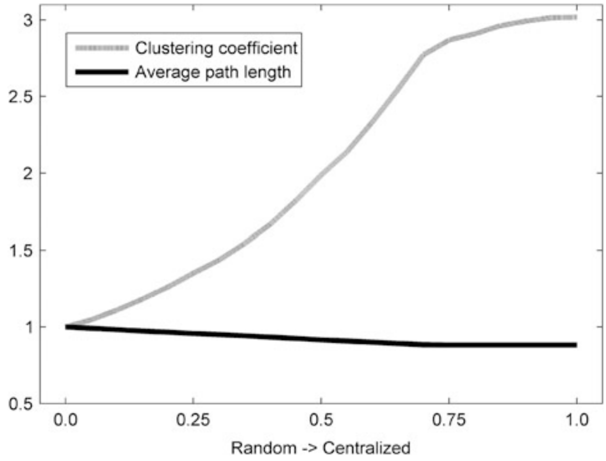


Fig. 5.2 Average distance and clustering coefficient in the modified preferential attachment model. Both measures are normalized to the value corresponding to the random network being unity

leaves no room for significant local densities. Second, in this specific model, for more centralized structures we have a densely connected core and all other nodes outside the core are connected to this group of nodes. The result is that the nodes in the core have relatively low clustering as their neighbours outside the core are not linked to each other, but the majority of the nodes outside the core have high clustering as their neighbours in the core are densely connected to each other. Average clustering rises because the high clustering of the many peripheral actors in the centralized structures dominates the low clustering of the few central ones.

Using the modified preferential attachment algorithm as a framework for our further analyses we can model different network structures along a well-defined interval ranging from random networks to centralized topologies. The random network on one extreme of the model can serve as a natural (and widely used) reference point for examining network structures, whereas the model moves towards more centralized topologies through scalefree structures which are empirically more relevant. See for example Barabási (2003) or Csermely (2006) for a general discussion.

As an empirical reference for the modified preferential attachment algorithm, we employ the patent-co-inventorship network of European NUTS2 regions, reported by Sebestyén and Varga (2013). This empirical network is situated approximately between a value close to 0.0 and around 0.6 on the horizontal axis of Fig. 5.2.[17]

[17] The empirical network has an average path length of 1.78 whereas the corresponding random network (with the same size and density) has a path length of 1.72 (not significantly different from the empirical number). As a consequence, with regards to the path lengths, this empirical network can be positioned on the left hand side of Fig. 5.2. The clustering coefficient of the empirical network is 0.66, 2.35 times higher than the coefficient of 0.28 characterizing the corresponding random network, thus from the clustering point of view, the empirical network is situated around 0.6 on the horizontal axis of Fig. 5.2. This shows that the network model can reflect empirically relevant topologies throughout its interval from random to centralized structures.

5.4.2 Accounting for Distance-Weighting in Random and Scalefree Structures

Although it is straightforward to state that the properties of the network located farther away from a given node is of less importance for that specific node, the question remains that exactly how much less this importance is. Technically speaking, the decay function for the distance weights W(d) must be determined. In general the choice of the decay function seems to be arbitrary. In this section we present some analysis to reveal how the ENQ index behaves under different weighting schemes and network structures. The aim is to provide a background for empirical analyses by giving a conceptual description of the ENQ index when using it for different network settings. The vehicle for analysing different structural settings is the modified preferential attachment model, and for the decay function we consider three basic and straightforward cases: the linear, the hyperbolic and the exponential decay.

Using linear weights we assume that moving one step farther away in the network, the absolute loss of information or knowledge is the same from neighborhood to neighborhood. In linear weighting we use the following formula:

$$W(d) = \frac{M - d}{M - 1} \tag{5.11}$$

where M is the size of the network. This specification has the property that at distance $d = 1$ its value is unity, whereas its value descends to zero when distance would cross the boundaries of the network, namely at $d = M$. In other words, the farthest possible node (at distance $d = M - 1$) has a small but positive weight. This form of the linear decay has different decay speeds for different network sizes, but rules out negative weights.

A hyperbolic decay can be defined simply as

$$W(d) = \frac{1}{d} \tag{5.12}$$

It is easy to prove that it satisfies the requirement that at distance 1 its value is unity. On the other hand, this option is independent of network size and gives positive values for any distance. However, due to the hyperbolically decreasing weights, in larger networks this method implies low values for Global Embeddedness especially if dense and highly knowledgeable nodes lay at relatively high distance from the node in question.

Exponential weighting is defined as

$$W(d) = e^{1-d} \tag{5.13}$$

It has similar characteristics as the hyperbolic decay, but the pace is faster. Exponential weighting corresponds to the situation when the information or knowledge lost from consecutive neighbourhoods decreases by a constant percentage.

Figure 5.3 depicts the results of a simulation experiment where we built networks using the modified preferential attachment model proposed before, moving in 21 steps from totally random networks to totally centralized ones. In all steps we generated 100 networks and then calculated for each network the ENQ values with linear, hyperbolic and exponential weights in the distance decay function. The lines in the figure represent average ENQ values for a given network structure and for different decay functions.[18]

It is clear from the picture that the different weighting methods end up with different ENQ values. The higher weights are given to more distant neighbourhoods, the higher will be the Global Embeddedness value and therefore the overall ENQ index. This is reflected by the figure: linear decay gives the highest weights from $d = 2$ on and this decay function gives the highest ENQ values. Hiperbolic decay lies in the middle, while exponential decay is the fastest leading to the smallest ENQ values.

In addition to this tendency, we also observe that ENQ values typically increase for higher centralization. This latter tendency can be easily explained by the interplay of two effects related to the changing characteristics of the network topology in the modified preferential attachment model. Decreasing average distances (see Fig. 5.2) lead to a higher ENQ ceteris paribus through the increase of the GE element (as path lengths shorten, initially more distant neighbourhoods come closer, thus the same KP and LC values are less discounted due to distance), irrespective of the choice of the decay function in the distance weights. On the other hand, the increasingly dense local neighbourhoods (reflected by the increasing clustering coefficient in Fig. 5.2) lead to higher LC values, not only in the direct neighbourhoods but also in more distant ones. These two effects reinforce each other as we move towards centralized structures, leading to higher ENQ indices for the centralized topologies.

It is also clear from the figure that the differences between the ENQ values calculated according to the different decay functions remain constant throughout the interval from random to centralized networks. This is explained by the fact that, as evidenced by Fig. 5.2, the increase in clustering (which is reflected by the LC element of ENQ) dominates the decrease in average distances. Therefore the choice of the distance decay function, is less relevant in the case of those network structures which are accounted for in the preferential attachment model. On the other hand, this constant difference means that the choice of the decay function has

[18] The figure illustrates the results of a simulation with networks of size 100 and density 30 %. For all structures 100 independent runs were executed and then averaged. The results shown are robust for networks with different sizes and densities.

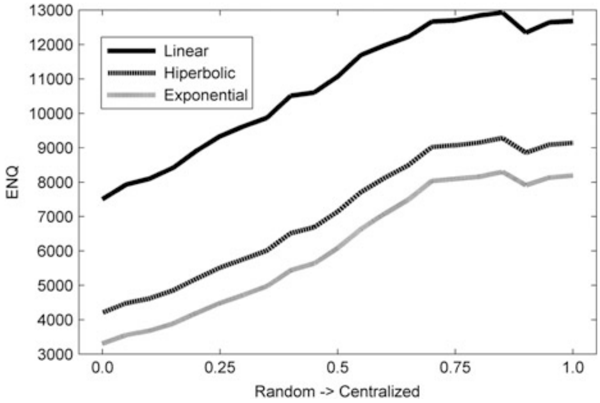


Fig. 5.3 Average ENQ values under different distance weighting schemes

only a shift effect on the ENQ in this specific interval between random and centralized structures.

These results with different distance weighting schemes and network structures indicate that the ENQ index is able to capture relevant structural differences, whereas (apart from the trivial shift effect), the choice of the decay function does not result in any bias when applying the index to different structural settings, meaning that it does not overestimate or underestimate ENQ under any weighting schemes compared to other topologies or weightings.[19]

5.4.3 Accounting for Structural Holes: Modifying the Structural Weights

As highlighted before, besides the weights attributed to indirect neighbourhoods at different distances, the other important weighting in the model is represented by the Local Connectivity value which weights the sum of knowledge levels at a given neighbourhood with the connectedness of that neighbourhood: the more connected the neighbourhood, the higher the weight. This kind of formulation seems intuitive and it relates to the notion of social capital as defined by Coleman (1986) who emphasized the role of cohesion, i.e. closed local structures as enhancing individual action. The more connected an individual's neighbourhood is, the more social capital he or she has and the better for him or her. On the other hand, Burt (1992) challenged this view emphasizing the role of structural holes in individual performance and as a source of social capital. Structural holes are present in a network

[19] Given a specific structural setting along the horizontal axis of Fig. 5.3 between random and centralized topologies, moving one step in either direction resulting in a different structural setting leads to the same absolute change in the ENQ index irrespective of the choice of the decay function.

where the Coleman-type cohesion is missing. In other words, nodes in structural holes fulfil the role of a gatekeeper or information broker among different groups. In this view, a node's position in the network is efficient, if its neighbourhood is not fully connected but consists of more, otherwise unconnected groups.

Although the concept of structural holes is intuitively appealing, its measurement leaves open questions. Once the analysis tries to capture the number of unconnected cliques in a neighbourhood, one immediately finds the problem of determining the threshold in connectivity from which a group of nodes are considered as a clique (and vice versa, the threshold from which groups are referred as distinct). There are several methods established in social network analysis from positional analysis to blockmodeling (see for example Wasserman and Faust 1994) which offer solutions to this question but all retain the crucial cornerstone of determining the threshold exogenously.

Linked to the previous problem, if one looks behind these concepts, it becomes clear that structural holes and cohesion (connectivity) are not independent structural characteristics. If the neighbourhood of a node is densely connected (cohesive), the chance for finding many unconnected groups in this neighbourhood is small. Conversely, if there are unconnected groups, the density must be lower.[20] As a result, we cannot construct independent metrics for the two concepts.

Taken all this together, the ENQ index developed in this paper provides a flexible framework to include these concepts. Although our definition of Local Connectivity in Eq. 5.4 reflects the cohesion approach a la Coleman and disregards the importance of structural holes, in a general sense we may define Local Structure as a weighting factor for Knowledge Potentials of the specific neighbourhoods and let this LS term account for different approaches depending on the actual investigation. In other words, Local Structure is a weighting factor capturing structural features of neighbourhoods, but these structural characteristics can be defined in different ways. Previously we defined and used Local Connectivity as a possible way to specify the Local Structure weight. In what follows, we implement an additional weighting for Local Structure in order to capture not only cohesion but also structural holes or the combination of these features.

Taken into account the previously mentioned problems of measuring structural holes, we propose a simple approach which captures the basic intuition behind the concept and provides an easy way to link this measurement to cohesion. We use a strict threshold for defining cliques in a neighbourhood, namely the number of connected components in the subgraph defined by the nodes in a given neighbourhood. This approach, although the threshold also comes from an exogenous source, can be labelled as a baseline solution to identifying cliques in the network.[21]

---

[20] See Fig. 5.5 and the explanation in the Appendix.

[21] It is known from graph theory that the number of connected components in a graph is given by the multiplicity of the zero eigenvalues of the Laplacian matrix of the graph. The Laplacian matrix is simply the difference of the diagonal degree matrix (with node degrees on the diagonal) and the

If we define CC_d^i as the number of Connected Components (or in other words the number of unconnected groups) in the neighbourhood at distance d from node i, the ENQ index can then be reformulated with structural holes being the weighting factor of Knowledge Potential in addition to the distance weights:

$$ENQ^i = \sum\nolimits_d W_d Q_d^i = \sum\nolimits_d W_d KP_d^i CC_d^i \tag{5.14}$$

This formulation though, puts the index to the other extreme, taking into account only structural holes and disregarding cohesion. On the other hand, it is also true that both Local Connectivity and Connected Components take a very strict view and measurement of the phenomena they intend to capture. Local Connectivity captures simply the intensity of cooperation by counting the links in different neighbourhoods, while Connected Components restricts the counting only to totally unconnected groups. However, by combining the two approaches, ENQ can reflect a more refined picture about the structure of local neighbourhoods. Let's redefine ENQ with the product of Local Connectivity and Connected Components as the weighting factor of Knowledge Potentials (the Local Structure component, defined before):

$$ENQ^i = \sum\nolimits_d W_d Q_d^i = \sum\nolimits_d W_d KP_d^i CC_d^i LC_d^i \tag{5.15}$$

This formulation refines the two extreme cases by positively weighting diverse groups and at the same time the strength of connectivity. In addition to the fact that the empirical literature is not conclusive on the relevance of the two approaches,[22] it is intuitively reasonable to think that an optimal network position combines these two structural features: too much cohesion is not good as the advantage of access to diverse information and knowledge is lost but the lack of cohesion can also be disadvantageous as the connections in the neighbourhood can contribute to learning and knowledge creation through fast knowledge transfers, collective learning and recombination of ideas. The trade-off between the two concepts (referred to before and detailed in the Appendix) provides a natural way to combine the two effects as

adjacency matrix of a graph. (see e.g. Godsil and Royle 2001). Taking then the node-generated subgraphs spanned by the nodes at specific distances from the node in question and using the Laplacian method, we can easily calculate the number of connected components, although closed formula cannot be given.

[22] Although many of the results in this field show that a position in structural holes contribute to better performance in a diversity of fields (e.g. Hopp et al. (2010), Kretschmer (2004), Donckels and Lambrecht (1997), Zaheer and Bell (2005), Powell et al. (1999), Tsai (2001), Burt et al. (2000), Burton et al. (2010)), there is still evidence on the opposite (Salmenkaita 2004; Cross and Cummings 2004). Rumsey-Wairepo (2006) argues that the two structural settings are complementary to each other rather than substitutes in explaining performance. In general, it seems that different structural dimensions can be important for different networks. When information flows and power is important, structural holes indeed provide better position, however, as in our case, if knowledge production is in the focus, exclusion resulting from structural holes may be harmful and cohesiveness meaning better interaction may have positive contribution.

the multiplication of Connected Components and Local Connectivity in Eq. 5.15 attach higher weights to structures which lay in between neighbourhoods with extreme structural holes and extreme connectivity.

In order to evaluate the modification of the structural weights in the ENQ index, we executed similar simulations with the modified preferential attachment model in the background as for the distance weights. With these simulations we can gain insight into the behaviour of the ENQ index under different structural settings and weighting schemes for local neighbourhood structures. The results are summarized in Fig. 5.4. The simulations were executed for sparse and dense networks of size 500, and in all cases hyperbolic distance weighting was used.[23] The figure uses a logarithmic scale on the vertical axis in order to keep the tendencies visible.

The results on the figure show that the cases for dense and sparse networks are qualitatively the same, but the observed tendencies are sharper for the dense one. The solid lines show the same path for ENQ as that for the hyperbolic distance weights in Fig. 5.3. These values are obtained if we consider Local Connectivity (or putting differently the density of neighborhoods or cohesion) as a weighting factor for Knowledge Potential. The dashed lines in comparison are obtained for the case when connected components (structural holes) are used as a weighting factor, whereas the dotted lines show the mixed case.

The overall picture shows that the cases when only CC is used as structural weight are significantly lower – this is explained by the fact that CC weights tend to be lower as they count connected groups whereas LC counts links in the networks. Even a small departure from the star topology (where neighbourhoods are disjoint), leads to a sharp decrease in the number of connected components in the neighborhoods. This rule is responsible for the sharp increase in the ENQ values when it includes CC (structural holes and mixed cases in the figure) for the extremely centralized topologies. The second observation is that the mixed and the pure connectivity-weighted case results in similar ENQ values for a large interval of the underlying network structures. This is due to the fact that for this interval the number of unconnected groups tends to be small, thus the additional weighting by CC in the mixed case rarely leads to significant departures from the simple LC-weighted values (this similarity is not present for the highly centralized structures for the reason mentioned before). Additionally, in the dense network, the pure connectivity-weighted and the mixed case result in identical ENQ values as the high density leaves no room for unconnected groups, whereas in the sparse network unconnected groups are more likely to be present.

Note, however, that these results and tendencies mark overall, aggregate features of the ENQ index. Its local, node-specific characteristics are not taken explicitly into account, but it is still true that even if on the aggregate level there is no marked

[23] Further simulations showed that the results are robust for altering the size of the network (the tendencies are better illustrated by larger networks – this is why we used size 500, but are qualitatively the same for smaller networks). Sparse networks mean 5 % density while dense networks 30 % density as before, and for each structure 100 independent simulations were executed and the results averaged.

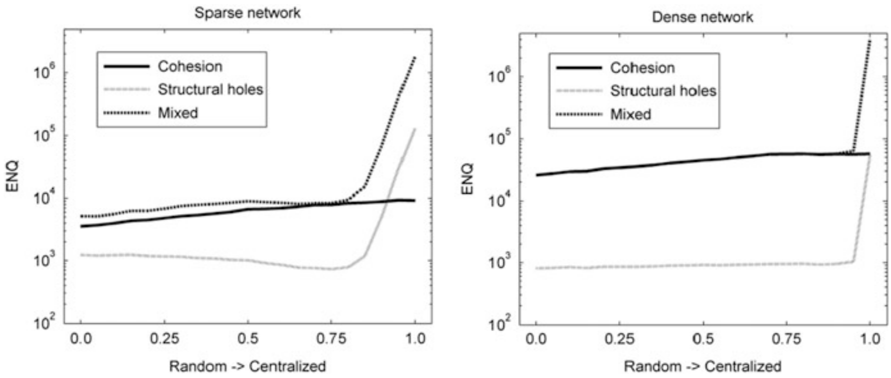


Fig. 5.4 ENQ values under different structural weighting methods

difference between the weighting methods, the local neighbourhoods can differ and node-specific ENQ indices can reflect these differences.

To sum up, we can see that there is no significant difference between the original (connectivity) weighting method and the augmented one (where both connectivity and structural holes are taken into account) in the networks characterized by the modified preferential attachment model, except for extremely centralized (star-like) topologies. On the other hand, whether structural holes or cohesion, or their combination provide an efficient network position in a general sense is still an open question and requires an empirical assessment. Our ENQ measure can accommodate both cases, and is flexible to account for different structural weighting methods. As a result, it can be used to test empirically the effect of structural settings on the efficiency of network positions.

5.5 Robustness of Different Weighting Schemes in ENQ: An Empirical Investigation with European Co-patenting Networks

Though simulations can gain insights into some important properties regarding cohesion, structural holes and distance in ENQ, the relative importance of the two structural settings and distance weighting schemes remains an open issue for empirical studies. In general, different environments and goals might favour different structural and distance decay settings. In this section a short empirical investigation is carried out in this respect. We use the ENQ index of co-patenting networks to explain R&D productivity in European regions and investigate if there is a variation in the extent to which different weighting methods affect regression results.

The analysis is based on the knowledge production function initially specified by Romer (1990) and parameterized by Jones (1995). In the interpretation of the parameters we follow Varga (2006). In this specification technological change is associated with contemporary R&D efforts and previously accumulated knowledge. We assume that the efficiency of R&D efforts is positively related to the quality of interregional knowledge networks measured by ENQ. We are interested in the explanatory power of the model specified with different structural weighting methods in the ENQ index.

In the empirical specification we follow Varga (2000) and Varga et al. (2013) and set out the knowledge output in region i (denoted by K_i) in function of R&D expenditures in that region (RD_i), national knowledge stocks (KS_N) and local agglomeration (AG_i):

$$\log(K_i) = \alpha_0 + \alpha_1 \log(RD_i) + \alpha_2 \log(KS_N) + \alpha_3 \log(AG_i) + \varepsilon_i \tag{5.16}$$

Then, we relate research productivity, measured by $\alpha_{1,i}$ in region i (the parameter of the R&D variable in Eq. 5.16) to the quality of the interregional knowledge network:

$$\alpha_{1,i} = \beta_0 + \beta_1 ENQ_i \tag{5.17}$$

Substituting Eq. 5.17 into Eq. 5.16 results in the following equation to be estimated:

$$\begin{aligned}\log(K_i) = \alpha_0 + \beta_0 \log(RD_i) + \beta_1 \log(RD_i) ENQ_i + \alpha_2 \log(KS_N) \\ + \alpha_3 \log(AG_i) + \varepsilon_i\end{aligned} \tag{5.18}$$

The analysis is based on a sample of 189 European regions (a mix of NUTS2 and NUTS1 regions) for which information was complete enough for our purposes. The network under question is a network of these regions and the links are patent co-inventorships between 1998 and 2002. By definition, the network is a weighted network with the number of patents co-invented by inventors from two regions being the weights of a link. The network was built using data from the REGPAT database of OECD (2009). From this network we calculated ENQ indices with different structural weighting methods with patent stocks playing the role of knowledge levels (k_i) in the calculation of ENQ. The knowledge output on the left hand side is proxied by new patents generated in 2002 (the end of the aggregation period for the knowledge network). Research effort is measured by annual R&D expenditures by the regions in 2000 (the time lag is included in order to account for timely effects of research efforts on innovative output). Agglomeration is measured by the size-adjusted location quotient of technology- and knowledge-intensive sectors. The source of the latter data is the Eurostat New Cronos database.[24]

[24] See Varga et al. (2013) and Sebestyén and Varga (2013) for further details on data and methodology.

Table 5.1 Main regression results with different distance and structural weighting schemes

		Structural weights	
		Cohesion	Mixed
Distance weights	Linear	0.2737**	0.0755
		(0.7534)	(0.7482)
	Hyperbolic	0.2570***	0.3366***
		(0.7599)	(0.7571)
	Exponential	0.2418***	0.3687***
		(0.7609)	(0.7607)

Note: Asterisks refer to significance levels (*10 %; **5 %; ***1 %). R-squared values of the corresponding regressions are in parentheses

Detailed estimation results of Eq. 5.18 under different structural and distance weighting methods can be found in Table 5.2 in the Appendix. Table 5.1 below presents the coefficients of the ENQ indices, i.e. the estimated contributions of network position to R&D efficiency (β_1 in Eq. 5.18) and regression fits (in parentheses). With the exception of linear distance weights regression results appear to be robust to the choice of distance and structural weights. Though estimated parameter values in the mixed case are somewhat higher compared to cohesion weighting, this difference does not show up in the respective equations' explanatory powers.

To sum up, as already signalled by the simulation analyses with regards the distance and structural weighting schemes, empirical findings reinforce that the ENQ index is able to robustly capture the position of a node in the network and the specific choice of distance and local neighbourhoods structure weighting are of secondary importance.

5.6 Summary

In this paper we introduced the Ego Network Quality (ENQ) index, which intends to capture the value of knowledge available from a node's immediate and indirect neighbourhood in a given network. The index integrates three aspects of the network position into one comprehensive measure. First, it is based around the concept of Knowledge Potential which sums the value of knowledge available at the neighbours. Second, this Knowledge Potential is weighted by the structure of the neighbourhood on the basis of the assumption that in addition to the individual knowledge level of partners, the structure of how they are related to each other also contributes to the value of knowledge available from one's network. Third, not only direct neighbourhoods are taken into account, but also indirect partners with their knowledge levels and structural characteristics.

Research on the impact of networks on knowledge production either concentrates on the characteristics of networks ("a-spatial studies") or on the characteristics of connected nodes (i.e., knowledge level in "spatial studies") but neither on

Regression Results

Table 5.2 Results of OLS regressions under different distance and structural weighting schemes in the ENQ index

	Local connectivity			Connected components			Mixed structural weights		
	Linear	Hyperbolic	Exponential	Linear	Hyperbolic	Exponential	Linear	Hyperbolic	Exponential
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
const	−1.347**	−0.8783	−0.8009	−1.592**	−1.283**	−1.013*	−1.595**	−1.024*	−0.7979
	(0.557)	(0.5866)	(0.5918)	(0.5484)	(0.5711)	(0.5902)	(0.5486)	(0.5818)	(0.5934)
l_GRD00	0.5163**	0.5035**	0.5125**	0.6846**	0.5596**	0.4804**	0.6905**	0.5075**	0.4745**
	(0.1297)	(0.1076)	(0.1037)	(0.1016)	(0.1194)	(0.1263)	(0.0999)	(0.1144)	(0.1123)
l_PATST00_N	0.1949**	0.1699**	0.1663**	0.2224**	0.2117**	0.2005**	0.2199**	0.1853**	0.1715**
	(0.04196)	(0.04286)	(0.04297)	(0.04026)	(0.04031)	(0.04056)	(0.04044)	(0.04188)	(0.0424)
l_AGGL	1.272**	1.176**	1.159**	1.335**	1.269**	1.196**	1.337**	1.231**	1.171**
	(0.3715)	(0.369)	(0.3687)	(0.3742)	(0.3726)	(0.3723)	(0.3743)	(0.3695)	(0.3684)
l_GRD00*ENQ	0.2737**	0.2570**	0.2418**	0.07624	0.2204*	0.3101**	0.0755	0.3366**	0.3687**
	(0.1311)	(0.08352)	(0.07552)	(0.1037)	(0.1149)	(0.1223)	(0.1124)	(0.1251)	(0.116)
n	189	189	189	189	189	189	189	189	189
R^2	0.7534	0.7599	0.7609	0.7483	0.7525	0.7561	0.7482	0.7571	0.7607
lnL	−277.6	−275.1	−274.7	−279.6	−278	−276.6	−279.6	−276.2	−274.8

Dependent variable is the log of regional patenting activity. The ENQ index in each model is calculated according to the distance and structural weight combination specified in the header. Standard errors are reported in parentheses

*indicates significance at the 10 % level

**indicates significance at the 5 % level both features. It is shown in this chapter that ENQ is an integrated measure of network position and node characteristics.

ENQ splits the network around a given node into consecutive neighbourhoods depending on the distance of other nodes from the node in question, then it weights the knowledge levels (Knowledge Potential) in each neighbourhood with the structure and the distance of the neighbourhood. This chapter focuses on the specific way these weightings are executed. In the case of distance, three possible decay functions (linear, hyperbolic and exponential) were analysed while following the literature on network position and social capital we proposed two distinct ways for structural weighting, namely one which attaches high weights for dense local structures (Local Connectivity) and one which weights structural holes (Connected Components). Led by intuition, we also proposed a weighting scheme combining these features.

Using simulation exercises we demonstrated that the ENQ is able to reflect the structural patterns of networks, without leading to significant bias resulting from the choice of the different weighting methods, especially on a specific interval of network structure which can be remarked as empirically relevant based on some stylized facts of empirical co-patenting networks. It is also demonstrated that the ENQ index can flexibly accommodate different definitions of structural weights, but under the same relevant structural interval of network topologies, this choice is of secondary importance. Empirical findings further reinforce that ENQ is able to robustly capture the position of a node in the network and the specific choice of distance and local neighbourhoods structure weighting are of secondary importance.

However, we must be aware of the limitations of the ENQ index and of this study. First, although we employed a network model which spans a range of different network structures, the analysis is still limited to the structures included in this model, namely from random to centralized topologies. As mentioned before, empirical networks can exhibit more refined characteristics than reproduced by simple models – for example hierarchy, multiple cores, etc. Second, our analysis is restricted to the global behaviour of the ENQ index under different structures and weighting schemes. How the node level characteristics behave under these settings was not tackled in the present paper.

We believe that the proposed measure of the ENQ index, although developed and applied for a specific type of network, namely interregional knowledge networks, is able to bear general acceptance across different fields through its flexibility. The general innovation is that not only the structural characteristics are captured at the local and global levels through accounting for direct and indirect neighbourhoods, but ENQ also accommodates and accounts for individual characteristics of the nodes in these neighbourhoods. Although in this study we labelled these characteristics as knowledge levels (Knowledge Potential), in a general sense any node-specific feature can be substituted here. In addition, the weighting factor for the structural features is also flexible to accommodate any structural property the researcher wishes to emphasize.

Acknowledgements The research leading to this paper has received funding from the Hungarian Academy of Sciences (n° 14121: MTA-PTE Innovation and Economic Growth Research Group) and OTKA (OTKA-K101160). The authors also wish to express their thanks to the useful comments by Nicolas Carayol, Claude Raynaut, Frank van Oort and Mario Maggioni.

Appendix

The Modified Preferential Attachment Model

The model is developed in order to provide a transition from random graphs of the Erdős-Rényi type through scale-free structures to highly centralized networks. The model starts from a network of M nodes connected randomly with average degree D. Then we increase the size of the network step by step from M to N, adding one new node to the network at a time. In each step the new node establishes exactly D links with the existing ones, on the basis of a probabilistic parameter, r. With probability r the new link is attached to the node with the highest degree in the network and with probability $1 - r$ the new link is attached randomly to any existing node. It is easy to see that using this method we have two parameters, namely M and r, which contribute to the scalefree characteristics of the underlying network. If r increases with a given M, the network moves towards a more centralized structure and vice versa. However, if r is zero, we still do not have a random network for an arbitrary M as the growth of the network in the algorithm still contributes to an underlying asymmetric degree distribution (older nodes tend to have more links than younger ones).

On the other hand, modifying M and r jointly, we can set up a one-dimensional interval from 0 to 1 which moves from random graphs to centralized graphs through scalefree networks. At one end of this scale we have $M = N$ and $r = 0$, which is a random graph by definition. Then we gradually increase r and at the same time decrease M. As a result, the network structure resulting from the previously described algorithm departs from being random and becomes more centralized. At the other end of the scale we reach the most centralized structure with $r = 1$ and $M = 1$. Note however, two things. First, we can express this process with one parameter, say z, ranging from 0 to 1. Then we have $r = z$ and $M = z + (1 - z)N$ as inputs to our model and the value of z expresses the position between random and centralized graphs. Second, the extreme case of $z = 1$ is not necessarily the star network as if the degree is higher than one, there is a connected core in the network, but it is true that the size of this core is D and all other nodes are linked only to this core.

The model thus has analogous logic to the Watts-Strogatz model (Watts and Strogatz 1998), with random and star-like topologies on the extremes and scalefree structures in between.

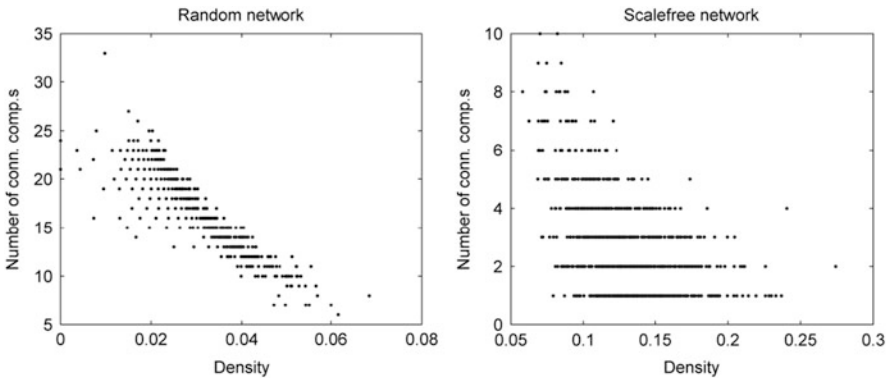


Fig. 5.5 Trade-off between structural holes and density in random and scalefree networks

Trade-Off Between Density and Connected Components

We executed a simple simulation on a random network and on a scalefree one (with the Erdős-Rényi and the Barabási algorithms respectively). The networks in these simulations are sparse networks with global density of 3 %. The sparsity is required from a presentational point of view as the higher the overall density, the more neighborhoods are connected and there is less possibility to find unconnected groups in the neighborhoods.

For both networks we calculated the density and the number of connected components in the direct neighbourhoods of every node (taking these neighbourhoods as subgraphs and calculating density and the connected components on these subgraphs). Figure 5.5 plots the calculated density values and connected components for each node. As connected components must be an integer, the data points are arranged in horizontal lines. The figure clearly shows that there is indeed a trade-off between the two values and this trade-off is stronger in the random network. In the scalefree case we rather observe a missing upper triangle in the diagram, which shows that there are no nodes with dense neighbourhoods *and* many unconnected groups in their neighbourhood, whereas the other three combinations are present. In addition to dense neighbourhoods with few groups and sparse neighbourhoods with many groups, there are nodes the neighbourhoods of which are sparse and characterized by a small number of unconnected groups, which are not present in the random network. The difference between the two network structures stems from the different degree distributions.

References

Anselin L, Varga A, Acs Z (1997) Local geographic spillovers between university research and high technology innovations. J Urban Econ 42:422–448
Barabási AL (2003) Linked: how everything is connected to everything else what it means for business, science and everyday life. Penguing Group, New York
Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286:509–512
Bonacich P (1972) Factoring and weighting approaches to clique identification. J Math Sociol 2:113–120
Bonacich P (2007) Some unique properties of eigenvector centrality. Soc Netw 29:555–564
Burt RS (1992) Structural holes. Harvard University Press, Cambridge, MA
Burt RS, Hogarth RM, Michaud C (2000) The social capital of French and American managers. Organ Sci 11:123–147
Burton P, Yu W, Prybutok V (2010) Social network position and its relationship to performance of IT professionals (Report). Informing Science: the International Journal of an Emerging Transdiscipline. Informing Science Institute. HighBeam Research
Coleman JS (1986) Social theory, social research, and a theory of action. Am J Sociol 91:1309–1335
Cross R, Cummings JN (2004) Tie and network correlates of individual performance in knowledge-intensive work. Acad Manage J 47:928–937
Csermely P (2006) Weak links: stabilizers of complex systems from proteins to social networks. Springer, Berlin
Diez R (2002) Metropolitan innovation systems – a comparison between Barcelona, Stockholm, and Vienna. Int Reg Sci Rev 25:63–85
Donckels R, Lambrecht J (1997) The network position of small businesses: an explanatory model. J Small Bus Manage 35(2):13–26
Erdős P, Rényi A (1959) On random graphs I. Publ Math 6:290–297
Fischer M, Varga A (2002) Technological innovation and interfirm cooperation. An exploratory analysis using survey data from manufacturing firms in the metropolitan region of Vienna. Int J Technol Manage 24:724–742
Godsil C, Royle GF (2001) Algebraic graph theory. Graduate texts in mathematics. Springer, New York
Hopp WJ, Iravani S, Liu F, Stringer MJ (2010) The impact of discussion, awareness, and collaboration network position on research performance of engineering school faculty. Ross School of Business paper no. 1164
Jones C (1995) R&D-based models of economic growth. J Polit Econ 103(4):759–784
Kretschmer H (2004) Author productivity and geodesic distance in bibliographic co-authorship networks, and visibility on the web. Scientometrics 60(3):409–420
Lundvall BA (1992) National systems of innovation. Pinter Publishers, London
Maggioni M, Uberti T (2011) Networks and geography in the economics of knowledge flows. Qual Quant 45(5):1031–1051
Nelson RR (ed) (1993) National innovation systems: a comparative analysis. Oxford University Press, Oxford
OECD (2009) REGPAT database. Oct 2009, Paris
Ponds R, van Oort F, Frenken K (2010) Innovation, spillovers and university–industry collaboration: an extended knowledge production function approach. J Econ Geogr 10:231–255
Powell WW, Koput KW, Smith-Doerr L, Owen-Smith J (1999) Network position and firm performance: organizational returns to collaboration in the biotechnology industry. In: Andrews SB, Knoke D (eds) Networks in and around organizations. JAI Press, Greenwich
Romer PM (1990) Endogenous technological change. J Polit Econ 5(98):S71–S102
Rumsey-Wairepo A (2006) The association between co-authorship network structures and successful academic publishing among higher education scholars. Brigham Young University

Salmenkaita JP (2004) Intangible capital in industrial research: effects of network position on individual inventive productivity. In: Bettis R (ed) Strategy in transition. Blackwell, Malden, pp 220–248

Sebestyén T (2011) Knowledge networks and economic performance. Approaches for modeling and empirical analysis. VDM Verlag, Saarbrücken

Sebestyén T, Varga A (2013) Research productivity and the quality of interregional knowledge networks. Ann Reg Sci. doi:10.1007/s00168-012-0545-x

Tsai W (2001) Knowledge transfer in intraorganizational networks: effects of network position and absorptive capacity on business unit innovation and performance. Acad Manage J 44 (5):996–1004

Varga A (2000) Local academic knowledge transfers and the concentration of economic activity. J Reg Sci 40(2):289–309

Varga A (2006) The spatial dimension of innovation and growth: empirical research methodology and policy analysis. Eur Plann Stud 9:1171–1186

Varga A, Pontikakis D, Chorafakis G (2013) Metropolitan Edison and cosmopolitan Pasteur? Agglomeration and interregional research network effects on European R&D productivity. J Econ Geogr 51(1):155–189

Wasserman S, Faust K (1994) Social network analysis – methods and application. Cambridge University Press, Cambridge, UK

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393 (6684):409–410

Zaheer A, Bell GG (2005) Benefiting from network position: firm capabilities, structural holes and performance. Strateg Manage J 26:809–825

Chapter 6
Network Autocorrelation and Spatial Filtering

Yongwan Chun

Abstract Geographical flows have been frequently modeled with gravity type spatial interaction models. The estimation of spatial interaction models is often achieved with regression techniques, including linear regression and Poisson/negative binomial regression based on the nature of the observations under the independence assumption among observations. Recent studies show, with a development of neighborhood structure among network flows, that geographical flows such as population migration tend to have a significant level of correlation. This phenomenon, called network autocorrelation, leads to a violation of the independence assumption and raises a necessity of a proper modeling method which can account for network autocorrelation. The eigenvector spatial filtering method furnishes a way to incorporate network autocorrelation in linear regression and generalized linear regression. Specifically, the eigenvector spatial filtering method can be utilized to describe positive autocorrelation in Poisson/negative binomial regression, whereas their counterpart auto models are able to describe only negative autocorrelation due to the integrability condition. This chapter discusses different specifications of eigenvector spatial filtering to model network autocorrelation in a spatial interaction modeling framework. These methods are illustrated with applications with interregional commodity flows and interstate migration flows in the U.S.

6.1 Introduction

Interest in the geography of R&D networks has substantially increased over the last two decades, and leads to theoretical and methodological advances in the literature (see Chap. 1 of this volume). Spatial interaction models, which are often used in modeling cross-region R&D collaboration activities, have been improved along

Y. Chun (✉)
University of Texas, Dallas, USA
e-mail: ywchun@utdallas.edu

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_6,

with these methodological advances. Since R&D networks can be considered as geographical flows, spatial interaction models can furnish an analytic method for the geography of R&D networks (see Barber and Scherngell 2013). Recent developed methods show that such geographical flow models can be remarkably improved by taking network autocorrelation into account in their model specifications.

Geographic flows can be referred to as movement of people, goods, or services on the earth surface. Statistical modeling for geographic flows has been commonly conducted in a spatial interaction modeling framework. In a gravity type spatial interaction model, the amounts of geographical flows are explained with three different types of variables, which tend to capture the characteristics of origins, destinations, and impedance between a dyad of an origin and a destination, respectively. Linear regression is frequently utilized in model estimation (e.g., Celik and Guldmann 2007; Greenwood 1985), and Poisson regression is also commonly applied to count type flow data such as population migration (Flowerdew and Aitkin 1982).

Spatial interaction models are further improved by taking the effects of spatial structure into consideration. Curry (1972), Griffith and Jones (1980), and Fotheringham (1981) discuss that parameter estimates of spatial interaction models may be unreliable or biased when the effects of spatial structure are not incorporated. Specifically, an estimate of global distance decay is likely to be biased due to its model misspecification in which localized spatial structure effects cannot be distinguished from the global distance decay effect. For example, when multiple destinations are closely located with each other, competition among the destinations may occur and accordingly each of them may have less inflow than ones without competition. That is, localized spatial arrangement affects decisions on a geographic space. One methodological improvement is achieved by introducing a variable capturing spatial structure effects in its model specification, which often has a form of accessibility measure (e.g., Kwan 1998) This approach includes competing destination models (Fotheringham 1983) and intervening opportunity models (Stouffer 1960), which capture spatial structure effects among origins and destinations, respectively. Studies show that spatial interaction models can further be improved by introducing these two effects simultaneously (e.g., Chun et al. 2012).

Recent research has developed model specifications to explicitly incorporate dependence structure among observations in spatial interaction models (Chun 2008). These model specifications include stochastic terms that are applicable to capture a dependence structure embedded in geographic flows. The dependence structure among flows is referred to as network autocorrelation (Black 1992). Studies (e.g., Chun 2008; Griffith 2009; Fischer and LeSage 2010; LeSage and Pace 2008) show that incorporation of network autocorrelation remarkably improves spatial interaction models. While these extended spatial interaction models can be specified in the spatial autoregressive model framework (Chun et al. 2012; LeSage and Pace 2008), the eigenvector spatial filtering method furnishes a flexible way to account for network autocorrelation in linear and

Poisson regression model specifications (Chun 2008; Griffith and Chun 2013; Fischer and Griffith 2008; Patuelli et al. 2011).

This chapter discusses how spatial interaction models can be improved by accounting for network autocorrelation. Specifically, it shows how eigenvector spatial filtering can be extended to accommodate network autocorrelation. The rest of this chapter is organized as follows. Section 6.2 presents how network neighborhood structure can be specified. Section 6.3 describes the eigenvector spatial filtering method to incorporate network autocorrelation in spatial interaction models. Section 6.4 illustrates the proposed method with two empirical data analyses. The final section presents conclusions and discussion.

6.2 Network Dependence Structure

Network autocorrelation can be defined as correlation among values in one variable which is attached to network flows. In an example of interregional migration flows, a network link is defined as a direct connection between an origin and a destination (which are considered as nodes), and the number of migrants from the origin and to the destination can be an attached variable. Investigations concern how observations attached to network flows are associated in a given network structure (i.e., similar or dissimilar tendency). This requires an operational framework to define network neighbors as spatial neighbors are used for a spatial autocorrelation measure such as Moran's *I*. That is, Moran's *I* can be utilized to measure network autocorrelation with a defined network neighbor structure. This can be implemented with a matrix, which is called network weights matrix here. Each element of a network weights matrix contains a non-zero value for network neighbors and zero otherwise. Generally, a network weights matrix has a larger dimension than a spatial weights matrix. For example, while a spatial weights matrix has n-by-n dimension for n regions in a study area, a network weights matrix can have n^2-by-n^2 dimension for n^2 flows among the n regions.

It is important to determine a structure in which the values of network flows are considered to be associated with each other. Chun and Griffith (2011) show that a network weights matrix can be generated from a spatial weights matrix as follows:

$$\mathbf{B}^{\mathrm{N}} = \mathbf{I} \bigotimes \mathbf{B} \tag{6.1}$$

$$\mathbf{B}^{\mathrm{N}} = \mathbf{B} \bigotimes \mathbf{I} \tag{6.2}$$

$$\mathbf{B}^{\mathrm{N}} = \mathbf{B} \bigoplus \mathbf{B} = \mathbf{B} \bigotimes \mathbf{I} + \mathbf{I} \bigotimes \mathbf{B} \tag{6.3}$$

$$\mathbf{B}^{\mathrm{N}} = \mathbf{B} \bigotimes \mathbf{B} \tag{6.4}$$

where $\mathbf{B}^{\mathrm{N}}$ is an n^2-by-n^2 binary network weights matrix, $\mathbf{B}$ is an n-by-n binary spatial weight matrix (e.g., 1 when spatial units share a boundary; otherwise 0), $\mathbf{I}$ is an identify matrix with n-by-n dimension, $\bigotimes$ denotes Kronecker product, and $\bigoplus$ denotes Kronecker sum. An extended network weight matrix can be generated with matrix addition between Eqs. 6.3 and 6.4 as:

$$\mathbf{B}^{\mathrm{N}} = \mathbf{B} \bigoplus \mathbf{B} + \mathbf{B} \bigotimes \mathbf{B}. \tag{6.5}$$

In Eq. 6.1, network flows whose origins are same and whose destinations are spatially neighbors are considered as network neighbors to each other. In Eq. 6.2, network flows with same destinations and spatially neighbored origins are considered as network neighbors. These two network weights matrices reflect the effects of spatial structure around destinations and origins, respectively. Chun (2008) discusses competing destination models and intervening opportunity models as the rationales for these two types of network weights matrix in the context of population migration. The third type network weight matrix can be constructed with the sum of the two network weights matrices in Eqs. 6.1 and 6.2, reflecting the spatial structure effects around both destinations and origins. In the network weight matrix in Eq. 6.4, a network flow is associated to other network flows from its origin's spatial neighbors to its destination's spatial neighbors.

The last type of network weights matrix in Eq. 6.5 can be generated by adding the network weights matrices in Eqs. 6.3 and 6.4. In this network weights matrix, network flow is associated to all network flows that possibly occur between spatially neighboring origins and spatially neighboring destinations including its origin and destination. Figure 6.1 illustrates network dependence structures with Eqs. 6.3 and 6.4. The solid line network flow is associated to the 12 dotted line network flows in Fig. 6.1a and 36 dotted line network flows in Fig. 6.1b. In a network weight matrix with Eq. 6.5, these 48 network flows are considered to be associated with the solid line (Fig. 6.1a, b). A network weights matrix based on Eqs. 6.3 or 6.4 has been frequently used in studies (e.g., Chun 2008; Griffith 2009; Fischer and Griffith 2008; Mitze 2012). However, a network weights matrix based on Eq. 6.5 has not yet been used. This weights matrix, in which all network flows between spatially neighboring origins and destinations are considered, may allow one to reflect a comprehensive network dependence structure.

6.3 Spatial Filtering in Spatial Interaction Models

Spatial interaction models have been one of most commonly used methods to model interregional flows. A simple gravity type spatial interaction model can be expressed as:

$$F_{ij} = k \cdot P_i^{\beta_O} \cdot P_j^{\beta_D} \cdot \exp\left(\beta_d \cdot d_{ij}\right), \qquad i, j = 1, \ldots, n \tag{6.6}$$

where F_{ij} is flow from i to j, P_i and P_j are population at i and j, respectively, and d_{ij} is

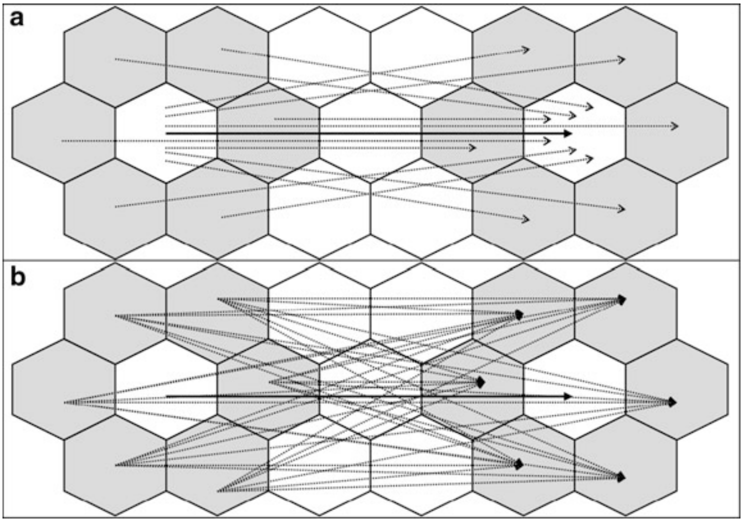


Fig. 6.1 Network dependence structures: (**a**) network flows from spatial neighbors of an origin to a same destination or network flows from a same origin to spatial neighbors of a destination, and (**b**) network flows from spatiallly neighboring origins to spatially neighboring destinations

the distance between i and j. The parameters ($k, \beta_O, \beta_D, \beta_d$) are often estimated with linear regression after taking natural logarithm on both sides of the equation, or generalized linear regression (e.g., Poisson regression). An augmented spatial interaction model contains additional independent variables which reflect the characteristics of origins, destination, and/or flows. In linear regression an augmented spatial interaction model with a log-linear specification can be rewritten in matrix form as:

$$\ln(\mathbf{F}) = \ln(\mathbf{X})\boldsymbol{\beta} + \boldsymbol{\xi}, \tag{6.7}$$

where $\mathbf{X}$ is a design matrix and $\boldsymbol{\xi}$ is the vector of residuals. Spatial interaction models can be extended to accommodate network autocorrelation. The residuals, $\boldsymbol{\xi}$, often have a significant level of network autocorrelation in an empirical network flow dataset. While spatial autoregressive (SAR) model approach provides a way to incorporate network autocorrelation in its specification (e.g., Chun et al. 2012; Fischer and LeSage 2010), eigenvector spatial filtering technique furnishes an alternative method. The eigenvector spatial filtering (ESF) utilizes eigenvectors extracted from a transformed spatial weight matrix, $(\mathbf{I} - \mathbf{11}^T/n)\mathbf{B}(\mathbf{I} - \mathbf{11}^T/n)$ where $\mathbf{I}$ is an identity matrix with n-by-n dimension, $\mathbf{1}$ is an n-by-1 vector of ones, and $\mathbf{B}$ is a spatial weights matrix. The eigenvectors are uncorrelated and orthogonal. Hence, the eigenvectors represent distinct map patterns when they are portrayed on the tessellation from which a spatial weight matrix is generated (see Griffith 2003 for details). In regression the ESF method includes a set of eigenvectors as independent variables to capture unexplained spatial autocorrelation which violates the independence assumption. Therefore, an ESF model does not suffer from spatial autocorrelation that often leads to biased parameter estimates in regression. In order to account for network autocorrelation in a spatial interaction model, eigenvectors can be generated from MB^NM where $\mathbf{M} = \left(\mathbf{I}_{n^2} - \mathbf{1}_{n^2}\mathbf{1}_{n^2}^T/n^2\right)$, $\mathbf{I}_{n^2}$ is an identity matrix n^2-by-n^2 dimension, and $\mathbf{1}_{n^2}$ is an n^2-by-1 vector of ones. That is, $\mathbf{M}$ is a modified matrix to match the dimension of $\mathbf{B}^N$. Hence, these eigenvectors, $\mathbf{E} = (\mathbf{E}_1,\mathbf{E}_2, \cdots,\mathbf{E}_n)$, in descending order of their corresponding eigenvalues denoted as $\lambda = (\lambda_1,\lambda_2, \cdots,\lambda_n)$, can be utilized to capture network autocorrelation among network flows. An ESF model specification of a spatial interaction model can be expressed as:

$$\ln(\mathbf{F}) = \ln(\mathbf{X})\beta + \mathbf{E}_k\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}, \tag{6.8}$$

where $\mathbf{E}_k$ denotes k selected eigenvectors, $\boldsymbol{\beta}_k$ are corresponding coefficients, and ε is random errors. Since $\mathbf{E}_k$ capture network autocorrelation, the residuals do not have a significant level of network autocorrelation and, hence, the parameter estimates become unbiased (Griffith and Chun 2013). An identification of a feasible set of eigenvectors can be conducted with the conventional stepwise regression technique from a candidate set of eigenvectors. Generally, a candidate set of eigenvectors is constructed by dropping eigenvectors which do not account for a substantial level of network autocorrelation. Alternatively, eigenvectors can be selected with minimizing network autocorrelation until network autocorrelation in residuals are close to its expected value or its z-score is close to zero (Tiefelsdorf and Griffith 2007).

6.4 Applications

In this section, two empirical interregional flow datasets are analyzed in a spatial interaction model framework. In the first application, interregional commodity flows in the U.S. (measured in million dollars) are analyzed in linear regression framework. In the second application, interstate migration flows in the U.S. are modeled with Poisson and negative binomial regression.

6.4.1 Interregional Commodity Flows in the U.S.

A dataset for interregional commodity flows in the U.S. were obtained from the 2002 Freight Analysis Framework (FAF)[1] which provides estimates of interregional freight movements in the U.S. by integrating mainly Commodity Flow

[1] http://www.ops.fhwa.dot.gov/freight/freight_analysis/faf/

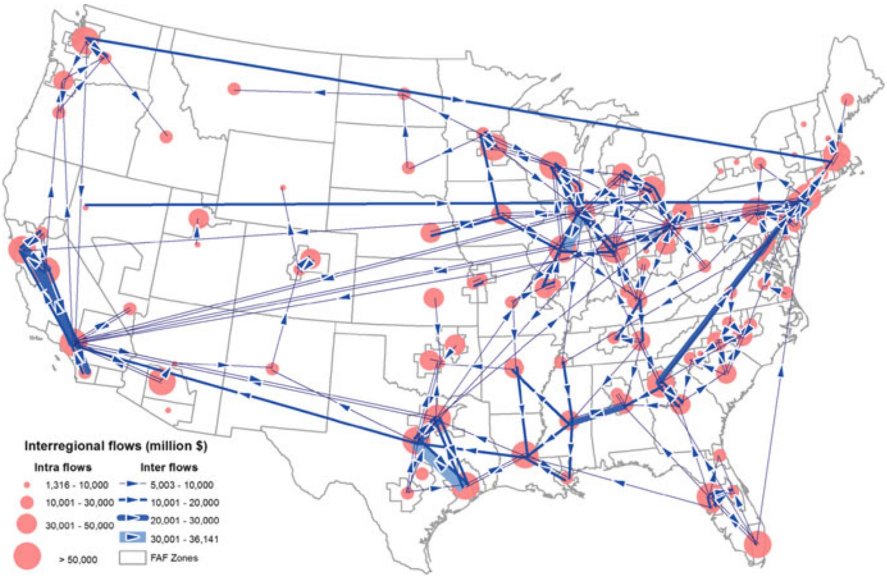


Fig. 6.2 The dominant interregional commodity flows among the 111 FAF zones in the U.S.

Survey and other resources. In this research, origin–destination (OD) flows among 111 FAF regions which reside in the conterminous U.S. are analyzed. Hence, the total number of flows is 12,321. Figure 6.2 displays dominant interregional commodity flows that are more than five billion dollars with line symbols. The point symbols in the figure represent internal commodity flows within individual FAF zones. The points are located at population weighted centers calculated with county level population from the 2000 U.S. Census.

Following the classical spatial equilibrium model (Enke 1951; Samuelson 1952; Bröcker 1989), a spatial interaction model for the interregional commodity flows is specified with nine independent variables in linear regression, similarly to Chun et al. (2012). Four variables reflect the characteristics of origins, including income per capita (o_inc), average plant size (o_plant), average production value (o_prod), and the number of employed people (o_emp). The three variables associated with destinations are population (d_pop), manufacturing (d_manuf), and income per capita (d_inc). The interregional distances (dist) are calculated with spherical distances between the population weighted centers. Finally, a dummy variable (intra) is included to capture the effects of large internal flows. These independent variables are transformed with natural logarithm except the intra dummy variable. Although natural logarithm is also commonly applied to a dependent variable, Box-Cox transformation is applied to make the transformed variable close to a normal distribution (e.g., Celik and Guldmann 2007).

Network autocorrelation of the transformed variable is measured with the five network weights matrix types defined in Eqs. 6.1, 6.2, 6.3, 6.4, and 6.5. As you can see in Table 6.1, the dependent variable has an extremely high level of network

Table 6.1 Network autocorrelation measure of the interregional commodity flows among the 111 FAF zones

Network weights matrix types		Type 1 (Eq. 6.1)	Type 2 (Eq. 6.2)	Type 3 (Eq. 6.3)	Type 4 (Eq. 6.4)	Type 5 (Eq. 6.5)
z-score of Moran's I (p value)	Dependent variable	96.5136 (0.0000)	103.6558 (0.0000)	141.5669 (0.0000)	140.0323 (0.0000)	195.2715 (0.0000)
	Residuals of linear regression	60.4844 (0.0000)	62.0667 (0.0000)	86.7295 (0.0000)	81.4931 (0.0000)	116.4794 (0.0000)

autocorrelation regardless of network weights matrix type. However, it has the highest z-score of Moran's I value (195.2715) with the network weights matrix in Eq. 6.5. Also network autocorrelation is measured with the residuals of the base linear regression model with the nine independent variables. As expected, the level of network autocorrelation decreases noticeably, but still a significant level of network autocorrelation remains among the residuals. The highest level of network autocorrelation is observed with the network weights matrix in Eq. 6.5. Hence, the network weight matrix in Eq. 6.5 is utilized in this research including ESF.

Table 6.2 shows the results of the base and the ESF models. The ESF model is estimated with 212 selected eigenvectors. There are three noticeable differences. First, eigenvector spatial filtering successfully accounts for network autocorrelation. While the residuals of the base model have a significant level of network autocorrelation (z-score of Moran's $I = 116.4794$), the ESF model does not have a significant level of network autocorrelation (z-score of Moran's $I = -0.1184$). Second, the ESF model has a better model fit. Its adjusted R^2 value (0.7612) is larger than that of the base model (0.6458). Also the ESF model has a smaller AIC value than the base model. The likelihood ratio test statistically supports that the ESF model has a better model fit (the test statistics is 5,061.98 with 202 degrees of freedom). Third, statistical significance changed for three independent variables by accounting for network autocorrelation. The origin income variable is not significant at the 5 % level in the base model but becomes significant in the ESF model. In contrast, two variables, plant size in origin and average production value, are significant at the 5 % level in the base model, but become insignificant in the ESF model. Unlike these differences, the estimates for the other variables are not significantly different from each other, and the estimates are significant with the expected negative sign in both models coefficients.

6.4.2 Interstate Migration Flows in the U.S.

The American Community Survey (ACS) has published state-to-state migration flows among the U.S. states. Here, the 2005–2009 ACS migration flows among the 48 states and Washington D.C. in the continental U.S. are analyzed, excluding Alaska and Hawaii due to their remote locations . Also, with a focus on interstate migration, internal migration flows within one state are excluded. This gives total 2,352 ($= 49^2 - 49$) interstate flows. Figure 6.3 shows dominant migration flows with more than 20,000 migrants, which are about top 3 % largest interstate migration flows in the 5-year period. Some noticeable large migration flows with more than 50,000 migrants are from California to Texas, from New Year to Florida, from California to Arizona, from Florida to Georgia, from Louisiana to Texas, from New York to New Jersey, and from California to Nevada. The large migration from Louisiana to Texas in this period may be explained by the effects of hurricane Katrina.

Table 6.2 The results of the base and eigenvector spatial filtering linear regression models

	Base model		Spatial filter model	
	Coefficient	Std. error	Coefficient	Std. error
Intercept	−29.4192	1.3004***	−37.3490	1.2496***
o_inc	0.1501	0.1187	0.4789	0.1136***
o_plant	−0.1714	0.0866*	0.0667	0.0868
o_prod	0.0694	0.0162***	0.0260	0.0155
o_emp	1.3445	0.0180***	1.3988	0.0164***
d_pop	0.3542	0.0306***	0.3702	0.0284***
d_manuf	0.9252	0.0253***	0.9906	0.0240***
d_inc	−0.0326	0.0762	0.2641	0.0745***
dist	−0.0014	0.0000***	−0.0013	0.0000***
Intra	5.1917	0.1295***	2.7393	0.1174***
z-score of Moran's I (p value)	116.4794 (0.0000)		−0.1184 (0.5471)	
R^2 (Adjusted R^2)	0.6461 (0.6458)		0.7653 (0.7612)	
AIC	42,222.79		37,564.8	
Log likelihood	−21,100.39 (df = 11)		−18,569.4 (df = 213)	
# of selected eigenvectors	–		202	

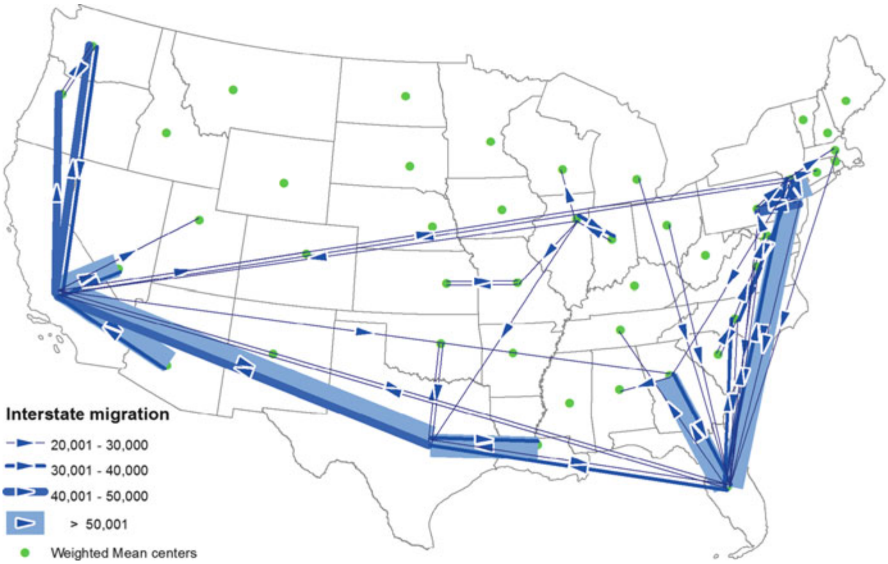


Fig. 6.3 The dominant interstate migration flows with more than 20,000 migrants during 2004–2009

The interstate migration flows are modeled with Poisson and negative binomial (NB) regression in a spatial interaction framework. As the numbers of migrants are count, Poisson and NB models can provide a more appropriate modeling approach (e.g., Flowerdew and Aitkin 1982; Abel 2010). Also, these models are estimated with the ESF method. In the models, seven independent variables are included, which were obtained also from the 5-year ACS in 2009. Following an extended gravity model of migration (e.g., Greenwood 1985; Hunt and Mueller 2004), three origin variables are total population (o_pop), unemployment rates (o_unemp), and income per capita (o_inc). The same three variables are included for destinations (i.e., d_pop, d_unemp, and d_inc). The distance-decay (dist) is model spherical distances between on weighted centers based on county populations using power function instead of exponential function in Eq. 6.6.

Table 6.3 reports the results of the Poisson and NB models. When the results of the base and the ESF Poisson models are compared, the ESF Poisson model has a significantly improved model fit with a smaller AIC and a large log-likelihood value than the base Poisson model. The log-likelihood ratio test confirms the increase of the model fit (the test statistics is 2,337,349.4 with 145 degrees of freedom). The dispersion parameter decreases to 530.38 in the ESF Poisson model from 1,819.62 of its counterpart base model. This considerable decrease of a scale parameter estimate has been constantly observed in empirical flow data modeling (e.g., Chun 2008; Fischer and Griffith 2008), although this still shows a high level of overdispersion. This might suggest a NB model specification to account for the overdispersion (e.g., Congdon 1989).

The estimation results of NB models show a similar pattern as the Poisson models when the base and ESF models are compared. The ESF NB model has a better model fit with a smaller AIC value and a larger log-likelihood value. The log-likelihood ratio test (its test statistics is 2,041.54 with 82 degrees of freedom) confirms the improvement of model fit. Figure 6.4 shows scatterplots of observed versus estimated values of the NB models. It shows that the estimated values of the EFS NB model are closer to the observed values than those of the base NB model. The ESF NB model has a larger estimate for θ parameter than the base NB model. As the variance of negative binomial given its mean (μ) is $\mu + \mu/\theta$, the result of the ESF NB model shows a less variability with a smaller estimate for dispersion (0.2870) than the base NB model (0.6407).

Statistical inferences for independent variables change by accounting for network autocorrelation in the Poisson and NB models. In the Poisson models, the statistical inference for o_unemp and o_inc variables changed at the 5 % level. In the base NB model, o_unemp and d_ unemp variables are significant at the 5 % level but become insignificant in the ESF NB model at the same level. With regards to the distance-decay effect, the ESF Poisson model produces a higher level of distance-decay effect than the base Poisson model. However, the estimated distance-decay effect of the ESF NB model is not statistically different from that of the base NB model.

Table 6.3 The results of the base and eigenvector spatial filtering Poisson and negative binomial regression models

	Poisson model						Negative binomial model	
	Base model		Spatial filter model		Base model		Spatial filter model	
	Coefficient	Std. error	Coefficient	Std. error	Coefficient	Std. error	Coefficient	Std. error
intercept	−3.3713	1.8117	−18.9807	1.7015***	−11.7273	1.8160***	−29.9591	1.6529***
o_pop	0.9248	0.0217***	1.0236	0.0147***	0.8463	0.0209***	0.9284	0.0148***
o_unem	−0.2493	0.1175*	0.1081	0.1029	−0.4024	0.1052***	0.0354	0.0868
o_inc	0.0831	0.1267	0.5829	0.1164***	0.4140	0.1221***	1.0758	0.1088***
d_pop	0.9189	0.0222***	1.0011	0.0150***	0.8564	0.0209***	0.9454	0.0148***
d_unem	−0.7765	0.1169***	−0.4050	0.1038***	−0.5791	0.1052***	−0.1191	0.0868
d_inc	−1.0819	0.1331***	−0.2945	0.1198*	−0.2993	0.1222*	0.4037	0.1088***
dist	−0.7839	0.0196***	−1.0495	0.0247***	−0.9177	0.0256***	−0.9379	0.0311***
Dispersion	1,819.6210		530.3807		–		–	
Dispersion	–		–		0.6407		0.2870	
AIC	3,447,202		1,110,143		38,341.12		36,463.57	
Log likelihood	−1,723,593		−554,918.3		−19,161.56		−18,140.79	
# of selected eigenvectors	–		145		–		82	

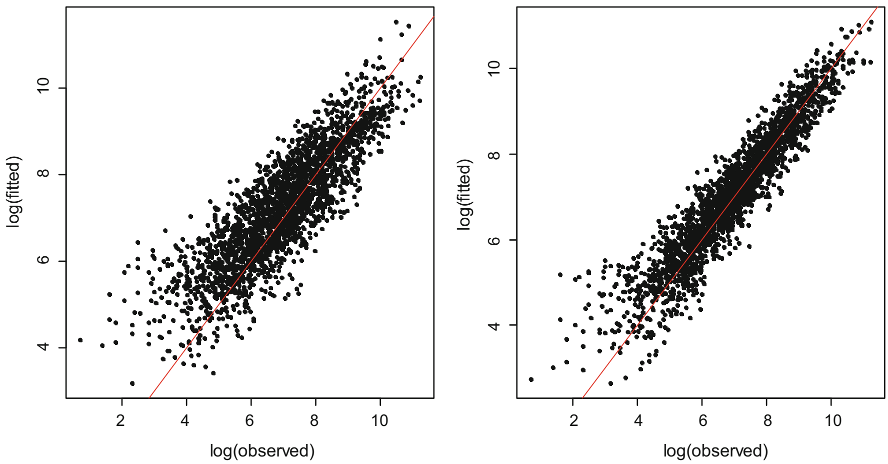


Fig. 6.4 The scatterplots of observed vs. estimated values of the base NB model (*left*) and the ESF NB model (*right*)

6.5 Conclusions

This research investigated network autocorrelation with the two empirical interregional flows in the U.S.: interregional commodity flows and interstate migration flows. These flow datasets were analyzed using the ESF technique in a spatial interaction modeling framework. The level of network autocorrelation in the interregional commodity flows is measured with five different types of network weights matrices. Although highly significant positive network autocorrelation was measured with all types of network weights matrices, the highest level of network autocorrelation was observed with one defined in Eq. 6.5. Since, in this network weights matrix, all network flows between spatially neighboring origins and destinations are considered as a neighbor, a more comprehensive network dependence structure is reflected. Although an advanced model specification allows more than one weights matrices simultaneously (e.g., LeSage and Pace 2008), many currently available functions allow only one weights matrix for spatial models. Hence, it is beneficial to reflect an appropriate network dependency structure in a weights matrix, and the empirical results show that the network weights matrix in Eq. 6.5 is possibly a good specification.

The two empirical analyses demonstrate that the ESF method successfully accounts for network autocorrelation and, consequently, leads to a better model fit in both linear and generalized linear (i.e., Poisson and NB) regression models. The ESF linear regression model does not have a significant level of network autocorrelation, while the base linear regression model suffers from an extremely high level of network autocorrelation in its residuals. One interesting point of the Poisson and NB regression results is that an estimate for extra variability decreases by accounting for network autocorrelation. This finding concurs with the fact that a variance increases in a data distribution when a significant level of spatial autocorrelation is present in a random variable (Griffith 2011).

Gravity type spatial interaction models have been commonly utilized in quantitative geographical flow modeling. These models are often estimated without considering network autocorrelation, although the effects of spatial structure or spatial autocorrelation in spatial interaction have been recognized in the literature. Recent development in modeling network autocorrelation improves spatial interaction modeling for network flows. Especially, the ESF method, with its flexible specification, furnishes a useful method to modeling network autocorrelation in linear and generalized linear models.

References

Abel GJ (2010) Estimation of international migration flow tables in Europe. J R Stat Soc Ser A 173 (4):797–825

Barber M, Scherngell T (2013) Is the European R&D network homogeneous? Distinguishing relevant network communities using graph theoretic and spatial interaction modelling approaches. Reg Stud. doi:10.1080/00343404.2011.622745

Black WR (1992) Network autocorrelation in transport network and flow systems. Geogr Anal 24 (3):207–222

Bröcker J (1989) Partial equilibrium theory of interregional trade and the gravity model. Pap Reg Sci Assoc 66:7–18

Celik HM, Guldmann J-M (2007) Spatial interaction modeling of interregional commodity flows. Socioecon Plann Sci 41:147–162

Chun Y (2008) Modeling network autocorrelation within migration flows by eigenvector spatial filtering. J Geogr Syst 10(4):317–344

Chun Y, Griffith DA (2011) Modeling network autocorrelation in space-time migration flow data: an eigenvector spatial filtering approach. Ann Assoc Am Geogr 101(3):523–536

Chun Y, Kim H, Kim C (2012) Modeling interregional commodity flows with incorporating network autocorrelation in spatial interaction models: an application of the U.S. interstate commodity flows. Comput Environ Urban Syst 36(6):583–591

Congdon P (1989) Modelling migration flows between areas: an analysis for London using the Census and OPCS Longitudinal Study. Reg Stud 23:87–103

Curry L (1972) A spatial analysis of gravity flows. Reg Stud 6:131–147

Enke S (1951) Equilibrium among spatially separated markets: solution by electric analogue. Econometrica 19(1):40–47

Fischer MM, Griffith DA (2008) Modeling spatial autocorrelation in spatial interaction data: an application to patent citation data in the European Union. J Reg Sci 48:969–989

Fischer MM, LeSage JP (2010) Spatial econometric method for modeling origin–destination flows. In: Fischer MM, Getis A (eds) Handbook of applied spatial analysis. Springer, Berlin, pp 435–460

Flowerdew R, Aitkin M (1982) A method of fitting the gravity model based on the Poisson distribution. J Reg Sci 22(2):191–202

Fotheringham AS (1981) Spatial structure and distance-decay parameters. Ann Assoc Am Geogr 71:425–436

Fotheringham AS (1983) A new set of spatial-interaction models: the theory of competing destinations. Environ Plann A 15:15–36

Greenwood MJ (1985) Human migration: theory, models, and empirical studies. J Reg Sci 25 (4):521–544

Griffith DA (2003) Spatial autocorrelation and spatial filtering. Springer, Berlin
Griffith DA (2009) Modeling spatial autocorrelation in spatial interaction data: empirical evidence from 2002 Germany journey-to-work flows. J Geogr Syst 11:117–140
Griffith DA (2011) Positive spatial autocorrelation impacts on attribute variable frequency distributions. Chil J Stat 2(2):3–28
Griffith DA, Chun Y (2013) Spatial autocorrelation and spatial filtering. In: Fischer MM, Nijkamp P (eds) Handbook of regional science. Springer, Berlin
Griffith DA, Jones KG (1980) Explorations into the relationship between spatial structure and spatial interaction. Environ Plann A 12:187–201
Hunt GL, Mueller RE (2004) North American migration: returns to skill, border effects, and mobility costs. Rev Econ Stat 86:988–1007
Kwan M-P (1998) Space-time and integral measures of individual accessibility: a comparative analysis using a point-based framework. Geogr Anal 30(3):191–216
LeSage J, Pace RK (2008) Spatial econometric modelling of origin–destination flows. J Reg Sci 48:941–967
Mitze T (2012) Empirical modelling in regional science, lecture notes in economics and mathematical systems 657. Springer, Heidelberg
Patuelli R, Griffith DA, Tiefelsdorf M, Nijkamp P (2011) Spatial filtering and eigenvector stability: space-time models for German unemployment data. Int Reg Sci Rev 34(2):253–280
Samuelson PA (1952) Spatial price equilibrium and linear programming. Am Econ Rev 42:283–303
Stouffer SA (1960) Intervening opportunities and competing migrants. J Reg Sci 2(1):1–26
Tiefelsdorf M, Griffith DA (2007) Semi-parametric filtering of spatial autocorrelation: the eigenvector approach. Environ Plann A 39:1193–1221

Chapter 7
Assortativity and Hierarchy in Localized R&D Collaboration Networks

Joan Crespo, Raphaël Suire, and Jérôme Vicente

Abstract One of the challenges of innovative clusters relies on their ability to overlap technological domains in order to maintain their growth path along the cycle of technological markets. The paper studies two particular structural properties of collaboration networks that provide new insights for understanding this overlapping process. On the one hand, the degree distribution of knowledge networks captures the level of hierarchy within networks. It gives a first measure of the ability of networked organisations to coordinate their actions. On the other hand, the degree correlation captures the level of assortativity of networks. It gives a measure of the ability of knowledge to flow between highly and poorly connected organisations. We propose to combine these simple statistical measures of network structuring in order to study the parameters window that allow localized knowledge networks combining technological lock-in with regional lock-out.

7.1 Introduction

The study of R&D collaboration networks has become a subject of a growing interest in spatial analysis and geography of innovation (Autant-Bernard et al. 2007; Scherngell and Barber 2011). In particular, clusters analysis have found through the identification of localized R&D collaboration networks new means for assessing regional performances (Owen-Smith and Powell 2004; Vicente et al. 2011;

J. Crespo (✉)
Political Science Institute, University of Toulouse, Toulouse, France
e-mail: juan.crespo-cepas@ut-capitole.fr

R. Suire
University of Rennes 1, CREM-CNRS, Rennes, France

J. Vicente
University of Toulouse, Toulouse Business School, and Political Science Institute, Toulouse, France

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_7,

Balland et al. 2013), beyond the simple co-location of innovative activities or the black box of local knowledge spillovers (Breschi and Lissoni 2001).

Our contribution fits with this research challenge, with a particular focus on the ability of localized R&D collaboration networks to maintain a long term performance in a context of rapid business and technological cycles. The aim is to capture the structural properties of collaboration networks that allow clusters performing in particular technologies without compromising their renewal capabilities when markets for these technologies decline. As a matter of fact, some clusters can have difficulties in coping with technological and market decline, even if they were leading places during the maturity stage of the industry. At the opposite some others can succeed in disconnecting their cycle to the cycle of technologies by reorganizing resources and networks towards a new stage of growth based on a new or related growing market. Literature provides some highlighting stylized facts of such patterns of cluster evolution. For instance, Saxenian (1990) describes the renewal of the Silicon Valley in the 1980s from the declining semiconductor industry towards the emerging computer industry. She stresses on the fact that such a renewal was more the consequence of a reorganisation of knowledge flows into the local organisational network rather than the consequence of market or national policy concerns. Tödling and Trippl (2004) converge towards the same conclusions in their study of the differentiated renewal capabilities of clusters in a sample of old industrial areas; while Cho and Hassink (2009) find evidences according to which some clusters reach their maturity through an increasing rigidity of their networks that plays against their ability to react to market cycles.

Then clusters life cycles (Suire and Vicente 2009, 2013; Menzel and Fornahl 2010; Crespo 2011; Boschma and Fornahl 2012) can find explanations in the structural organisation of collaboration networks and their evolving patterns along the cycle of technologies and markets. Do successful clusters in a mature industry necessarily locked into a rigid trajectory and then to decline, or are there particular structural properties of localized collaboration networks that enable clusters to combine performance in mature industries and renewal capabilities towards emerging ones?

In order to disentangle this question, we propose in a second section to discuss the micro-motives of organisations for joining a network and building knowledge relations, and the resulting consequences on the emerging structural properties of knowledge networks. This section will show that network hierarchy and assortativity appear as two salient topological and structural properties that play together in the long term performance of localized R&D collaboration networks. Section three proposes to associate these structural properties to two statistical signatures of collaboration networks that provide tools for developing new evidences on the critical factors of the long term dynamics of clusters.

7.2 Clusters as R&D Collaboration Networks

7.2.1 Clusters Growth and Structuring

A cluster can now be defined as a local relational structure that results from the identification of a set of nodes of various institutional forms (the organisational demography) and the ties between them (the relational structure). Inter-organisational ties in a cluster can be of different nature (productive, commercial, cognitive or social) and of different geographical length. Our discussion focuses on localized R&D collaboration networks, and then organisational relations locally constructed to exchange knowledge in high-tech technological domains.

Network theory is very useful for analysing cluster properties, since it has identified several drivers of network formation (Ahuja et al. 2012) that can be founded on micro-economic behaviours. In particular, these micro-foundations are necessary to understand how new entrants join a cluster, and (re)shape its relational structure.

Firstly, networks can evolve through the entry of new nodes that do not connect to any other node (isolates), or through the entry of new nodes that connect to others by purely random attachment mechanism. It means that entering nodes connect to others with no particular preference for their position in the structure. Isolate entrants and random attachment mechanism will give rise to a rather flat hierarchy of degrees in the collaboration network. In terms individual strategies, both kinds of processes can be associated with a locational cascade (Suire and Vicente 2009). In locational cascades, new entrants draw pay-offs from belonging to the structure as a whole, not from targeted connections to particular nodes in the structure. Locational cascades have been largely evidenced for clusters that attract new organisations because of an external audience and a geographical charisma (Romanelli and Khessina 2005; Appold 2005). Organisations converge to a “locational norm” since the charisma displayed by one place in terms of R&D productivity provides a signal of quality and a strong incentive for being located there, whatever the position in the relational structure.

Secondly, entries can occur through a process of preferential attachment. In this opposite case, nodes with many ties at a given moment of time have a higher probability to receive new ties from new entering nodes. The higher the degree of an organisation in the collaboration network, the more this organisation is attractive for receiving new ties, so that the network grows through an increasing hierarchy (Albert and Barabási 2002). This behavioural pattern of nodes can be associated to a network effect in location decision externalities. This means that the more new entrants are connected to highly connected nodes, the more their payoffs increase, due to the benefits of reciprocal knowledge accessibility and technological connections to an emerging and growing standard. This branching process is now linked to targeted connections in the structure rather than random ones, and is consistent with the relational constraints that typify the production and diffusion of technological standards in high-tech industries and markets (Farell and Saloner 1985; Arthur

1989). It is also consistent with the relational behaviour of spinoffs that tend to connect to their often highly connected parent's company (Klepper 2010).

Beyond node entry, clusters structure themselves through the construction and dissolution of ties. The literature acknowledges two categories of individual incentives that shape social structures, and dissociate closure from bridging network strategies (Baum et al. 2012). Triadic closure implies that a node with links to two other nodes increases the probability for these two nodes to have a tie between them. Such an argument is grounded on the process of trust construction that grows between two related nodes, because it fosters cooperation and knowledge integration within groups of nodes. Closure in collaboration networks strengthens the mutual monitoring capability of organisations. Indeed, on one hand, it decreases the possibilities of opportunistic behaviours (Coleman 1988). On the other hand, it increases the effects of conformity required by technological standardization processes: without such closure, organisations can be tempted to play the battle of standards and accept the risk of a payoff decrease. As this process develops, the clustering coefficient of the network increases, and triadic closure tends to shape a core-component in the collaboration network (Borgatti and Everett 1999), in particular when closure prevails for highly connected organisations. The second category of individual incentives relates to bridging strategies and introduces the idea of a more disruptive relational behaviour. For a given network, bridging ties will be shaped when one organisation finds an opportunity to connect disconnected organisations or groups of organisations. Such an agency behaviour (Burt 2005) is more entrepreneurial than the former, since bridging provides access to new and non-redundant knowledge and new opportunities for improving innovation capabilities (Ahuja et al. 2009).

7.2.2 Structural Properties of R&D Collaboration Networks

According to the mechanisms of network formation and structuring at work in clusters, they will display a high degree of variability in the structural and topological properties of their collaboration network. Previously captured using different methodological approaches (Markusen 1996; Iammarino and McCaan 2006), this variety of cluster relational structures can be assessed using network theory through a set of simple key-indexes that echo important features of collaborative process of innovation.

The first property relies on the degree of connectedness of the collaborative network. A cluster will be fully connected if there is no isolates in the population of nodes and all the nodes can be reached by the other nodes. The second one is the density of the collaborative network. Clusters can have a very weak level of relational density if organisations value isolated strategies over knowledge partnerships. In that case, the clusters are no more than the simple result of a co-location process, as for the well-known satellite platform of Markusen (1993). On the contrary, clusters can display a high level of density when knowledge complementarities, trust and social proximity (Boschma 2005) lead to high levels of local cohesiveness into the collaboration network.

More importantly, even for a full connectedness and a fixed level of density, other structural properties matter and provides relevant information on the collaboration process. Considering the degree centrality of each organisation, i.e. direct interaction neighbourhood, the distribution of degree can vary from a flat distribution to an asymmetric one. To put it differently, the shape of the degree distribution refers to the hierarchy of positions in the web of relationships, and can be captured by ranking organisations in a network according to their degree and putting into relation with their own actual degree. Some organisations can have many relations due to a high relational capacity (König et al. 2010). This is generally linked to the size of the organisations, their absorptive capacities or the openness of their model of knowledge valuation. On the contrary, some others remain poorly connected due to their newness, their small size or their closed model of knowledge valuation.

Moreover, considering again the degree of each organisation, clusters can vary in their structure according the shape of the degree correlation. Indeed, clusters can display various levels of structural homophily, which is generally captured by an index of assortativity (Newman 2003; Watts 2004; Rivera et al. 2010). Here again, the assortativity of a network can be captured by the relation between the degree of each organisation and the mean degree of the organisations in its direct neighbourhood. The structure of relationships will be assortative when highly (poorly) connected nodes tend to be connected disproportionately to other high (weak) degree nodes. In that case, the degree correlation of the network is positive. At the opposite, the structure of relationships will be disassortative when highly (poorly) connected nodes tend to be connected disproportionately to other weak (high) degree nodes. In that case, the degree correlation of the network will be negative. Therefore, the level of network assortativity gives a formal representation of the way knowledge flows between central and more peripheral nodes.

How these properties can play together for that localized R&D collaboration networks perform of global markets without compromising their ability to adapt to business and market cycles? Recall that some successful clusters can decline when the market for their products decline, while some others succeed in disconnecting their cycle from the cycle of markets and develop renewal capabilities towards emerging ones.

The properties of hierarchy and assortativity provide new insights for that purpose. As a matter of fact, successful clusters at a moment in time and in a particular technological field are the ones that have succeeded in going from the exploration of new ideas to the exploitation of a technological standard or dominant design on a mass market, with in between, a collective process of knowledge integration between complementary organisations along the knowledge value chain (Cooke 2005). Beyond the traditional scheme of exploration/exploitation that typifies the innovation process of a single organisation, the knowledge integration phase is at the heart of the cluster's purpose. Indeed, the success of many products results from their degree of compositeness (Antonelli 2006), the variety of uses and applications supported by the products, scientific as well as symbolic knowledge (Asheim et al. 2011), and the compatibility and easy interoperability between elements that are the rule of a dominant design diffusion (Frenken 2006). The chasm that sometimes prevents some products from reaching the mass market (Moore 1991) is more often the consequence of a failed integration process, i.e. a problem of industrial organisation, rather than a problem of the product quality in itself. Successful clusters are therefore the ones that achieve the imposition of well-integrated and performing complete technological systems on mass markets. As the literature shows (Klepper and Simmons 1997; Audretsch et al. 2008), these clusters evolve from an initial scattered structure of burgeoning organisations in the early market stages to a structure with a limited number of hub and oligopolistic organisations in mature markets. Along the life cycle of products, and especially composite ones, such a network dynamic produces path dependence and technological lock-in. The more the technologies generate increasing returns to adoption, the more markets for these technologies become locked-in and resist to other competing technologies (Arthur 1989).

But are clusters producing these technologies necessarily locked-in too? The answer depends on the way in which their relational structure evolves along the life cycle of products. First, recall that R&D collaboration networks can grow through preferential attachment. This means that the more nodes display a high degree, the more newcomers connect to these nodes, engendering a high level of hierarchy in the degree distribution of organisations. But secondly, recall that beyond network growth through node entry, networks can also evolve by the addition and rewiring of ties between existing nodes through closure or bridging (Baum et al. 2012). When closure prevails, the cluster evolves towards a high level of transitivity between nodes which is the mark of isomorphic and conformist relational behaviours. In that case, the structure of the cluster exhibits tight couplings into a core-component and a loosely connected periphery of nodes. The ossification of the cluster goes with the formation of an assortative collaboration network, in which highly connected nodes are tied predominantly with other highly connected nodes, and poorly connected nodes remains connected between themselves. On the contrary, a structure with a disassortative web of knowledge relationships can emerge as the entry of newcomers and rewiring process go. For that, the node bridging strategy has to prevail over the closure strategy. Consequently, highly connected organisations spend a share of their relational capacity towards peripheral organisations, and the network as a whole displays more paths between highly and poorly connected nodes than for the assortative network.

The patterns governing the entry dynamics into networks and the structuring process that follows are at the heart of the lock-in/lock-out debate. Academics acknowledge that preferential attachment is a natural pattern of social and human networks that contributes to fostering the legitimacy of social norms and conformist effects in Sociology (Watts 2004), or technological standards and dominant designs in Business Studies (Frenken 2006). But the debate between closure and bridging is more controversial, and it is also controversial for cluster studies (Eisingerich et al. 2010). Indeed, closure favours technological lock-in and thus the ability of the relational structure to perform in markets. The tight coupling between high degree organisations favours conformism and trust in a stable and cohesive structure that prevents opportunism and promotes an efficient integration of knowledge in a context of weak environmental uncertainty. But closure favours network assortativity, and then prevents regional lock-out, since the low connectivity between the core nodes and the peripheral ones limits the re-organisation of knowledge flows when uncertainty grows or when the market starts to decline. So when preferential attachment and closure interact, the ability of clusters to deal with a positive technological lock-in goes against the collaboration network to produce the conditions for a regional lock-out (Simmie and Martin 2010).

In order to foster adaptability, clusters also have to develop bridging strategies in order to open more disruptive relations, preserving minimal cohesiveness in the core, while multiplying the channels for potential or latent flows of fresh and new ideas coming from peripherals nodes (Grabher and Stark 1997; Cattani and Ferriani 2008). Such a mix of patterns does not undermine the hierarchy of degrees that emerges when the technology goes towards exploitation. But to be disassortative, the oligopoly structure of hub-organisations that appears as the technology reaches maturity has to maintain a not too low amount of entrepreneurial connections with the periphery, in order to overlap exploitation in a particular knowledge domain and exploration in another related one (Cohen and Klepper 1992; Almeida and Kogut 1997; Schilling and Phelps 2007). Such a structural property of clusters is consistent with the behaviour of firms according to their maturity and age. Indeed, Baum et al. (2012) develop evidence on the predisposition of organisations to deal with closure or bridging strategies according to their age. Supposing that the age of organisations is positively related to their hub position and high degree, then the renewal capabilities of local knowledge structures can be weakened by an insufficient level of connectivity with newcomers, as shown by Saxenian (1990) for the semiconductor collaboration network in the Silicon Valley. If it is supposed that the capacity constraints in the amount of ties an organisation can maintain is related to its size and age, as König et al. (2010) do, then the high capacities of hub and central organisations can be a strong source of renewal if they go against the natural tendency to reproduce existing and conformist ties. Ahuja et al. (2009) find empirical evidence on that by capturing the micro motives for more disassortative behaviours. They highlight a threshold and non-monotonic effect in the strategy of embeddedness and closure between central nodes. According to them, the growing benefits in terms of trust and knowledge acquisition can go with an increasing rigidity and conformity that produces disincentives for new collaborations. Likewise, in spite of risks of knowledge hold-up and contract incompleteness, they find that peripheral organisations succeed in connecting to central nodes, through a "creeping" strategy facilitated by the ability of mature organisations to find sometimes new and disruptive opportunities to connect to peripheral newcomers.

7.3 Two Simple Statistical Signatures of Collaboration Networks

The level of hierarchy of node degree and the level of assortativity therefore appear as two simple statistical signatures of the ability of clusters to perform but also to avoid negative lock-in through their endogenous renewal capabilities. The following definition of statistical signatures of localized R&D collaboration networks aims to discuss the parameters space that allows clusters overlapping exploitation of technologies on mature markets and exploration of new or related technologies for emerging markets.

7.3.1 Degree Distribution and Correlation

Hierarchy and assortativity can be measured through two simple statistical signatures. The first corresponds to the degree distribution of the network. By degree distribution, we mean the relation between the ranking of nodes in a network according to their degree and their actual degree.[1] The more sloped the distribution is, the more the network displays hierarchy in the degree of nodes. From weakly connected nodes to highly connected nodes, the degree distribution exemplifies the level of heterogeneity in the network in terms of actual relational capacity. The second property corresponds to the degree correlation. Here, degree correlation is defined as the relation between the degree of each node and the mean degree of nodes in its neighbourhood. Networks can be characterized as assortative or disassortative to the extent that they display a positive or negative degree correlation. A network is assortative when high degree nodes are connected to other high degree nodes, and low degree nodes are preferentially connected to low degree nodes, so that the degree correlation is positive. And a network is disassortative when high degree nodes tend to connect to low degree nodes, and vice versa, so that the degree correlation is negative. For a given amount of nodes and ties in a particular network, one can easily capture these two salient properties.

Consider a fixed number of nodes and ties in a network N.[2] If we note k the degree of a particular node h, we can then write two simple equations to characterize the network topology. By referring to a rank-size rule, we can classify node degrees from the largest to the smallest[3] and then draw the distribution on a *log-log* scale. Such that:

[1] Another traditional representation consists in mapping degree distribution using frequencies of degree values.

[2] Then we only focus on the structuring of the network. Entries are considered as exogenous, or occurring in previous periods.

[3] If two nodes have the same degree, we arbitrarily rank them as long as it has no incidence on the slope on the power law.

$$k_h = C\left(k_h^*\right)^a,$$

with k_h^* being the rank of the node h in the degree distribution, C a constant and $a < 0$ the slope of the distribution or equivalently,

$$\log(k_h) = \log(C) + a\log\left(k_h^*\right)$$

Secondly, we can calculate for each node h, the mean degree of the relevant neighbourhood (V_h), i.e.,

$$\overline{k_h} = \frac{1}{k_h}\sum\nolimits_{i \in V_h} k_i,$$

where k_i is the degree of node i belonging to the interaction neighbourhood of node h.

Then we estimate a linear relationship between $\overline{k_h}$ and k_h, such that

$$\overline{k_h} = D + bk_h,$$

with D a constant and b a coefficient capturing the degree correlation.

If $b > 0$, the network N exhibits assortativity with a positive degree correlation, whereas if $b < 0$, the network N is disassortative with a negative degree correlation.

Finally, thanks to the ordinary least squares method, the joint estimation of parameters a and b enables us to characterize useful structural network properties.

$$\begin{cases} \textit{degree distribution} : \log(k_h) = \log(C) + a\log\left(k_h^*\right) \\ \textit{degree correlation} : \overline{k_h} = D + bk_h \end{cases} \tag{7.1}$$

7.3.2 Discussion

Using Eq. 7.1, and considering a fully connected network N with a fixed number of nodes ($n = 33$) and ties ($t = 64$),[4] Fig. 7.1 summarizes this proposition, giving more details on three typical topologies and their statistical signatures.

(i) The so-called "flat" network presents a relatively flat degree distribution | a| = 0,37 with a degree correlation b ~ 0. This type of collaboration network displays a strong potential for knowledge flows re-organisation and diffusion since the nodes are linked by many paths. But such a random network does not succeed in generating conformity effects and the emergence of technological standards. Indeed, the lack of cohesiveness in to the network and the absence

[4] In such a way that the density remains the same for the three networks $2t/n(n - 1) = 0.1212$, where t is the number of actual links and n the number of nodes).

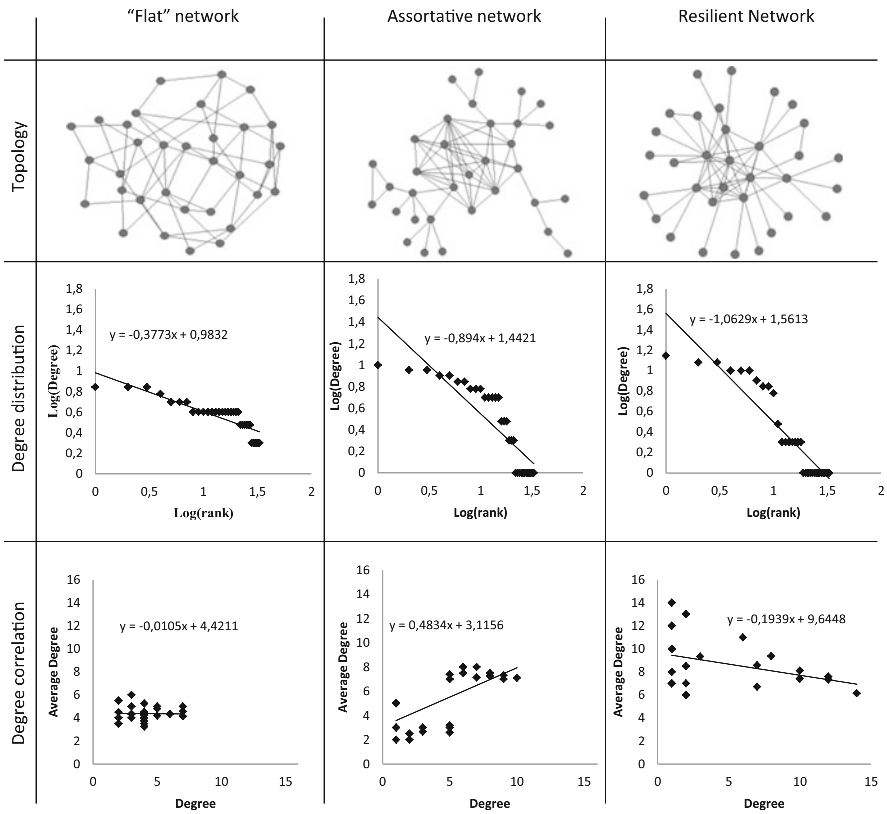


Fig. 7.1 Network topology, degree distribution and degree correlation

of a core group weaken the control of collective behaviours that would exploit products on the market by efficiently gathering pieces of knowledge.

(ii) On the contrary, the assortative network presents a strong slope in the degree distribution $|a| = 0{,}89$ so that the cohesiveness of the core promotes a conformity effect, and, from a technological perspective, a high probability of the emergence of a standard. Nevertheless, its strong assortative structure ($b > 0$) weakens its renewal properties since peripheral nodes are loosely connected to the central ones. This excess of assortativity will reduce the ability of the existing structure to activate new explorative ties when markets for the exploited technology decline, due to a weak level of bridging between the oligopoly structure and the peripheral ones. Therefore the assortative knowledge network favours technological lock-in without maintaining regional lock-out conditions because of its relative inability to overlap exploitation links on mature markets and explorative ones on emerging related ones.

(iii) Finally, the resilient network exhibits here again a high sloped degree distribution with $|a| = 1{,}06$, but the degree correlation is now negative ($b < 0$), so

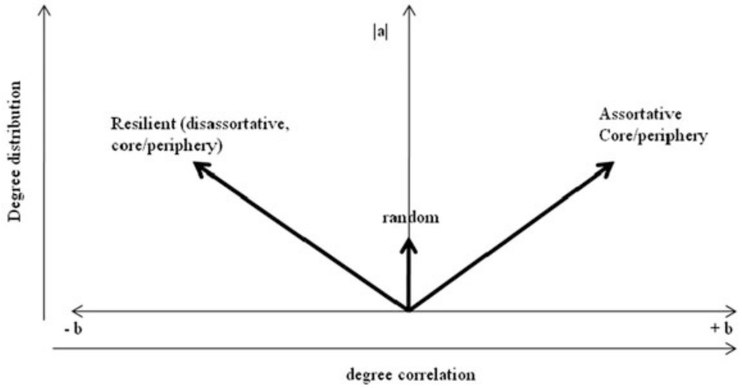


Fig. 7.2 Statistical signatures of cluster structural properties

that the network presents a certain level of disassortativity. In other words, this negative correlation indicates a high level of connections between the core and the periphery of the collaboration network, so that information and knowledge can circulate through many structural bridges between highly and poorly connected nodes. Thus targeted shocks on core members do not weaken the whole structure to the same extent as in the previous structure. Similarly, innovative or explorative behaviour can diffuse more easily from peripheral to central members, due to the ability of the oligopolistic organisations to combine closure and bridging and overlap explorative and exploitive phases in their relational patterns.

Figure 7.2 provides a more abstracted representation of these critical structural properties of local knowledge networks.

By representing degree distribution and degree correlation in the same layout, one can have a better understanding of how the structure and properties of local clusters can together improve aggregate performance and structural conditions for renewal along the cycles of markets. The further up in the layout a cluster is, the more the structural hierarchy of its collaboration network enables it to impose standards and dominant designs on markets. And the further left in the layout it is, the more the disassortative patterns of relationships in the network increase regional renewal capabilities. The emerging oligopolistic structure that arises when the technology reaches maturity has to remain sufficiently linked to fresh and new ideas coming from peripheral but promising nodes for future collaborations. On the other hand, when closure strategies in the mature oligopolistic structure exceed a certain threshold, then redundancy of knowledge flows and conformity effects prevail and the possibilities for regional resilience fall unavoidably. Then if some clusters decline when their dedicated markets decline, the reasons are not necessarily to find in an ossification of the structure of the network or in an excess of rigidity due to the firm growing size, but in the relational strategies of hub and leading organisations, and a decreasing degree of openness towards peripheral but strategic newcomers.

7.4 Conclusion

In spite of its high level of abstraction and complexity, the science of networks applied to geography of innovation provides promising perspectives for static as well as dynamic analysis of clusters. Here we have tried to show that it was possible to reduce this complexity to two simple statistical signatures of collaboration networks. Degree distribution and degree correlation highlight the critical structural properties that increase the performance of clusters in a particular technological field, without decreasing their renewal properties. If the hierarchy of degrees is a more or less common pattern of social and organisational networks, disassortativity is less manifest. Indeed, human and social behaviours are generally characterized by structural homophily, so that the more an agent increases its relational capacity, the larger is his tendency to interact with other highly connected agents. However, this property of assortativity of local knowledge networks weakens the ability of clusters to combine market exploitation and absorption of fresh and new ideas, and then, can be a source of negative regional lock-ins.

The combined measures of degree distribution and degree correlation confirm that a window of parameters exists, for which clusters can display performance in the short run, and renewal capabilities in the long run. Capturing this window more precisely requires an additional effort of modelling. But at this stage, such a framework furnishes new perspectives to highlight empirical evidence on the ability of regional systems of innovation to resist and adapt to turbulent macroeconomic environments, new growing consumer paradigms and the shortening of market cycles.

References

Ahuja G, Polodiro F, Mitchell W (2009) Structural homophily or social asymmetry? The formation of alliances by poorly embedded firms. Strateg Manag J 30:941–958

Ahuja G, Soda G, Zaheer A (2012) The genesis and dynamics of organisational networks. Organ Sci 23(2):434–448

Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47–97

Almeida P, Kogut B (1997) The exploration of technological diversity and geographic localization in innovation: start-up firms in the semiconductor industry. Small Bus Econ 9(1):21–31

Antonelli C (2006) The business governance of localized knowledge: an information economics approach for the economics of knowledge. Industry Innov 13(3):227–261

Appold S (2005) The location patterns of U.S. Industrial research: mimetic isomorphism, and the emergence of geographic charisma. Reg Stud 31(1):17–39

Arthur WB (1989) Competing technologies, increasing returns and lock-in by historical events. Econ J 99:116–131

Asheim BT, Boschma R, Cooke P (2011) Constructing regional advantage: platform policies based on related variety and differentiated knowledge bases. Reg Stud 45(7):893–904

Audretsch D, Falck O, Feldman M, Heblich S (2008) The life cycle of regions, CEPR Discussion paper DP6757

Autant-Bernard C, Billand P, Frachisse D, Massard N (2007) Social distance versus spatial distance in R&D cooperation: empirical evidence from European collaboration choices in micro and nanotechnologies. Pap Reg Sci 86(3):495–520

Balland PA, Suire R, Vicente J (2013) Structural and geographical patterns of knowledge networks in emerging technological standards: evidence from the European GNSS industry. Econ Innov New Tech 22(1):47–72

Baum JAC, McEvily B, Rowley TJ (2012) Better with Age? Tie longevity and the performance implications of bridging and closure. Organ Sci 23(2):529–546

Borgatti SP, Everett MG (1999) Models of core/periphery structures. Soc Networks 21:375–395

Boschma RA (2005) Proximity and innovation: a critical assessment. Reg Stud 39(1):61–74

Boschma R, Fornahl D (2012) Cluster evolution and a roadmap for future research. Reg Stud 45 (1):1295–1298

Breschi S, Lissoni F (2001) Knowledge spillovers and local innovation systems: a critical survey. Ind Corp Chang 10(4):975–1005

Burt RS (2005) Brokerage and closure. Oxford University Press, New York

Cattani G, Ferriani S (2008) A core/periphery perspective on individual creative performance: social networks and cinematic achievements in the Hollywood film industry. Organ Sci 19 (6):824–844

Cho R, Hassink R (2009) Limits to locking-out through restructuring: the textile industry in Daegu, South Korea. Reg Stud 43(9):1183–1198

Cohen WM, Klepper S (1992) The tradeoff between firm size and diversity in the pursuit of technological progress. Small Bus Econ 4(1):1–14

Coleman JS (1988) Social capital in the creation of human capital. Am J Sociol 94:95–120

Cooke P (2005) Rational drug design, the knowledge value chain and bioscience megacentres. Camb J Econ 29(3):325–341

Crespo J (2011) How emergence conditions of technological clusters affect their viability? Theoretical perspectives on cluster lifecycles. Eur Plan Stud 19(12):2025–2046

Eisingerich AB, Bell SJ, Tracey P (2010) How can clusters sustain performance? The role of network strength, network openness, and environmental uncertainty. Res Policy 39:239–253

Farrell J, Saloner G (1985) Standardization, compatibility, and innovation. RAND J Econ 16 (1):70–83

Frenken K (2006) Technological innovation and complexity theory. Econ Innov New Tech 15 (2):137–155

Grabher G, Stark P (1997) Organizing diversity: evolutionary theory, network analysis, and post-socialism. Reg Stud 31(5):533–544

Iammarino S, McCann P (2006) The structure and evolution of industrial clusters: transactions, technology and knowledge spillovers. Res Policy 35(7):1018–1036

Klepper S (2010) The origin and growth of industry clusters: the making of silicon valley and Detroit. J Urban Econ 67:15–32

Klepper S, Simons KK (1997) Technological extinctions of industrial firms: an inquiry into their nature and causes. Ind Corp Chang 6(2):379–460

König MD, Tessone CJ, Zenou Y (2010) From assortative to dissortative networks: the role of capacity constraints. Adv Complex Syst 13(4):483–499

Markusen A (1996) Sticky places in slippery space: a typology of industrial districts. Econ Geogr 72:293–313

Menzel MP, Fornahl D (2010) Cluster life cycles-dimensions and rationales of cluster evolution. Ind Corp Chang 19(1):205–238

Moore G (1991) Crossing the chasm: marketing and selling high-tech products to mainstream customers. HarperBusiness, New York

Newman MEJ (2003) Mixing patterns in networks. Phys Rev E 67:026126

Owen-Smith J, Powell WW (2004) Knowledge networks as channels and conduits: the effects of spillovers in the Boston biotechnology community. Organ Sci 15(1):5–21

Rivera M, Soderstrom S, Uzzi B (2010) Dynamics of dyads in social networks: assortative, relational, and proximity mechanisms. Annu Rev Sociol 36:91–115

Romanelli E, Khessina O (2005) Regional industrial identity: cluster configurations and economic development. Organ Sci 16(4):344–358

Saxenian AL (1990) Regional networks and the resurgence of silicon valley. Calif Manage Rev 33 (10):89–112

Scherngell T, Barber M (2011) Distinct spatial characteristics of industrial and public research collaborations: evidence from the 5th EU framework programme. Ann Reg Sci 46:247–266

Schilling MA, Phelps CC (2007) Interfirm collaboration networks: the impact of large-scale network structure on firm innovation. Manag Sci 53(7):1113–1126

Simmie J, Martin R (2010) Regional economic resilience: towards an evolutionary approach. Camb J Reg Econ Soc 3(1):29–51

Suire R, Vicente J (2009) Why do some places succeed when others decline? A social interaction model of cluster viability. J Econ Geogr 9(3):381–404

Suire R, Vicente J (2013) Clusters for life or life cycles of clusters: In Search for the critical factors of clusters resilience, entrepreneurship and regional development (forthcoming)

Todtling F, Trippl M (2004) Like phoenix from the ashes? The renewal of clusters in old industrial areas. Urban Stud 41:1175–1195

Vicente J, Balland PA, Brossard O (2011) Getting into networks and clusters: evidence from the Midi-Pyrenean GNSS collaboration network. Reg Stud 45(8):1059–1078

Watts DJ (2004) The "new" science of networks. Annu Rev Sociol 30:243–270

Part III
Structure and Spatial Characteristics of R&D Networks

Chapter 8
Observing Integration Processes in European R&D Networks: A Comparative Spatial Interaction Approach Using Project Based R&D Networks and Co-patent Networks

Rafael Lata, Thomas Scherngell, and Thomas Brenner

Abstract This study focuses on integration processes in European R&D by analyzing the spatio-temporal dimension of two different R&D collaboration networks across Europe. These networks cover different types of knowledge creation, namely co-patent networks and project based R&D networks within the EU Framework Programmes (FPs). Integration in European R&D – one of the main pillars of the EU Science Technology and Innovation (STI) policy – refers to the harmonisation of fragmented national research systems across Europe and to the free movement of knowledge and researchers. The objective is to describe and compare spatio-temporal patterns at a regional level, and to estimate the evolution of separation effects over the time period 1999–2006 that influence the probability of cross-region collaborations in the distinct networks under consideration. The study adopts a spatial interaction modeling perspective, econometrically specifying a panel generalized linear model relationship, taking into account spatial autocorrelation among flows by using Eigenfunction spatial filtering methods. The results show that geographical factors are a lower hurdle for R&D collaborations in FP networks than in co-patent networks. Further it is shown that the geographical dynamics of progress towards more integration is higher in the FP network.

R. Lata • T. Scherngell (✉)
Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria
e-mail: thomas.scherngell@ait.ac.at

T. Brenner
Economic Geography and Location Research, Philipps University Marburg, Marburg, Germany

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_8,

8.1 Introduction

Today it is widely recognised that *first*, innovation processes are increasingly based on interaction, research collaborations and networks of various actors (see, for instance, Powell and Grodal 2005),[1] and, *second*, innovation is the key element for sustained economic growth of firms, industries, regions and countries (see, for example, Romer 1990).[2] Based on these arguments, the main focus of the Europe 2020 Strategy is on research and innovation in order to achieve a new growth path that leads to a smart, sustainable and inclusive economy (European Commission 2011). In this context, the concept of the *Innovation Union* – one of the seven flagships scheduled in the Europe 2020 Strategy – is intended to improve conditions for innovation and knowledge diffusion to ensure that innovative ideas are efficiently turned into new products and services that create growth and employment (European Commission 2010). One of the main pillars of the *Innovation Union* is the realisation of an integrated European Research Area (ERA), defined as one explicit principal purpose to fulfil progress towards the Innovation Union.

The concept of the European Research Area (ERA) refers to the objective to enable and facilitate “free circulation of researchers, knowledge and technology” across the countries of the EU, and, by this, stimulating integration processes in European R&D (see Commission of the European Union (CEU) 2008, p. 6). This policy goal is to be addressed by improving coherence of the European research landscape, thus removing barriers – such as geographical, cultural, institutional and technological impediments – for knowledge flows, knowledge diffusion and researcher mobility by a European-wide coordination of national and regional research activities and policy programmes, including a considerable amount of jointly-programmed public research investment (see Delanghe et al. 2009).

To gain insight into the nature of integration processes in European R&D, there is urgent need for analysing the geographical dimension of R&D networks across

[1] The literature on R&D networks underlines the crucial importance of cooperative agreements between universities, companies and governmental institutes, for developing and integrating new knowledge in the innovation process (see Powell and Grodal 2005). This is explained by considerations that innovation nowadays takes place in an environment characterised by uncertainty, increasing complexity and rapidly changing demand patterns in a globalised economy. Organisations must collaborate more actively and more purposefully with each other in order to cope with increasing market pressures in a globalizing world, new technologies and changing patterns of demand. In particular, firms have expanded their knowledge bases into a wider range of technologies (Granstand 1998), which increases the need for more different types of knowledge, so firms must learn how to integrate new knowledge into existing products or production processes. It may be difficult to develop this knowledge alone or acquire it via the market. Thus, firms form different kinds of co-operative arrangements with other firms, universities or research organisations that already have this knowledge to get faster access to it.

[2] The theory of endogenous growth, and the geography-growth synthesis both consider that economic growth and spatial concentration of economic activities emanate from localised knowledge diffusion processes, in particular transferred via network arrangements between different actors of the innovation system.

Europe from a longitudinal perspective. The geography of such networks has – from a static perspective – attracted increasing interest in Regional Science and Economic Geography in the recent past. While from its beginning, the measurement of such phenomena has faced numerous problems, the empirical investigation of knowledge flows and R&D collaborations has significantly improved during the 1990s by using new indicators such as patent citations (see, for instance, the pioneering study by Jaffe et al. 1993; Fischer et al. 2006), co-publications (see, for instance, Hoekman et al. 2010) or project based R&D networks within the FPs (see Scherngell and Barber 2009, 2011), and by introducing new methods, in particular new spatial econometric techniques (see, for instance, LeSage et al. 2007). Recent studies focus on the structure of knowledge flows by adopting a spatial interaction modelling perspective (see, for instance, Scherngell and Barber 2009), employing a social network analysis perspective (see, for instance, Breschi and Cusmano 2004) or a combination of both (see Barber and Scherngell 2011).

However, as these studies just provide a static picture on the geography of R&D collaborations, novel questions arise – both in theoretical and empirical research as well as in a European policy context – regarding R&D network structures and its dynamics. Concerning the particular focus on integration processes in European R&D, the evolution of different kinds of separation effects over time – such as geographical, technological, institutional or cultural barriers – that determine R&D collaboration networks is of crucial interest. Thus, this study shifts emphasis to the investigation of the geographical dynamics of two different types of R&D collaboration networks across Europe, namely co-patent networks and project based R&D networks within the European Framework Programmes (FPs). We take these types of R&D collaboration networks to analyse integration processes in European R&D over time from two different angles, shifting attention to a comparison of European integration processes in these networks.

By this, the study addresses one of the major drawbacks of the current empirical literature: the lack of a longitudinal and comparative perspective on distinct R&D collaboration networks. Some exceptions are the studies of Maggioni and Uberti (2009), Hoekman et al. (2010, 2013), and Scherngell and Lata (2013).[3] The current study intends to complement the picture drawn in these studies, by shifting attention to a longitudinal and comparative perspective on two different R&D networks across Europe. The objective is to identify and compare the evolution of

[3] Hoekman et al. (2010) and Scherngell and Lata (2013) investigate the ongoing process of European integration by determining the impact of geographical distance and territorial borders on the probability of research collaborations between European regions. By analysing co-publication and FP network patterns and trends, the authors show that geographical distance has a negative effect on co-publication activities and FP cooperation, while for the FP networks this effect decreases over time. The study of Maggioni and Uberti (2009) focuses on the structure of knowledge flows by analysing four distinct collaboration networks, including co-patenting. Hoekman et al. (2013) focus on the effect of participation in FP networks on subsequent international publications, showing that the FPs indeed positively influence international co-publications, and, by this, seem to enhance integration across Europeans research systems.

geographical, technological, institutional or cultural effects that influence the probability for collaboration activities in the different collaboration networks, and provide direct evidence on integration processes in European R&D from different angles. We adopt a regional perspective that is an appropriate approach to observe different R&D collaboration networks in geographical space (see, for instance, Hoekman et al. 2010; Scherngell and Barber 2009) over the time period 1999–2006. The study employs a Poisson spatial interaction modelling perspective to address these research questions. We adjust the spatial interaction models by accounting for spatial autocorrelation issues of flows by means of Eigenvector spatial filtering (see Chun 2008; Scherngell and Lata 2013).

The paper is organised as follows. Section 8.2 sets forth the conceptual background of the study with a special focus on ERA, before Sect. 8.3 reflects on the different types of R&D collaboration networks under consideration. Section 8.4 describes the empirical setting and the data, accompanied by some descriptive statistics and exploratory spatial data analysis. Section 8.5 specifies the empirical model in form of a panel version of the spatial interaction modelling framework that is used to identify the evolution of separation effects influencing the probability of cross-region collaboration activities in the distinct networks. Section 8.6 presents the modelling results, before Sect. 8.7 closes with a summary of the main results and some conclusions in a European policy context.

8.2 The ERA Goal of Progress Towards More Integration in European R&D

One significant turning point in the EU Science, Technology and Innovation (STI) policy was the design of the concept of the European Research Area (ERA) presented at the Lisbon Council in the year 2000, rooted in the increasing awareness that European research activities suffer from diverse and fragmented national research systems (Boyer 2009). The overall goal of ERA is to overcome fragmentation in the European research system and to address the establishment of an 'internal market' for research across Europe, where researchers, technology and knowledge are supposed to circulate freely (see Delanghe et al. 2009; European Council 2000). The ERA green paper (CEC 2007) underlines the overall objectives of the Lisbon strategy, emphasising that the future European science and research landscape should be characterised by an adequate flow of competent researchers with high levels of mobility between institutions by integrated and networked research infrastructures and effective knowledge sharing, notably between public research and industry. This requires the reduction of geographical, cultural, institutional, and technological obstacles in order to generate research collaboration across European regions and countries (see, for instance, Hoekman et al. 2013; Scherngell and Lata 2013).

The Framework Programmes (FPs) of the European Commission (EC) constitute the main instrument to achieve this goal, shifting emphasis on supporting and stimulating collaborative R&D activities between innovating organisations across Europe, in particular firms and universities. At the same time, regional and national research policies deal with similar issues as reflected by a growing awareness among national policy makers that national efforts are often insufficient to keep pace in the international innovation competition. In this context the European Council underlined the importance of cross border cooperation for the achievement of these objectives and put collaborative R&D activities at the centre of its strategy (Guzzetti 2009). Svanveldt (2009) highlights the crucial importance of cross-border cooperation as instrument for adequately dealing these challenges.

During the last decade, the ERA concept has been developed further, becoming strong political support in the context of the conception of the so-called *Innovation Union* (European Commission 2010). As one of the seven flagships scheduled in the *Europe 2020 Strategy*, the Innovation Union is intended to improve framework conditions for innovation and knowledge diffusion. Moreover one of the main objectives of the Innovation Union is to "... quickly taking all measures necessary for a well functioning and coherent European Research Area in which researchers, scientific knowledge and technology circulate freely, in which RDI investments are less fragmented and the intellectual capital across Europe can be fully exploited" (European Commission 2010, p. 7). In order to tackle these challenges, specific commitments have been introduced. One of these commitments is to complete the ERA by 2014 with the goal to remove the remaining obstacles for collaborative knowledge production and consequently to foster the integration in the European research landscape (European Commission 2010).

With this in mind, the present study aims to evaluate the progress towards more integration in European R&D – as formulated in the concept of ERA and the Innovation Union. To gain empirical insight into the nature of such integration processes across Europe, the study focuses on a broad spectrum of R&D collaboration activities, namely co-patent networks and project based R&D networks within the FPs. In estimating the evolution of separation effects that capture the above mentioned obstacles for collaborative knowledge production across Europe, the analysis will show distinct mechanisms of integration processes corresponding to the different types of R&D networks. The section that follows reflects on the two different network types under consideration in some detail.

8.3 A Network Perspective on Integration in European R&D

R&D networks – defined as sets of organisations performing joint R&D activities – have attracted burst of attention in the recent past as essential element of modern knowledge production and innovation processes (see, for instance, Castells 1996).

In the current study, we take such network arrangements across Europe to analyse integration processes in European R&D, focusing on two different types of networks that capture different types of knowledge production processes. We focus on R&D networks in the form of joint patenting, resulting in co-patents, and project based R&D networks within the FPs.

Co-patent networks mainly reflect research collaborations that are related to applied knowledge generation focusing on the development of marketable innovations and industry research activities (Maggioni and Uberti 2009). Patents represent a well established indicator of knowledge generation activities and are widely used in empirical studies on knowledge flows (see, for instance, Jaffe et al. 1993; Fischer et al. 2006). A co-patent is defined as a patent invented by at least two inventors from two different organisations. Therefore, it represents knowledge exchange across actors within an inventor network in the process of patenting an invention (see, for instance, Ejermo and Karlsson 2006).

The second type of R&D networks refers to project based R&D collaboration within the FPs. While co-patent networks mainly reflect applied research, project based FP networks involve basic and applied research aspects, given by the fact that publications and patents may be outputs of FP networks. In the FP network, the research collaboration is constituted by joint R&D projects conducted by organisations distributed across Europe. The FPs are the main political instrument to support pre-competitive collaborative R&D within the European Union. The key objectives are, *first*, to strengthen the scientific research and technological development in the scientific landscape, and, by this, to foster the European competitiveness, and, *second*, to promote research activities in support of other EU policies (Maggioni et al. 2009).[4] FP projects share specific characteristics (see for example Roediger-Schluga and Barber 2006). *First*, they are all promoted by self-organised consortia and have distinct partners – for instance individuals, industrial and commercial firms, universities, research organisations, etc. – that are located in different EU members and associated states. *Second,* they focus primarily on pre-competitive R&D projects. *Third*, they are characterised by less market orientation and longer development periods (Polt et al. 2008).

Given the properties of the two different network types under consideration, it may be hypothesised that integration processes for these network types differ. This may, on the one hand, be related to the different knowledge generation processes in these networks, on the other hand, to governance rules and policy programmes implemented by the EC influencing the resulting network structures. Spatial interaction models (see Sect. 8.5) will enable us to proof this hypothesis, and disclose distinct spatial characteristics and collaboration patterns in the networks under

[4] Since their introduction in 1984, different thematic aspects and issues of the European scientific landscape have been addressed by the FPs. Although the FPs have undergone different changes in their orientation during the past years, their fundamental rational remained unchanged (Roediger-Schluga and Barber 2006).

consideration, and, by this, drawing a more detailed picture on integration processes in European R&D.

8.4 Data and Descriptive Statistics

In our empirical analysis we aim to investigate integrations processes in European R&D networks focusing on two different types of collaboration networks, that is FP collaboration networks and co-patent networks. The EUPRO database is used to capture project based R&D networks within the FPs, while the Regpat database is taken to construct co-patent networks. The *EUPRO database* currently comprises information on more than 60,000 research projects funded by the EU FPs and all participating organisations. A network link is given between two organisations when they conduct a joint research project in the FPs. We use information on the geographical location in form of the city to trace the geographical dimension of the network. The *Regpat database* contains information on patent applications from various patent offices worldwide. It is provided by the OECD and contains, among many others, all patent applications issued at the European Patent Office (EPO), and the national patent offices of the European countries. A network link between two organisations is given when inventors from two different organisations appear on a patent application. We use information on the inventor address of an EPO patent application to trace the origin of the invention.

The European coverage is achieved by using $i, j = 1, \ldots, n$ NUTS-2 regions[5] of the 25 pre-2007 EU member-states as well as Norway and Switzerland. We extract n-by-n collaboration matrices for each time period $t = 1, \ldots, T$, both for the FP- and for the co-patent network, by aggregating the number of individual collaborative activities at the organisational level in time period t to the regional level. This leads to the observed number of R&D collaborations y_{ijt} between two regions i and j in time period t in the respective network, that is the FP and the co-patent network. The resulting regional collaboration matrix $\boldsymbol{Y}_t$ for the two networks[6] for a given year t contains the collaboration intensities between all (i, j)-region pairs, given the $i = 1, \ldots, n$ regions in the rows and the $j = 1, \ldots, n$ regions in the columns.[7] Figure 8.1 illustrates the spatial distribution of the cross-region R&D collaborations in the FP- (Fig. 8.1a) and the co-patent network (Fig. 8.1b) across Europe. In the

[5] Although substantial size differences and interregional disparities of some regions exist, these units are widely recognized to be an appropriate level for modelling and analysis purposes (see, for example, LeSage et al. 2007).

[6] Note that we do not distinguish between the FP network and the co-patent in the formal description of data as well as the modelling approach in the section that follows.

[7] We use a full counting procedure for the construction of our collaboration matrices (see, for example, Katz 1994). For a project with, for example, three different participating organizations a, b and c, which are located in three different regions, we count three links (from a to b, from b to c and from a to c).

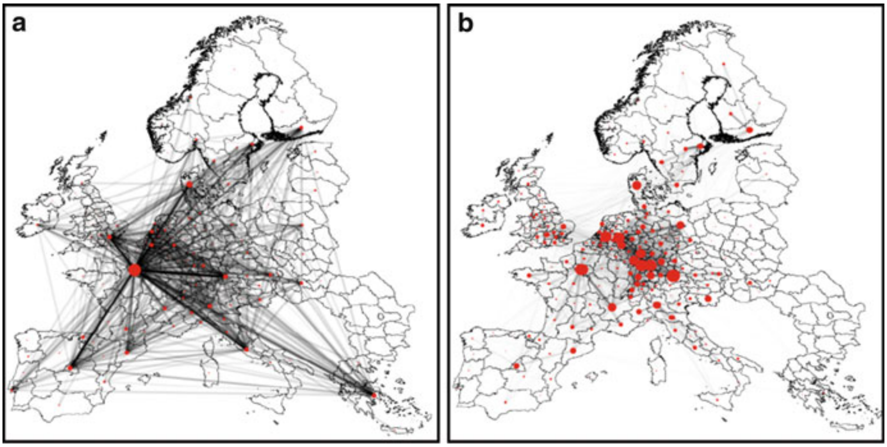


Fig. 8.1 Spatial distribution of the cross-region R&D networks for the year 2006. (**a**) R&D collaborations within the FP-network. (**b**) R&D collaborations within the co-patent network

spatial network maps, the sizes of the nodes are proportional to the number of regional participations in the two distinct networks. The darkness of the lines corresponds with the number of joint R&D collaborations between two regions, i.e. the darker the higher the interaction intensity. It is shown that the spatial structures of the distinct networks differ markedly. The most striking difference concerns the fact that the international collaboration activity is much higher in the FP network than in the co-patent network. In the latter, R&D collaborations are widely confined within national boundaries, while such boundaries seem to play a minor role for the structure of the FP network. Furthermore, the intra-regional collaboration intensity seems much higher in the co-patent network than in the FP network, pointing to the geographical localisation of the co-patents within NUTS-2 regions, while the cross-region collaboration intensity is much higher in the FP network.

Concerning the spatial distribution of the regions with high intra-regional co-patent activities, a high intensity can be found for regions belonging to the traditional industrial core of Europe (see Hoekman et al. 2012), also referred to as the European ‘blue banana’ (Brunet 2002), while the participation within the FP network seems to be spatially more dispersed. However, both networks seem to be spatially concentrated in some European regions that show high collaboration intensity. In this context the question arises, whether a spatial clustering of interaction patterns in the two networks can be observed, and which network shows a higher degree of spatial clustering, also referred to as spatial autocorrelation of flows (see, for instance, Berglund and Karlstrom 1999). Spatial autocorrelation of flows is, for example, when flows from a particular origin may be correlated with other flows that have the same origin, and, similarly, flows into a particular destination may be correlated with other flows that have the same destination (Scherngell and Lata 2013). In our case, this means that the intensity of R&D

collaborations from an origin region i to a destination region j may be correlated with the intensity of R&D collaborations from the same origin i to another destination j, or vice versa. Such a situation is specifically interesting from the perspective of our research question on integration in European R&D, namely by assessing whether such R&D collaborations are statistically concentrated to a geographical core of regions that are located nearby to each other.[8]

In order to test for the existence of spatial autocorrelation of flows, we calculate a Moran's I test for spatial dependence as widely used in exploratory spatial data analysis (see Griffith 2003), given by

$$I_t = \frac{\boldsymbol{y}_t{}' \boldsymbol{W}^* \boldsymbol{y}_t}{\boldsymbol{y}_t{}' \boldsymbol{y}_t} \tag{8.1}$$

where $\boldsymbol{y}_t$ is a vector of our observed collaboration flows at time t with $N = n^2$ elements $(y_{ijt}) = (y_{11t}, \ldots, y_{1nt}, y_{21t}, \ldots, y_{2nt}, \ldots, y_{n1t}, \ldots, y_{nnt})$, and $\boldsymbol{W}^*$ is defined by $\boldsymbol{W} \otimes \boldsymbol{W}$ where $\boldsymbol{W}$ is the n-by-n spatial weights matrix and $\otimes$ denotes the Kronecker product. For $\boldsymbol{W}$, we set

$$w_{ij} = \begin{cases} 1 \text{ if } s_{ij}^{(1)} \leq s_{ig_{(i)}}^{(1)} \\ 0 \text{ otherwise} \end{cases} \tag{8.2}$$

where $s_{ij}{}^{(1)}$ measures the great circle distance between the economic centers of two regions i and j, and g_i denotes the g-nearest neighbour of i. We define $g = 5$, as used in various empirical studies dealing with European regions (see, for instance, Scherngell and Lata 2013). The respective Moran's I statistics for the years 1999–2006 are reported in Table 8.1. The results are most often significant pointing to substantial spatial autocorrelation of R&D collaborations in both networks under consideration, i.e. a high number of flows is correlated with flows that come from nearby origins, and going into nearby destinations. However, the degree of spatial dependence is much higher for the co-patent network as has been expected considering the spatial distribution of the flows that are visualised in Fig. 8.1. Furthermore, the Moran's I for the FP network shows a decreasing trend, while for the co-patent network no time trend can be observed, pointing to differences in

[8] From a theoretical perspective the spatial autocorrelation of R&D collaboration flows may be explained by the assumption that the collaboration behaviour of one region influences the collaboration behaviour of neighbouring regions because – as described in various empirical studies – contiguity of regions may induce knowledge flows between them, to them, and from them, and, thus, evoke the transfer of information on potential collaboration partners that are located further away (Scherngell and Lata 2013). To give an example, if region A has many collaborations with region B (that is no neighbour of region A), region A may influence a neighbouring region C also to collaborate with region B due to information flows between region A and region C, in particular flows of 'know who' type information (see Cohen and Levinthal 1990).

Table 8.1 Spatial autocorrelation of R&D collaboration in two distinct networks

	Moran' I							
	1999	2000	2001	2002	2003	2004	2005	2006
FP-network	0.016*	0.006*	0.003*	0.000	–0.001	0.007*	–0.009	–0.001
Co-patent network	0.136*	0.120*	0.132*	0.144*	–0.139*	0.153*	0.146*	–0.147*

*significant at the 0.001 significance level

integration processes for the two network types. In this context, the existence of spatial autocorrelation also bears important implications in a modeling context, since estimates may be biased neglecting spatial autocorrelation issues of flows (see, for instance, Fischer and Griffith 2008; Scherngell and Lata 2013).

8.5 The Empirical Model

This section shifts direct attention to the modelling approach used to estimate how specific separation effects influence the variation of cross-region R&D collaborations in two distinct collaboration networks over time, and, by this, providing direct evidence on distinct integration processes in different types of R&D. We employ a spatial interaction modelling approach.[9] In implementing a panel version of the spatial interaction model, we are able to identify time effects that are necessary to observe potential integration processes of the networks over the time period 1999–2006. In what follows we will specify the panel version of the spatial interaction model, an extension accounting for spatial autocorrelation issues of flows, and describe the independent variables of the model.

8.5.1 The Panel Version of the Spatial Interaction Model to Be Estimated

Let us denote Y_{ijt} as a random dependent variable corresponding to observed R&D collaborations y_{ijt} within the FP- or the co-patent network between origin i ($i = 1, \ldots, n$) and destination j ($j = 1, \ldots, n$) at time t ($t = 1, \ldots, T$). As in the previous section, we do not distinguish between the two networks in the formal model presentation; our basic model is given by

[9] Spatial interaction models are widely used for modelling origin-destination flows data and were used to explain different kinds of flows, such as migration, transport or communication flows, between discrete units in geographical space (see, for instance, Fischer and LeSage 2010 among many others).

$$Y_{ijt} \middle| y_{ijt} = \mu_{ijt} + \varepsilon_{ijt} \quad i,j = 1, \ldots, n; \;\; t = 1, \ldots, T \tag{8.3}$$

where μ_{ijt} denotes some mean expected interaction frequency between origin i and destination j at time t, ε_{ijt} some disturbance term about the mean with the property $E[\varepsilon_{ijt}|y_{ijt}] = 0$. As in classical spatial interaction theory (see, for instance, Fischer and LeSage 2010), we model the mean interaction frequencies μ_{ijt} between origin i and destination j at time t by some origin function O_{it} which characterizes the origin i of interaction in time period t, some destination function D_{jt} which describes the destination j of interaction in time period t, and some separation function S_{ijt} which accounts for the separation between an origin region i and a destination region j in time period t. Then we use a multiplicative relationship for our basic model, given by

$$\mu_{ijt} = O_{it}\, D_{jt}\, S_{ijt} \quad i,j = 1, \ldots, n; \;\; t = 1, \ldots, T \tag{8.4}$$

where

$$O_{it} = o_{it}^{\alpha_1} \quad i,j = 1, \ldots, n; \;\; t = 1, \ldots, T \tag{8.5}$$

$$D_{jt} = d_{jt}^{\alpha_2} \quad i,j = 1, \ldots, n; \;\; t = 1, \ldots, T \tag{8.6}$$

$$S_{ijt} = \exp\left[\sum_{k=1}^{K} \beta_k\, s_{ijt}^{(k)}\right]. \quad i,j = 1, \ldots, n; \;\; t = 1, \ldots, T \tag{8.7}$$

o_{it} and d_{jt} are origin and destination variables, $s_{ijt}^{(k)}$ are K ($k = 1, \ldots, K$) separation variables that are introduced below. α_1, α_2 and β_k are parameters to be estimated.

As has come into fairly wide use for spatial interaction models, we assume $(Y_{ij}) \sim$ *Poisson* due to the true integer non-negative count nature of our R&D collaboration flows (see, for instance, Cameron and Trivedi 1998; Fischer et al. 2006). The resulting panel version of the Poisson spatial interaction model is given by,

$$\mu_{ijt} = \exp\left[\alpha_1 \log(o_{it}) + \alpha_2 \log\left(d_{jt}\right) + \sum_{k=1}^{K} \beta_k s_{ijt}^{(k)} + \gamma_{ij}\right] \tag{8.8}$$

where γ_{ij} denotes the unobserved individual specific effect, also referred to as the one-way error component model (see Baltagi 2008). The random term γ_{ij} is time invariant but varies across all (i, j)-region pairs. In our case γ_{ij} accounts for region-pair specific effects that are not included in the model. We assume the γ_{ij} to be correlated across our time periods for the same (i, j)-region pair, i.e. we follow a *random effects* specification, and integrate out the random effect γ_{ij} of the joint probability $\prod_{t=1}^{T} \Pr(y_{ij1}, \ldots, y_{ijT})$ by obtaining

$$\Pr\left(y_{ij1}, \ldots, y_{ijT}\right) = \int \Pr\left(y_{ij1}, \ldots, y_{ijT}, \gamma_{ij}\right) d\gamma_{ij} = \int \Pr\left(y_{ij1}, \ldots, y_{ijT} \middle| \gamma_{ij}\right) g\left(\gamma_{ij}\right) d\gamma_{ij}. \tag{8.9}$$

Note that this is the same approach used in models for event counts to condition the heterogeneity out of the Poisson model to produce the Negative Binomial model (see Baltagi 2008), i.e. when $(Y_{ij}) \sim$ *Poisson* with mean μ_{ijt} as given by Eq. 8.8, and $\exp(\gamma_{ij}) \sim$ *Gamma*, then our *random effects Negative Binomial spatial interaction model* to be estimated is

$$\Pr\left(y_{ij1}, \ldots, y_{ijT}\right) = \frac{\left(\prod_{t=1}^{T} \mu_{ijt}{}^{y_{ijt}}\right) \Gamma\left(\theta + \sum_{t=1}^{T} y_{ijt}\right)}{\left(\Gamma(\theta) \prod_{t=1}^{T} y_{ijt}!\right) \left[\left(\sum_{t=1}^{T} \mu_{ijt}\right)^{\sum_{t=1}^{T} y_{ijt}}\right]} Q_i (1 - Q_i)^{\sum_{t=1}^{T} y_{ijt}} \tag{8.10}$$

with

$$Q_i = \frac{\theta}{\theta + \sum_{t=1}^{T} \mu_{ijt}} \tag{8.11}$$

where $\Gamma(.)$ denotes the Gamma distribution and θ its variance. Parameter estimation is achieved via maximum likelihood estimation procedures (see Cameron and Trivedi 1998).

8.5.2 *Accounting for Spatial Autocorrelation and Time Effects*

Given the results of the spatial autocorrelation analysis of the previous section, it can be assumed that spatial dependence among our collaboration flows may lead to biased estimates. Thus, we re-specify our panel version of the Negative Binomial spatial interaction model by accounting for spatial autocorrelation issues as well as by introducing time effects enabling us to infer on time trends concerning the evolution of collaboration patterns in the two networks.

As noted by Chun (2008), maximum likelihood estimation assumes that all observations, in our case collaboration flows in our two networks under consideration, are mutually independent. A violation of this assumption may be in particular induced by spatial autocorrelation of flows leading to incorrect inferences due to inconsistence of the standard errors, and, thus, unrealistic significances (Chun 2008;

Griffith 2003).[10] We follow Scherngell and Lata (2013) who apply a spatial filtering method to filter out spatial autocorrelation of residual flows in a Negative Binomial spatial interaction context. The essence of the spatial filtering approach is to extract eigenvectors from a modified spatial weights matrix that serve as spatial surrogates for omitted spatially autocorrelated origin and destination variables (see Fischer and Griffith 2008). These proxy variables are extracted as n eigenvectors[11] from the modified spatial weights matrix of the form $\left(\boldsymbol{I}-\boldsymbol{1}\boldsymbol{1}^{T}\frac{1}{n}\right)\boldsymbol{W}\left(\boldsymbol{I}-\boldsymbol{1}\boldsymbol{1}^{T}\frac{1}{n}\right)$ with $\boldsymbol{I}$ denoting the n-by-n identity matrix, $\boldsymbol{1}$ is an n-by-1 vector of one's, $\boldsymbol{1}^T$ its transpose, and $\boldsymbol{W}$ the n-by-n spatial weights matrix, as defined by Eq. 8.2. The eigenvectors can be interpreted as synthetic map variables that represent specific natures and degrees of potential spatial autocorrelation (Chun 2008; Griffith 2003).

As noted by Griffith (2003) it is not appropriate to use the full set of $\boldsymbol{E}_n$ eigenvectors for the construction of the spatial filter variables. Further, we face a situation where Eigenvectors have to be selected for each time period due to the panel version of the spatial interaction model (Patuelli et al. 2011). As in Patuelli et al. (2011) we select in a *first* step a subset of distinguished eigenvectors on the basis of their Moran's I values. Then, we follow Fischer and Griffith (2008) and extract those Eigenvectors $\boldsymbol{E}_m$ that show a higher Moran's I value than 0.25. In a *second* step, it is necessary to adapt these Eigenvectors to our spatial interaction framework; origin candidate eigenvectors are drawn from $\boldsymbol{1} \otimes \boldsymbol{E}_m$ and the destination candidate eigenvectors are obtained from $\boldsymbol{E}_m \otimes \boldsymbol{1}$. In a *third* step, these Eigenvectors are added as explanatory variables to $T = 9$ cross-section versions of the Negative Binomial spatial interaction model, from which statistically significant Eigenvectors are identified. In a *fourth* step, we determine those eigenvectors that are significant over all time periods and define the resulting set of common origin and destination eigenvectors, $\boldsymbol{E}_q$ and $\boldsymbol{E}_r$, respectively, as our time invariant

[10] One way to capture spatial autocorrelation of flows is the use of spatial autoregressive techniques (LeSage and Pace 2008). An alternative approach is the use of spatial filtering methods. The key advantage of the spatial filtering approach is that it can be applied to any functional form and thus, does not depend on normality assumptions (Patuelli et al. 2011). Consequently, we prefer the spatial filtering approach over spatial autoregressive model as we are dealing with a Poisson spatial interaction framework.

[11] The extracted eigenvectors have several characteristics. First, as shown by Griffith (2003), each extracted eigenvector relates to a distinct map pattern that has a certain degree of spatial autocorrelation. Second, the selected eigenvectors are centered at zero due to the pre and post multiplication of $\boldsymbol{W}$ by the standard projection Matrix $\left(\boldsymbol{I}-\boldsymbol{1}\boldsymbol{1}^{T}\frac{1}{n}\right)$. Third, the modification of $\boldsymbol{W}$ ensures that the eigenvectors provide mutually orthogonal and uncorrelated map patterns ranging from the highest possible degree of positive spatial correlation to highest possible degree of negative spatial correlation as given by the Moran's I (MI). (Griffith 2003). Hence, the first extracted eigenvector is the one showing the highest degree of positive spatial autocorrelation that that can be achieved by any spatial recombination; the second eigenvector has the largest achievable degree of spatial autocorrelation by any set that is uncorrelated with until the last extracted eigenvector will maximize negative spatial autocorrelation (Griffith 2003).

spatial filter.[12] The time invariant spatial filter covers the total number of space-time observations, and accounts for spatial dependence of flows in our origin and destination data.

We add the selected origin filters $\boldsymbol{E}_q$ and destination filters $\boldsymbol{E}_r$ as regressors to our panel version of the Negative Binomial spatial interaction model. Further we introduce the subset of $\boldsymbol{Z}_t$ time dummies in order to capture aggregate year effects (Woodridge 2008).[13] This leads to the spatially filtered panel version of the Negative Binomial spatial interaction model accounting for time effects, given by re-specifying the conditional mean μ_{ijt} so that

$$\mu_{ijt} = \exp\left[\sum_{q=1}^{Q} E_q \psi_q + \alpha_1 \log(o_{it}) + \sum_{r=1}^{R} E_r \varphi_r + \alpha_2 \log\left(d_{jt}\right) + \sum_{k=1}^{K} \beta_k s_{ijt}^{(k)} + \sum_{t=1}^{T} Z_t \nu_t + \gamma_{ij}\right] \tag{8.12}$$

The coefficients to be estimated for the spatial filters are ψ_q and φ_r, ν_t is the associated parameter for the time dummy at time t.

8.5.3 Independent Variables

We use one origin measure, and one destination measure for the FP network model and the co-patent network model. For the model on the FP networks, the origin variable o_{it} is measured in terms of organizations participating in joint FP projects in region i, while the destination variable d_{it} denotes the number of organizations participating in joint FP projects in region j. For the co-patent network model, the origin variable o_{it} is measured in terms of the number of co-patents in region i, while the destination variable d_{it} denotes the number of co-patents in region j.

From the background of our research focus our interest is on $K = 5$ separation measures: $s_{ijt}^{(1)}$ measures the geographical distance between the economic centres of two regions i and j in time period t, by using the great circle distance.[14] $s_{ijt}^{(2)}$ is a neighbouring region dummy variable that takes a value of one if the regions i and

[12] We use an time invariant specification of the spatial filter as we assume an time invariant underlying spatial process.

[13] In order to determinate changes of our separation variables we include interaction terms (see, for an overview, Wooldridge 2008). In this procedure, variables of interest, for example R&D (see, Griliches 1984), interact with time dummy variables and illustrate if effects changed over a certain time period or not. In our case (time) interaction terms represent the interaction between our separation variables and the time dummies and determinate how separation effects have changed over time. These interaction terms pick up the inter-temporal variation of our separation effect and remain only cross-sectional variation.

[14] Note further that according to Bröcker (1989), we calculate the intraregional distance as $s_{ii}^{(1)} = (2/3)\ (A_i/\pi)^{0.5}$, where A_i denotes the area of region i, i.e. the intraregional distance is two third the radius of an presumed circular area.

j in time period t are direct neighbours, and zero otherwise. $s_{ijt}^{(3)}$ is a country border dummy variable that we use as a proxy for institutional barriers. The variable takes a value of zero if two regions i and j in time period t are located in the same country, and one otherwise. $s_{ijt}^{(4)}$ is a language dummy variable accounting for cultural barriers that takes a value of zero if two regions i and j in time period t are located in the same language area, and one otherwise.[15] $s_{ijt}^{(5)}$ captures technological distance by using regional patent data from the European Patent Office (EPO). The application date is used to extract the data for each year of our time frame. We follow Moreno, Paci and Usai (2005) and construct a vector for each region i that contains region i's share of patenting in each of the technological subclasses of the International Patent Classification (IPC). Technological proximity between two regions i and j in time period t is given by the uncentred correlation between their technological vectors.

8.6 Estimation Results

Table 8.2 reports the results from the estimation of the spatially filtered random effects Negative Binomial spatial interaction models as specified in the previous section. Standard errors are given in brackets. The first column presents the results for the FP network, while the second column contains the estimates for the co-patent network. As can been seen, the estimates for the origin, destination and separation variables are most often statistically significant. The bottom of the table presents some model diagnostics that are of methodological interest.[16]

The results are interesting in the context of the geography of innovation literature, but also very relevant and insightful from a European STI policy perspective. Geographical distance, as evidenced by the estimate of β_1, exerts in both networks, the FP network and the co-patent network, a negative effect on collaboration probability, i.e. in both networks R&D collaboration intensity between two regions significantly decreases when they are located further away in geographical distance, and this effect seems only to differ slightly in magnitude. However, concerning other geographical factors, we find a much stronger negative effect in the co-patent network than in the FP network. One striking result concerns the high negative effect of country borders, as evidenced by the estimate for β_3, for the co-patent network as compared to the FP network, showing that for R&D collaborations in the FPs country borders constitute only a low hurdle.

[15] Language areas are defined by the region's dominant language. However, in most cases the language areas are combined countries, as for instance Austria, Germany and Switzerland (one exception is Belgium, where the French speaking regions are separated from the Flemish speaking regions).

[16] The dispersion parameter is statistically significant in both model versions, indicating that the Negative Binomial specification is essential to account for overdispersion in the data. A likelihood ratio test which compares the panel estimator with the pooled estimator confirms the appropriateness of the random effects specification.

Table 8.2 Estimation results of the spatially filtered random effects negative binomial spatial interaction models

	FP-network	Co-patent network
Origin and destination variable $[\alpha_1] = [\alpha_2]$	0.955^{***} (0.001)	0.354^{***} (0.003)
Geographical distance $[\beta_1]$	-0.209^{***} (0.005)	-0.266^{***} (0.005)
Neighbouring region $[\beta_2]$	0.229^{***} (0.021)	0.710^{***} (0.017)
Country border effects $[\beta_3]$	-0.063^{***} (0.016)	-1.058^{***} (0.016)
Language area effects $[\beta_4]$	-0.164^{***} (0.013)	-0.740^{***} (0.014)
Technological distance $[\beta_5]$	-0.305^{***} (0.018)	-1.536^{***} (0.023)
Number of significant time effects	7	2
Number of origin spatial filters	32	39
Number of destination spatial filters	29	47
Constant $[\alpha_0]$	-9.799^{***} (0.045)	-2.426^{***} (0.041)
Dispersion parameter	19.804^{***} (0.253)	2.722^{***} (0.045)
LR test (spatial filters)	$1{,}335.17^{***}$	$4{,}932.10^{***}$
LR test (random effects)	$190{,}354.7^{***}$	$30{,}634.3^{***}$
LR test (overdispersion)	$281{,}497.1^{***}$	$2{,}232{,}645.8^{***}$
Log likelihood	−879,642.1	−435,630.7

Notes: ***significant at the 0.001 significance level; The LR Test (spatial filter) is a Likelihood Ratio test that compares the model fit of the spatially filtered model against the unfiltered model versions. The test statistic is significant for both models. Thus the spatially filtered model specification is appropriate. The LR Test (random effects) is a Likelihood Ratio test that compares the panel estimator with the pooled estimator. The significant values confirm the importance of a random effects specification. The LR Test (overdispersion) is the Likelihood Ratio Test that compares the random effects negative binomial model to the random effects Poisson specification. A significant value points to the existence of overdispersion, namely, the negative binomial specification is to be preferred to the Poisson specification

In addition, co-patent networks seem to be to a high degree focused on neighbouring regions, i.e. the collaboration significantly increases when two organisations are located in regions that share a common border (β_2). This effect is much higher than in the FP network, pointing to a stronger spatial concentration and geographical localisation of R&D collaborations reflected by co-patents. Concerning language area effects (β_4), we also find considerable differences between the FP network and the co-patent network. The negative effect of language is much higher for the co-patent network than for the FP network, i.e. the probability that organisations located in two different language areas collaborate is much lower in the co-patent network. This may be explained by the fact that the co-patent networks are much more subject to the industry sector, where such language barriers may – as suggested by results provided from Scherngell and Barber (2011) – constitute a lower hurdle than for research including public research organisations, in particular universities. Technological distance (β_5) is the most important determinant for cross-region R&D collaborations in both networks, and, by this, earlier results by Scherngell and Barber (2009, 2011) or Fischer et al. (2006) are confirmed.

However, the effect is much stronger in the co-patent network, which is to be expected since co-patent networks are more application oriented, where specific technologies and technological devices are more important. Furthermore, the FPs are intended to support in particular interdisciplinary knowledge production. Overall, in the context of our focus on integration in European R&D, we can infer that integration is much higher in the FP network than in the co-patent network, as most of the separation variables exert a higher negative effect. This result has been expected, since more applied oriented, competitive research is subject to a minor group of actors often located within one region. The precompetitive character of knowledge production in the FPs may lead to a higher propensity to share this knowledge with partners, while patenting is to a larger degree subject to strategic considerations of the innovating organisation. However, having in mind the ERA goal of progress towards more integration in European R&D, covering different phases of R&D, one may conclude that barriers hampering collaborations in the co-patent network – for instance language barriers or country borders – should be addressed more thoroughly. This may be done by education programs for overcoming language barriers or policy initiatives that remove institutional hurdles for collaborations in patenting, though, one have to be clear that due to the competitive character of this type of research, such patterns may never fully disappear.

However, in order to be able to gain empirical insight into progress towards more integration, we need to reflect on time trends. For this reason we look at interaction terms between selected separation variables and our time dummies. Table 8.3 presents the results for these interaction terms in the two networks for the years 2000–2005.

The most striking result is that all separation variables accounting for spatial effects significantly decline in the FP-network, i.e. the FP network becomes more geographically integrated over the observed time period. This cannot be observed for the co-patent network. In particular for the years 2004 and 2005 we cannot identify a significant interaction effect between time and spatial separation variables, i.e. progress towards more integration cannot be observed, while this progress can be clearly observed for the FP network.

8.7 Conclusions

The focus of this study has been on the nature of integration processes in European R&D. More specifically we have shifted emphasis to the investigation of the geographical dynamics of two different types of R&D collaboration networks across Europe, namely co-patent networks and project based R&D networks within the EU Framework Programmes (FPs). Adopting a spatially filtered panel version of the Negative Binomial spatial interaction model, we have identified and compared geographical, technological, institutional and cultural effects that influence

Table 8.3 Time trends for identifying distinct geographical integration patterns in the networks

Time interaction terms	FP-network					
	2000	2001	2002	2003	2004	2005
Geographical distance	−0.057***	−0.042***	−0.039***	−0.033***	−0.010***	−0.003***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Neighbouring region	0.171***	0.130***	0.089***	0.088***	0.029**	0.002
	(0.012)	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)
Country border effects	−0.083***	−0.073***	−0.074***	−0.060***	−0.018***	−0.000***
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)

Time interaction terms	Co-patent network					
	2000	2001	2002	2003	2004	2005
Geographical distance	−0.030***	−0.016***	−0.014***	−0.007	−0.006	−0.008
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Neighbouring region	0.104**	0.057**	0.113***	0.043	0.003	−0.024
	(0.029)	(0.029)	(0.029)	(0.024)	(0.028)	(0.028)
Country border effects	−0.087***	−0.059***	−0.065***	−0.064**	−0.003	0.018***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.022)	(0.022)

***significant at the 0.001 significance level, **significant at the 0.01 significance level

the probability for collaboration activities in the different collaboration networks over time, and, by this, have provided novel evidence on integration processes in European R&D.

The most elemental and important result, both in the context of the literature on the geography of innovation as well as in a European policy context, is that integration in FP networks seems to be much higher than in the co-patent network. This is underpinned by the strong intra-national character of the co-patent network in contrast to the FP network, as well as the higher geographical localisation of co-patent collaboration activities within narrow geographical boundaries. These results may on the one hand be explained by the different nature of the knowledge creation process in the two networks, but also by policy related circumstances, in that the FP programmes explicitly foster integration processes, and at the same time more policy efforts should be envisaged that ease collaboration in more applied oriented research.

Methodologically, the study is interesting as it breaks new ground by estimating a panel version of the Negative Binomial spatial interaction model accounting for spatial autocorrelation of flows. Though robustness of the model may be tested further, the methodological approach seems to be an important contribution to the debate on spatial autocorrelation issues of flows, applied to a panel data structure posing additional modelling requirements that have been applied in this study.

Some ideas for future research come to mind. *First*, the estimation of time trends, for instance by means of a dynamic version of the spatial interaction model, is a core subject for future research, requiring both theoretical as well as computational advancements. *Second*, the inclusion of other types of R&D

networks in the comparative analysis, in particular co-publication networks, is essential to complement the results provided by the current study.

Acknowledgments This work has been funded by the FWF Austrian Science Fund (Project No. P21450). We are grateful to Manfred M. Fischer (Vienna University of Economics) and Michael Barber (AIT) for valuable comments on an earlier version of the manuscript.

References

Baltagi B (2008) Econometric analysis of panel data, 4th edn. Wiley, Chichester

Barber MJ, Scherngell T (2011) Is the European R&D network homogeneous? Distinguishing relevant network communities using graph theoretic and spatial interaction modelling approaches. Reg Stud 46:1–16

Berglund S, Karlström A (1999) Identifying local spatial association in flow data. J Geogr Syst 1 (3):219–236

Boyer R (2009) From the Lisbon agenda to the Lisbon treaty: national research systems in the context of European integration and globalization. In: Delanghe H, Muldur U, Soete L (eds) European science and technology policy. Towards integration or fragmentation? Edward Elgar, Cheltenham, pp 101–126

Breschi S, Cusmano L (2004) Unveiling the texture of a European research area: emergence of oligarchic networks under EU framework programmes. Int J Technol Manag 27(8):747–772, Special Issue on Technology Alliances

Bröcker J (1989) Partial equilibrium theory of interregional trade and the gravity model. Pap Reg Sci 66:7–18

Brunet R (2002) Lignes de force de l'espace Européen. Mappemondo 66:14–19

Cameron AC, Trivedi PK (1998) Regression analysis of count data. Cambridge University Press, Cambridge/New York

Castells M (1996) The rise of the network society. Blackwell, Malden

CEC (Commission of the European Communities) (2007) Green paper the European research area: new perspectives. {SEC(2007) 412}, COM(2007)161 final, Brussels, 4 April 2007

CEU (Commission of the European Union) (2008) Council conclusions on the definition of a "2020 vision for the European research area". 16012/08 RECH 379 COMPET 502, Brussels, 9 Dec 2008

Chun Y (2008) Modeling network autocorrelation within migration flows by eigenvector spatial filtering. J Geogr Syst 10:317–344

Cohen WM, Levinthal DA (1990) Absorptive capacity: a new perspective on learning and innovation. Adm Sci Q 35(1):128–152

Delanghe H, Muldur U, Soete L (eds) (2009) European science and technology policy. Towards integration or fragmentation? Edward Elgar, Cheltenham

Ejermo O, Karlsson C (2006) Interregional inventor networks as studied by patent coinventorships. Res Policy 35(3):412–430

European Council (2000) Presidency conclusions, Lisbon European Council, 23–24 March 2000

European Council (2010) Conclusions on innovation union for Europe. Brussels, 26 Nov

European Commission (2011) Progress report on the Europe 2020 strategy. Brussels, European Commission, Com (2011) 815 final

Fischer MM, Griffith D (2008) Modeling spatial autocorrelation in spatial interaction data: an application to patent citation data in the European Union. J Reg Sci 48:969–989

Fischer MM, LeSage J (2010) Spatial econometric methods for modeling origin destination flows. In: Fischer Manfred M, Getis A (eds) Handbook of applied spatial analysis. Springer, Berlin/Heidelberg, pp 409–432

Fischer MM, Scherngell T, Jansenberger E (2006) The geography of knowledge spillovers between high-technology firms in Europe. Evidence from a spatial interaction modelling perspective. Geogr Anal 38:288–309
Granstand O (1998) Towards a theory of the technology-based firm. Res Policy 27:465–489
Griffith D (2003) Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. Springer, Berlin Heidelberg
Griliches Z (1984) R&D, Patents, and Productivity, NBER Books, National Bureau of Economic Research. Chicago
Guzzetti L (2009) The European research area idea in the history of community policy-making. In: Delanghe H, Muldur U, Soete L (eds) European science and technology policy. Towards integration or fragmentation? Edward Elgar, Cheltenham, pp 64–77
Hoekman J, Frenken K, Tijssen RJW (2010) Research collaboration at a distance: changing spatial patterns of scientific collaboration in Europe. Res Policy 39:662–673
Hoekman J, Scherngell T, Frenken K, Tijssen R (2013) Acquisition of European research funds and its effect on international scientific collaboration. J Econ Geogr 13:23–52
Jaffe AB, Trajtenberg M, Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. Q J Econ 108(3):577–598
Katz JS (1994) Geographical proximity and scientific collaboration. Scientometrics 31:31–43
LeSage J, Pace K (2008) Spatial econometric modeling of origin-destination flows. J Reg Sci 48:941–967
LeSage J, Fischer MM, Scherngell T (2007) Knowledge spillovers across Europe. Evidence from a Poisson spatial interaction model with spatial effects. Pap Reg Sci 86:393–421
Maggioni MA, Uberti TE (2009) Knowledge networks across Europe: which distance matters? Ann Reg Sci 43:691–720
Moreno R, Paci R, Usai S (2005) Spatial spillovers and innovation activity in European regions. Environ Plann A 37:1793–1812
Patuelli R, Griffith D, Tiefelsdorf M, Nijkamp P (2011) Spatial filtering and eigenvector stability: space-time models for German unemployment data. Int Reg Sci Rev 34(2):253–280
Polt W, Vonortas N, Fisher R (2008) Innovation impact, Final report to the European commission, Brussels, DG Research
Powell WW, Grodal S (2005) Networks of innovators. In: Fagerberg J, Mowery DC, Nelson RR (eds) The Oxford handbook of innovation. Oxford University Press, Oxford, pp 56–85
Roediger-Schluga T, Barber M (2006) The structure of R&D collaboration networks in the European Framework Programmes. Unu-MERIT Working Paper Series 2006-36, Economic and Social Research and Training Centre on Innovation and Technology, United Nations University, Maastricht
Romer PM (1990) Endogenous technological change. J Polit Econ 98:71–102
Scherngell T, Barber MJ (2009) Spatial interaction modelling of cross-region R&D collaborations. Empirical evidence from the 5th EU framework programme. Pap Reg Sci 88:531–645
Scherngell T, Barber MJ (2011) Distinct spatial characteristics of industrial and public research collaborations: evidence from the 5th EU framework programme. Ann Reg Sci 46:247–266
Scherngell T, Lata R (2013) Towards an integrated European research area? Findings from eigenvector spatially filtered spatial interaction models using European framework programme data. Paper Reg Sci 92:555–577
Svanfeldt C (2009) A European research area built by the member states? In: Delanghe H, Muldur U, Soete L (eds) European science and technology policy: towards integration or fragmentation? Edward Elgar, Cheltenham/Northampton, pp 44–63
Wooldridge JM (2008) Econometric analysis of cross section and panel data. MIT Press, Cambridge

Chapter 9
The Community Structure of European R&D Collaboration

Michael J. Barber and Thomas Scherngell

Abstract We characterize the geography of communities in the European R&D network using data on R&D projects funded by the fifth European Framework Programme. Communities are subnetworks whose members are more tightly linked to one another than to other members of the network. We characterize the communities by means of spatial interaction models, and estimate the impact of separation factors on the variation of cross-region collaboration activities in a given community at the level of 255 NUTS-2 regions. The results demonstrate that European R&D networks are made up of distinct, relevant substructures characterized by spatially heterogeneous community groups.

9.1 Introduction

Today it is widely believed that interaction between firms, universities and research organizations is crucial for successful innovation in the knowledge-based economy, in particular in knowledge-intensive industries. This gives rise to the notion of R&D networks, defined as a set of organizations performing joint R&D, for instance in the form of collaborative research projects, joint conferences and workshops, or shared R&D resources in the form of labor and capital (see, for instance, Powell and Grodal 2005). By acknowledging that R&D networks are crucial for innovation and that innovation is crucial for sustained economic growth (see Romer 1990), it is natural that modern STI policies emphasize supporting and fostering linkages between innovating actors. The principal European example of such STI policy instruments are the European Framework Programmes (FPs), which support pre-competitive R&D projects, creating a pan-European network of actors performing joint R&D.

M.J. Barber (✉) • T. Scherngell
Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria
e-mail: michael.barber@ait.ac.at

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_9,

Therefore, the investigation of the structure and dynamics of R&D networks is of great current interest, both in a scientific and in a policy context, and currently receives much attention in theoretical and empirical research of different scientific disciplines (see Ozman 2009). Here, we can distinguish between empirical research focusing on knowledge transfer in formalized joint research activities, as given by joint R&D projects or joint publications, and empirical studies using networks as measured by different indicators, such as co-patenting or patent citations, to trace knowledge flows or knowledge spillovers between organizations, regions, or countries (see Ejermo and Karlsson 2006).

There are two major approaches taken to analyze R&D networks: a regional science or geography of innovation perspective and a social network analysis perspective. In a regional science or geography of innovation context, the investigation of the geographical dimension of R&D collaborations is the central research objective. This follows from the assumption that geographical space is crucial for the localization of R&D collaborations and knowledge flows. The pioneering empirical study of Jaffe et al. (1993) provides evidence for the localization hypothesis of knowledge diffusion processes, in general confirmed by more recent empirical studies using different indicators and new spatial econometric techniques (see, for instance, Maurseth and Verspagen 2002; Fischer et al. 2006; Maggioni et al. 2007; Hoekman et al. 2009; Scherngell and Barber 2009, 2011). In a social network analysis context, the focus shifts to the analysis of network structures and dynamics using the mathematics of graph theory, under the assumption that structural relations are often more important for understanding observed behaviors than are attributes of the actors (see, for instance, Zucker and Darby 1998a, b; Singh 2005; Thompson 2006; Vicente et al. 2010). Ter Wal and Boschma (2009) provide an overview of the increasing importance of social network analysis techniques in the fields of regional science and economic geography.

In this chapter, we combine the two research traditions by taking a social network analysis perspective when identifying substructures of European R&D networks constituted under the FPs, followed by taking a regional science perspective when analyzing the geographical dimension of identified substructures. In this context, previous work of and empirical studies by Scherngell and Barber (2009, 2011) are central starting points for the current study. employ a social network perspective to analyze R&D collaborations with the objective of unveiling the texture of the European Research Area (ERA) using data on joint research projects of the fifth EU Framework Programme (FP), while Scherngell and Barber (2009, 2011) focus on the geography of R&D collaborations across European regions.

However, results of these previous empirical works may differ across relevant substructures or communities of the whole FP network. Stated informally, a community is a subnetwork whose members are more tightly linked to one another than to other members of the network. A variety of approaches have been taken to explore this concept (see Fortunato 2010 for a useful review). Since network edges often indicate relationships of interest, detecting community groups can be used to partition the network vertices into meaningful sets, enabling quantitative investigation of relevant subnetworks. Properties of the subnetworks may differ from the aggregate properties of the network as a whole, e.g., modules in the World Wide Web are sets of topically related web pages.

The research approach applied in this chapter is relevant in both a scientific as well as in a European policy context. It describes a way of looking into R&D network structures in Europe that combines social network analysis with a geography of innovation perspective. As noted by Autant-Bernard (2007), the geographical dimension of innovation and knowledge diffusion deserves closer attention by analyzing such phenomena as R&D collaborations. Such analyses are also of crucial interest for European STI policy, in particular for the integration and cohesion objective outlined in the concept of the European Research Area (ERA): improved coherence of the European research landscape and the removal of barriers to knowledge diffusion in a European system of innovation (see CEC 2007). Of course, insight into the status of integration in different thematic areas is a particularly valuable new view on this topic.

Further, the analysis provides important policy implications. By lending crucial insight into real-world topical structures of R&D networks constituted under earlier FPs, the analysis can inform the design of future FPs. Complementarily, a rich picture for regional policy actors is provided at the regional level on leading European regions with respect to cooperative research activities in specific thematic areas.

The objectives of the current study are: first, to detect communities in European R&D networks; second, to describe the spatial patterns of the identified communities; and, third, to identify determinants of the observed spatial patterns. We use data on joint research projects funded by the European Framework Programmes to capture European R&D networks. The identification of thematically distinct communities in these networks is realized using graph theoretic techniques described by Barber and Clark (2009). Further, we employ spatial analysis techniques to identify and describe spatial patterns of identified FP communities at a regional level. By means of a Poisson spatial interaction model, we estimate the impact of various separation factors on cross-region collaboration activities in a given community. In particular, we focus on how geographical distance impacts cross-region collaboration intensities across different FP communities. The results demonstrate that European R&D networks are not homogeneous, instead showing distinct, relevant substructures characterized by thematically homogeneous and spatially heterogeneous communities.

9.2 Background and Main Hypotheses

R&D networks inducing knowledge transfer between firms, universities and research organizations are considered to be crucial for successful innovation in the knowledge-based economy in general, and in knowledge-intensive industries in particular. In fact, we face a considerable increase – and we have done so for decades – in the number of inter-organizational R&D collaborations (Hagedoorn and van Kranenburg 2003). The main reasons for this have been alleged to include the increasing need to access external knowledge – characterized by complementarity and tacitness – and the high degree of strategic flexibility in collaborative agreements (Kogut 1988; Teece 1992). Another reason may be the growing complexity of technology and the existence of converging technologies (see Pavitt 2005). In particular, firms have expanded their knowledge bases into a wider range of technologies (Granstrand 1998), increasing the need for distinct types of knowledge, so firms must learn how to integrate new knowledge into existing products or production processes (Cowan 2004). It may be difficult to develop this knowledge alone or acquire it via the market. The importance of R&D networks for innovation is also stressed by the various systems of innovation concepts that focus on interactions between different actors in a specific region, country or sector (see Lundvall 1992, among others). The main argument is that the sources of innovation are often distributed between firms, universities, suppliers and customers, giving rise to the notion of networks being the locus of innovation. Networks create incentives for interactive organizational learning, leading to faster knowledge diffusion within the innovation system and stimulating the creation of new knowledge or new combinations of existing knowledge.

The EU follows this view in its science and technology policy, mainly reflected in the concept of the European Research Area (ERA), whose aim is to improve coherence of the European research landscape and remove barriers for knowledge diffusion in a European system of innovation (see CEC 2007). The cornerstone of corresponding EU policy instruments is formed by the Framework Programmes (FPs) on Research and Technological Development. By means of this policy initiative, the EU has co-funded thousands of trans-national collaborative R&D projects. The main objectives of the instrument from a European technology policy view are to integrate national and regional research communities and to coordinate national research policies. Empirical studies such as the one of provide evidence for the establishment of a pan-European network of firms, universities, public research organizations, consultants and government institutions performing joint research funded by the FPs (see Roediger-Schluga and Barber 2006 for a comprehensive discussion of the EU FPs).

Previous empirical studies usually focused on complete FPs to describe networks of European R&D cooperation as captured by data on joint FP projects. However, empirical results of these studies may differ across relevant, thematically distinct community groups of the whole FP networks, and these differences may be of crucial interest in a European policy context. Stated informally, a community is a portion of the network whose members are more tightly linked to one another than to other members of the network. Precise formulation of the problem presents two main challenges. First, the notion of communities is somewhat vague, requiring a definition to be provided for what formally constitutes a community. Second, it must be possible to identify community solutions for networks of real-world scientific or policy interest given limitations on time and computational resources. The interplay between these challenges allows a variety of community definitions and community identification algorithms suited to networks of different sizes (for useful overviews, see Fortunato and Castellano 2008; Fortunato 2010; Porter et al. 2009).

Meaningful communities have been identified in many networks of diverse character, corresponding to specialized research areas in co-authorship networks, topically related pages on the World Wide Web, and functional modules in cellular or genetic networks, amongst many others. Following the pioneering work of Girvan and Newman (2002) and Newman and Girvan (2004), many researchers, particularly in statistical physics, have investigated methods for detecting communities in large networks. Similarly, we hypothesize first that the European FP network consists of relevant, thematically distinct subnetworks that show distinct thematic and spatial characteristics.

Second, we hypothesize that geographic localization effects of knowledge flows are significantly smaller within identified communities than for the whole FP5 network, since the transfer of tacit knowledge may be easier in thematically relatively homogenous community groups. As mentioned above, the geography of innovation literature argues that knowledge flows among knowledge producing agents may be geographically bounded, since important parts of new knowledge have some degree of tacitness. Though the cost of transmitting codified knowledge may be invariant to distance, presumably the cost of transmitting non-codified knowledge across geographic space rises with geographic distance (see Jaffe et al. 1993; Audretsch and Feldman 1996). Scherngell and Barber (2009) provide evidence for the geographical localization of FP5 networks. In this study, we anticipate that localization effects decrease for an identified, thematically homogenous community. Due to a more homogeneous thematic focus of a community, the transfer of non-codified knowledge may not be as costly as would be the case for thematically more dispersed actors.

9.3 Empirical Setting and Data

Our core data set to capture collaborative activities in Europe is the EUPRO database, which presently comprises data on funded research projects of the EU FPs (complete for FP1-FP6) and all participating organizations. It contains systematic information on the participating organizations including the full name, the full address, the type of the organization, and, where appropriate and possible, the organizational subentity involved in the project. For a full description of the EUPRO database and its contents, see Roediger-Schluga and Barber (2008).[1]

[1] The version of the EUPRO database used for this study contains information on 61,169 projects funded from FP1 to FP6, yielding 323,638 participations by 60.034 organizations (status: December 2010).

9.3.1 Constructing FP5 Research Networks

The study at hand draws on information concerning joint R&D projects funded in FP5.[2] We selected FP5 as it has the greatest number of projects and, at the time of the computations, the greatest processing of organizational data. Other FPs also show strong community structure (Barber et al. 2008).

Using the EUPRO database, we construct a graph or network containing the collaborative projects from FP5 and all organizations that are participants in those projects; no other forms of collaboration (e.g., co-publication or co-patents) are used here. An organization is linked to a project if and only if the organization is a member of the project. Since an edge never exists between two organizations or two projects, the network is bipartite. The network edges are unweighted; in principle, the edges could be assigned weights to reflect the strength of the participation, but the data needed to assign such network weights is not available.

Previous investigations of the FPs often have made use of one-mode networks (Almendral et al. 2007; Barber et al. 2006; Roediger-Schluga and Barber 2008), typically by (possibly implicitly) projecting the bipartite network onto a network of organizations that are linked based on co-participation in projects. While the one-mode networks can be useful, their construction discards information available in the bipartite networks, which can lead to incorrect community structures (Guimerà et al. 2007). In the present work, we thus focus exclusively on representation of FP5 as a bipartite network.

9.3.2 Detecting Communities in European Collaboration Networks

Community identification in networks is the assignment of the network vertices to a smaller number of clusters. These clusters are hopefully relevant, and thus, drawing on the context of social networks, called communities. Recent community identification methods are based on analyzing the network structure, identifying communities as groups of vertices that are internally strongly connected but only weakly connected to the rest of the network. In empirical networks, vertices within communities are often found to be usefully related by content: edges reflect underlying processes relevant to the entities corresponding to vertices, so communities consist of entities with similar properties.

Community identification methods have been developed that are efficient enough to be suitable for large networks containing thousands or millions of vertices and edges. One such method is the label propagation algorithm (LPA) of Raghavan et al. (2007). Each vertex is assigned a label; a community is the set of all

[2] FP5 had a total budget of 13.7 billion EUR and ran from 1998 to 2002 (CORDIS 1998). See Scherngell and Barber (2009) and CORDIS (1998) for further details on FP5.

vertices with a particular label. The vertices are initialized with distinct labels, thus beginning with all vertices in distinct communities. Vertices are repeatedly updated, replacing their labels with ones that better match the labels of their neighbors. Within tightly interlinked subnetworks, common labels reinforce one another, encouraging uniform labels to be adopted. In contrast, weak linking between tightly interlinked subnetworks means that relatively few neighbors will differ in labels, hindering the propagation of labels between the subnetworks. These two properties accord with the above idea of community, so the LPA proves to be quite effective in practice (Leung et al. 2009).

Two properties of community solutions found by LPA warrant comment. *First*, since each vertex has a single label, the communities are disjoint; no vertex belongs to two communities. *Second*, community solutions are not generally unique; more than one label may be satisfactory for a vertex. Both of these properties suggest that some portion of the vertices may fit well in more than one community, so some care should be taken in interpreting specific community memberships. In this work, we consider statistical properties of the communities, which are more robust against reassignment of a few labels.

In determining the communities, we make use of modest extensions to the LPA (Barber and Clark 2009). The specifics of the algorithms are detailed in Appendix 3. Since we investigate bipartite networks, the communities will include vertices from the two parts of the network, i.e. communities will contain both projects and organizations.

9.3.3 Observing Spatial Collaboration Patterns of Communities Across European Regions

To analyze the spatial patterns of the identified communities we first geocode each organization to a specific European region. We use a concordance scheme provided by Eurostat between postal codes and NUTS regions to trace the specific NUTS-2 region of an organization. The European coverage is achieved by using 255 NUTS-2 regions (NUTS revision 2003) drawn from the 25 pre-2007 EU member-states, Norway and Switzerland. The detailed list of regions is given in Appendix 1.[3] Next we construct a region-by-region collaboration matrix $\boldsymbol{P}^{(c)}$ for each community c, aggregating collaborative activities at the organizational level to the regional level, giving the observed number of R&D collaborations $p_{ij}^{(c)}$ between two regions i and j $(i, j, = 1, \ldots, n)$ for each community c.

[3] We follow previous similar empirical work and rely on a NUTS2 disaggregation of the European territory (see Fischer et al. 2006; LeSage et al. 2007; Scherngell and Barber 2009, 2011). The NUTS2 level provides the basis for the provision of structural funds by the EU, as well as for the evaluation of regional growth processes across Europe (see Fischer et al. 2009).

Following Scherngell and Barber (2009), we use a full counting method. For a project with three participating organizations in three different regions – say regions a, b, and c – we count three links: from region a to region b, from b to c and from a to c. When all three participants are located in one region we count three intraregional links. We exclude self loops to eliminate spurious self collaborations. The resulting regional collaboration matrix $\boldsymbol{P}^{(c)}$ then contains the collaboration intensities $p_{ij}^{(c)}$ between all (i, j)-region pairs for community c. The n-by-n matrix for each community is symmetric by construction ($p_{ij}^{(c)} = p_{ji}^{(c)}$).

9.4 Community Structure in European R&D Networks

Using the label propagation approach described in the previous section, we partitioned the network into 3,482 communities. The communities vary greatly in size, as measured either by the number of organizations in the community or by the number of projects in the community. Most (2,878) communities are small, consisting of just a single project with some or all of the organizations participating in it; these offer little insight into collaboration patterns. In contrast, nine communities are large, containing 20 or more projects; these communities contain over a third of the organizations and over half of the projects present in FP5. Here, we consider the eight of these nine largest communities that are concerned with R&D; the ninth is of different character than the others, focusing instead on international cooperation. We do not further consider the smaller communities here, preferring instead to investigate the large communities in greater detail.

Table 9.1 shows the sizes of the identified communities. We manually assign names to the communities based on consideration of their constituent projects and organizations.

The largest community (2,366 organizations), *Life Sciences*, shows a broad selection of topics in biotechnology and the life sciences, including health, medicine, food, molecular biology, genetics, ecology, biochemistry, and epidemiology. The second largest (2,307 organizations), *Electronics*, focuses principally on information technology and electronics, with projects in related fields dealing with materials science, often related to integrated circuits; projects on algorithms, data mining, and mathematics; and a definite subset of projects concerning atomic, molecular, nuclear, and solid state physics. The third largest community (1,855 organizations), *Environment*, is focused on environment topics, including environmental impact, environmental monitoring, environmental protection, and sustainability.

As communities become smaller, they also become more focused. We see, for example, three distinct transportation related communities. The largest of these (1,146 organizations), *Aerospace*, is focused on aerospace, aeronautics and related topics, including materials science, manufacturing, fluid mechanics, and various energy topics. The next (686 organizations), *Ground Transport*, is focused on land

Table 9.1 Sizes of FP5 communities

Community	Number of organizations	Number of projects
FP5	25,839	9,490
Aerospace	1,146	576
Aquatic resources	81	69
Electronics	2,307	1,447
Enviroment	1,855	971
Ground transport	686	374
Information processing	40	20
Life sciences	2,366	1,468
Sea transport	218	73

transport, with the projects dominated by railroad and, especially, automotive topics; notable subtopics include manufacturing, fuel systems, concrete, and pollution. The smallest transportation community (218 organizations), *Sea Transport*, focuses specifically on sea transport; virtually all project titles are shipping-related. The remaining communities, *Aquatic Resources* and *Information Processing*, are the smallest and most uniform thematically. Their thematic contents are fisheries and statistics.

Figure 9.1 visualizes the network of key FP5 communities. We determine the position for the communities using methods from spectral graph analysis, so that communities that are strongly interconnected are positioned nearer to each other (for a practical overview see Higham and Kibble 2004). The node size corresponds to the number of organizations of the respective community, with the widths of the connection links corresponding to the number of inter-community project participations.

Due to the strong inter-community links, the *Electronics* community appears to have the highest collaboration intensity with other communities, i.e. competences relevant to this field are used intensively in other fields. The *Life Sciences* community shows a strong connection to the third largest community, *Environment*. The three transport-related communities are positioned near one another, i.e. they show relatively high inter-community collaboration intensity. The largest of these is *Aerospace*, and shows a stronger interaction with *Ground Transport* than with *Sea Transport*. The community *Aquatic Resources* has the strongest connection to *Environment*, while *Information Processing* shows comparably low collaboration intensities to all other communities.

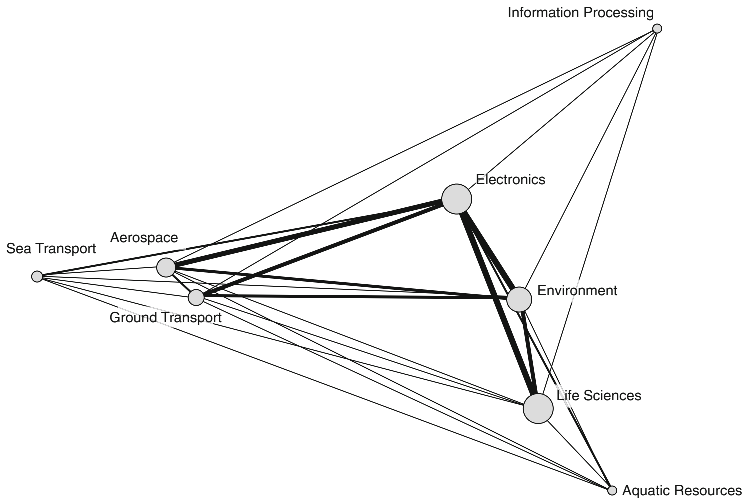


Fig. 9.1 Community groups in the network of FP5 R&D cooperation

9.5 Spatial Structure of Communities in European R&D Networks

We next consider the spatial distribution of the eight FP5 communities. In Fig. 9.2, we illustrate the spatial networks of the communities by aggregating individual observations on the organizations of a community to the regional level. Note that the region-by-region networks are undirected graphs from a network analysis perspective. The nodes represent regions; their size is relative to the number of organizations in the region that belong to the community.

The spatial network maps in Fig. 9.2 reveal considerable differences among the collaboration patterns of the eight FP5 communities. One immediate result is that the region Île-de-France takes an important position in all communities. Furthermore, the visualization clearly reveals the different spatial patterns of the transport-related communities, *Aerospace*, *Ground Transport*, and *Sea Transport*. Though the region Île-de-France appears to be the central hub in the three transport related communities, the directions of the highest collaboration flows from Île-de-France differ markedly. For the *Sea Transport* community we observe intensive collaborations to important sea ports in the north (Zuid Holland, Agder og Rogaland, Danmark, Hamburg) and the south (Liguria, Lisboa, Attiki), while, for the *Ground Transport* community, collaborations to the east and south are dominant (Lombardia, Oberbayern, Stuttgart).

In the *Aerospace* community we can observe a strong localization of collaborations within France and its neighboring countries. In the largest community, *Life Sciences*, the highest number of collaborations is observed between the regions of Île-de-France and Piemonte (174), while the second largest community,

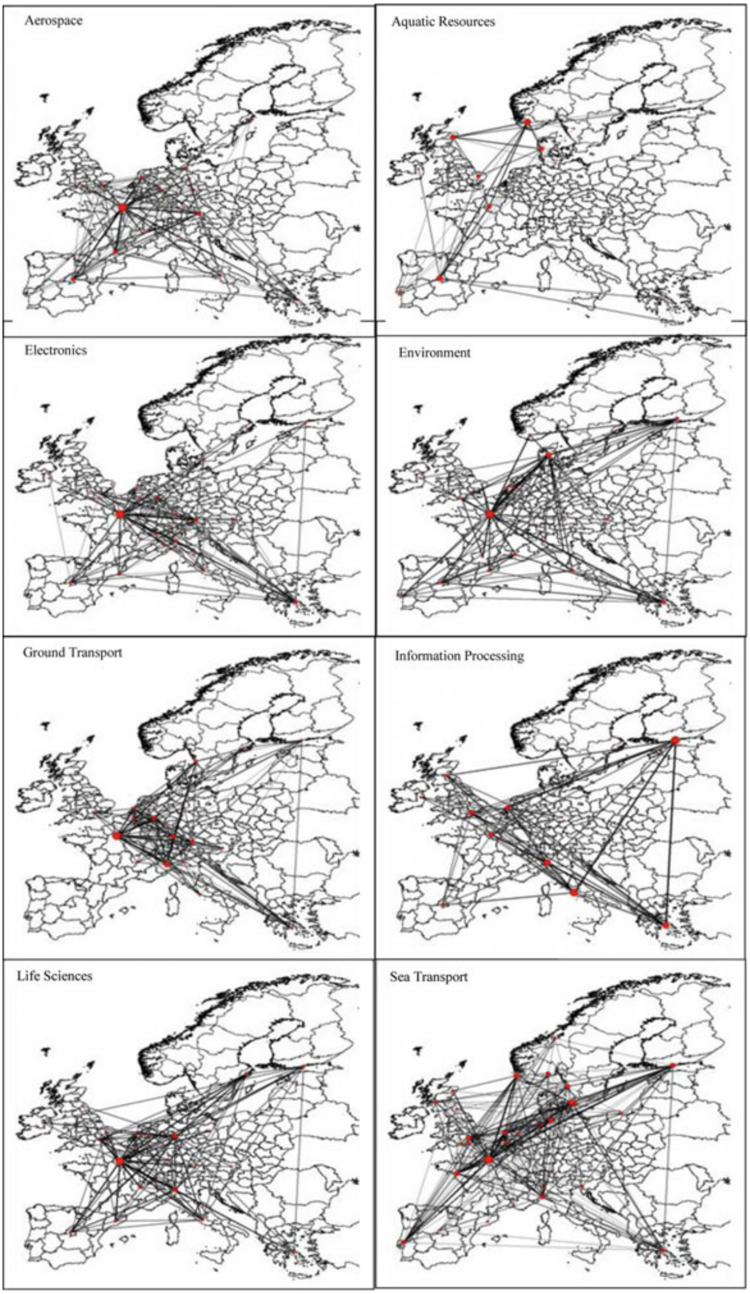


Fig. 9.2 Spatial patterns of eight FP5 communities

Electronics, is characterized by a very high collaboration intensity between the regions of Île-de-France and Oberbayern (474 collaborations), followed by Île-de-France and Köln (265 collaborations), and Oberbayern and Köln (157 collaborations). In the *Environment* community we find the strongest collaboration intensity between Danmark and Etelä-Suomi (131 collaborations). In the community *Aquatic Resources* the regions Danmark and Agder og Rogaland (Norway) show the highest collaboration intensity, not only between them (21 collaborations) but also to other regions, while for the community *Information Processing* we identify Etelä-Suomi as the central region, featuring intensive collaboration with Attiki, Lazio and Lombardia.

To complement the maps shown in Fig. 9.2, the numbers of project participations by organizations in each region for each community are also of interest; we tabulate the most active participants in Appendix 2. This provides insight into which regions are most active for each community, in contrast to which regions are best connected, as described above. Interestingly, well connected regions may markedly differ from the most active regions.

9.6 Identifying Determinants of Spatial Community Patterns

Our objective in this paper is not only to detect communities in European FP networks and describe their spatial configurations, but also to investigate determinants that influence the spatial community patterns. In particular, whether the influence of geographical distance differs across communities is of crucial importance in the context of an aspired European Research Area. Thus, we measure separation effects on the constitution of cross-region R&D collaborations in all detected communities. The spatial interaction model of the type used by Scherngell and Barber (2009, 2011) in a similar context serves again as an appropriate basis. Spatial interaction models incorporate a function characterizing the origin i of interaction, a function characterizing the destination j of interaction and a function characterizing the separation between two regions i and j. The model is characterized by a formal distinction implicit in the definitions of origin and destination functions on the one hand, and separation functions on the other (see, for example, Sen and Smith 1995). Origin and destination functions are described using weighted origin and destination variables, respectively, while the separation functions are postulated to be explicit functions of numerical separation variables. The general model in our case is given by

$$P_{ij}^{(c)} = A_i\, B_j\, S_{ij} \quad i, j = 1, \ldots, n \tag{9.1}$$

with

$$A_i = A(a_i, \alpha_1) = a_i^{\alpha_1} \quad i,j = 1, \ldots, n \tag{9.2}$$

$$B_j = B(b_j, \alpha_2) = b_j^{\alpha_2} \quad i,j = 1, \ldots, n \tag{9.3}$$

$$S_{ij} = \exp\left[\sum_{k=1}^{K} \beta_k \, d_{ij}^{(k)}\right] \quad i,j = 1, \ldots, n \tag{9.4}$$

where $P_{ij}^{(c)}$ denotes a stochastic dependent variable that is realized by the number of observed collaboration flows $p_{ij}^{(c)}$ between region i and region j for each community c.[4] A_i denotes the origin function, B_j denotes the destination function, while S_{ij} represents a separation function. The a_i and b_j are measured in terms of the number of organizations participating in EU FP5 projects in the regions i and j, while α_1 and α_2 are scalar parameters to be estimated. Note that due to the symmetry of the origin and destination variables, we have a special case with $\alpha_1 = \alpha_2$, i.e. numerical results for α_1 and α_2 should be equal up to numerical precision. The $d_{ij}^{(k)}$ are K separation measures, the β_k are corresponding parameters to be estimated that will show the relative strengths of the separation measures. We rely on separation measures used in similar studies (see, for instance, Fischer et al. 2006; Scherngell and Barber 2009). We can group these separation variables into three categories:

(i) Variables accounting for *spatial effects*: $d_{ij}^{(1)}$ denotes geographical distance between two regions i and j as measured by the great circle distance between the economic centers of the regions, while $d_{ij}^{(2)}$ is a dummy variable that controls for neighboring region effects. We set $d_{ij}^{(2)}$ to one if two organizations are located in neighboring regions and zero otherwise, where neighboring regions are defined to share a common border.

(ii) Variables accounting for *institutional and cultural effects*: $d_{ij}^{(3)}$ is a country border dummy variable that takes a value of zero if two regions i and j are located in the same country and one otherwise, while $d_{ij}^{(4)}$ is a language area dummy variable that takes a value of zero if two regions i and j are located in the same language area and one otherwise.

(iii) Variables accounting for *technological effects*: $d_{ij}^{(5)}$ measures technological distance by using regional patent data from the European Patent office (EPO). The variable is constructed (see Scherngell and Barber 2009) as a vector $\boldsymbol{t}_i$ that measures region i's share of patenting in each of the technological subclasses of the International Patent Classification (IPC). Technological subclasses correspond to the third-digit level of the IPC systems. We use the Pearson correlation coefficient between the technological vectors of two regions i and j to define how close they are to each other in technological space. Though we focus on spatial, cultural and institutional effects in this study, we include technological distance, mainly as a control variable to allow for the possibility that geographical distance may just be a proxy for technological distance.

[4] Note that we do not exclude zero-flows or intraregional flows.

At this point, we are interested in estimating the parameters $\alpha_1 = \alpha_2$ and β_k for each community c. OLS estimation procedures are not appropriate for modeling research collaborations, due to their true integer nature and due to the assumption of non-normal errors. This suggests a Negative Binomial density distribution, i.e. a Poisson specification with heterogeneity, allowing for the overdispersion often observed for real world count data (see Cameron and Trivedi 1998). The Negative Binomial density distribution in our case is given by

$$f\left(P_{ij}^{(c)}\right) = \frac{\Gamma\left(p_{ij}^{(c)} + \delta^{-1}\right)}{\Gamma\left(p_{ij}^{(c)} + 1\right)\Gamma\left(\delta^{-1}\right)} \left(\frac{\delta^{-1}}{A_i\, B_j\, S_{ij} + \delta^{-1}}\right)^{\delta^{-1}} \left(\frac{A_i\, B_j\, S_{ij}}{A_i\, B_j\, S_{ij} + \delta^{-1}}\right)^{p_{ij}^{(c)}} \tag{9.5}$$

Here, $\Gamma(\cdot)$ denotes the gamma function and δ is the dispersion parameter. Model estimation is done by Maximum Likelihood procedures (see Long and Freese 2001).

Table 9.2 presents the sample estimates of the spatial interaction models, with standard errors given in brackets. We use the Negative Binomial model specification as given by Eq. 9.5. The dispersion parameter δ is significant for all model versions, indicating that the Negative Binomial version is the right specification, i.e. the standard Poisson specification would be biased due to unobserved heterogeneity between the region pairs (see Scherngell and Barber 2009). The existence of unobserved heterogeneity that cannot be captured by the covariates leads to overdispersion and, thus, to biased model parameters for the standard Poisson model.

The models produce quite interesting results in the context of the literature on European R&D networks on the one hand, and in the context of the literature on the geographic localization of knowledge flows on the other hand. The second column contains, for the purpose of comparison, the sample estimates for total FP5. The main conclusion of this model is that geographical distance between two organizations has a significant negative effect on the likelihood that they collaborate. However, technological distance between regions shows a larger negative effect on cross-region collaborative activities.

The impact of the different separation effects varies considerably across observed FP5 communities, both with respect to the magnitude of the estimates and to statistical significance. The most important result is that the negative effect of geographical distance is significantly weaker in any given FP5 community than for all FP5 collaborations taken as a whole. This indicates that geographical integration in European research is better developed in thematically more homogenous communities than between communities. In the *Aquatic Resources* community, the *Sea Transport* community and the *Information Processing* community, the effect of geographical distance is even insignificant – within these communities there is no observable effect of geographical distance on the probability of

Table 9.2 Estimation results of the Negative Binomial spatial interaction models [65,025 observations, asymptotic standard errors given in brackets]

	Negative Binomial spatial interaction models								
	Total FP5	Life sciences	Aquatic resources	Electronics	Environment	Sea transport	Ground transport	Aerospace	Information processing
$\alpha_1 = \alpha_2$	0.706***	0.865***	0.777***	0.794***	0.659***	0.771***	1.055***	0.808***	1.202***
	(0.003)	(0.005)	(0.024)	(0.005)	(0.005)	(0.004)	(0.010)	(0.006)	(0.008)
Geo	−0.278***	−0.110***	−0.072	−0.038**	−0.036**	−0.020	−0.224***	−0.103***	−0.017
[β_1]	(0.008)	(0.011)	(0.051)	(0.012)	(0.012)	(0.038)	(0.020)	(0.017)	(0.016)
Neig	0.184***	0.043	0.312	0.051	0.274***	0.201	0.033	0.253***	0.186
[β_2]	(0.036)	(0.051)	(0.248)	(0.052)	(0.057)	(0.019)	(0.048)	(0.013)	(0.283)
Count	−0.008	0.009	−0.588***	−0.148***	0.119	−0.558***	−0.121	−0.342***	−0.141
[β_3]	(0.023)	(0.0006)	(0.168)	(0.039)	(0.099)	(0.143)	(0.081)	(0.055)	(0.216)
Lang	−0.098***	−0.004	−1.118***	−0.002	−0.123***	0.326	−0.088	−0.023	−0.798***
[β_4]	(0.024)	(0.034)	(0.152)	(0.035)	(0.038)	(0.271)	(0.057)	(0.019)	(0.113)
Tech	−1.413***	−1.437***	−6.312***	−1.577***	−2.421***	−3.797***	−0.606**	−2.511***	−1.533*
[β_5]	(0.115)	(0.167)	(0.657)	(0.171)	(0.181)	(0.544)	(0.283)	(0.257)	(0.763)
Cons.	−2.539***	−7.042***	−8.557***	−6.197***	−4.148***	−6.175***	−10.124***	−5.303***	−16.851***
	(0.128)	(0.175)	(0.654)	(0.179)	(0.185)	(0.545)	(0.297)	(0.260)	(0.855)
(δ)	1.047***	0.969***	2.835***	0.341***	0.530***	2.943***	2.044***	0.982***	6.115***
	(0.009)	(0.016)	(0.743)	(0.018)	(0.013)	(0.025)	(0.044)	(0.015)	(0.526)
LL	−135,234.21	−65,657.63	−8,537.42	−75,200.45	−73,257.76	−18,829.43	−31,445.52	−54,124.21	−3,723.43
SS	6.523	5.212	4.979	5.001	4.213	5.176	6.712	5.732	5.174
M-R^2	0.173	0.224	0.128	0.196	0.155	0.133	0.251	0.171	0.249
BIC	−35,455.09	−37,935.68	−24,44.31	−36,647.48	−26,853.30	−4,169.42	−21,006.38	−22,283.32	−2,424.85

The dependent variable is the cross-region collaboration intensity between two regions *i* and *j* in a given community. The independent variables are defined as given in the text.

LL denotes the log-likelihood, *SS* sum of squares, *M-R^2* McFadden's R-sqaured, *BIC* Bayesian Information Criterion

***significant at the 0.001 significance level, **significant at the 0.01 significance level, *significant at the 0.05 significance level collaboration between two organizations in Europe. The highest negative effect of geographical distance within a community is identified for the *Ground Transport* community ($\beta_1 = -0.224$).

While geographical distance effects are generally lower for the communities than for all FP5 collaborations, the neighboring region effects are more variable. Neighboring regions effects cannot be identified for most communities, with the exception of the *Environment* community and the *Aerospace* community, which are subject to stronger neighboring region effects than the average of all FP5 collaborations, i.e. there is considerable significant spatial clustering of research collaborations in these communities at the regional level. Also institutional and cultural effects vary considerably across communities. The modeling results point to the existence of institutional barriers at the national level for collaboration in the *Aquatic Resources* community, the *Electronics* community, the *Sea Transport* community, and the *Aerospace* community, even though FP5 as a whole shows no such barriers. Language area effects are generally lower or insignificant, but the *Aquatic Resources* community and the *Information Processing* community are characterized by quite high negative language area effects, i.e. collaboration probability significantly decreases between organizations located in different language areas.

Concerning technological distance, we find that, in each community, the negative effect of technological distance is higher than for the whole FP network, except for *Ground Transport*; the collaboration probability with 'technologically distant' regions in a thematically homogenous community is lower than the average collaboration probability in FP5. For the outlier *Ground Transport*, one may speculate that the thematic area uses rather mature and/or widely used technologies prevalent in all regions, leading to a lower negative effect of technological distance. Additional background information on the composition and configuration of the communities would be needed for further interpretations of the sample estimates. Most importantly, the results demonstrate that separation effects for collaboration depend on the FP communities; this may provide a starting point for further research, in particular concerning the interpretation of the parameter estimates.

9.7 Conclusion

Using data on joint research projects funded by FP5, we have in this chapter analyzed European R&D collaborations, investigating the hypotheses (1) that the collaborative network consists of communities with distinct thematic and spatial characteristics and (2) that geographical localization effects of knowledge flows are smaller in these communities that for the network as a whole. We have used techniques described by Barber and Clark (2009) to identify network communities, subnetworks whose members are more tightly linked to one another than to other members of the network. The determinants of the spatial patterns in eight of the largest identified communities are examined by means of Negative Binomial spatial interaction models, estimating how various separation factors – such as geographical distance – affect the variation of cross-region collaboration activities in a given community.

The results of the analysis are supportive of our hypotheses and of interest both from a scientific point of view and in a European policy context. First, we detected relevant, thematically relatively homogenous FP5 communities, providing a new view on the R&D collaboration landscape in Europe. The largest communities identified are Life Sciences, Electronics, and Environment; these may contain further substructures of equal relevance. As communities become smaller, they also become more focused. We identified three transport-related communities: Aerospace, Ground Transport, and Sea Transport. The remaining communities, Aquatic Resources and Information Processing, are the smallest and most uniform thematically of those we have considered. Second, the spatial analysis of the large communities clearly reveals that the spatial configuration varies across communities. However, the region of Île-de-France plays a central role in each of the large communities. Third, the estimation results of the spatial interaction model show that the spatial integration of collaboration activities within the analyzed communities is more developed than for FP5 collaborations as a whole. The negative impact of geographical distance on the probability that two organizations collaborate is much lower when these organizations belong to the same community, while the negative impact of technological differences is generally more pronounced.

From a policy perspective, the identification and characterization of the spatial patterns of these thematically relevant substructures is of crucial interest. First, our analysis may serve as a starting point for analyzing the empirical thematic landscape of European R&D collaboration, which is of strategic interest for the design of future European policy programs supporting collaborative R&D, in particular concerning the orientation of thematic foci. Second, the simple but essential spatial characterization of the large communities may serve as an important source of information for regional and national policy makers to identity their main peers for benchmarking exercises or stimulation of specific collaborations; this is tabulated in Appendix 2. Third, in the context of the European policy goal of an integrated and coherent research area, the results indicate that the degree and evolution of integration may differ across technological areas and that specific technological characteristics should be considered when assessing progress towards that goal.

The study suggests several directions for future research. First, the interpretation of the spatial configuration of the largest identified communities was confined to the descriptive level, as in-depth interpretations of the different separation effects would require further background information about the actors involved in a specific community. Further work could focus on interpretation of separation effects, building on the results presented here. Second, the (spatial) evolution of the detected communities over time could be investigated, providing a deeper understanding on the dynamics of community formation and their spatial integration in the European R&D collaboration landscape. Third, while we have focused on large communities that cover the majority of the projects, there are thousands of smaller communities that we have not considered. Thus, strategies for analyzing these smaller communities could be explored, as could policy implications such as how to encourage integration of the small communities into the larger ones. Finally, alternative community identification methods could be used, for example to consider overlapping or hierarchical communities, accounting for the subthemes recognized in the larger communities.

Acknowledgments This work has been partly funded by the Austrian Science Fund (FWF): [I 886-G11] and the Multi-Year Research Grant (MYRG) – Level iii (RC Ref. No. MYRG119(Y1-L3)-ICMS12-HYJ) by the University of Macau.

Appendix 1

NUTS is an acronym of the French for the "nomenclature of territorial units for statistics", which is a hierarchical system of regions used by the statistical office of the European Community for the production of regional statistics. At the top of the hierarchy are NUTS-0 regions (countries) below which are NUTS-1 regions and then NUTS-2 regions. This study disaggregates Europe's territory into 255 NUTS-2 regions located in the EU-25 member states (except Cyprus and Malta) plus Norway and Switzerland. We exclude the Spanish North African territories of Ceuta y Melilla, the Portuguese non-continental territories Azores and Madeira, and the French Departments d'Outre-Mer Guadeloupe, Martinique, French Guayana and Reunion.

Appendix 2

We list here the most active regions for the eight communities considered in depth in this paper. For each community, we give the 20 regions with the highest number of participations in projects from the community. The number of participations is shown parenthetically. Regions are given in descending order of the number of participations.

Aerospace:	Île de France (1232), Comunidad de Madrid (691), Oberbayern (581), Danmark (526), Noord-Holland (440), Köln (365), Attiki (320), Inner London (306), Lombardia (285), Greater Manchester (276), Bedfordshire & Hertfordshire (271), Etelä-Suomi (269), Campania (266), Midi-Pyrénées (248), Dytiki Ellada (247), Outer London (243), Lazio (241), Liguria (239), Hampshire & Isle of Wight (225), País Vasco (224)
Aquatic Resources:	Agder og Rogaland (97), North Eastern Scotland (93), Danmark (91), Comunidad de Madrid (73), Flevoland (67), Noord-Holland (67), Hamburg (57), Algarve (55), Kriti (49), Attiki (47), Northern Ireland (39), Southern and Eastern (38), East Anglia (31), Andalucía (26), País Vasco (25), Galicia (24), Prov. West-Vlaanderen (22), Etelä-Suomi (21), Eastern Scotland (18), Vestlandet (17)

(continued)

Electronics:	Île de France (3537), Oberbayern (1390), Attiki (1182), Rhône-Alpes (1012), Comunidad de Madrid (863), Köln (831), Lombardia (768), Lazio (728), Zuid-Holland (578), Danmark (563), Berkshire, Buckinghamshire & Oxfordshire (559), Berlin (540), Région lémanique (531), Noord-Brabant (523), Inner London (519), Cataluña (509), Prov. Vlaams-Brabant (483), Southern and Eastern (471), Stuttgart (433), Outer London (430)
Environment:	Île de France (1020), Danmark (782), Αττική/ Attiki (627), Etelä-Suomi (580), Lazio (565), Zuid-Holland (526), Noord-Holland (479), Comunidad de Madrid (426), East Anglia (414), Lombardia (395), Southern and Eastern (378), Cataluña (373), Stockholm (357), Gelderland (355), Wien (350), Andalucía (326), Utrecht (306), Karlsruhe (305), Agder og Rogaland (295), Hampshire & Isle of Wight (294)
Ground Transport:	Île de France (846), Stuttgart (698), Piemonte (587), Köln (385), Zuid-Holland (346), Lombardia (323), Oberbayern (293), Västsverige (290), Etelä-Suomi (226), Berkshire, Buckinghamshire & Oxfordshire (218), Kentriki Makedonia (200), Lazio (177), Hannover (175), País Vasco (168), Comunidad de Madrid (144), Steiermark (141), Noord-Holland (127), Prov. Vlaams-Brabant (123), Rhône-Alpes (119), Darmstadt (118)
Information Processing:	Eastern Scotland (40), Lombardia (21), Etelä-Suomi (20), Lazio (18), Zuid-Holland (16), Hampshire & Isle of Wight (14), Île de France (12), Attiki (11), Outer London (11), Stockholm (10), Sør-Østlandet (10), Danmark (7), Darmstadt (7), Southern and Eastern (7), Noord-Holland (5), Comunidad de Madrid (4), Essex (4), Limburg (NL) (4), Luxembourg (Grand-Duché) (4), Espace Mittelland (3)
Life Sciences:	Île de France (1860), Danmark (1055), Gelderland (843), Outer London (703), Lombardia (658), East Anglia (637), Comunidad de Madrid (636), Inner London (605), Cataluña (569), Zuid-Holland (547), Utrecht (538), Lazio (529), Stockholm (521), Karlsruhe (519), Prov. Vlaams-Brabant (495), Rhône-Alpes (494), Southern and Eastern (481), Oberbayern (458), Région de Bruxelles-Capitale/Brussels Hoofdstedelijk Gewest (442), Eastern Scotland (396)
Sea Transport:	Danmark (190), Liguria (144), Hamburg (137), Île de France (135), Outer London (115), South Western Scotland (105), Agder og Rogaland (99), Zuid-Holland (88), Attiki (76), Pays de la Loire (61), Bremen (58), Surrey, East & West Sussex (48), Västsverige (43), Comunidad de Madrid (40), Etelä-Suomi (36), Friuli-Venezia Giulia (35), Gelderland (35), Hampshire & Isle of Wight (33), Trøndelag (32), Région de Bruxelles-Capitale/Brussels Hoofdstedelijk Gewest (30)

Appendix 3

Raghavan et al. (2007) proposed a label propagation algorithm (LPA) for identifying communities in networks. Community membership is tracked by labels assigned to the graph vertices; a community is a set of all vertices with a particular label. Each vertex is assigned a single label, and thus belongs to a single community.

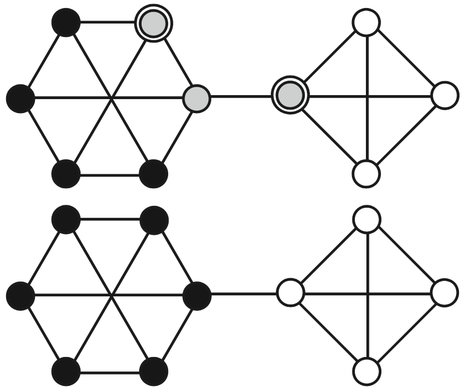


Fig. 9.3 Community identification with label propagation

Call a label *satisfactory* for a vertex when no other label occurs more frequently among its neighbors. The core of the LPA is a process of replacing unsatisfactory labels with satisfactory ones, continuing until all vertices have satisfactory labels. This idea is illustrated in Fig. 9.3 using a toy network with visually apparent community structure. In Fig. 9.3a, there are three different labels, shown by the vertex shading. The black and white labels are all satisfactory for their vertices. Of the three gray labels, two are unsatisfactory for their vertices, shown by double borders on the vertices: one neighbors a single gray vertex and two black vertices, the other neighbors a single gray vertex and three white vertices. The third gray label is satisfactory: the vertex neighbors two gray vertices and two black vertices. In Fig. 9.3b, all vertices have satisfactory labels.

The algorithm begins from a state where all vertices have different labels (and thus are generally all unsatisfactory). Taken in random order, the vertices are considered to see whether their labels are satisfactory and updated to be satisfactory when not; if multiple labels would be satisfactory, one is chosen at random. For the example network shown in Fig. 9.3a, the two vertices with gray labels must then be updated, one to have a black label, the other to have a white label; note that changing these two gray labels will cause the third gray label to become unsatisfactory. Multiple relabeling passes are made through the vertices, with the algorithm halting when all vertices have a satisfactory label, such as in Fig. 9.3b.

The LPA offers a number of desirable qualities. As described above, it is conceptually simple, being readily understood and quickly implemented. The algorithm is efficient in practice. Each relabeling iteration through the vertices has a computational complexity linear in the number of edges in the graph. The total number of iterations is not a priori clear, but relatively few iterations are needed to assign the final label to most of the vertices (typically over 95 % of vertices in 5 iterations, see Raghavan et al. 2007; Leung et al. 2009).

The LPA defines communities procedurally, rather than as optimization of an objective function, and thus provides no intrinsic measure for the quality of communities found. To assess community quality, we can introduce an auxiliary measure, such as the popular modularity measure (Newman and Girvan 2004); in this work, more suitable is a version of modularity specialized to bipartite networks (Barber 2007). Using modularity, communities found using LPA are seen to be of high quality (Raghavan et al. 2007): label propagation is both fast and effective. Indeed, Leung et al. (2009) have proposed extensions to the label propagation algorithm that make it comparable to the best algorithms for community detection in quality and efficient enough to analyze very large networks.

Barber and Clark (2009) have elucidated the connection between label propagation and modularity, showing that modularity can be maximized by propagating labels subject to additional constraints and proposing several variations of the LPA. In this paper, we make use of a hybrid, two-stage label propagation scheme, consisting of the LPAr variant followed by the LPAb variant (see Barber and Clark 2009 for details). LPAr is defined similarly to the original LPA presented above, but with additional randomness to allow the algorithm to avoid premature termination. In practice, this produces better communities as measured by modularity than does LPA. LPAb imposes constraints on the label propagation so that the algorithm identifies a local maximum in the bipartite modularity. The overall hybrid algorithm thus belongs to the recent class of algorithms based on modularity maximization (for a survey, see Fortunato 2010).

References

Almendral JA, Oliveira JG, López L, Sanjuán MAF, Mendes JFF (2007) The interplay of universities and industry through the FP5 network. New J Phys 9(6):183–198

Audretsch DB, Feldman MP (1996) R&D spillovers and the geography of innovation and production. Am Econ Rev 86:630–640

Autant-Bernard C, Mairesse J, Massard N (2007) Spatial knowledge diffusion through collaborative networks. Pap Reg Sci 86:341–350

Barber MJ (2007) Modularity and community detection in bipartite networks. Phys Rev E Stat Nonlin Soft Matter Phys 76(6):066102

Barber MJ, Clark JW (2009) Detecting network communities by propagating labels under constraints. Phys Rev E Stat Nonlin Soft Matter Phys 80(2):026129

Barber MJ, Faria M, Streit L, Strogan O (2008) Searching for communities in Bipartite networks. In: Bernido C, Carpio-Bernido MV (eds) Proceedings of the 5th Jagna international workshop: stochastic and quantum dynamics of biomolecular systems. Springer, New York

Barber MJ, Krueger A, Krueger T, Roediger-Schluga T (2006) Network of European Union-funded collaborative research and development projects. Phys Rev E Stat Nonlin Soft Matter Phys 73(3):036132

Breschi S, Cusmano L (2004) Unveiling the texture of a European research area: emergence of oligarchic networks under EU framework programmes. Int J Technol Manag 27(8):747–772, Special Issue on Technology Alliances

Cameron AC, Trivedi PK (1998) Regression analysis of count data. Cambridge University Press, Cambridge/New York

CEC (Commission of the European Communities) (2007) Green paper 'The European research area: new perspectives'. {SEC(2007) 412}, COM(2007)161 final, Brussels, 4 April 2007

CORDIS (1998) Fifth framework programme (1998–2002) http://cordis.europa.eu/fp5/about.htm

Cowan R (2004) Network models of innovation and knowledge diffusion, Unu-MERIT working paper series, 2004-016

Ejermo O, Karlsson C (2006) Interregional inventor networks as studied by patent coinventorships. Res Policy 35(3):412–430

Fischer MM, Scherngell T, Jansenberger E (2006) The geography of knowledge spillovers between high-technology firms in Europe. Evidence from a spatial interaction modelling perspective. Geogr Anal 38:288–309

Fischer MM, Scherngell T, Reismann M (2009) Knowledge spillovers and total factor productivity: evidence using a spatial panel data model. Geogr Anal 41:204–220

Fortunato S (2010) Community detection in graphs. Phys Rep 486(3–5):75–174

Fortunato S, Castellano C (2008) Community structure in graphs. In: Robert A Meyers (ed) Encyclopedia of complexity and system science. Springer, New York

Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci U S A 99(12):7821–7826

Granstrand O (1998) Towards a theory of the technology-based firm. Res Policy 27(5):465–489

Guimerà R, Sales-Pardo M, Amaral LAN (2007) Module identification in bipartite and directed networks. Phys Rev E Stat Nonlin Soft Matter Phys 76(3):036102

Hagedoorn J, van Kranenburg H (2003) Growth patterns in R&D partnerships: an exploratory statistical study. Int J Ind Organ 21:517–531

Higham DJ, Kibble M (2004) A unified view of spectral clustering. Mathematics Research Report 02, University of Strathclyde

Hoekman J, Frenken K, van Oort F (2009) The geography of collaborative knowledge production in Europe. Ann Reg Sci 43:721–738

Jaffe AB, Trajtenberg M, Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. Q J Econ 108(3):577–598

Kogut B (1988) Joint ventures: theoretical and empirical perspectives. Strateg Manag J 9:319–332

LeSage J, Fischer MM, Scherngell T (2007) Knowledge spillovers across Europe. Evidence from a Poisson spatial interaction model with spatial effects. Pap Reg Sci 86:393–421

Leung IXY, Hui P, Lio' P, Crowcroft J (2009) Towards real-time community detection in large networks. Phys Rev E Stat Nonlin Soft Matter Phys 79(6):066–107

Long JS, Freese J (2001) Regression models for categorical dependent variables using Stata. Stata Corporation, College Station

Lundvall BA (1992) National innovation systems. Pinter, London

Maggioni MA, Nosvelli M, Uberti TA (2007) Space versus networks in the geography of innovation: a European analysis. Pap Reg Sci 86:471–493

Maurseth PB, Verspagen B (2002) Knowledge spillovers in Europe: a patent citation analysis. Scand J Econ 104:531–545

Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E Stat Nonlin Soft Matter Phys 69(2):026113

Ozman M (2009) Inter-firm networks and innovation: a survey of literature. Econ Innovation New Tech 18:39–67

Pavitt K (2005) Innovation processes. In: Fagerberg J, Mowery DC, Nelson RR (eds) The Oxford handbook of innovation. Oxford University Press, Oxford, pp 86–114

Porter MA, Onnela J-P, Mucha PJ (2009) Communities in networks. Not Am Math Soc 56(9): 1082–1097

Powell WW, Grodal S (2005) Networks of innovators. In: Fagerberg J, Mowery DC, Nelson RR (eds) The oxford handbook of innovation. Oxford University Press, Oxford, pp 56–85

Raghavan UN, Albert R, Kumara S (2007) Near linear time algorithm to detect community structures in large-scale networks. Phys Rev E Stat Nonlin Soft Matter Phys 76(3):036106

Romer PM (1990) Endogenous technological change. J Polit Econ 98:71–102

Roediger-Schluga T, Barber MJ (2006) The structure of R&D collaboration networks in the European framework programmes, Unu-MERIT working paper series, 2006-36

Roediger-Schluga T, Barber MJ (2008) R&D collaboration networks in the European framework programmes: data processing, network construction and selected results. Int J Foresight Innov Policy 4:321–347

Scherngell T, Barber MJ (2009) Spatial interaction modelling of cross-region R&D collaborations. Empirical evidence from the 5th EU framework programme. Pap Reg Sci 88:531–546
Scherngell T, Barber M (2011) Distinct spatial characteristics of industrial and public research collaborations: evidence from the 5th EU framework programme. Ann Reg Sci 46:247–266
Sen A, Smith TE (1995) Gravity models of spatial interaction behaviour. Springer, Berlin/New York/Heidelberg
Singh J (2005) Collaborative networks as determinants of knowledge diffusion patterns. Manag Sci 51(5):756–770
Teece D (1992) Competition, co-operation, and innovation. Organisational arrangements for regimes of rapid technological progress. J Econ Behav Organ 18:1–25
Ter Wal ALJ, Boschma R (2009) Applying social network analysis in economic geography: framing some key analytic issues. Ann Reg Sci 43(3):739–756
Thompson P (2006) Patent citations and the geography of knowledge spillovers: evidence from inventor- and examiner-added citations. Rev Econ Stat 88(2):383–388
Vicente V, Balland PA Brossard O (2010) Getting into networks and clusters: evidence from the Midi-Pyrenean global navigation satellite systems (GNSS) collaboration network. Reg Stud. doi:10.1080/00343401003713340
Zucker LG, Darby MR, Armstrong J (1998a) Geographically localized knowledge: spillovers or markets? Econ Inq XXXVI:65–86
Zucker LG, Darby MR, Brewer M (1998b) Intellectual human capital and the birth of U.S. biotechnology enterprises. Am Econ Rev 88(1):290–306

Chapter 10
Determinants of International R&D Activities: Evidence from a Gravity Model

Sandra Leitner, Robert Stehrer, and Bernhard M. Dachs

Abstract Firms not only produce or sell their products and services abroad, but increasingly also conduct research and development (R&D) at locations outside their home countries – a phenomenon referred to as the 'internationalization of business R&D'. This chapter analyses the internationalization of business R&D for OECD countries and identifies specific home and host country characteristics that are conducive or obstructive to R&D expenditure of foreign affiliates. The analysis employs a recently compiled novel data set on R&D expenditure of foreign-owned firms in the manufacturing sectors of a set of OECD countries. The results point to the pivotal role of market size and of cultural, physical and technological proximity for R&D efforts of foreign-owned firms. Moreover, the analysis demonstrates that sufficient human capital and strong indigenous technological capabilities in the host country tend to be conducive to R&D activities of foreign affiliates. In contrast, a rich human capital base in the home country is obstructive to the process of R&D internationalization. Geographic distance turns out to be a strong deterrent.

10.1 Introduction

Firms not only produce or sell their products and services abroad, but increasingly also conduct research and development (R&D) at locations outside their home countries – a phenomenon referred to as the 'internationalization of business R&D' (Narula and Zanfei 2005; OECD 2008b; Hall 2010).

The internationalization of business R&D is more of a recent phenomenon. The international economics as well as the international business literature long

S. Leitner (✉) • R. Stehrer
Vienna Institute for International Economic Studies, Vienna, Austria
e-mail: sleitner@wiiw.ac.at

B.M. Dachs
Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_10,

regarded R&D and the accumulation of knowledge as activities that are bound to the home countries of multinational firms. In their seminal paper on R&D in large multinational enterprises, Patel and Pavitt (1991, p. 17) concluded that the production of technology remained ‘far from globalized’, but was concentrated in the home countries. Hence in the 1990s, R&D was still ‘an important case of non-globalization’ (Patel and Pavitt 1991, p. 17). Theories of the multinational firm following Hymer’s (1976 [1960]) seminal contribution stress that the international expansion of R&D is a means to exploit existing intangible assets and knowledge capital of the firm in foreign markets (Dunning 1988; Markusen 2002; Helpman 2006; Forsgren 2008).

However, during the last two decades, the internationalization of business R&D activities has accelerated strikingly. Specifically, as highlighted by the OECD (2008a), between 1995 and 2003, R&D expenditure of foreign affiliates increased twice as rapidly as their turnover or their host countries’ aggregate imports. This renders R&D activities of foreign affiliates one of the most dynamic elements of the process of globalization. Until recently, the main actors and recipients of cross-border R&D expenditure were developed countries. Lately, some new players emerged, giving rise to new patterns of R&D internationalization. Especially in Asia, emerging economies gained importance as host countries of R&D internationalization activities but developing countries also increasingly engaged in outward R&D activities. Despite these developments, the largest part of international R&D still takes place between the triad area, comprising the US, the EU and Japan (OECD 2008b).

Given the benefits that accrue from the presence and activities of R&D intensive foreign-owned firms, attracting them has been high on the political agenda of many economies. R&D expenditure of foreign-owned firms may increase aggregate R&D and innovation expenditure of the country. It may give rise to substantial information and knowledge spillovers (Blomström and Kokko 2003), foreign-owned firms may boost the demand for skilled personnel including R&D staff, or R&D efforts and the presence of foreign-owned firms may lead to structural change and agglomeration effects (Young et al. 1994).

The ensuing analysis investigates determinants of the process of internationalization of business R&D. It uses a novel and unique database of bilateral business R&D expenditure of foreign affiliates in the manufacturing sector of selected OECD countries for the period from 2001 to 2007. Given the type and quality of the data, the analysis contributes greatly to the ongoing discussion as to key determinants of the process of R&D internationalization as previous data-related shortcomings are remedied. Specifically, since the analysis uses R&D expenditure data instead of patent data, some of the potential biases and limitations patent data suffer from are bypassed and avoided (Cohen et al. 2000; Hinze and Schmoch 2004; Nagaoka et al. 2010). Methodologically, an extended gravity approach is taken which helps shed light on the roles of standard gravitational forces like market size, distance, cultural or physical proximity for the internationalization of R&D, extended to include additional technology and innovation related drivers of R&D internationalization.

The results highlight the essential role of market size, cultural, physical and technological proximity for the process of R&D internationalization. Moreover, it finds evidence that additional scientific or technological capabilities matter strongly: abundant human capital in the host country is conducive to R&D activities of foreign-owned firms, while a lack of human capital in the home country appears to encourage the relocation of innovative activities abroad. Similarly, strong and internationally competitive R&D capabilities in the host country turn out to be conducive to R&D efforts of foreign-owned firms. They can exploit these capabilities for own research activities. Finally, the analysis finds that R&D expenditure of foreign-owned firms is regionally decentralized and not concentrated within the EU.

The remainder of the paper is structured as follows. Section 10.2 presents related literature and previous empirical evidence on important determinants of cross-border R&D activities while Sect. 10.3 discusses the data used in the analysis and provides some general patterns of R&D internationalization. Furthermore, some hypotheses are formulated that will be tested empirically in the ensuing analysis. The econometric specifications tested are outlined in Sect. 10.4 while Sect. 10.5 presents and discusses the results. Finally, Sect. 10.6 concludes.

10.2 Related Literature

Empirical evidence is quickly mounting: the process of the internationalization of R&D is the product of a number of different key factors and drivers. In that respect, an ever growing body of empirical literature consistently points at the pivotal role played by economic size of countries in fostering cross-border R&D activities. Specifically, foreign-owned firms may have to adapt their products and production processes to suit local demand patterns, consumer preferences or to comply with legal regulations and laws. In view of that, these firms may find it easier to cover the cost of adaptive R&D in larger markets with higher demand for their goods and services, better sales prospects and consequently larger revenues. In the same way, foreign-owned firms may have stronger incentives to develop new products or processes from scratch in faster growing markets. As highly uncertain and risky activities, innovative activities gobble up immense resources that can easier and faster be recovered on larger markets with more promising market potentials. Dachs and Pyka (2010) use EPO patents for the period 2000–2005 to identify essential determinants of cross-border patents. They show that cross-border patenting activities are significantly higher if both home and host economies are larger.

Moreover, empirical studies have stressed that cross-country differences in the quality and size of a skilled workforce are an important determinant of the process of R&D internationalization: Lewin et al. (2009) demonstrate that a shortage of high skilled science and engineering talent in the US explains the relocation of product development to other parts of the world while Hedge and Hicks (2008) stress that innovative activities of overseas US subsidiaries are strongly related to the scientific and engineering capabilities of the host countries. A similar pull-effect of human capital is identified by Erken and Kleijn (2010) who show that strong human resources in science and technology in the host country are strong location factors for international R&D activities.

In addition, technological proximity which captures similarities in technological specialization among countries is found to be conducive to cross-border innovative activities. Guellec and van Pottelsberghe de la Potterie (2001) find that countries with similar patterns of technological specialization tend to more strongly cooperate in patenting activities.

Similarly, stronger R&D efforts in terms of higher R&D intensities in both home and host countries foster the internationalization of R&D (Dachs and Pyka 2010). Moreover, effects tend to differ across countries as the technological strength of the home country appears to exert a stronger push effect than the technological strength of the host country. In a similar vein, Erken and Kleijn (2010) show that the stock of private R&D capital in a country represents an essential driver of the process of R&D internationalization, either as a guarantee for sizeable knowledge spillovers, or as a so-called ‘place-to-be effect’.

The attractiveness of countries for overseas R&D activities is also shaped by public policy intervention. Specifically, as highlighted by Steinmueller (2010), science, technology and innovation (STI) policy measures like public subsidies for R&D performing firms or measures to foster cooperation among firms or between firms and universities or research institutes may remove obstacles to innovation and strengthen the capabilities of national innovation systems. An innovation-friendly environment, in turn, may be a considerable locational advantage and influence internationalization decisions of firms in R&D. Related to that, Dachs and Pyka (2010) emphasize that strong IPR mechanisms also matter for cross-border patenting. As such, they highlight that systematic policies aimed at the strengthening of prevailing IPR mechanisms help render cross-border patenting activities more attractive.

Moreover, while differences in labour cost between the home country and locations abroad are one of the most important motives for the internationalization of production, empirical evidence that differences in the cost of R&D personnel are a major driver for the internationalization of R&D is weak, however: compared to other factors, cost advantages of R&D location are found to be pretty modest (Booz Allen Hamilton and INSEAD 2006; Thursby and Thursby 2006; Kinkel and Maloca 2008; Belderbos et al. 2009; European Commission 2010). However, cost differences appear to gain importance when firms consider to locate R&D and innovation activities in emerging economies, or when firms have to choose between two similarly attractive locations (Booz Allen Hamilton and INSEAD 2006; Thursby and Thursby 2006; Cincera et al. 2009).

The negative relationship between distance and any bilateral flows of either goods, capital or people is one of the most robust findings in the rich strand of literature emerging from the gravity model tradition. Traditionally, as emphasized by Tinbergen (1962), distance is interpreted as a proxy for transportation costs or an index of uncertainty and information costs firms have to shoulder when penetrating foreign markets. In the case of overseas R&D, these costs include additional costs of coordinating geographically dispersed R&D activities, the costs of transferring knowledge over distance, and a loss of economies of scale and scope when R&D becomes more decentralized (Sanna-Randaccio and Veugelers 2007; Gersbach and Schmutzler 2011). Related evidence is provided by Castellani et al. (2011) who throw light on the specific role of distance for cross-border R&D FDI relative to manufacturing investments. They emphasize that once social, cultural and institutional factors like shared language or membership in the same regional trade agreement are accounted for, the location of R&D labs abroad is independent of geographic distance and therefore equally likely to be found close by or farther away. This is taken as conclusive evidence for the limited role of transportation costs but the pivotal role of uncertainty and prevailing informational barriers and costs in deterring cross-border R&D FDI. In contrast, however, geographic distance remains an important determinant for FDI in manufacturing or other types of FDI.

Supportive evidence also emerges for the importance of both cultural and physical proximity between countries for cross-border flows and activities, as typically proxied by common language or common borders, respectively. Such proximity effects potentially counteract the effects of pure geographical distance and thus have to be taken into account separately. In particular, lower cultural barriers between culturally similar countries as well as shared borders between countries often facilitate the flow of goods, capital or people. Strong cultural ties between countries ease communication and the exchange of information and knowledge across borders, rendering cross-border flows and activities easier and less costly. Physical proximity reduces transportation and travel costs and therefore further enhances cross-border flows. Various authors stress that foreign-owned firms have to overcome additional institutional and cultural barriers, a disadvantage that is known as the ‘liability of foreignness’ (Zaheer 1995; Eden and Miller 2004). This concept captures foreign-owned firms’ lack of market knowledge but also their lower degree of embeddedness in informal networks in their host countries, decisive elements for foreign-owned firms when devising innovation strategies in terms of whether and how to develop new or adapt existing products and/or processes to local preferences and what resources to allot to these innovative activities. Supportive empirical evidence is provided by Guellec and van Pottelsberghe de la Potterie (2001) who use patent data for 29 OECD member countries to explain prevailing patterns of cross-border ownership of inventions as well as of research cooperation in the mid-1980s and the mid-1990s. They stress that both cross-border ownership of patent inventions are more widespread among countries that share common borders. Moreover, Guellec and van Pottelsberghe de la Potterie (2001) also demonstrate that cross-border patenting and cooperation is significantly stronger among culturally similar countries.

Finally, empirical evidence also points at the regional concentration or scientific integration of cross-border inventive activities. As such, cross-border patenting is higher among EU-15 countries (Dachs and Pyka 2010), while probably due to the shared history and broad cultural similarities, cross-border ownership of inventions as well as of research cooperation was stronger among Nordic countries (Guellec and van Pottelsberghe de la Potterie 2001).

From this survey a couple of hypotheses concerning R&D expenditure decisions can be extracted which will be explored and tested below. First, market size as proxied by GDP and GDP per capita of the host and home countries is an important determinant of bilateral R&D activities. Second, concerning the quality and size of skilled workforce both push and pull factors are at play with a lack of such workers forcing firms to invest abroad whereas a skill workforce might attract R&D activities in the host countries. Third, existing R&D efforts in both the host and home countries are conducive to further bilateral R&D spending. Finally, there is a set of variables capturing issues of distance and proximity: particularly, geographical distance is expected to correlate negatively with bilateral R&D expenditures whereas factors like technological, cultural and physical proximity (measured e.g. by language and border effects) are expected to correlate positively.

Some potential additional determinants emerge from the literature survey which however could not explicitly be taken into account either due to high correlation with other independent variables or a lack of data. These variables are labour costs (which are highly correlated with GDP per capita) and measures of public policy intervention. Instead, a number of dummies will be included to capture such effects. The next section presents descriptive patterns of bilateral R&D expenditures and discusses the sources of data that will be used for the econometric analysis.

10.3 The Role of Gravitational Forces

The ensuing analysis is based on a recently compiled database of bilateral business R&D expenditure of foreign affiliates in the manufacturing sector of selected OECD countries.[1] Bilateral R&D expenditure of firms from country A in country B will be referred to as inward R&D expenditure or R&D expenditure of foreign affiliates throughout the text.

Data on inward R&D expenditure cover the period from 2001 to 2007 and was collected from national sources and compiled by the Austrian Institute of Technology (AIT) and the Vienna Institute for International Economic Studies (wiiw) in 2011.[2] This data set was complemented by additional data from different sources: standard gravity indicators such as distance ($DIST_{ij}$), common language

[1] The following OECD countries are covered: Austria (AUT), Belgium (BEL), Bulgaria (BUL) Canada (CAN), the Czech Republic (CZE), Denmark (DNK), Estonia (EST), Finland (FIN), France (FRA), Germany (GER), Greece (GRC), Hungary (HUN), Ireland (IRL), Japan (JPN), the Netherlands (NLD), Norway (NOR), Poland (POL), Portugal (PRT), Romania (ROM), Spain (ESP), the Slovak Republic (SVK), Slovenia (SVN), Sweden (SWE), Turkey (TUR), the UK (GBR) and the US (USA).

[2] Data was collected as part of the project 'Internationalisation of business investments in R&D and analysis of their economic impact' and have been slightly revised and updated for this paper.

($COMLANG_{ij}$) or common boarder ($COMBORD_{ij}$) are taken from databases created by CEPII. Information on real GDP, tertiary school enrolment rates, high-technology exports and patent applications of resident and non-residents and total populations in country i and j come from the World Bank's World Development Indicators (WDI). Finally, information on the technology distance between country i and j was calculated with patent data provided by the EPO PATSTAT database. This index measures correlations in the technological specialisation between countries. It is designed as a matrix of correlation coefficients such that the technology distance proxy increases with a decreasing technological distance between two countries. Descriptive statistics of all variables used in the estimations are provided in Tables 10.4 and 10.5 in the Appendix.

Figures 10.1, 10.2, 10.3 and 10.4 below give a general picture of the magnitudes of R&D internationalization, identify key players (Fig. 10.1) and attractive locations for R&D efforts of foreign affiliates (Figs. 10.2 and 10.3) and show the spatial structure of the network of bilateral R&D expenditure between European countries (Fig. 10.4). As such, they reveal important phenomena and underpin the hypotheses that will be tested in the ensuing analysis.

A general picture of inward R&D expenditure in the manufacturing sector by country of origin for key global players (that is the EU, the USA, Japan, China and Switzerland) is drawn in Fig. 10.1 below. The size of each pie chart captures the total amount of inward R&D expenditure in a country, while pie slices represent the volume of inward R&D expenditure by country of origin. Arrows illustrate major relations in inward R&D expenditure between countries. Figure 10.1 emphasizes that, as major recipients of inward R&D expenditure, both, the USA as well as the EU are the two key players in the process of internationalization of R&D. Specifically, in 2007, inward R&D expenditure of US firms in the EU and inward R&D expenditure of EU firms in the US together accounted for two-third of total inward R&D expenditure in manufacturing worldwide.[3]

Moreover, Fig. 10.1 points at the strong mutual importance of both key players for their respective inward R&D expenditure volumes: in 2007, US firms accounted for more than 65 % of total inward R&D expenditure in manufacturing in the EU. Similar, around 62 % of EU inward R&D expenditure in the manufacturing sector stem from US firms located in the EU. In addition, Switzerland was the second most important country of origin with around 16 % of all inward R&D expenditure coming from Swiss firms located in the EU and around 22 % located in the USA. In contrast, Japanese firms located either in the EU or the US accounted for a comparatively small fraction of inward R&D expenditure only.

More recently, China emerged as a new attractive location for R&D efforts of foreign-owned firms. While Chinese data is incomplete and plagued by methodological issues which render a comparison with data from OECD countries difficult,

[3] The European Union is considered as one entity, and intra-EU relationships (for example R&D of German firms in France) are not taken into account.

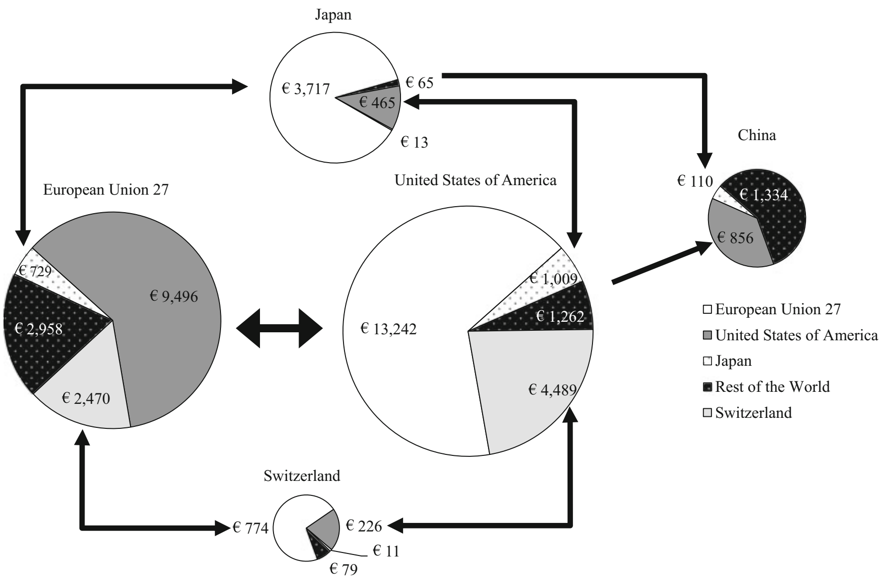


Fig. 10.1 Inward R&D expenditure between the EU, the US, Japan, China and Switzerland: manufacturing only (2007, in EUR million at current prices). *Reading note*: Firms from the European Union spent EUR 774 million on R&D in Switzerland in 2007; Swiss firms spent EUR 2.470 million on R&D in the EU-27 in 2007. Swiss data include also the service sector; data for China is estimated based on national sources and US and Japanese outward data (Source: OECD, Eurostat, national statistical offices, own calculations)

data on R&D expenditure of wholly foreign-owned firms that operate in China suggest around EUR 2.5 billion for the year 2007.

Next, Fig. 10.2 takes a closer look at R&D expenditure of foreign affiliates in the US, by country of origin (between 1998 and 2010) and therefore identifies the importance of inward R&D efforts of single EU countries in the US.[4] Specifically, it depicts the simple country penetration, as the ratio of inward R&D expenditure from a specific EU country to total inward R&D expenditure from the EU in the US and points at the dominance of three EU countries only. As far back as 1998 and up to 2006, affiliates of German, French and British firms accounted for around 80 % of total inward R&D expenditure by EU firms in the US. Throughout, Germany ranked first, followed by the UK and France. Only in 2006 did the UK overtake Germany as the most important investor in R&D in the US. Hence, given that the US is the world's largest economy with a huge market and attractive sales potentials, this supports the hypothesis that market size matters.

The opposite perspective is taken in Fig. 10.3 which depicts R&D expenditure of US foreign affiliates located in the EU, by country of destination (between 1998 and

[4] Due to lacking data on outward R&D expenditure for most EU countries, Fig. 10.2 is based on US inward data.

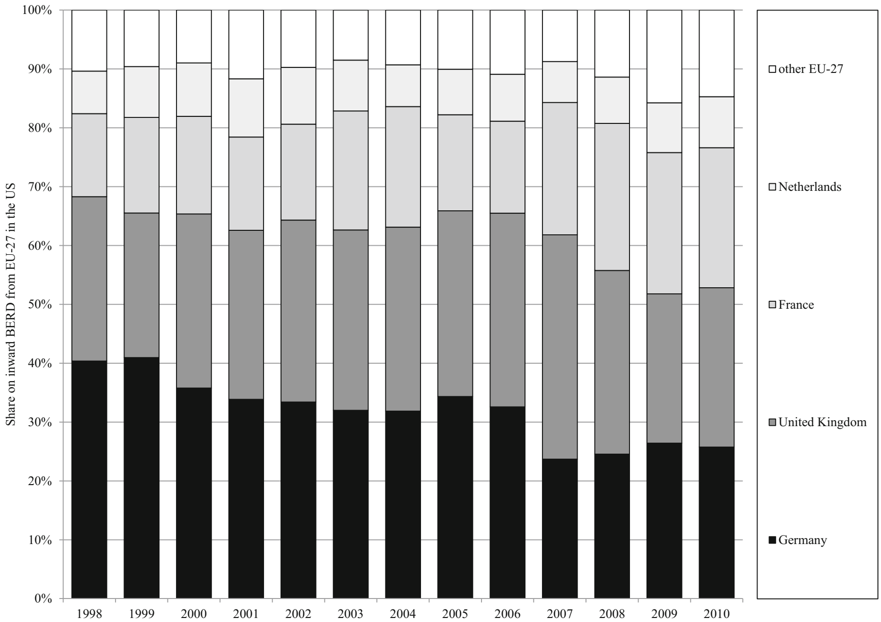


Fig. 10.2 Countries of origin of inward R&D expenditure by EU firms in the US, 1998–2010. *Note:* Total EU-27 includes all European companies except Swiss companies. (Source: OECD based on US data by the US Bureau of Economic Analysis, own calculations)

2010) as the ratio of US outward R&D expenditure in a particular EU country to total US outward R&D expenditure in the EU. It demonstrates that throughout the period from 1998 to 2010, the UK and Germany were the two most important and attractive individual EU countries for US R&D efforts, together absorbing more than 50 % of all US outward R&D expenditure in the EU. However, starting in 2005, France, Italy and Spain appear to have lost some ground while other, smaller Member States have become more attractive locations for US R&D efforts. The importance of the two largest EU economies as key locations for US R&D efforts in the EU underscores above hypothesis that 'the size of the market matters'.

In addition, a comparison of Figs. 10.2 and 10.3 shows that US inward R&D expenditure in the EU is much less concentrated in a few economies only than EU inward R&D expenditure in the US, as small and medium-sized EU economies (like Belgium, Ireland, the Netherlands or Austria) are comparatively more important locations for R&D efforts of US companies than the US is for foreign affiliates from small and medium-sized EU economies in the US.

Finally, Fig. 10.4 zooms in on the EU and depicts the spatial structure of the network inward R&D expenditure among European countries. The edge size (that is the link between countries) corresponds to the sum of inward R&D expenditure of firms from country A in country B and vice versa[5] while the node size of each

[5] This measure corresponds to weighted degree centrality in the social network analysis literature.

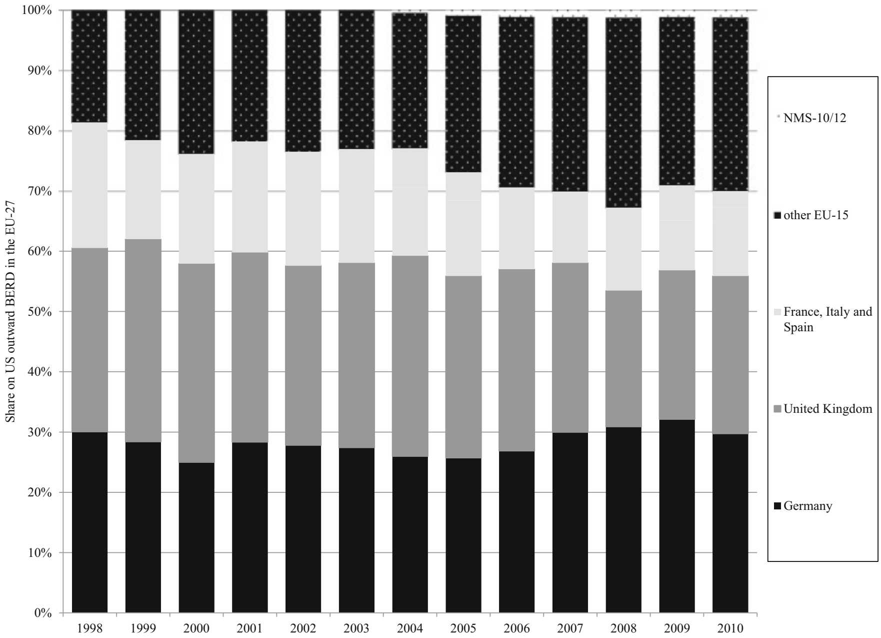


Fig. 10.3 Location of inward R&D expenditure of US firms in the EU,1999–2010. *Note:* * NMS-10/12 comprises the Czech Republic, Estonia, Cyprus, Latvia, Lithuania, Hungary, Malta, Poland, Slovenia and Slovakia (all from 2004 to 2007) and in 2007 Bulgaria and Romania also. (Source*:* OECD based on US outward data by the US Bureau of Economic Analysis, own calculations)

country corresponds to the total sum of inward R&D expenditure in the country. Nodes are located at the capital cities of each country.

The spatial network map for 2007 reveals a strong regional clustering of inward R&D expenditure in the centre of Europe while the periphery is participating to a lower degree. There are strong neighbouring effects between some countries, in particular Germany, the Netherlands, Switzerland and Austria. Moreover, Germany appears as the central hub, showing high interaction intensity, particularly with its direct neighbours the Netherlands, Switzerland, Austria or France. Similar neighbourhood effects are apparent for the UK or Spain, which show particular high interaction intensity with Sweden and France or France and Belgium, respectively. In contrast, Finland has a diverse and big set of partner countries, in terms of absolute size, however, the interactions are comparatively low.

All in all, while New EU Member States (NMS) are in general connected to the system of R&D investment in Europe, the magnitudes are comparatively low, with the Czech Republic and Hungary showing the strongest R&D-based embeddedness. This peripheral position of NMS may mainly be due to the low number of multinational firms originating from there. Interestingly, business R&D investment of NMS appears far less integrated than public research (including universities and research institutions): Scherngell and Barber (2011) use information on international collaboration patterns in the European Framework Programmes (FPs) and

Fig. 10.4 Inward R&D expenditure flows between European countries (2007). *Note:* The strength of lines between country A and B corresponds to the sum of R&D expenditure of firms from country A which operate in country B, and vice versa. The size of the node per country corresponds to the sum of R&D expenditure of all foreign-owned firms in the country (Source*:* OECD, Eurostat, national statistical offices, own calculations)

demonstrate that NMS seem to be rather well integrated in pan-European research collaborations, while Fig. 10.4 highlights that this is less so for R&D efforts in the industry sector.

10.4 Econometric Specification

In order to identify both home and host country characteristics that are either conducive or obstructive to the process of R&D internationalization, a gravity model approach is pursued. Generally, in the empirical literature, gravity models are popular and well known for their success in explaining international trade flows (see Anderson 1979 or Deardorff 1984 for a theoretical discussion and Breuss and Egger 1999 or Helpman et al. 2008 for some empirical results).

In essence, the gravity equation for trade says that trade flows between two countries are proportional to the two country's size (as proxied by GDP) but inversely related to the distance between them. Moreover, these models also often account for physical or cultural proximity in terms of shared border, common language or colonial history, respectively. Increasingly, gravity models are also used to explain FDI flows (Brainard 1997; Jeon and Stone 1999 or Bergstrand and Egger 2007), migration flows (Lewer and Van den Berg 2008) or flows of workers' remittances (Lueth and Ruiz Arranz 2006) between countries.

More recently, gravity models also found their way into the analysis of cross-border inventive activities (see, for example Guellec and van Pottelsberghe de la Potterie 2001; Dachs and Pyka 2010 or Castellani et al. 2011). In some cases simple gravity specifications might suffer from interdependencies such that FDI or also R&D expenditures in one destination are not independent from activities in other destinations (see e.g. Bloningen 2005, for a survey of FDI determinants). Furthermore in some cases more complex spatial interdependencies might matter as e.g. market size of neighbouring countries or regions affect FDI or R&D decisions. Given the limitations of the data at hand such effects can however not be considered in the specification used in this paper.

Hence, following the tradition of the gravity literature, the following econometric specifications are estimated to shed light on the roles of home and host country characteristics in driving inward R&D expenditure:

$$\begin{aligned} \ln RD_{ijt} = {} & \lambda_t + \alpha_i + \alpha_j + \beta_1 \ln DIST_{ij} + \beta_2 COMLANG_{ij} + \beta_3 COMBORD_{ij} + \ldots \\ & \ldots + \beta_4 \ln GDP_{it} + \beta_5 \ln GDP_{jt} + \delta_z X_{zijt} + \varepsilon_{ijt}. \end{aligned} \tag{10.1}$$

And, if account is also taken of the level of economic development:

$$\begin{aligned} \ln RD_{ijt} = {} & \lambda_t + \alpha_i + \alpha_j + \beta_1 \ln DIST_{ij} + \beta_2 COMLANG_{ij} + \beta_3 COMBORD_{ij} + \ldots \\ & \ldots + \beta_4 \ln GDP_{it} + \beta_5 \ln GDP_{jt} + \beta_6 \ln\left(\frac{GDP_{it}}{POP_{it}}\right) + \beta_7 \ln\left(\frac{GDP_{jt}}{POP_{jt}}\right) + \delta_z X_{zijt} + \varepsilon_{ijt}, \end{aligned} \tag{10.2}$$

where $\ln RD_{ijt}$ is the log of business R&D expenditure of foreign affiliates from country j located in the host country i at time t.

$\ln DIST_{ij}$ is the log of the geographical distance between country i and j, measured as the simple distance between most populated cities (in km). As an index of uncertainty and additional information costs (like additional costs of coordinating geographically dispersed R&D activities or of transferring knowledge over distance), R&D expenditure of foreign-owned firms is expected to decline with growing distance.

$COMLANG_{ij}$ and $COMBORD_{ij}$ are dummies taking the value 1 if the two countries i and j share a common language or border, respectively. Both are included to capture cultural and physical proximity between country i and j and are expected to foster R&D activities of foreign-owned firms. Specifically, strong cultural ties between countries ease communication and the exchange of information and knowledge across borders, while physical proximity reduces transportation costs, together rendering cross-border R&D activities comparatively easier and less costly.

Furthermore, $\ln GDP_{it}$ and $\ln GDP_{jt}$ refer to the log of real gross domestic product in country *i* and *j*, respectively and are proxies for the economic size of countries *i* and *j*. Positive effects are expected, since, given their superior market potentials and sales prospects that allow for an easy and quick recovery of sizeable R&D outlays, larger markets are more attractive and conducive to R&D efforts of foreign-owned firms.

Account is also taken of the role a country's level of economic development has in attracting business R&D expenditure of foreign-owned firms. As such, wealthier economies (as proxied by their respective real GDPs per capita, namely $\ln(GDP_{it}/POP_{it})$ for country *i* and $\ln(GDP_{jt}/POP_{jt})$ for country *j*) may not only have a higher purchasing power, but may also be home to consumers with a more pronounced 'love for variety' (see Dixit and Stiglitz 1977) so that foreign-owned firms which develop or produce novel products or processes consider economies with higher standards of living more attractive markets with better profit perspectives.

In addition to above standard gravity model indicators, innovation related indicators are included to throw light on their roles in driving the internationalization of R&D. X_{zijt} is a matrix of *z* additional innovation related variables that are expected to affect R&D expenditure of foreign affiliates to different degrees. In particular, the analysis includes gross tertiary school enrolment rates in country *i* and *j* to account for the pivotal role the quality of human capital plays for any successful R&D efforts (ENR_TER). Specifically, empirical evidence highlights that cross-country differences in the quality and size of a skilled workforce are an important determinant of R&D internationalization: Lewin et al. (2009) show that firms relocate product development to other parts of the world if faced with a shortage of skilled science and engineering talent, while Hedge and Hicks (2008) highlight that an abundance of graduates in science and technology and strong scientific and engineering capabilities in a host country are able to attract business R&D into a host country.

Moreover, to capture a country's general level of inventiveness, the ratio of patent applications of residents to patent applications of non-residents in country *i* and *j* is included (PA_RATIO). Specifically, more inventive host countries are attractive for foreign-owned firms seeking to harness prevailing local technology and innovation capabilities for the development of new products or processes.

R&D activities of foreign-owned firms may also crucially depend on differences in countries' abilities to develop and produce internationally competitive high-technology products. In particular, countries with strong indigenous R&D and technological capabilities tend to specialize in high-technology industries and to generate high-technology products (and services) that more easily withstand fierce competition in the global arena. Hence, a high share of high-technology exports in GDP is indicative of an internationally competitive indigenous R&D base foreign-owned firms can harness to successfully develop new products and processes or to adapt products and processes to local conditions and preferences. Therefore, high-technology exports of country *i* and *j* (defined as the share of high-technology exports that are produced with high R&D intensity in total GDP) are included to capture the quality of indigenous R&D and technological capabilities (HTX_SH).

Additionally, cross-country differences in the levels of technological development may also affect the internationalization of R&D. Specifically, there has been a long-standing debate in the FDI literature on the existence and extent of technological spillovers from foreign direct investments with, however, lacking consensus. Some empirical studies lend support to the catching-up hypothesis put forward by Findlay (1978) and find that technological spillovers increase with a widening of the technology distance (e.g. Castellani and Zanfei 2003 or Peri and Urban 2006). Others suggest the opposite such that only a narrow technology distance is conducive to technological spillovers (e.g. Kokko et al. 1996 or Liu et al. 2000) as closer levels of technological development across countries renders them technologically more compatible, with sufficient absorptive capacities to benefit from each other's research efforts and successes. Hence, the technological distance between country i and j is included, in terms of a correlation coefficient which, by construction, lies between [0, 1] (TDIS). A high value of the coefficient indicates a narrow technological distance and similar specialization patterns between two countries.

Furthermore, dummies for EU membership are included which capture whether only country i is a member of the EU, whether country j is a member of the EU only, or whether both i and j are EU-member countries. This will show whether R&D expenditure of foreign-owned firms is higher between EU member countries or between EU and non-EU countries. Boschma (2005) refers to institutional proximity to capture that a common institutional set-up of two countries may facilitate business activities of firms abroad.

Finally, Eq. 10.1 also includes host and home country fixed effects (α_i and α_j for country i and j, respectively) to account for country heterogeneity and year fixed effects (λ_t) to take account of common macroeconomic shocks.

10.5 Results

Results are presented in Table 10.1 for different econometric specifications (see Eqs. 10.1 and 10.2) and estimation techniques: columns (1) to (3) provide results for the basic specification as given in Eq. 10.1, while columns (4) to (6) also account for the effect of the level of economic development on R&D expenditure of foreign-owned firms as specified in Eq. 10.2. Moreover, from a methodological point of view, columns (1) and (4) provide results for pooled OLS, columns (2) and (5) for fixed effects for receiving and sending countries and columns (3) and (6) for random effects specific for bilateral country pairs. The main shortcoming of the pooled OLS approach lies in its inability to allow for heterogeneity of host and home countries since it assumes that all countries are homogeneous. This is remedied by fixed effects (column (2)) and random effects approaches (column (3)) which explicitly account for the heterogeneity of both individual host and home countries as well as for heterogeneity of host-home country pairs, respectively.

Table 10.1 Host and home country determinants of R&D internationalization (2001–2007)

Dep.Var.: log of R&D expenditure of foreign-owned firms from country j in country i

Estimation technique	Pooled OLS	Country FE	Country-pair RE	Pooled OLS	Country FE	Country-pair RE
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Constant	−21.499***	−85.323**	−18.047***	−35.198***	−110.343**	−31.121***
	(18.55)	(2.00)	(10.88)	(21.18)	(2.13)	(13.25)
Log distance	−0.725***	−0.276***	−0.819***	−0.558***	−0.278***	−0.612***
	(7.70)	(3.05)	(5.66)	(6.11)	(3.07)	(4.37)
Common language	0.645***	−0.134	1.159***	0.091	−0.137	0.585
	(2.72)	(0.64)	(3.13)	(0.39)	(0.65)	(1.64)
Common border	0.399*	1.346***	0.292	0.873***	1.352***	0.761**
	(1.88)	(7.09)	(0.83)	(4.26)	(7.11)	(2.26)
Log real GDP HOST	1.082***	0.905	1.041***	0.832***	−0.938	0.770***
	(21.65)	(0.47)	(13.96)	(12.57)	(0.13)	(8.05)
Log real GDP HOME	0.896***	5.754***	0.841***	0.790***	9.946*	0.748***
	(17.54)	(2.62)	(11.12)	(16.04)	(1.80)	(10.33)
Log real GDP per capita HOST				0.666***	1.868	0.772***
				(4.86)	(0.29)	(4.13)
Log real GDP per capita HOME				1.139***	−4.851	0.938***
				(10.22)	(0.84)	(6.25)
Tertiary enrolment rate HOST	0.044***	0.011	0.029***	0.023***	0.009	0.009
	(9.04)	(0.45)	(4.61)	(4.11)	(0.38)	(1.35)
Tertiary enrolment rate HOME	0.002	0.007	−0.005	−0.008**	0.009	−0.011**
	(0.43)	(0.37)	(1.05)	(2.11)	(0.50)	(2.38)
Ratio patent applications residents HOST	−0.050***	0.009	−0.003	−0.050***	0.010	−0.003
	(2.95)	(0.26)	(0.18)	(3.12)	(0.31)	(0.22)
Ratio patent applications residents HOME	−0.081***	−0.050	−0.021	−0.096***	−0.053	−0.023
	(4.21)	(1.21)	(1.17)	(5.25)	(1.27)	(1.28)

(continued)

Table 10.1 (continued)

Dep.Var.: log of R&D expenditure of foreign-owned firms from country j in country i						
Estimation technique	Pooled OLS	Country FE	Country-pair RE	Pooled OLS	Country FE	Country-pair RE
Share high-tech exports HOST	0.036*	0.039	0.049**	0.033*	0.039	0.045**
	(1.80)	(0.45)	(2.06)	(1.68)	(0.46)	(1.97)
Share high-tech exports HOME	0.021	−0.051	−0.020	0.016	−0.035	−0.023
	(1.19)	(1.21)	(1.07)	(0.96)	(0.76)	(1.27)
Technological distance	−0.250	1.388**	−0.318	0.779*	1.362**	0.510
	(0.55)	(2.49)	(0.47)	(1.78)	(2.43)	(0.79)
Dummy: HOST EU-member	1.031***		0.434	0.694**		0.348
	(3.27)		(0.87)	(2.32)		(0.74)
Dummy: HOME EU-member	1.797***		1.504***	1.610***		1.347***
	(5.60)		(2.80)	(5.30)		(2.66)
Dummy: HOST and HOME EU-member	1.259***		0.346	1.270***		0.518
	(3.73)		(0.66)	(3.96)		(1.05)
Year fixed effects	YES	YES	YES	YES	YES	YES
Observations	1054	1054	1054	1054	1054	1054
Adj. R^2	0.587	0.779	0.580	0.631	0.779	0.624
Number of i			362			362

Note: t-statistics in parentheses
All regressions include time fixed effects. Estimation results in columns (1) and (4) are based on pooled OLS, results in columns (2) and (5) use country fixed effects for both receiving and sending countries while results in columns (3) and (6) use random effects specific for bilateral country-pairs
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

As expected, the size of both home and host countries emerges as one key determinant of R&D expenditure of foreign-owned firms. In particular, a 1 % increase in the both host and home country's market size is associated with a rise in R&D expenditure of foreign affiliates by between 0.8 % and 1 %. However, size effects slightly differ across countries and tend to be stronger in the host country. This again provides supportive evidence of the 'size matters' hypothesis.

The analysis also demonstrates that apart from size, prevailing levels of economic development matter for the scale of cross-border R&D expenditure. In particular, cross-border R&D expenditure tends to be higher in wealthier economies: a 1 % rise in the host country's GDP per capita increases R&D expenditure of foreign-owned firms by around 0.7–0.8 % while a similar 1 % increase in the home country's GDP per capita has a slightly higher effect and is associated with an around 1 % increase in R&D efforts of foreign-owned firms.

Moreover, light is shed on the particular roles additional innovation-related indicators play for the process of R&D internationalization. Results in Table 10.1 highlight that human capital emerges as a non-negligible determinant of cross-country R&D expenditure of foreign-owned firms. However, results also reveal that underlying dynamics appear to differ across specifications. Specifically, column (1) to (3) show that, in line with findings by Hedge and Hicks (2008), there is evidence that a strong human capital base in the host country attracts business R&D: a 1 % point increase in the host country's tertiary enrolment rate is associated with a 2.9 % increase in inward R&D expenditure. In contrast, results presented in columns (4) to (6) stress that, once levels of economic development of both host and home country are also taken into account, an abundance of human capital in the home country appears to discourage R&D internationalization activities of foreign-owned firms. This is in line with findings by Lewin et al. (2009) who emphasize that firms tend to relocate product development to other parts of the world if faced with a shortage of skilled science and engineering talent at home. However, diverging results on the role of human capital for the process of R&D internationalization are not – as it may seem – contradictory but suggest that, once levels of economic development are also controlled for, the host country's endowment with human capital becomes of secondary importance while its level of development (together with its economic size) assumes the role of main driver of the process of R&D internationalization.

Similarly, there is evidence that a strong and internationally competitive indigenous R&D base in the host country is conducive to R&D expenditure of foreign-owned firms. Hence, host countries that specialize in and generate internationally competitive high-technology products are attractive R&D locations for foreign-owned firms as they possess indigenous technological capabilities foreign-owned firms can exploit for their innovative activities. In contrast, no decisive role can be attributed to a country's general level of inventiveness in fostering R&D expenditure of foreign affiliates.

Finally, the results support the hypothesis concerning distance and proximity related determinants. The analysis finds consistent evidence for the pivotal role geographic distance between countries plays in curbing the process of R&D internationalization. Specifically, inward R&D expenditure falls by between

0.3 % and 0.8 % in response to a 1 % increase in distance between countries, where distance captures additional coordinative costs of regionally dispersed R&D activities or diseconomies of scale and scope as a result of more decentralized R&D activities.

Moreover, cultural proximity tends to be a conducive determinant of R&D expenditure of foreign affiliates. This supports the 'liability of foreignness' hypothesis formulated above: lower cultural barriers improve market knowledge and the understanding of customer needs and facilitate communication and the exchange of information and knowledge across borders. In a similar vein, physical proximity also fosters the internationalization of R&D such that foreign affiliates located in neighbouring countries tend to spend significantly more on R&D activities than affiliates located farther away.

In line with results by Guellec and van Pottelsberghe de la Potterie (2001), the analysis also emphasizes that technological distance matters. In particular, R&D expenditure of foreign-owned firms appears to be higher between countries with similar technological specializations which may indicate that R&D activities of foreign-owned firms are attracted by potential spillovers in technological domains similar to their own specialization. Finally, the analysis also demonstrates that cross-border R&D expenditure tend to be regionally dispersed across EU as well as non-EU member countries.

10.6 Summary and Conclusion

In the course of the last two decades, R&D expenditure of foreign-owned firms increased tremendously, an indication that firms increasingly conduct research and development outside their home countries. Against that backdrop, the analysis identified important determinants of this more recent process of increased R&D internationalization. It used a novel data set on R&D expenditure of foreign-owned firms in the manufacturing sector of a set of OECD countries, spanning the period from 2001 to 2007.

Generally, the results attribute a pivotal role to geographic distance in curbing R&D expenditure of foreign-owned firms. This may be explained by the costs of R&D internationalization (like additional costs of coordinating geographically dispersed R&D activities or of transferring knowledge over distance) which tend to noticeably increase with distance which, in turn, renders highly dispersed R&D activities more costly and consequently less attractive. Moreover, cultural proximity which facilitates communication and the exchange of knowledge as well as physical proximity which turns neighbouring countries attractive R&D hubs emerge as important determinants of the process of R&D internationalization. Furthermore, as expected, economic size and wealth of host and home countries alike are key determinants which – in the light of larger markets with more favourable sales prospects as well as wealthier consumers with a stronger and more pronounced 'love for variety' – stimulate R&D efforts of foreign affiliates.

In addition, R&D efforts of foreign-owned firms also respond to additional scientific or technological capabilities. In particular, while some indication is found that a strong human capital base in the host country attracts business R&D of foreign-owned firms, there is additional evidence that an abundance of human capital in the home country tends to curtail the relocation of innovative activities to other parts of the world. Similarly, a strong and internationally competitive indigenous R&D base in the host country which foreign-owned firms can harness and exploit for their own research activities is conducive to R&D expenditure of foreign affiliates. Furthermore, R&D expenditure of foreign-owned firms is also significantly stronger among countries with similar levels of technological development, which renders technological compatibility among countries a non-negligible driver of the process of R&D internationalization. Finally, some indication is found that R&D expenditure of foreign-owned firms is regionally decentralized and not concentrated within the EU.

These results have important implications for science, technology and innovation policy. They point at areas where governments can take concerted action to render their countries more attractive for R&D activities of foreign-owned firms. These critical areas are science and education. Governments that succeed in strengthening domestic research and development capabilities and in raising tertiary enrolment rates may also succeed in attracting R&D of foreign-owned firms (Veugelers et al. 2005; OECD 2008a; De Backer and Hatem 2010). This study provides empirical evidence on how proximity among countries and country-specific attributes like economic size, wealth, inventiveness, etc. affects the intensity of cross-country R&D flows.

Though this sheds a first light on determinants on this increasingly important phenomenon, analyses in this field still suffer from severe data limitations and inconsistencies which have to be addressed and resolved in future research. Other potentially important factors capturing R&D and innovation systems, interaction with public R&D and institutions like universities and research institutions, market structures and FDI flows, etc. would also have to be considered to give a more complete picture of R&D flows across countries. Methodologically a comprehensive panel data set should allow to further account for spatial dependencies and spatial lag structures incorporating effects of neighbouring countries performance and market potentials (see, e.g., Chap. 6 of this volume by Chun). Finally, R&D patterns are largely determined by a few, potentially large, enterprises suggesting that firm level data and firm as well as country case studies would be enlightening though challenging avenues for future research (see Dachs et al. 2014, for some detailed evidence).

Acknowledgements The analysis of this paper is based on data collected by the project 'Internationalisation of business investments in R&D and analysis of their economic impact', commissioned by the European Commission, DG Research and Innovation (Contract Nr. RTD/DirC/C3/2010/SI2.563818). Special thanks go to Matthieu Delescluse, who supported the project as project officer. Moreover, we acknowledge the support by EUROSTAT in collecting the data. The authors also thank the participants of special session on the internationalisation of business R&D held at the Annual ERSA conference 2012 in Bratislava for their useful comments.

Appendix

Table 10.2 Correlation matrix for host and home country determinants of R&D internationalization

	Log DIST	COMLANG	COMBORD	Log RGDP HOST	Log RGDP HOME	ENR_TER HOST	ENR_TER HOME	PA_RAT HOST	PA_RAT HOME	HTX_SH HOST	HTX_SH HOME	TDIST
Log DIST	1											
COMLANG	−0.096	1										
COMBORD	−0.454	0.300	1									
Log RGDP HOST	0.230	0.094	0.000	1								
Log RGDP HOME	0.308	0.096	0.001	−0.026	1							
ENR_TER HOST	0.157	−0.052	−0.061	0.197	−0.004	1						
ENR_TER HOME	−0.074	−0.053	−0.010	−0.003	0.064	0.040	1					
PA_RAT HOST	−0.048	−0.032	0.010	0.074	0.001	0.335	0.055	1				
PA_RAT HOME	−0.113	−0.032	0.030	0.000	0.087	0.052	0.414	0.042	1			
HTX_SH HOST	−0.033	0.085	−0.014	−0.144	0.005	−0.067	−0.014	−0.012	−0.011	1		
HTX_SH HOME	−0.053	0.077	0.000	0.004	−0.040	−0.010	0.034	−0.004	0.056	−0.020	1	
TDIST	−0.009	0.175	0.100	0.288	0.347	0.189	0.156	0.077	0.112	0.033	0.070	1

Table 10.3 Correlation matrix for host and home country determinants of R&D internationalization – with levels of economic development

	Log DIST	COMLANG	COMBORD	Log RGDP HOST	Log RGDP HOME	Log RGDP pc HOST	Log RGDP pc HOME	ENR_ TER HOST	ENR_ TER HOME	PA_ RAT HOST	PA_ RAT HOME	HTX_ SH HOST	HTX_ SH HOME	TDIST
Log DIST	1													
COMLANG	−0.096	1												
COMBORD	−0.454	0.300	1											
Log RGDP HOST	0.230	0.094	0.000	1										
Log RGDP HOME	0.308	0.096	0.001	−0.026	1									
Log RGDP pc HOST	0.133	0.138	−0.045	0.502	−0.012	1								
Log RGDP pc HOME	−0.092	0.102	0.022	−0.011	0.288	−0.015	1							
ENR_TER HOST	0.157	−0.052	−0.061	0.197	−0.004	0.256	−0.018	1						
ENR_TER HOME	−0.074	−0.053	−0.010	−0.003	0.064	−0.029	0.506	0.040	1					
PA_RAT HOST	−0.048	−0.032	0.010	0.074	0.001	0.068	−0.007	0.335	0.055	1				
PA_RAT HOME	−0.113	−0.032	0.030	0.000	0.087	−0.021	0.223	0.052	0.414	0.042	1			
HTX_SH HOST	−0.033	0.085	−0.014	−0.144	0.005	0.131	0.004	−0.067	−0.014	−0.012	−0.011	1		
HTX_SH HOME	−0.053	0.077	0.000	0.004	−0.040	0.004	0.189	−0.010	0.034	−0.004	0.056	−0.020	1	
TDIST	−0.009	0.175	0.100	0.288	0.347	0.164	0.200	0.189	0.156	0.077	0.112	0.033	0.070	1

Table 10.4 Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Log RDij	1,054	2.47	2.89	−4.61	8.78
Log distance	1,054	7.34	1.09	4.09	9.32
Common language	1,054	0.09	0.29	0.00	1.00
Common border	1,054	0.15	0.36	0.00	1.00
Log RGDP HOST	1,054	12.40	1.72	8.73	16.23
Log RGDP HOME	1,054	13.03	1.56	9.04	16.23
Tertiary enrolment rate HOST	1,054	60.70	13.22	24.50	93.80
Tertiary enrolment rate HOME	1,054	61.28	16.55	9.94	96.10
Ratio patent applications residents HOST	1,054	4.91	3.97	0.03	23.92
Ratio patent applications residents HOME	1,054	4.83	3.56	0.04	28.75
Share high-tech exports HOST	1,054	4.43	3.24	0.24	16.19
Share high-tech exports HOME	1,054	4.83	3.59	0.14	32.76
Technological distance	1,054	0.65	0.17	0.10	0.93

Table 10.5 Descriptive statistics – with levels of economic development

Variable	Obs	Mean	Std. Dev.	Min	Max
Log RDij	1,054	2.47	2.89	−4.61	8.78
Log distance	1,054	7.34	1.09	4.09	9.32
Common language	1,054	0.09	0.29	0.00	1.00
Common border	1,054	0.15	0.36	0.00	1.00
Log RGDP HOST	1,054	12.40	1.72	8.73	16.23
Log RGDP HOME	1,054	13.03	1.56	9.04	16.23
Log RGDP pc HOST	1,054	9.51	0.75	7.46	10.62
Log RGDP pc HOME	1,054	9.97	0.56	6.12	10.87
Tertiary enrolment rate HOST	1,054	60.70	13.22	24.50	93.80
Tertiary enrolment rate HOME	1,054	61.28	16.55	9.94	96.10
Ratio patent applications residents HOST	1,054	4.91	3.97	0.03	23.92
Ratio patent applications residents HOME	1,054	4.83	3.56	0.04	28.75
Share high-tech exports HOST	1,054	4.43	3.24	0.24	16.19
Share high-tech exports HOME	1,054	4.83	3.59	0.14	32.76
Technological distance	1,054	0.65	0.17	0.10	0.93

References

Anderson JE (1979) A theoretical foundation for the gravity equation. Am Econ Rev 69:106–116
Belderbos R, Leten B, Suzuki S (2009) Does excellence in academic research attract foreign R&D? UNU-MERIT working paper series no. 066, Masstricht
Bergstrand JH, Egger P (2007) A knowledge-and-physical-capital model of international trade flows, foreign direct investment, and multinational enterprises. J Int Econ 73(2):278–308
Blomström M, Kokko A (2003) The economics of foreign direct investment incentives. NBER working paper no. 9489, Cambridge, MA
Bloningen BA (2005) A review of the empirical literature on FDI determinants. Atlantic Econ J 33:383–403
Booz Allen Hamilton, INSEAD (2006) Innovation: is global the way forward? Paris
Boschma R (2005) Proximity and innovation: a critical assessment. Reg Stud 39:61–74
Brainard SL (1997) An empirical assessment of the proximity concentration tradeoff between multinational sales and trade. Am Econ Rev 87(4):520–544
Breuss F, Egger P (1999) How reliable are estimations of east–west trade potentials based on cross-section gravity analyses? Empirica 26(2):81–95
Castellani D, Zanfei A (2003) Technology gaps, absorptive capacity and the impact of inward investments on productivity of European firms. Econ Innov New Technol 12(6):555–576
Castellani D, Palmero AJ, Zanfei A (2011) The gravity of R&D FDI. Working papers series in economics, mathematics and statistics WP-EMS No. 2011/06. Facoltà di Economia, Università degli Studi di Urbino "Carlo Bo", Urbino
Cincera M, Cozza C, Tübke A (2009) The main drivers for the internationalization of R&D activities by EU MNEs. In: Draft for the 4th annual conference of GARNET network, IFAD, Rome, 11–13 Nov 2009
Cohen WM, Nelson RR, Walsh J (2000) Protecting their intellectual assets: appropriability conditions and why U.S. manufacturing firms patent (or not). NBER working paper no. 7552, Cambridge, MA
Dachs B, Pyka A (2010) What drives the internationalisation of innovation? Evidence from European patent data. Econ Innov New Technol 19:71–86
Dachs B, Stehrer R, Zahradnik G (eds) (2014) The internationalisation of business R&D. Edward Elgar, Cheltenham
De Backer K, Hatem F (2010) Attractiveness for innovation. Location factors for international investment. Organization for Economic Co-operation and Development, Paris
Deardorff A (1984) Testing trade theories and predicting trade flows. In: Kenen PB, Jones RW (eds) Handbook of international economics, vol 1. Elsevier, Amsterdam
Dixit AK, Stiglitz JE (1977) Monopolistic competition and optimum product diversity. Am Econ Rev 67:297–308
Dunning J (1988) Explaining international production. Harper Collins, London
Eden L, Miller S (2004) Distance matters: liability of foreignness, institutional distance and ownership strategy. In: Hitt MA, Cheng JLC (eds) The evolving theory of the multinational firm, advances in international management. Elsevier, Amsterdam
Erken H, Kleijn M (2010) Location factors of international R&D activities: an econometric approach. Econ Innov New Technol 19:203–232
European Commission (DG Research and Innovation, JRC-IPTS) (2010) Monitoring industrial research: the 2009 EU survey on R&D investment business trends. European communities, Luxembourg
Findlay R (1978) Relative backwardness, direct foreign investment and the transfer of technology: a simple dynamic model. Q J Econ 92:1–16
Forsgren M (2008) Theories of the multinational firm. Edward Elgar, Cheltenham
Gersbach H, Schmutzler A (2011) Foreign direct investment and R&D offshoring. Oxf Econ Paper 63:134–157

Guellec D, van Pottelsberghe de la Potterie B (2001) The internationalisation of technology analysed with patent data. Res Policy 30:1253–1266

Hall BA (2010) The internationalization of R&D. University of California, Berkeley. http://elsa.berkeley.edu/users/bhhall/papers/BHH10_RND_international_August.pdf

Hedge D, Hicks D (2008) The maturation of global corporate R&D: evidence from the activity of U.S. foreign subsidiaries. Res Policy 37(3):390–406

Helpman E (2006) Trade, FDI and the organisation of firms. J Econ Lit 44(3):589–631

Helpman E, Melitz M, Rubinstein Y (2008) Estimating trade flows: trading partners and trading volumes. Q J Econ 23(2):441–487

Hinze S, Schmoch U (2004) Opening the black box. In: Moed HF, Glänzel W, Schmoch U (eds) Handbook of quantitative science and technology research. Kluwer Academic, Dordrecht

Hymer SH (1960 [1976]) The international operations of national firms: a study of direct foreign investment. Ph.D. Dissertation, The MIT Press, Cambridge, MA

Jeon BN, Stone SF (1999) Gravity-model specification for foreign direct investment: a case of the Asia-Pacific economies. J Bus Econ Stud 5(1):33–42

Kinkel S, Maloca S (2008) FuE-Verlagerungen in Ausland – Ausverkauf deutscher Entwicklungs-kompetenz? Fraunhofer ISI, Karlsruhe

Kokko A, Tansini R, Zejan M (1996) Local technological capability and productivity spillovers from FDI in Uruguayan manufacturing sector. J Dev Stud 32(4):602–611

Lewer JJ, van den Berg H (2008) A gravity model of immigration. Econ Lett 99(1):164–167

Lewin AY, Massini S, Peeters C (2009) Why are companies offshoring innovation? The emerging global race for talent. J Int Bus Stud 40(6):901–925

Liu X, Siler P, Wang C, Wei Y (2000) Productivity spillovers from foreign direct investment: evidence from UK industry level panel data. J Int Bus Stud 31(3):407–425

Lueth E, Ruiz Arranz M (2006) A gravity model of workers' remittances'. IMF working paper no. 06/290, Washington, DC

Markusen JR (2002) Multinational firms and the theory of international trade. MIT Press, Cambridge, MA/London

Nagaoka S, Motohashi K, Goto A (2010) Patent statistics as an innovation indicator. In: Hall BA, Rosenberg N (eds) Handbook of the economics of innovation, vol 2. Elsevier, Amsterdam

Narula R, Zanfei A (2005) Globalisation of innovation: the role of multinational enterprises. In: Fagerberg J, Movery DC, Nelson RR (eds) The Oxford handbook of innovation. Oxford University Press, Oxford

OECD (2008a) Recent trends in the internationalisation of R&D in the enterprise sector, special session of globalisation', DSTI/EAS/IND/SWP(2006)1/Final, working paper on statistics. Organisation for Economic Co-operation and Development, Paris

OECD (2008b) The internationalisation of business R&D: evidence, impacts and implications. Organisation for Economic Co-operation and Development, Paris

Patel P, Pavitt K (1991) Large firms in the production of the world's technology: an important case of 'non-globalisation. J Int Bus Stud 22(1):1–21

Peri G, Urban D (2006) Catching-up to foreign technology? Evidence on the 'Veblen-Gerschenkron' effect of foreign investments. Reg Sci Urban Econ 36:72–98

Sanna-Randaccio F, Veugelers R (2007) Multinational knowledge spillovers with decentralised R&D: a game-theoretic approach. J Int Bus Stud 38:47–63

Scherngell T, Barber M (2011) Distinct spatial characteristics of industrial and public research collaborations: evidence from the 5th EU framework programme. Ann Reg Sci 46(2):247–266

Steinmueller WE (2010) Economics of technology policy. In: Hall BA, Rosenberg N (eds) Handbook of economics of innovation. Elsevier, Amsterdam

Thursby J, Thursby M (2006) Here or there? A survey of factors in multinational R&D location. National Academies Press, Washington, DC

Tinbergen J (1962) The world economy: suggestions for an international economic policy. Twentieth Century Fund, New York

Veugelers R, Dachs B, Falk R, Mahroum S, Nones B, Schibany A (2005) Internationalisation of R&D: trends, issues and implications for S&T policies. Background report for the OECD forum on the internationalization of R&D. Organisation for Economic Co-operation and Development, Brussels

Young S, Hood N, Peters E (1994) Multinational enterprises and regional economic development. Reg Stud 28(7):657–677

Zaheer S (1995) Overcoming the liability of foreignness. Academy Manag J 38:341–364, References

Chapter 11
Joint Knowledge Production in European R&D Networks: Results from a Discrete Choice Modeling Perspective

Florian Reinold, Manfred Paier, and Manfred M. Fischer

Abstract The objective of this study is to explore the determinants of inter-organizational knowledge generation within European networks of R&D collaboration. It is argued that social capital is a key determinant for successful knowledge generation. Thus, factors that influence the development of social capital like geographical separation, or collaboration duration and intensity are expected to have an impact on inter-organizational knowledge generation. Determinants of inter-organizational knowledge generation are investigated by casting a binary response model in the form of a latent regression – index function model. Units of analysis are dyads of organizations that jointly participated in projects of the Fifth EU Framework Programme [FP5]. The data used in this study derives from a survey among FP5 participants and the EUPRO database.

Our findings suggest that crossing national border has a significantly positive rather than negative effect on scientific knowledge generation [measured in terms of reported co-publication activity]. This can be attributed to the participation rules and proposal selection procedures of the Framework Programmes. Another important result is that university dyads have the highest probability not only to generate scientific knowledge jointly, but also to jointly generate knowledge that is commercially relevant. In contrast, industry dyads show a low probability for both types

F. Reinold (✉)
Institute for Economic Geography and GIScience, Vienna University of Economics and Business (WU Wien), Vienna, Austria

Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria
e-mail: florian.reinold@gmx.at

M. Paier
Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria

M.M. Fischer
Institute for Economic Geography and GIScience, Vienna University of Economics and Business (WU Wien), Vienna, Austria

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_11,

of knowledge generation. This result is probably due to the fact that inter-organizational knowledge generation entails disclosure of knowledge, which is actually a task of universities but problematic for industry organizations.

11.1 Introduction

New growth theory suggests that innovation is the major engine of economic growth and competitiveness (see, for instance, Romer 1990). Since scientific and technological knowledge is regarded as the major input for innovation, the competiveness of an economy depends on its ability to generate new knowledge. Generation of knowledge is a social process and, therefore, the performance of an economy to generate knowledge crucially depends on successful cooperation between involved actors not only on the individual, but also on the organizational level (see, for instance, Lundvall 1992). Since markets lack the necessary long-term commitment for the transfer of tacit knowledge, networks are an increasingly important mode of cooperation for inter-organizational R&D activities (DeBresson and Amesse 1991; Powell and Grodal 2005). A major R&D network in Europe is the network created by the European Framework Programmes [FPs]. The FPs are the main instrument of the EU's R&D policy and are designed to support collaborative R&D projects including actors from distinct organizational types and different countries. Recently, several studies have been published in regard to R&D partner choices (Autant-Bernard et al. 2007; Paier and Scherngell 2011; Scherngell and Barber 2009) and joint knowledge generation (Hoekman et al. 2009) in Europe and in the FPs.

This study contributes to the existing literature by investigating the determinants of inter-organizational knowledge generation within the FP network. By using dyads of organizations that jointly participated in a project of the Fifth Framework Programme [FP5] as units of analysis, this study distinguishes itself from previous studies by focusing on the organizational level and not on the regional level. The data for carrying out this study is taken from a survey among FP5 participants and the EUPRO database. Determinants of inter-organizational knowledge generation are investigated by employing a binary response model derived from a latent regression.

Although FP projects are supposed to generate scientific knowledge as a direct output of the project, it is stipulated by the participation rules that the results should be exploitable for commercial purposes. Thus, this study distinguishes between two types of inter-organizational knowledge generation: scientific knowledge and commercially relevant knowledge generation. Scientific knowledge generation is measured in terms of whether or not co-authored publications exist, while commercially relevant knowledge generation refers to the fact that co-owned commercially relevant outcome [e.g. co-owned patents] is reported.

The remainder of this chapter is organised as follows. Section 11.2 provides an overview of the goals, participation rules and proposal selection procedures of the

FPs because it can be assumed that they have an influence on the pattern of inter-organizational knowledge generation. Section 11.3 identifies social capital and the ability to coordinate researchers from different organizations as key determinants of inter-organizational knowledge generation. Considerations are made how geographical separation and different organizational types of cooperation affect social capital and coordination problems and thereby influence inter-organizational knowledge generation. Section 11.4 describes the sample and the construction of variables in detail. Section 11.5 outlines the econometric model and presents the estimation results. Section 11.6 concludes.

11.2 The EU Framework Programmes

The FPs are recurrent mid-term research programmes that subsidy collaborative R&D projects linking partners from different countries and organizational types. The overall goal of the FPs is to strengthen the scientific and technological bases of European industry and to enhance its international competitiveness. Moreover, the FPs aim at fostering European market integration and regional income convergence by establishing common technological standards, increasing the mobility of researchers and promoting the dissemination of knowledge. Thus, the FPs can be seen as an important instrument for the implementation of EU policy beyond the area of science and technology (Stajano 2006, pp. 289–305). Since its establishment in 1984, seven FPs have been launched. Despite shifting thematic areas and instruments, the fundamental rationale of the FPs has remained unchanged, namely to support collaborative, pan-European research that involves different actors from scientific and the private sector (Roediger-Schluga and Barber 2007). This study relies on the Fifth Framework Programme [FP5], 1999–2002.

There is a set of participation rules stipulated by the European Commission, which shapes the structure of collaboration within FP5. The majority of proposals were subject to the following participation rules (European Council 1998). First, proposals had to be handed in by self-organised consortia. Second, the consortia had to consist of at least two mutually independent legal entities. Third, the consortia had to include legal entities from at least two different member states or one member state and one associated state.

Proposals handed in were evaluated by a panel of independent experts on the basis of a set of criteria defined by European Council decision. The final decision about which projects were funded and which were rejected rested with the European Commission. Proposals should have met following criteria (European Commission 2001). First, high quality of research and high degree of innovation; second, added-value by carrying out the project at the European level and by combining complementary expertise of different organizational types; third, contribution to one or more EU policies, e.g. cohesion, or the integration of new member states into the European Research Area; fourth, the usefulness and range of applications, the quality of the exploitation plans and dissemination strategies for the expected results; and finally the quality of the partnership, i.e. adequate complementarity of the partners and a reasonable division of tasks within the consortium.

Since the FPs involve subsidies for organizations from the private sector, there is a potential for thwarting the competition policy of the EU. In order to avoid distortion of the internal market, the FPs are restricted to pre-competitive research, i.e. research that is sufficiently distant to the market in order to avoid distortion of competition on product markets (Guzetti 1995, pp. 77–78). Some studies came to the conclusion that organizations can materialise commercially relevant outcome from participating in the FPs already in a short time after the termination of a FP project because they link the FP project with other in-house projects (Guy et al. 2005; Luukkonen and Hälikkä 2000; Matt and Wolff 2003). Moreover, exploitation-related goals were the major motivation of industry organizations for participating in FP5 (Guy et al. 2005). Thus, this study will not only focus on scientific knowledge as an outcome of explorative research but also on commercially relevant knowledge as an outcome of exploitative research.

11.3 Potential Determinants of Inter-Organizational Knowledge Generation

Inter-organizational knowledge generation primarily involves sharing and combining knowledge that is held by [at least two] different organizations (Moran and Ghoshal 1996). Two conditions have to be fulfilled in order that inter-organizational knowledge generation in networks can take place. First, the organizations must decide that they want to enter into a network relation in order to share and combine knowledge. Second, knowledge has to be successfully shared and combined so that novel knowledge [or a novel combination of already existing pieces of knowledge] may be generated. The first condition boils down to the question about determinants of collaboration choices, which has been already investigated for the FPs (see, for instance, Autant-Bernard et al. 2007; Paier and Scherngell 2011), the investigation of the second condition is the topic of this study.

Successful sharing and combining of knowledge depends on the willingness of organizations to share knowledge and on the capacities of organizations to absorb knowledge. The willingness to share and the capacity to absorb knowledge is positively influenced by social capital. Social capital refers to resources that evolve from networks of relationships over time by repeated interactions (Nahapiet and Ghoshal 1998). Since social capital exists only between individuals, it cannot be appropriated by one individual but is collectively owned (Coleman 1988). Social capital is conducive for sharing and absorbing knowledge by providing resources like trust, shared norms, shared goals, shared language and shared mental models. Von Hippel (1987) observed in his qualitative study about US steel mini-mill producers that knowledge was shared even with competitors because it was trusted that this will be rewarded in the long run by reciprocal behaviour. A quantitative analysis about R&D consortia in Taiwan conducted by Lin et al. (2009) provides evidence that trust, shared norms and shared goals influence knowledge transfer positively.

Since transfer of tacit knowledge is costly, not all knowledge that might be necessary for inter-organizational knowledge generation is shared (Grant 1996). A great part of the necessary knowledge is combined by coordinating people, in whom tacit knowledge is embedded, to build up inter-organizational capabilities for knowledge generation. Building up inter-organizational capabilities for knowledge generation is difficult since this requires complex modes of coordination. Simple modes of coordination like coordination by rules and standards or coordination by planning are not feasible because generation of knowledge involves high uncertainty and task interdependence (Kline and Rosenberg 1986; van de Ven et al. 1976). Thus employees have to be coordinated by complex modes of cooperation like mutual adjustment and group meetings (Grant 1996). Resources derived from social capital like shared goals or shared understandings facilitate complex coordination problems (Hämäläinen and Schienstock 2001).

11.3.1 Collaboration Duration and Intensity

Since social capital and common capabilities are built up by repeated interactions, it can be expected that duration of collaboration and the intensity of collaboration are crucial determinants for successful inter-organizational knowledge generation.

11.3.2 Geographical Separation

It is widely believed that geographical separation is detrimental to inter-organizational knowledge generation for three reasons (see Boschma 2005). First, geographical separation complicates repeated face-to-face communication which is regarded as important for the development of social capital. Second, geographical separation is often negatively correlated with cultural proximity which provides potential research partners with an already existing stock of social capital in the form of shared languages and shared norms. Third, geographical separation also makes complex coordination more difficult since it complicates mutual adjustment and group meetings (van de Ven et al. 1976). The majority of studies confirm the negative relationship between geographical separation and the occurrence of R&D collaboration (Katz 1994; LeSage et al. 2007; Maggioni and Uberti 2009; Paier and Scherngell 2011; Scherngell and Barber 2009).

Some authors question that the proposition about the negative relationship between geographical separation and inter-organizational knowledge generation is universally valid. Bathelt et al. (2004) argue that firms are only innovative in the long run if they maintain a balance between geographically separated and geographically close R&D collaborations because geographical separated collaborations are necessary to acquire new knowledge while close collaborations are necessary to exploit new knowledge. Torre and Rallet (2005) point to the fact that organizations need not be co-localised for close R&D collaboration since people are mobile. Often, co-localisation is not necessary for the whole duration of a joint research project and short- or medium term visits are sufficient. Moreover, large organizations can afford to relocate a part of the R&D staff for the duration of joint collaboration projects. Another differentiated view was presented by Moodysson et al. (2008). They distinguish between two modes of inter-organizational knowledge generation: synthetic knowledge generation and analytical knowledge generation. While geographical separation has a negative influence on synthetic knowledge generation, it is less detrimental to analytical knowledge generation. Analytical knowledge generation is highly formalized and is mainly carried out by a process of theory-led deduction and subsequent hypothesis testing. Since the primary type of knowledge involved is know-why, primarily codified knowledge is exchanged. Often, activities related to analytical knowledge generation are only of sequential interdependence which entails only simple coordination problems. An example for inter-organizational analytical knowledge generation is the conducting of a clinic study by a research hospital on behalf of a pharmaceutical research company.

Although geographical separation might complicate the development of social capital and inter-organizational capabilities, we argue that the design of the FPs offset the negative influence of geographical separation on inter-organizational knowledge generation for four reasons. First, the division of labour in the FPs is highly formalized because of pre-defined work packages, ex ante agreements on meetings and milestones (Matt and Wolff 2003). Thus, it can be expected that research conducted within the FPs resembles an analytical mode of knowledge generation. Second, the participation rules and goals of the FPs ensure that the FPs are an explorative and an international research network (see Sect. 11.2). Third, since the support of the mobility of researchers is one of the main instruments of the FPs, it can be expected that increased mobility of researchers substitute for a lack of co-location of organizations. Fourth, the legal framework provided by the FPs partly substitutes for a lack of cultural proximity and social capital (Luukkonen 2001).

11.3.3 Organizational Types of Cooperation

One objective of the FPs is to stimulate collaborations between the scientific sector [in particular universities and public research organizations] and the private sector [in particular R&D laboratories of industry organizations]. Since the scientific sector and the private sector carry out complementary tasks within the innovation process, interaction between the scientific sector and the private sector is regarded as conducive for innovation and economic development (see, for instance, Mowery and Rosenberg 1993). However, collaboration between the scientific sector and the private sector is often difficult since the two sectors pursue different goals and share different cultures (Ponds et al. 2007). A major aim of scientific organizations is to generate new knowledge and share this knowledge with the scientific community by publishing in order to increase reputation. Private organizations, by contrast, regard knowledge generation as a means to generate profit by reaping Schumpeterian rents, and are, therefore, highly interested in keeping knowledge secret. Moreover, they are to a lesser degree than scientific organizations interested in explorative research activities and are more interested in exploiting existing knowledge. Thus, although collaborations between scientific and private organizations are important for innovation and economic development, these collaborations can be expected to have a low productivity for inter-organizational knowledge generation because of differences in goals and culture.

11.4 Variables and Data

Two data sources are used in this study, namely, the EUPRO database and a survey among FP5 participants conducted by Austrian Institute of Technology in 2007. The EUPRO database is constructed and maintained by revising and standardizing raw data obtained from the CORDIS project database. It contains detailed information on funded projects and project participants of the EU Framework Programmes (for the first six see Barber et al. 2008). The survey restricted its population to projects involving less than 21 participants, which applies to roughly 97 % of all collaborative projects in FP5. 12,892 questionnaires were sent by email, from which 8,534 were received. The survey resulted in 1,686 valid questionnaires. Because a full data set in the EUPRO database is missing for 472 cases, only 1,214 questionnaires are used in this study.

Since the objective of this study is to explore the factors that are responsible for the fact that collaboration results in successful inter-organizational knowledge generation, the units of analysis has to be a form of inter-organizational collaboration. Following previous studies, collaboration is considered if two organizations participate in the same FP5 project (see, for instance, Autant-Bernard et al. 2007; Paier and Scherngell 2011). Thus, the units of analysis in this study are dyads of organizations that jointly participated in a FP5 project. The full sample consists of 7,776 dyads, which are formed by a set 3,343 distinct organizations that collaborated in 861 distinct FP5 projects.

The area of analysis is formed by 23 countries. All EU members at the time of the FP5 [i.e. the EU15] as well as the Central East European candidate countries that joined EU in 2004 are included. Table 11.4 in the Appendix gives an overview about the distribution of distinct organizations and participations, disaggregated by country.

11.4.1 Dependent Variables

Measuring knowledge generation is difficult since generated knowledge exists initially in the mind of those who generated it and is thus not directly observable (Fischer 2001). However, if the generated knowledge is sufficiently valuable, one can expect that it materialises in observable outcomes. This study relies on survey questions to capture outcomes of inter-organizational knowledge generation through a dichotomous variable. Thus, joint scientific knowledge generation is measured in terms of the occurrence of co-authored publications, and joint commercially relevant knowledge generation is measured in terms of co-owned commercially relevant outcome (e.g. co-owned patents). Each fifth dyad reported co-authored publications; joint commercially relevant knowledge generation is by far less common.

11.4.2 Independent Variables

In Sect. 11.3, we have argued that collaboration duration, collaboration intensity, geographical separation and organizational types of cooperation influence inter-organizational knowledge generation. Two variables are constructed to account for the duration that is necessary for developing social capital: project duration and previous collaboration. Project duration is measured in terms of the duration of the FP5 project [in months] in which the members of the dyad jointly participated. Previous collaboration is taken as a dummy variable into account that equals one if the partners of the dyad had collaborated together in a previous FP project. Intensity of collaboration is represented by the variable important research collaboration, with information from the survey, and is designed as a dummy variable that equals one if at least one dyad partner classified the other as an important research partner.

Two types of geographical barriers are included as independent variables: the existence of national borders and of EU's external border, designed as dummy variables. The variable national border equals one if the organizations forming the dyad are located in different countries. The variable EU's external border equals one if one organization of the dyad is located in the EU15 and the other in a Central East European candidate country.

The sample includes four organizational types: industry organizations [including consulting firms], universities, public research organizations and government organizations. Since there are few government organizations, only dummy variables for combinations of universities, research organizations and industry organizations were created. Thus, there are six dummy variables: university – university, university – research organization, industry organization – university, industry organization – industry organization, industry organization – research organization, research organization – research organization. Dyads that include government organizations take on the role of a default dummy.

11.4.3 Control Variables

Collaboration is measured in terms of joint FP project participation. This measurement approach works well for small FP projects, but in large FP projects it is unlikely that every participant collaborated directly with every participant (Fürlinger 2010). In order to control for this shortcoming, the variable project size, measured in terms of number of project participants, is included as a control variable.

Table 11.1 summarises descriptive statistics about the variables. See also the Appendix for the definition of the variables. Correlation analysis of the independent variables revealed a phi coefficient of −0.45 between the intent to generate scientific knowledge and the intent to generate commercially relevant knowledge. All other correlations were far less problematic.

Although the focus of this study is on relational characteristics, internal capacities of the organizations forming the dyad might also have an influence on knowledge generation. Since no information like budget or R&D personnel is available, proxy variables had to be used. EU funding devoted to the FP5 project, in which the organizations of the dyad jointly participated, serves as proxy for the monetary resources available for generating scientific or commercially relevant knowledge. The commitment of an organization to scientific or commercially relevant knowledge generation may have also an impact on the resources available for these activities. Thus, we include two further dummy variables that take the organizations' motive for participating in FP5 into account. The first dummy variable equals one if the intent of at least one member of the dyad was to generate scientific knowledge. The second dummy variable equals one if the intent of at least one member of the dyad was to generate commercially relevant knowledge. Both dummy variables were taken from the survey.

11.5 The Econometric Model and Estimation Results

Since the dependent variable $\boldsymbol{y^*}$ [inter-organizational knowledge generation] is measured in terms of its dichotomous realisations $\boldsymbol{y}$ [observable outcomes], the appropriate econometric model is a binary response model that can be derived from a latent regression – index function model (Verbeek 2004, pp. 190–193). By assuming a linear additive relationship between inter-organizational knowledge generation and a set of explanatory variables we obtain the following latent regression:

$$\boldsymbol{y}^* = \boldsymbol{\beta X} + \boldsymbol{\varepsilon} \tag{11.1}$$

where $\boldsymbol{y^*}$ denotes a n-by-1 vector of latent indices of knowledge generation for $n = 7{,}776$ dyad observations, $\boldsymbol{X}$ denotes a n-by-K matrix including a constant and

Table 11.1 Descriptive statistics for the variables used

	Project duration [in months]	Project size [number of project members]	EU project funding [in million €]
Minimum	4	2	0.01
First quartile	24	8	0.41
Median	36	10	0.76
Third quartile	36	13	1.03
Maximum	60	20	3.23
Mean	31.34	10.74	0.80
Standard deviation	8.49	3.58	0.48
Frequency of the dummy variables in the sample			
Previous collaboration (yes = 1)			22 %
Important research collaboration (yes = 1)			33 %
National border (yes = 1)			81 %
EU's external border (yes = 1)			8 %
University – university (yes = 1)			13 %
University – research organization (yes = 1)			15 %
Research organization – research organization (yes = 1)			7 %
Industry organization – university (yes = 1)			18 %
Industry organization – research organization (yes = 1)			18 %
Industry organization – industry organization (yes = 1)			22 %
Intent to generate scientific knowledge (yes = 1)			58 %
Intent to generate commercially relevant knowledge (yes = 1)			46 %

K-1 explanatory variables, $\boldsymbol{\beta}$ denotes a K-by-1 vector of parameters to be estimated, and ε a n-by-1 random error term symmetrically distributed about the mean. In this context, $\boldsymbol{\beta X}$ is called the index function (Greene 2008, p. 776).

Inter-organizational knowledge generation is not directly observable but its outcomes. Thus, we define a link between inter-organizational knowledge generation $\boldsymbol{y}^*$ and the binary outcomes $\boldsymbol{y}$.

$$\boldsymbol{y} = \begin{cases} 1 \text{ if } \boldsymbol{y}^* > \alpha \\ 0 \text{ if } \boldsymbol{y}^* \leq \alpha \end{cases} \tag{11.2}$$

where α is a threshold that has to be surpassed in order that the generated knowledge results in an observable outcome. Since the value of the threshold has only an influence on the value of the intercept in the regression model, the threshold value is set equal zero for sake of simplicity (Greene 2008, p. 776).

Binary response models derived from a latent regression explain the probability of an event occurring dependent on the explanatory variables of the latent regression [$\boldsymbol{X}$].

$$P\left(\boldsymbol{y} = 1 \,\middle|\, \boldsymbol{X}\right) = \mathrm{F}(\boldsymbol{\beta X}) \tag{11.3}$$

where F(.) denotes the cumulative distribution function of $\boldsymbol{\varepsilon}$. Consequently, the latent variable approach leads to a binary choice model whose form depends upon the distribution that is assumed for $\boldsymbol{\varepsilon}$ (Verbeek 2004, p. 192). Since we assume $\boldsymbol{\varepsilon} \sim \mathrm{N}(0,1)$, a probit model is specified:

$$P\left(\boldsymbol{y}=1 \mid \boldsymbol{X}\right)=\Phi(\boldsymbol{\beta X})=\int_{-\infty}^{\boldsymbol{\beta X}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\mathrm{t}^2\right\} \mathrm{dt} \tag{11.4}$$

where $\Phi(.)$ denotes the standard normal cumulative distribution function. The parameter estimates are derived by maximum-likelihood estimation (Greene 2008, pp. 777–779).

11.6 Estimation Results

Table 11.2 presents the maximum likelihood [ML] parameter estimates for inter-organizational knowledge generation; asymptotic standard errors are given in parentheses. For scientific knowledge generation three different models were estimated. The basic version [model 1] includes the full sample of 7,776 dyads, while model 2 uses a sample of 2,627 dyads, which consist only of universities and research organizations, and model 3 uses a sample of 4,729 dyads involving at least one industry organization. The model for commercially relevant knowledge generation uses the full sample of 7,776 dyads. The bottom of Table 11.2 provides various model fit measures. The likelihood ratio statistic that compares the estimated models with the constant-only null model indicates the significance of all models at the 0.01 significance level.

As expected, intensity and duration of collaboration increase the probability of both types of knowledge generation. Holding all other variables at their sample mean, previous collaboration increases the probability that inter-organizational scientific [commercially relevant] knowledge generation occurs by 8.3 [0.6] percentage points. An increase of the project duration from 4 months [the minimum in the sample] to 60 months [the maximum in the sample] increases the probability that inter-organizational generation of scientific knowledge and commercially relevant knowledge occurs by 13 and 5.3 percentage points, respectively. Thus, relative to project duration, previous collaboration is less important for generating commercially relevant knowledge than for generating scientific knowledge.

The classification of a dyad as an important research collaboration by at least one member of the dyad was used as a proxy for collaboration intensity. Table 11.2 shows that important research collaboration is by far the most important determinant for scientific knowledge generation. If a dyad is classified to indicate important collaboration the probability to generate scientific knowledge increases by 23.9 percentage points, on average. Important research collaboration has also a strong influence on commercially relevant knowledge generation. The classification of a

Table 11.2 ML estimates of the models for inter-organizational knowledge generation

Variables	Scientific knowledge generation: Model 1 [$n = 7{,}776$]	Model 2 [$n = 2{,}627$]	Model 3 [$n = 4{,}729$]	Commercially relevant knowledge generation [$n = 7{,}776$]
Constant	-1.560^{***} (0.129)	-0.875^{***} (0.190)	-2.012^{***} (0.219)	-3.196^{***} (0.246)
Collaboration duration and intensity				
Previous collaboration	0.360^{***} (0.041)	0.324^{***} (0.065)	0.430^{***} (0.057)	0.136^{*} (0.077)
Project duration [in months]	0.011^{***} (0.002)	0.006^{*} (0.004)	0.015^{***} (0.003)	0.021^{***} (0.004)
Important research collaboration	0.924^{***} (0.037)	1.055^{***} (0.059)	0.862^{***} (0.051)	0.391^{***} (0.069)
Geographical separation				
National border	0.087^{*} (0.048)	0.195^{**} (0.092)	0.025 (0.060)	−0.059 (0.083)
EU's external border	0.176^{***} (0.065)	0.119 (0.092)	0.304^{***} (0.098)	0.130 (0.118)
Organizational types of cooperation				
University-university	0.573^{***} (0.087)	0.184^{**} (0.081)		0.417^{**} (0.165)
University-research	0.404^{***} (0.086)	0.031 (0.080)		0.325^{**} (0.162)
Research -research	0.382^{***} (0.099)			−0.101 (0.220)
Industry– university	0.272^{***} (0.085)		0.539^{***} (0.172)	0.208 (0.160)
Industry– research	0.156^{*} (0.086)		0.428^{***} (0.173)	0.097 (0.163)
Industry– industry	−0.007 (0.086)		0.276 (0.173)	0.152 (0.159)
Control variables				
Project size [number of participants]	-0.080^{***} (0.006)	-0.085^{***} (0.009)	-0.080^{***} (0.008)	-0.020^{*} (0.010)
Intent for scientific knowledge	0.345^{***} (0.045)	0.274^{***} (0.076)	0.339^{***} (0.059)	0.167^{**} (0.077)
Intent for commercial knowledge	-0.080^{*} (0.043)	-0.143^{**} (0.074)	−0.047 (0.057)	0.532^{***} (0.079)
EU project funding [in mEUR]	0.187^{***} (0.043)	−0.013 (0.078)	0.295^{***} (0.053)	−0.040 (0.080)
Log-likelihood	−3084.364	−1262.162	−1640.875	−781.831
BIC	0.812	0.997	0.722	0.220
Likelihood ratio test (df = 15)	1599.091^{***}			140.351^{***}
Likelihood ratio test (df = 11)		637.894^{***}		
Likelihood ratio test (df = 12)			773.350^{***}	

Probit transformation of the dependent variable was used. The default dummy for organizational types of cooperation in the scientific knowledge generation models 1 and 3 as well as in the commercially relevant knowledge model are dyads involving government organizations. The default dummy for organizational types of cooperation in the scientific knowledge generation model 2 are dyads of the type research organization – research organization. Asymptotic standard errors are given in parentheses; *** significant at the 0.01 significance level, ** significant at the 0.05 significance level, * significant at the 0.1 significance level dyad partner as an important research partner increases the probability that the dyad generates commercially relevant knowledge by 1.8 percentage points.

University–university dyads have the highest probability of generating scientific knowledge, while dyads that involve only industry organizations the lowest. Switching from a university–university dyad to a dyad that includes a research organization and an industry organization decreases the probability of co-publishing by 10.5 percentage points. The only two dyad types that have a significant [and positive] impact on generation of commercially relevant knowledge are university-university dyads and dyads that consist of a university and a research organization. Thus, dyads involving industry organizations do not have a significant influence on commercially relevant knowledge generation.

How can it be explained that industry organizational cooperation is not at the forefront of commercially relevant knowledge generation? A likely explanation is that industry organizations do generate commercially relevant knowledge, as found by several studies (Guy et al. 2005; Luukkonen 2001), but not inter-organizationally because they fear negative knowledge spillovers of critical commercially relevant knowledge. Since a priori it is not known which knowledge will be useful to generate new knowledge, more knowledge is inevitably shared than necessary. The goal of industry organizations may be not to generate knowledge inter-organizationally in FP projects, but instead to pursue unilateral learning strategies to reduce knowledge spillovers and to maximise the benefit from FP participation (see Matt and Wolff 2003).

In Sect. 11.3 it has been argued that crossing national border and EU's external border should not have a significant impact on inter-organizational knowledge generation since the negative influence of geographical separation is offset by the participation rules of the FPs. This serves to be valid for commercially relevant knowledge generation, but not for scientific knowledge generation because both border dummies are significantly positive, but the effect is relatively small. Holding all other covariates at their sample means, crossing national border increases the probability of a dyad to generate scientific knowledge inter-organizationally only by two percentage points. Nevertheless, how can this small but significantly positive influence of crossing national border be explained?

One possible explanation is that the factors described in Sect. 11.3 that were expected to offset the negative influence of geographical separation on inter-organizational knowledge generation appear stronger than expected for scientific knowledge generation. Another possible explanation is that the significant and positive national border dummy can be attributed to collaborations within the scientific sector, while it is expected that industry organizations do not show an inclination to co-author publications with foreign organizations. This explanation is based on the consideration that researchers of universities and research organizations are more accustomed to work internationally and bound together by a common culture and shared mental models. As can be seen by the different significance of the coefficients of national border in models 2 and 3, inter-organizational generation of scientific knowledge is less sensitive to the presence of national borders within the scientific sector than between the scientific and the industrial sector.

An unexpected result is that the coefficient of crossing EU's external border is insignificant in model 2 but significantly positive in model 3. In order to shed more light on this result we have run a regression for scientific knowledge generation including only dyads as observations that cross EU's external border. The empirical results of this model are presented in Table 11.3. It is striking that a dyad that includes an industry organization located in a candidate country and a research organization located in a member state has the highest impact on inter-organizational generation of scientific knowledge among all organizational types of cooperation since this kind of cooperation has only a medium impact in the full sample model. This result can probably be attributed to the desire of industry organizations in candidate countries to catch up with their counterparts in member states. A research organization as partner can be regarded as a good choice since they are more applied oriented than universities but less reluctant to share their knowledge than industry organizations.

11.7 Summary and Conclusions

The objective of this study was to explore the determinants of inter-organizational knowledge generation in the network created by the FPs. It was argued that social capital is a key determinant for inter-organizational knowledge generation since social capital provides necessary resources [e.g. trust, shared language, shared mental models and shared goals] for knowledge exchange and facilitates the development of inter-organizational capabilities for knowledge generation. Thus, it was considered that factors influencing social capital are key determinants of inter-organizational knowledge generation. In Sect. 11.3, four possible determinants were identified: duration of collaboration, intensity of collaboration, geographical separation and the organizational types involved in inter-organizational knowledge generation.

A binary response model was derived from a latent regression in order to measure the impact of the above determinants on inter-organizational knowledge generation. Dyads of organizations that jointly collaborated in a FP5 project were used as units of analysis. Since inter-organizational knowledge generation is a latent process that is not directly measurable, observable outcomes of inter-organizational knowledge generation were used as proxies. The occurrence of a co-authored publication was used to measure scientific knowledge generation while generation of commercially relevant knowledge was measured in terms of co-owned commercially relevant outcome, like co-owned patents.

As expected, the results show that project duration and previous collaboration have a positive and significant impact on inter-organizational generation of

Table 11.3 ML estimates of the model for inter-organizational generation of scientific knowledge across EU's external border

Variables	Coefficient estimates [standard error in parentheses]
Constant	-1.265^{***} (0.456)
Collaboration duration and intensity	
Previous collaboration	0.294^{*} (0.164)
Project duration [in months]	0.005 (0.010)
Important research collaboration	1.017^{***} (0.135)
Organizational types of cooperation	
Industry organization (C) – research organization (EU)	0.740^{**} (0.325)
University (C) – university (EU)	0.576^{**} (0.290)
University (C) – industry organization (EU)	0.546^{*} (0.297)
University (C) – research organization (EU)	0.543^{*} (0.312)
Research organization (C) – university (EU)	0.300 (0.304)
Research organization (C) – research organization (EU)	0.041 (0.313)
Research organizations (C) – industry organization (EU)	−0.148 (0.309)
Industry organization (C) – university (EU)	−0.102 (0.413)
Industry organization (C) – industry organization (EU)	−0.430 (0.345)
Control variables	
Project size [number of participants]	-0.078^{***} (0.022)
Intent to generate scientific knowledge	0.297^{*} (0.156)
Intent to generate commercially relevant knowledge	−0.052 (0.154)
EU project funding [in millions of euros]	0.421^{***} (0.158)
Log-likelihood	−262.700
BIC	0.978
Likelihood ratio test (df = 16)	157.041^{***}

Dependent variable is inter-organizational scientific knowledge generation. Probit transformation of the dependent variable was used. The model includes a sample of 650 dyads that cross EU's external border. The default dummy for organizational types of cooperation are dyads including government organizations. (C) denotes that the corresponding organization was located in a candidate country; (EU) denotes that the corresponding organization was located in the EU; *** significant at the 0.01 significance level, ** significant at the 0.05 significance level, * significant at the 0.1 significance level

scientific and commercially relevant knowledge. Intensity of collaboration has the strongest positive impact on scientific knowledge generation and has also a strong and positive influence on the generation of commercially relevant knowledge.

Typically, geographical separation is expected to have a negative influence on inter-organizational knowledge generation because it curbs the development of social capital. In this study we expected that national border and EU's external border have an insignificant influence on inter-organizational knowledge generation because the negative effect of geographical separation on inter-organizational knowledge generation is offset by the participation rules and proposal selection procedures of the FPs. This assumption was confirmed for commercially relevant knowledge generation but not for scientific knowledge generation.

An unexpected result is that dyads involving industry organizations are not significant in regard to generation of commercially relevant knowledge. This result can probably be explained by the fact that inter-organizational knowledge generation entails disclosure of knowledge, which is problematic for industry organizations. As expected, dyads involving universities and research organizations are at the forefront in regard to inter-organizational scientific knowledge generation.

The results of this study are in accordance with the goals of the Framework Programmes. Inter-organizational knowledge generation is not curbed by national border. On the contrary, universities and research organizations use the FPs rather for international than national scientific knowledge generation. Moreover, as intended by the European Commission, the FPs are an appropriate instrument to introduce new members into the European Research Area. Fears that the FPs contradicts the competition rules of the internal market can be allayed since industry-industry collaborations in the FPs do not have a significant influence on inter-organizational commercially relevant knowledge generation.

Acknowledgements This chapter reports results of research carried out in the framework of the Innovation Economics Vienna – Knowledge and Talent Development Program.

Appendix

Table 11.4 Distribution of organizations and participants included in the sample by country

Country	Organizations (in %)	Participants (in %)
Germany	14.9	17.5
Italy	13.6	13.4
United Kingdom	13.2	13.1
Spain	11.9	10.2
France	11.7	11.9
Greece	4.4	4.4
Netherlands	4.2	5.1
Belgium	3.7	3.4
Sweden	3.3	3.7
Portugal	3.1	2.5
Austria	2.8	3.0
Denmark	2.6	2.4
Finland	2.1	2.6
Poland	1.8	1.6
Ireland	1.6	1.3
Czech Republic	1.6	1.2
Hungary	1.3	1.0
Slovenia	0.6	0.5
Slovakia	0.4	0.4
Lithuania	0.3	0.3
Latvia	0.3	0.2
Luxembourg	0.2	0.2
Estonia	0.2	0.1

Table 11.5 Definitions of variables

Variable name	Scale of measurement	Description	Data source
Dependent variables			
Generation of scientific knowledge	Dichotomous	1 if the members of the dyad co-authored a scientific publication	Survey
Generation of commercial knowledge	Dichotomous	1 if the members of the dyad co-own commercially relevant outcome	Survey
Collaboration duration and intensity			
Project duration	Ordinal	Duration of the project in which the members of the dyad jointly participated (months)	EUPRO
Previous collaboration	Dichotomous	1 if the two organizations of the dyad have already collaborated in previous FPs	Survey
Important research collaboration	Dichotomous	1 if at least one member of the dyad stated that the other was an important collaborator	Survey
Geographical separation			
National border	Dichotomous	1 if the organizations forming the dyad are located in different countries	EUPRO
EU's external border	Dichotomous	1 if one organization is located in the EU15 and the other in a CEE candidate country	EUPRO
Combinations of organization types			
University – university	Dichotomous	1 if both organizations of the dyad are universities	EUPRO
University – research	Dichotomous	1 if one organization is a university and the other is a research organization	EUPRO
Industry– university	Dichotomous	1 if one organization is a university and the other is an industry organization	EUPRO
Industry– industry	Dichotomous	1 if both organizations are industry organizations	EUPRO
Industry– research	Dichotomous	1 if one organization is an industry organization and the other a research organization	EUPRO
Research– research	Dichotomous	1 if both organizations of the dyad are research organizations	EUPRO
Control variables			
Project size	Ordinal	Number of participants of the project	EUPRO
Intent to generate scientific knowledge	Dichotomous	1 if the motivation of at least one member of the dyad was scientific research	Survey
Intent to generate commercial knowledge	Dichotomous	If the motivation of at least one member of the dyad was commercial knowledge	Survey
EU project funding	Continuous	EU funds (measured in million EUR) allocated to the project	EUPRO

References

Autant-Bernard C, Billand P, Frachisse D, Massard N (2007) Social distance versus spatial distance in R&D cooperation: empirical evidence from European collaboration choices in micro and nanotechnologies. Pap Reg Sci 86(3):495–519

Barber MJ, Heller-Schuh B, Roediger-Schluga T, Scherngell T (2008) The sysres EUPRO database manual. Deliverable D4.1, Vienna

Bathelt H, Malmberg A, Maskell P (2004) Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. Prog Hum Geogr 28(1):31–56

Boschma RA (2005) Proximity and innovation: a critical assessment. Reg Stud 39(1):61–74

Coleman JS (1988) Social capital in the creation of human capital. Am J Sociol 94(Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure):95–120

DeBresson C, Amesse F (1991) Networks of innovators: a review and introduction to the issue. Res Policy 20(5):363–379

European Commission (2001) Manual of proposal evaluation procedures. ftp://ftp.cordis.europa.eu/pub/fp5/docs/em_en_200201.pdf

European Council (1998) Council decision of 22 December 1998 concerning the rules for the participation of undertakings, research centres and universities and for the dissemination of research results for the implementation of the fifth framework programme of the European Community (1998–2002). http://www.cordis.europa.eu/fp5/management/particip/r-ecdec-part.htm. Accessed 19 June 2012

Fischer MM (2001) Innovation, knowledge creation and systems of innovation. Ann Reg Sci 35 (2):199–216

Fürlinger G (2010) Structures of scientific collaboration in EU R&D networks: a comparison of network models. Innovation economics Vienna – knowledge and talent development program Ph.D.- & Master-theses series 1, Vienna University of Economics and Business

Grant RM (1996) Toward a knowledge-based theory of the firm. Strateg Manage J 17(Winter Special Issue):109–122

Greene WH (2008) Econometric analysis, 6th edn. Pearson Prentice Hall, Upper Saddle River

Guy K, Amanatidou E, Psarra F (2005) Framework programme 5 (FP5) impact assessment: a survey conducted as part of the five-year assessment of European Union research activities (1999–2003). Sci Public Policy 32(5):349–366

Guzetti L (1995) A brief history of European union research policy. Office for Official Publications of the European Commission, Luxembourg

Hämäläinen TJ, Schienstock G (2001) The comparative advantage of networks in economic organization: efficiency and innovation in highly specialised and uncertain environments. In: OECD (ed) Innovative networks. Co-operation in national innovation systems. OECD, Paris, pp 17–49

Hoekman J, Frenken K, van Oort F (2009) The geography of collaborative knowledge production in Europe. Ann Reg Sci 43(3):721–738

Katz JS (1994) Geographical proximity and scientific collaboration. Scientometrics 31(1):31–43

Kline SJ, Rosenberg N (1986) An overview of innovation. In: Landau R, Rosenberg N (eds) The positive sum strategy. Harnessing technology for economic growth. National Academy Press, Washington, DC, pp 275–305

LeSage JP, Fischer MM, Scherngell T (2007) Knowledge spillovers across Europe: evidence from a Poisson spatial interaction model with spatial effects. Pap Reg Sci 86(3):393–421

Lin JL, Fang SC, Fang SR, Tsai FS (2009) Network embeddedness and technology transfer performance in R&D consortia in Taiwan. Technovation 29(21):763–774

Lundvall BA (ed) (1992) National systems of innovation. Towards a theory of innovation and interactive learning. Pinter, London

Luukkonen T (2001) Old and new strategic roles for the European Union Framework Programme. Sci Public Policy 3(28):205–218

Luukkonen T, Hälikkä S (2000) Knowledge creation and knowledge diffusion networks. Impacts in Finland of the EU's Fourth Framework Programme for Research and Development. Finnish Secretariat for EU R&D, Helsinki

Maggioni MA, Uberti TE (2009) Knowledge networks across Europe: which distance matters? Ann Reg Sci 43(3):691–720

Matt M, Wolff S (2003) "E.U. sponsored" versus "spontaneous" R&D collaborations: towards a micro-analysis and policy implications. Bureau d'Economie Théorique et Appliquée, Université Louis Pasteuer

Moodysson J, Coenen L, Asheim B (2008) Explaining spatial patterns of innovation: analytical and synthetic modes of knowledge creation in the Medicon Valley life-science cluster. Environ Plan A 40(5):1040–1056

Moran P, Ghoshal S (1996) Value creation by firms. In: Keys JB, Dosier LN (eds) Academy of management best paper proceedings. Fifty-six annual meeting of the academy of management. Academy of Management, Cincinnati/Briarcliff Manor, pp 41–45, 11–14 Aug 1996

Mowery DC, Rosenberg N (1993) The U.S. national innovation system. In: Nelson RR (ed) National innovation systems. A comparative analysis. Oxford University Press, Oxford

Nahapiet J, Ghoshal S (1998) Social capital, intellectual capital and the organizational advantage. Acad Manage Rev 23(2):242–266

Paier M, Scherngell T (2011) Determinants of collaboration in European R&D networks: empirical evidence from a discrete choice model. Ind Innov 18(1):89–104

Ponds R, van Oort F, Frenken K (2007) The geographical and institutional proximity of research collaboration. Pap Reg Sci 86(3):423–443

Powell WW, Grodal S (2005) Networks of innovators. In: Fagerberg J, Mowery DC, Nelson RR (eds) The Oxford handbook of innovation. Oxford University Press, Oxford, pp 56–85

Roediger-Schluga T, Barber MJ (2007) R&D collaboration networks in the European Framework Programmes: data processing, network construction and selected results. UNU-Merit working paper series 032, United Nations University, Maastricht Economic and Social Research and Training Centre on Innovation and Technology

Romer PM (1990) Endogenous technological change. J Polit Econ 98(5):71–102

Scherngell T, Barber MJ (2009) Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU framework programme. Pap Reg Sci 88(3):531–547

Stajano A (2006) Research, quality, competitiveness. European union technology policy for the information society. Springer, Boston

Torre A, Rallet A (2005) Proximity and localization. Reg Stud 39(1):47–59

van de Ven AH, Delbecq AL, Koenig RJ (1976) Determinants of coordination modes within organizations. Am Sociol Rev 41(2):322–338

Verbeek M (2004) A guide to modern econometrics, 2nd edn. Wiley, Chichester

von Hippel E (1987) Cooperation between rivals: informal know-how trading. Res Policy 16 (6):291–302

Chapter 12
Multilateral R&D Collaboration: An ERGM Application on Biotechnology

Çilem Selin Hazir

Abstract This chapter presents an empirical study on formation of multilateral R&D collaboration networks among organizations. The objective of the study is to investigate how geography and heterogeneity in institutional types affect the way organizations come together around consortiums to perform R&D. It makes use of data on project proposals submitted to the 7th Framework Program (FP) in the field of biotechnology to construct a two-mode network. It employs extensions of exponential random graph models (ERGM) (Frank and Strauss, J Am Stat Assoc 81(395):832–842, 1986; Wasserman and Pattison, Psychometrika 61(3):401–425, 1996, for affiliation networks (Wang et al., Soc Netw 31:12–25, 2009). The results show that higher education institutions and research institutions tend to show higher connectivity and hence bridge learning across consortiums. Furthermore, organizations located in the core European countries tend to participate in the same consortium and these consortiums tend to be small in size. Finally, homophily in institutional types and network effects do not affect the formation process.

12.1 Introduction

Increasing tendency to collaborate in research and development (R&D) (Hagedoorn 2002; Wuchty et al. 2007) created an interest in network formation processes in the field of geography of innovation to achieve a better understanding on knowledge flows in space. R&D collaborations sometimes involve more than two parties, which come together in the form of a consortium to perform R&D, and give rise to multilateral R&D collaboration networks. So far these networks have been

Ç.S. Hazir (✉)
Université de Lyon, Lyon F-69007, France

Université Jean Monnet, Saint-Etienne F-42000, France

CNRS, GATE Lyon Saint-Etienne, Ecully F-69130, France
e-mail: cilem.selin.hazir@univ-st-etienne.fr

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_12,

analysed assuming that they can be considered as a collection of independent bilateral interactions (Autant-Bernard et al. 2007; Paier and Scherngell 2008; Scherngell and Barber 2009; Scherngell and Lata 2013). Nevertheless, multilateral interactions – i.e. the dependencies among bilateral interactions- might also be important for the formation process, and hence for understanding how these networks modify spatial diffusion of knowledge.

First of all, in multilateral R&D collaboration networks (multilateral networks from now on) when a new consortium is created, a number of organizations get connected to each other all at once. So far, while some empirical evidence has shown that pairwise collaboration decisions are positively related to spatial proximity; how it affects composition of consortiums is not obvious. In other words, whether consortiums are created among proximate organizations or they bring very distant and very proximate organizations together such that the spatial advantages and disadvantages are offset is not clear.

Second, within consortiums organizations start exchanging and co-creating knowledge not only pairwise, but also group-wise. From a spatial point of view, on the one hand this implies that at least some part of knowledge flows simultaneously in a geography defined by the location of consortium members. On the other hand, it means that through multilateral collaboration some locations gain the ability to learn collectively. Then, it is a matter of interest to know whether there are any spatial limits for simultaneous knowledge flows/collective learning to occur.

Third, consortium members, who benefit from these knowledge flows, differ in terms of the mix of functions (Hekkert et al. 2007) that they perform in their local innovation system. Hence they differ in their aims and ways of doing research. How they come into groups and how they differ in their networking activity might inform policy-makers in developing tools to promote regions to get involved in such networks.

Therein, in this chapter we depart from the prior work by focusing on the following two questions: (1) how do different types of organizations (with different functions) come together around research consortiums in different combinations? (2) Is the creation process of these consortiums free of spatial constraints or not? Mainly two strands in the literature suggest some explanations to these questions. Network formation literature suggests that benefits that are obtained through partners of partners affect collaboration decisions (Jackson and Wolinsky 1996). Hence it emphasizes the role of network effects that result from network configuration and position of an agent in the network. On the other hand the proximity literature (Boschma 2005) suggests that the degree of similarity in exogenous attributes of agents affect collaboration decisions. While both strands of work provide a basis to answer to the above-mentioned question, none of them particularly address how group wise collaboration emerges. Thus, there exists a challenge to bridge pairwise explanations to the consortium level.

We try to handle this challenge by employing a random graph approach and working on a two-mode representation of the network, and we address the above-mentioned questions in the field of biotechnology. The chapter will proceed with a

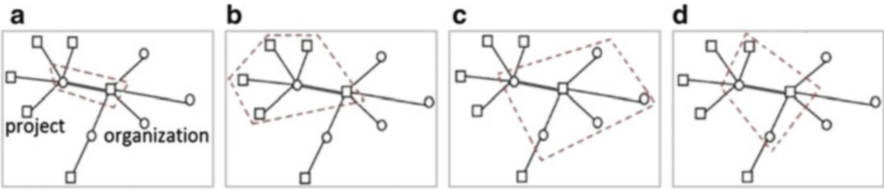


Fig. 12.1 Four perspectives to study participation decisions

discussion on processes that underlie formation of multilateral networks. Then, an empirical application will be presented based on data on project proposals submitted to the 7th Framework Program (FP) in the field of biotechnology.

12.2 Formation of Multilateral R&D Collaboration Networks

The smallest building block of a multilateral network is a tie that is realized when an organization decides to participate in a consortium. The determinants of this decision might be considered through four perspectives. First, as illustrated in Fig. 12.1a this decision might be taken into account in isolation from other participation decisions that make up the network, and the determinants that affect the interest of an organization in engaging in collaborative research might be studied. Second, a decision maker might be considered in isolation with other decision makers and a single decision made by the decision maker might be addressed within its portfolio of participations (Fig. 12.1b). Third, a consortium might be studied in isolation with other consortiums and attention might be accorded on determinants of co-participations (Fig. 12.1c). Finally, the intersection of the perspectives adopted in Fig. 12.1b, c might be considered, and hence the network effects that stem from interconnectedness of projects is not neglected (Fig. 12.1d).

12.2.1 The General Interest in Collaborative Research

An organization might engage in multilateral research collaborations for a variety of reasons, not all of which are necessarily technology related, like improving its business network or brand reputation. However, due to the fact that the context of collaboration is to perform R&D and this context is framed by a project plan, one could assume a stronger role for technology related aspects and describe the main motivation of organizations as accession to information, knowledge, skills, ideas, financial capacity to realize a research that they could not achieve on their own, etc.

The level of this interest, however, may vary across different types of organizations as the mixes of functions (Hekkert et al. 2007) that they perform in their local innovation system and their primal roles are different. To illustrate, the interest of a public organization in engaging in research consortiums as a user or a regulator is different than a higher education institution which seeks scientific or technological advancements. Furthermore, profit-seeking organizations have appropriability concerns as accession has some associated risks about control on the knowledge (Cassiman and Veugelers 2002).

Another reason why organizations might differ in their interest in participating in research consortiums might be the specific role played by some local features. On the one hand, a high level of industrialization and a well-organized local innovation system in a place might promote systemic learning and interactive innovation (Cooke et al. 1997) and hence foster the absorption capacity of organizations (Cohen and Levinthal 1990). On the other hand, the development of all regional forms of information services, technological transfer institutions and communication infrastructure may enhance the circulation of information and hence favour the ability of agents to be aware of potential consortiums.

12.2.2 Organization's Portfolio of Participations

Each participation decision individually offers an organization some change in the scope of knowledge and in absorption capacity. However, a portfolio of participations suggests more than the bare sum of these individual benefits. The reason is that organizations do not learn in each consortium in an isolated manner but sometimes cross-learning occurs across a number of projects (Powell et al. 1996). Hence, participation decisions made by an organization may depend on each other regarding the cross-learning opportunities that they suggest.

They depend on each other also because they are competing for the limited amount of resources that an organization can allocate for (collaborative) research. Even, they may compete on the same unique resource like staff. Hence, one can consider two main forces operating on the formation of an organizations portfolio of participations. While the former promotes multiple participations due to cross-learning opportunities, the latter dampens it due to cost/resource considerations.

Following the discussion in the previous section, it may be argued that the net effect of these two forces might differ for different types of organizations. On the one hand the tendency for multiple participations might be weaker for organizations with appropriability concerns as cross-learning might also have a negative impact on the control over knowledge. One the other hand, different types of organizations might differ in their financial or cross-learning capacity.

12.2.3 Determinants of Co-participations

In the two preceding sections determinants of participation decisions have been addressed as if organizations make their participation decisions in isolation from other organizations and hence independently. However, proximities or competition motivations may drive organizations to make their participation decisions together and hence join a consortium in pairs, triples, quadruples, etc. These effects could be considered in at least two headings.

First, some proximity dimensions (Boschma 2005) may breed co-participations via facilitating acquaintance or awareness on possible partners and consortiums. One such dimension is the social proximity, as social ties can play a role to convey information on possible consortiums, ease getting into contact with them, and hence result in socially proximate organizations participating in the same consortium. For the case of bilateral collaboration networks this role has been studied theoretically and empirically for different types of networks (Van der Leij et al. 2006; Jackson and Rogers 2007; Fafchamps et al. 2010; Autant-Bernard et al. 2007). In the same manner, organizational proximity in the form of hierarchies or in weaker forms like supply-chain relations or business networks may facilitate co-participations. Finally, geographical proximity may give rise to co-participations as organizations nearby can be identified more easily.

However, regions differ in terms of the extent that the systemic mechanisms that foster circulation of information are developed and the extent that their constituents interact (Cooke et al. 1997). Hence, local features, an integral part of geographical proximity, might matter in generation of co-participations. In this regard, space might also be related to network formation as a "*setting structure*", which refers to an exogenous constraint on possible tie dependencies (Pattison and Robins 2002).

Second, all types of proximities may breed co-participations as optimal levels of proximity between organizations may enhance joint learning (Boschma and Frenken 2009). For instance, some level of institutional proximity means closeness in standards, routines, values, goals, languages, etc., which in turn act as enabling mechanisms that provide stable conditions for interactive learning (Boschma 2005). Similarly, geographical proximity may promote transmission of knowledge via facilitating face-to-face contacts (Feldman and Florida 1994; Anselin et al. 2000). It may also facilitate cross-fertilization of ideas (Feldman and Florida 1994), pointing out a higher potential of knowledge that could be co-created. In addition to that, it may enable timely inflows of information (Feldman 1993) or reduce the cost of collaboration (Hoekman et al. 2009). Beside these, social proximity can enhance joint learning as social ties may involve trust. Trust is argued to be a factor that enables the exchange of ideas more openly (Zand 1972), reduces the cost of negotiations and conflicts (Zaheer et al. 1998), allows transmission of more private and tacit knowledge as compared to the information exchanged at arm's-length (Uzzi 1996). Concerning that a consortium may itself create/reinforce social ties and/or trust, social proximity may also result in new co-participations with old partners in new projects. Finally, some level of cognitive proximity might suggest a reason for co-participations as sharing common knowledge is a pre-requisite for understanding each other and benefit from collaboration (Frenken et al. 2007; Nooteboom et al. 2007). In the case that consortium members preserve their medium of interaction by engaging collectively in new consortiums, this may increase their cognitive proximity. While this may improve “relative absorptive capacity” (Lane and Lubatkin 1998) among members, it may also decrease the level of heterogeneity in knowledge levels and reduce the propensity to generate innovation (Cowan et al. 2007).

12.2.4 Network Effects

In a multilateral collaboration network, consortium members’ portfolios of participations enable organizations to access information created in other consortiums, in which they are not directly involved. In the literature, there are several models that relate these network effects to the network formation. One of these models is the connections model (Jackson and Wolinsky 1996; Bala and Goyal 2000), which shows that, for different parameter values, these network effects lead to different stable and/or efficient network configurations.

Similarly, the preferential attachment model by Barabási and Albert (1999) based on degree affinity as the driver of network formation, meaning that an agent prefers establishing a link with the agent who has the largest number of direct connections (i.e. degree). They show that degree affinity is capable of explaining the formation of the networks defined by the World Wide Web or patent citations.

12.3 Data

The empirical application is based on the European Commission records on project proposals submitted to the 7th Framework Program (FP7). The raw data is obtained from the French Ministry of Higher Education and Research and processed by EuroLIO (European Localized Innovation Observatory). This processing is called “disambiguation” since the variety in the way an organization is registered in the database results in ambiguity in organizations that is to be corrected for.

The empirical application that will be presented in the sequel is based on a sample that is obtained by selecting the information on small or medium-scale focused research project proposals in the field of “life sciences, biotechnology and biochemistry for sustainable non-food”. The sample includes 237 project proposals which have been proposed to the Commission in response to five different calls issued yearly from 2007 to 2011, and 1313 unique participants.

12.4 The Model

Exponential random graph models (known also as p* models and in short ERGM) (Frank and Strauss 1986; Wasserman and Pattison 1996) are based on the idea that the observed network is just one realization of all possible pattern of connections among a given set of nodes. The closed form of the model can simply be considered as a probability density function which expresses the probability of a network configuration in terms of some sub-structures, called local configurations or neighbourhoods, it contains. These local configurations may be as simple as edges or they can be more complex sub-structures resulting from dependencies among ties (Frank and Strauss 1986; Pattison and Robins 2002). Table 12.1 provides illustrations of some of these local configurations.

In this study, we employed the extension of ERGM for two-mode networks (Wang et al. 2009). A two mode network consists of two types of nodes and ties among them. The first set of nodes ($A = \{1,2,3, \ldots,n\}$) refers to organizations, and the second set ($P = \{1,2,3, \ldots,m\}$) refers to projects. Hence in such a network, the set (Ω) of all possible ties connecting each organization in A to each project in P is of size $n \times m$. We denote a possible tie between an organization $i \in A$ and a project $j \in P$, with the random variable Y_{ij}, which takes a value of 1 if the tie is realized (meaning that organization i participated in project j), and 0 otherwise. Then we express the overall network as a random vector ($\mathbf{Y}$), which is a collection of tie variables, i.e. $\mathbf{Y} = \{Y_{ij}\}$. We denote a realization of this vector with $\mathbf{y} = \{y_{ij}\}$. Then, the general form of ERGM can be expressed as follows (Robins et al. 2007)[1]:

$$P(\boldsymbol{Y} = \boldsymbol{y}) = \left(\frac{1}{\varkappa}\right) exp\left\{\sum_{Q} \eta_Q \, g_Q(\boldsymbol{y})\right\}$$

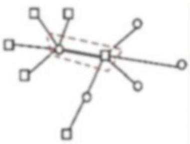
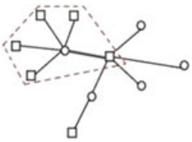
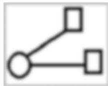
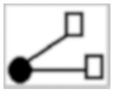
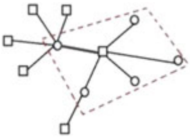
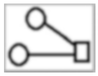
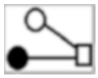
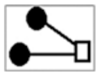
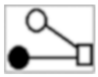
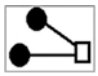
Where the following definitions hold:

- $P(\boldsymbol{Y} = \boldsymbol{y})$ is the probability of observing a particular network $\boldsymbol{y}$.
- $\varkappa$ is a normalizing constant assuring the probabilities given by this distribution adds up to 1:

$$\varkappa = \sum_{\boldsymbol{y} \in \boldsymbol{Y}} \left(\exp \sum_{Q} \eta_Q \, g_Q(\boldsymbol{y}) \right)$$

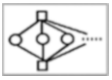
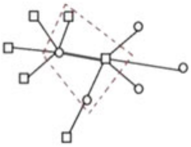
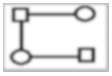
[1] For statistical and mathematical foundations of ERGM, readers are referred to the joint probability of a Markov field or the extensions of statistical mechanics of Gibbs to the study of networks by Park and Newman (2004), and to the Hammersely Clifford Theorem (Besag 1974) proving the Gibbs-Markov equivalence.

Table 12.1 Summary of hypothesis and variable definitions

Perspective	Variables		Local configuration
 The general interest in collaborative research	Edges	Number of participations	
	Edges by HES	Number of participations by higher education institutions (HES)	
	Edges by PRC	Number of participations by private institutions (PRC)	
	Edges by REC	Number of participations by research institutions (REC)	
	Edges by EU15	Number of participations by EU15 members	
	Edges by core	Number of participations by the core	
 Organization's portfolio of participations	Organization 2-stars	Number of 2 paths connecting two projects	
	HES 2-stars	Number of 2 paths connecting two projects and centred at HESs	
	PRC 2-stars	Number of 2 paths connecting two projects and centred at PRCs	
	REC 2-stars	Number of 2 paths connecting two projects and centred at RECs	
 Determinants of co-participations	Project 2-stars	Number of 2 paths connecting two organizations	
	Co-participations with HES	Number of 2 paths connecting two organizations, one of which is a HES	
	Co-participations with PRC	Number of 2 paths connecting two organizations, one of which is a PRC	
	Co-participations with REC	Number of 2 paths connecting two organizations, one of which is a REC	
	Homophily HES	Number of 2 paths connecting two HESs	
	Homophily PRC	Number of 2 paths connecting two PRCs	
	Homophily REC	Number of 2 paths connecting two RECs	
	Co-participations with core	Number of 2 paths connecting two organizations, one of which is in core regions	
	Homophily core	Number of 2 paths connecting two organizations located in core regions	

(continued)

Table 12.1 (continued)

Perspective	Variables		Local configuration
	Continuity of consortiums	Actor centred alternating k-two paths	
 Network effects	3-paths	Number of three paths	

- η_Q is the parameter corresponding to the local configuration (neighborhood) Q.
- $g_Q(\boldsymbol{y})$ is the network statistic corresponding to the local configuration (neighborhood) Q. In a homogeneous model, for a given type of neighborhood Q, which is a collection of isomorphic neighborhoods q, $g_Q(\boldsymbol{y})$ is given by:

$$g_Q(\boldsymbol{y}) = \sum_{q \epsilon Q} \left(\prod_{Y_{ij} \epsilon\ q} y_{ij} \right)$$

There are two main techniques suggested to estimate ERGM: Pseudo-Likelihood Estimation (PLE) (Straus and Ikeda 1990) and Markov Chain Monte Carlo Maximum Likelihood Estimation (MCMCMLE) (Snijders 2002). Wang et al. (2009) provide empirical evidence on the performance of the two estimation techniques for two-mode networks and propose that MCMCMLE should be the preferred method for two-mode networks.

12.5 The Variables

The four perspectives discussed in Sect. 12.2 are brought together in an ERGM specification that includes 21 variables (local configurations) (Table 12.1). The general interest in collaborative research is reflected by the edge configuration, which has been differentiated with respect to the institutional types of organizations and their locations. The effect of an organization's portfolio of participations is represented by organization 2-stars. The effect of co-participation drivers is measured by two different local configurations. The first one refers to project 2-stars, which has been differentiated with respect to institutional types and location. The second one refers to the continuity of consortiums that stems from the effect of trust and network learning. Finally, 3-paths are included to account for the network effects.

In studying the geographical dimension of the network a core-periphery perspective is adopted due to some data limitations and data characteristics. On the one hand, regional information conforming to NUTS classification is not available for all countries. Even for countries for which NUTS classification is available, regional information is not available for finer levels like NUTS3 level for all. On the other hand, co-participation decisions are not free of the design of FP program, since the Commission sets the minimum conditions[2] on the consortium size and location of participants. Hence, an analysis of the geographical dimension of co-participations at a fine regional level is imperfect. For this reason, the following neighbouring countries, which include regions with very high concentrations of people, finance, and industry, is called the "core": Austria, Belgium, France, Germany, Italia, the Netherlands, Switzerland, and United Kingdom. Being located in the core has been introduced as an attribute on organizations.

Finally, EU-15 membership has been introduced as an attribute on organizations as a control variable to account for the fact that over time the FP participation rules have changed, as well as set of countries that are called member states and associate states. While countries that have long been a member of the European Union have not been affected from these changes in participating in FP, those countries that have more recently joined to EU have been concerned with these changes.

12.6 Estimation Results and Discussion

The model specification that includes all the variables defined above has failed to converge.[3,4] According to Handcock (2003) this might result from two reasons. First, it may be the case for this specification that MLE does not exist at all. Letting $g(\boldsymbol{y})$ be the network statistics corresponding to the neighborhoods used to express the model, letting C be the convex hull of $\{g(\boldsymbol{y}): \mathbf{y} \in \mathbf{Y}\}$, and letting rint(C) and rbd (C) be the relative interior and relative boundary of C, respectively; Handcock (2003) states that a necessary and sufficient condition for the MLE not to exist is that $(g(\boldsymbol{y}^{obs})) \in rbd(C)$. Second, the Monte Carlo process used to approximate the MLE might have a difficulty to produce realizations that cover the observed values of the network statistics. In our case, manipulations on the chain length or step size

[2] Regulation (EC) No 1906/2006; Article 5/(1) states that "at least three legal entities must participate, each of which must be established in a Member State or associated country, and no two of which may be established in the same Member State or associated country".

[3] All estimations are carried out using "BPNet", which is an extension of the PNet programme (Wang et al. 2006) and bases on MCMCMLE technique.

[4] Convergence is measured by t-ratios calculated to check whether the estimate of the parameter vector is capable of producing a graph distribution centered at the observed network (Wang et al. 2009). Snijders (2002) suggests that if the absolute value of t-values for all local configurations ($|t_Q|$) are less than or equal to 0.1 convergence is excellent; if $0.1 < |t_Q| \leq 0.2$, it is good, else if $0.2 < |t_Q| \leq 0.3$ convergence is fair.

did not improve convergence. MCMLE process kept yielding different parameter vectors which generate networks far from the observed one.

Table 12.2 presents four specifications that converged excellently. Among these, Model 4, which includes all variables except for continuity of consortiums and shows the highest goodness of fit, will be interpreted in the sequel. The rest of the models given in the same table are provided for comparison purposes as they represent gradually increasing the number of perspectives adopted to study the participation decisions (recall Fig. 12.1).

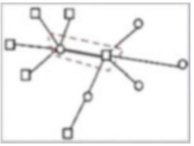
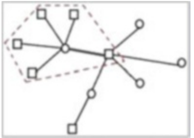
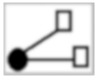
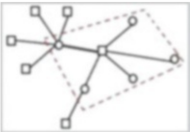
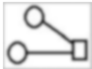
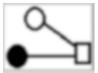
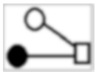
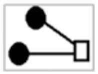
Model 4 shows that among five different types of organizations the general interests of higher education institutions (HES), private enterprises (PRC), and research institutions (REC) in collaborative research differ from each other and differ from that of the reference categories; i.e.; public organizations (PUB) and other types of organizations (OTH). The parameter estimates for these variables are all negative indicating that network configurations with fewer edges are more probable. Equivalently, complete or very dense network configurations are not probable. This is in line with the fact that in the network under study only 0.69 % of all possible ties are realized.

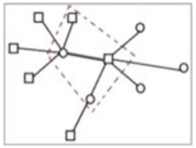
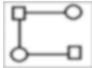
Model 4 reveals that the behaviour of organizations located in EU-15 countries is different than that of others. The positive and statistically significant parameter estimate for edges by EU-15 members indicates that network configurations with more edges by EU-15 agents are more probable. Accounting for the effect of EU-15 membership, being located in the core does not have an additional effect on the general interest in collaborative research.

Model 4 also suggests statistical evidence on the effect of dependence among two participation decisions given by the same organization on network formation. While these dependencies are statistically significant for all types of organizations, their magnitudes and directions are different. The parameter estimate for organization 2-stars shows the effect of the dependence among two participation decisions for the reference categories; i.e.; public organizations (PUB) and other types of organizations (OTH). The negative sign of the estimate for 2-stars and the negative parameter estimates for edges jointly show that organizations labelled as the reference category tend to create fewer edges at the aggregate level and also at the individual level they tend not to have star behaviour. On the other hand, as compared to the reference category a more negative estimate for edges by HES and a more positive estimate for HES 2-stars indicate that networks, where higher education institutions have single participations, are less likely; in contrast networks, where higher education institutions behave like stars, are more likely. Similar arguments hold for research institutions and private enterprises as well.

In addition to these Model 4 suggests that location of participants plays a role in composition of the consortiums. The negative and statistically significant parameter estimate for “co-participations with core” suggests that the more crowded that a consortium with a participant located in the core gets, the less likely the resulting network configuration. However, the positive and statistically significant parameter estimate for “homophily core” reveals that more likely networks are those including a higher number of co-participations by organizations that are located in the core.

Table 12.2 Parameter estimates (standard deviations in parenthesis)

Perspective	Variables		Model 1	Model 2	Model 3	Model 4
 The general interest in collaborative research	Edges		−5.541* (0.103)	−2.836* (0.489)	−2.345* (0.602)	−2.364* (0.586)
	Edges by HES		0.485* (0.100)	−2.423* (0.491)	−2.538* (0.599)	−2.460* (0.587)
	Edges by PRC		0.032 (0.103)	−1.037* (0.512)	−1.360* (0.634)	−1.343* (0.622)
	Edges by REC		0.340* (0.109)	−2.437* (0.497)	−2.795* (0.620)	−2.691* (0.624)
	Edges by EU15		0.275* (0.062)	0.297* (0.069)	0.289* (0.070)	0.286* (0.063)
	Edges by core		0.051 (0.053)	0.049 (0.057)	0.134 (0.125)	0.141 (0.122)
 Organization's portfolio of participations	Organization 2-stars			−3.401* (0.552)	−3.624* (0.568)	−3.634* (0.565)
	HES 2-stars			3.496* (0.553)	3.719* (0.569)	3.690* (0.563)
	PRC 2-stars			1.552* (0.575)	1.749* (0.598)	1.736* (0.594)
	REC 2-stars			3.438* (0.554)	3.660* (0.569)	3.631* (0.566)
 Determinants of co-participations	Project 2-stars				0.031 (0.071)	0.031 (0.069)
	Co-participations with HES				−0.046 (0.038)	−0.052 (0.039)
	Co-participations with PRC				0.026 (0.043)	0.025 (0.043)
	Co-participations with REC				0.026 (0.043)	0.017 (0.044)
	Homophily HES				0.045 (0.043)	0.039 (0.044)
	Homophily PRC				0.009 (0.052)	0.006 (0.053)
	Homophily REC				0.013 (0.060)	0.003 (0.060)
	Co-participations with core				−0.119* (0.017)	−0.121* (0.017)
	Homophily core				0.090* (0.017)	0.091* (0.017)

	Continuity of consortiums			
 Network effects	3-paths		0.002	(0.002)

*significant at 95 % level

Hence, these two parameters point out to a process where organizations in the core tend to collaborate with each other and in small consortiums.

In Model 4, all variables investigating the institutional aspect of consortium composition are found to be statistically insignificant. This result complies with the fact that in the observed network almost all consortiums are heterogeneous in terms of institutional types.

Also, the parameter estimate for 3-cycles, which are included in the model to test for network effects, is found to be statistically insignificant. This may be due to the fact that an organization's access to information on another's portfolio of participations is rather limited in real life. A study by Lhuillery and Pfister (2011), which investigates the awareness of firms of potential ties among their main direct partners by using French data, supports this by revealing that firms are aware of less than half of the potential indirect ties among their direct partners.

Finally, the goodness of fit of the models is assessed through a simulation study. According to results Model 4 performs better than the other models and reproduces many network properties successfully or almost successfully according to the heuristic criteria suggested by Wang et al. (2009). However, it suffers in replicating the clustering in the network. Despite the fact that the level of clustering is very low, it is a matter of fact that the model is not well-performing in this aspect.

12.7 Conclusions

In this study, multilateral R&D collaboration is considered as a particular context to study network formation since the nature of tie formation and knowledge flows in such networks differ from those in bilateral collaboration networks. The two main goals of the study was to explain how organizations come into groups to conduct research given that they are institutionally different and whether these grouping are free of spatial constraints or not. To answer these questions, mainly, insight provided by network formation theory and the proximity literature is moulded with social network analysis approach. Some empirical results are obtained based on the data on proposals submitted to FP7 on a specific sub-theme by using a two-mode representation of the multilateral network and exponential random graph models.

One set of conclusions that can be derived from this study helps understanding how different types of organizations behave in terms of connectivity/multi-connectivity and how they come into groups. The results suggest that higher education institutions and research institutions tend to participate in a higher number of consortiums. On the one hand this means that they constitute the main bridges for learning across consortiums. On the other hand, it may also point out to a difference in capacity or interest to create and maintain a portfolio of participations. Although the statistical evidence brought by this study is not sufficient to draw general policy recommendations, both of these interpretations might be informative for regional policy makers in designing customized policy tools to increase their regions' involvement in collaboration networks.

Apart from that, the findings show that homophily in institutional types does not play a role in the construction of consortiums. This points to a favourable situation in FP7 biotechnology network, as it means that knowledge can possibly diffuse among different parts of the economy. Hence, the underlying network formation processes in FP7 in biotechnology result in a network configuration that permits collective learning by different economic actors.

Another set of conclusions comprise the spatial dimension of these networks. The findings suggest that organizations located in the core European countries tend to participate in the same consortium and these consortiums tend to be small in size. This means that collective learning tend to localize in a continuous corridor in the Western Europe and joint learning capabilities of organizations located in this corridor is being reinforced. Furthermore, the results reveal that the interest in multilateral collaboration is higher for organizations located in EU-15 as compared the others. From a spatial point of view, this means a difference between Western and Eastern Europe in benefiting from flows in multilateral R&D collaboration.

From an analytical and methodological point of view, the results obtained through a core-periphery perspective – i.e. participation decisions depend on each other in the core, and the effect of this dependence is statistically significant-highlights an important point. On the one hand this illustrates the additional explanatory capacity suggested by models relaxing the tie independence assumption. On the other hand, it shows that geography might not only play a role by affecting the utility out of collaboration but also as a delimiter/facilitator of tie dependence.

Nevertheless, the results obtained in this study lack further research effort in several aspects. First, due to data limitations the geographical dimension has been addressed at a broad scale. Second, in this study the network is studied with a static approach as if all consortiums are created simultaneously. Obviously, an organization makes some of its decisions simultaneously, and some at different time instances. Hanneke et al. (2010) and Krivitsky and Handcock (2010) proposed temporal extensions to ERGM to study the evolution as a discrete time Markov process. In this respect, we think that integrating a temporal aspect to study the evolution of a multilateral cooperation network is promising. Third, the study draws empirical evidence from a single technological field. Whether the network formation dynamics in biotechnology apply to other fields remains as an issue to be addressed.

References

Anselin L, Varga A, Acs Z (2000) Geographical spillovers and university research: a spatial econometric perspective. Growth Change 31:501–515

Autant-Bernard C et al (2007) Social distance versus spatial distance in R&D cooperation: empirical evidence from European collaboration choices in micro and nanotechnologies. Pap Reg Sci 86(3):495–519

Bala V, Goyal S (2000) A non-cooperative model of network formation. Econometrica 68: 1181–1229

Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286(5439): 509–512

Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. J R Stat Soc Ser B (Methodol) 36(2):192–236

Boschma RA (2005) Proximity and innovation: a critical assessment. Reg Stud 39(1):61–74

Boschma RA, Frenken K (2009) The spatial evolution of innovation networks: a proximity perspective. In: Boschma RA, Martin R (eds) Handbook of evolutionary economic geography. Edward Elgar, Cheltenham

Cassiman B, Veugelers R (2002) R&D cooperation and spillovers: some empirical evidence from Belgium. Am Econ Rev 92(4):1169–1184

Cohen WM, Levinthal DA (1990) Absorptive capacity: a new perspective on learning and innovation. Adm Q 35(1):128–152, Special Issue: Technology, Organizations, and Innovation

Cooke P, Uranga MG, Etxebarria G (1997) Regional innovation systems: institutional and organisational dimensions. Res Policy 26:475–491

Cowan R, Jonard N, Zimmermann JB (2007) Bilateral collaboration and the emergence of innovation networks. Manage Sci 53(7):1051–1067

Fafchamps M, Goyal S, van der Leij MJ (2010) Matching and network effects. J Eur Econ Assoc 8 (1):203–231

Feldman MP (1993) An examination of the geography of innovation. Ind Corp Change 2(3): 451–470

Feldman MP, Florida R (1994) The geographic sources of innovation: technological infrastructure and product innovation in the United States. Ann Assoc Am Geogr 84(2):210–229

Frank O, Strauss D (1986) Markov graphs. J Am Stat Assoc 81(395):832–842

Frenken K, Van Oort F, Verburg T (2007) Related variety, unrelated variety and regional economic growth. Reg Stud 41(5):685–697

Hagedoorn J (2002) Inter-firm R&D partnerships: an overview of major trends and patterns since 1960. Res Policy 31:477–492

Handcock MS (2003) Assessing degeneracy in statistical models of social networks. Working paper 39, Centre for Statistics and Social Sciences, University of Washington

Hanneke S, Fu W, Xing EP (2010) Discrete temporal models of social networks. Electron J Stat 4:585–605

Hekkert MP et al (2007) Functions of inovation systems: a new approach for analyzing technological change. Technol Forecast Soc Change 74:413–432

Hoekman J, Frenken K, van Oort F (2009) The geography of collaborative Knowledge production in Europe. Ann Reg Sci 43(3):721–738

Jackson OM, Rogers BW (2007) Meeting strangers and friends of friends: how random are social networks? Am Econ Rev 97(3):890–915

Jackson MO, Wolinsky A (1996) A strategic model of social and economic networks. J Econ Theory 71:44–74

Krivitsky PN, Handcock MS (2010) A separable model for dynamic networks. arXiv:1011.1937v1

Lane PJ, Lubatkin M (1998) Relative absorptive capacity and inter-organizational learning. Strateg Manage J 19:461–477

Lhuillery S, Pfister E (2011) Do firms know the scope of their R&D network? An empirical investigation of the determinants of network awareness on French survey data. Ind Innov 18 (1):105–130

Regulation (EC) No 1906/2006 of the European Parliament and of the Council of 18 December 2006 laying down the rules for the participation of undertakings, research centres and universities in actions under the Seventh Framework Programme and for the dissemination of research results (2007–2013)

Nooteboom B et al (2007) Optimal cognitive distance and absorptive capacity. Res Policy 36(7): 1016–1034

Paier M, Scherngell T (2008) Determinants of collaboration in European R&D networks: empirical evidence from a binary choice model perspective. NEMO working paper 10

Park J, Newman MEJ (2004) The statistical mechanics of networks. Phys Rev E 70(6), 066117
Pattison P, Robins G (2002) Neighborhood-based models for social networks. Sociol Methodol 32:301–337
Powell WW, Koput KW, Smith-Doerr L (1996) Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology. Adm Sci Q 41(1):116–145
Robins G et al (2007) An introduction to exponential random graph (p*) models for social networks. Soc Netw 29:173–191
Scherngell T, Barber MJ (2009) Spatial interaction modeling of cross-region R&D collaborations: empirical evidence from the 5th EU Framework Program. Pap Reg Sci 88(3):531–546
Scherngell T, Lata R (2013) Towards an integrated European Research Area? Findings from Eigenvector spatially filtered spatial interaction models using European Framework Programme data. Pap Reg Sci 92:555–577
Snijders TAB (2002) Markov chain Monte Carlo estimation of exponential random graph models. J Soc Struct 3, Article 2
Straus D, Ikeda M (1990) Pseudolikelihood estimation for social networks. J Am Stat Assoc 85 (409):204–212
Uzzi B (1996) The sources and consequences of embeddedness for the economic performance of organizations: the network effect. Am Sociol Rev 61(4):674–698
Van der Leij M, Fafchamps M, Goyal S (2006) Matching and network effects. Discussion paper 611, University of Essex
Wang P et al (2006) PNet: a program for the simulation and estimation of exponential random graph models. School of Behaviour Science, University of Melbourne. http://www.sna.unimelb.edu.au/pnet/pnet.html
Wang P et al (2009) Exponential random graph (p*) models for affiliation networks. Soc Netw 31:12–25
Wasserman S, Pattison P (1996) Logit models and logistic regression for social networks: I. An introduction to Markov graphs and p*. Psychometrika 61(3):401–425
Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. Science 316:1036–1039
Zaheer A, McEvily B, Perrone V (1998) Does trust matter? Exploring the effects of interorganizational and interpersonal trust on performance. Organ Sci 9(2):141–159
Zand DE (1972) Trust and managerial problem solving. Adm Sci Q 17(2):229–239

Chapter 13
The Structure and Geography of Collaboration Networks in the European GNSS Industry

Jérôme Vicente, Pierre-Alexandre Balland, and Raphaël Suire

Abstract The concentration and dispersion of innovative activities in space have been largely evidenced by the nature of knowledge and the geographical extent of knowledge spillovers. One of the empirical challenges is to go beyond this by understanding how the geography of innovation is shaped by particular structural properties of R&D collaboration networks. This paper contributes to this challenge focusing on the case of global navigation satellite systems at the European level. We exploit a database of R&D collaborative projects based on the fifth and sixth EU Framework Programs, and apply social network analysis. We study the properties both of the network of organisations and the network of collaborative projects. We show that the nature of the knowledge involved in relationships influences the geographical and structural organisations of the technological field. The observed coexistence of a relational core/periphery structure with a geographical cluster/pipeline one is discussed in the light of the industrial and geographical dynamics of technological standards.

13.1 Introduction

Technological innovations emerge according to micro–macro dynamics in which networks and geography shape the process that turns new ideas into dominant designs. This paper aims to evidence this process, focusing on the structural dimensions of R&D collaboration networks. The literature on geography of

J. Vicente (✉)
University of Toulouse, Toulouse Business School, and Political Science institute, Toulouse, France

P.-A. Balland
University of Utrecht, Eindhoven Centre for Innovation Studies, Utrecht, Netherlands

R. Suire
University of Rennes 1, CREM-CNRS, Rennes, France

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_13,

innovation has provided important empirical evidence showing that firms learn more easily from each other when they are located within the same place (Feldman 1999). The economics of innovation literature has also early recognized the central role of networks in the development of new products, new processes and new knowledge (Freeman 1991; Hagedoorn 2002). And recent applications of concepts and tools originally developed in network science have pushed further our understanding of the role played by network structure on innovation processes (Ter Wal and Boschma 2009).

To understand better the role of geography and networks in innovation activities, scholars have often investigated the type of knowledge which is actually exchanged between actors. The conceptualization of the nature of knowledge has been a central debate in the field (Cowan et al. 2000) and especially the reference to tacit knowledge has increasingly been used to explain the spatial patterns of innovative activities. Despite this strong interest, empirical studies that investigate how geographical and structural patterns of technological fields are affected by the nature of knowledge remain scarce. Indeed, knowledge spills over both network structures and geography (Breschi and Lissoni 2001), and little is said about the links between the nature of knowledge and the structural organisations of technological fields. Noticeable exceptions come from Broekel and Graf (2012), who investigate how the structure of R&D networks varies depending on the fundamental or applied nature of knowledge ties.

This paper aims at contributing to this challenge by investigating empirically how geographical and structural patterns of technological fields change according to the different stages of technological development. To deal with this challenge, we focus on the particular case of Global Navigation Satellite Systems (GNSS) in Europe. GNSS is a set of satellite systems that provide positioning and navigation solutions. The diffusion of these technologies, as for many information technologies and technological standards, depends on the level of interoperability at the infrastructure level, as well on the level of technological integration between infrastructures, materials (receivers, chipsets) and applications.

The paper is organized as follows: Sect. 13.2 recalls the challenging introduction of structural properties of collaboration networks into the traditional parameters of the geography of innovation. Section 13.3 discusses a set of testable propositions that link cognitive and geographical dimensions of networks to their purely structural dimensions, stressing on the particular case of technological fields in which standardization influences the structuring of networks. Section 13.4 presents the data set of R&D collaborations in the European technological field of GNSS. Section 13.5 proposes an original network analysis developed for both identifying the nature of knowledge involved in relationships and the structural properties of the R&D collaboration network. Section 13.6 tests separately each proposition and discusses the formal results. Section 13.7 combines these results, emphasizing how and why the knowledge process at work in the European GNSS technological field matches geographical cluster/pipeline and network core/periphery structures in a way that permits an emerging idea to be turned into a mass market standard.

13.2 Theoretical Context

The geography of innovation exhibits structures that result from localization and knowledge externalities. One of the main results is that innovation activities tend to be concentrated since tacit knowledge limits the diffusion of knowledge and geographical dispersion occurs as far as knowledge grows in codification (Audretsch and Feldman 1996). But for Breschi and Lissoni (2001), what is hidden behind knowledge externalities could be more the result of the intentional effort of organisations to exchange and combine knowledge than a simple corridor effect. The geographical extent of knowledge spillovers does not depend only on distance but also on the ability of knowledge to flow across relational structures.

The literature in knowledge economics has addressed the micro-motives for shaping knowledge relations, showing that these relations partly involve opportunities to access missing knowledge and partly involve risk of weakening knowledge appropriability (Antonelli 2006). The key parameters for the valuation of these risks and opportunities are the degree with which the knowledge bases of partners complement each other and the degree of openness of their model of knowledge valuation. Organisations decide to form a knowledge partnership only when each one assumes that the benefits of knowledge accessibility will exceed the risks of under appropriation. Structural properties will be not purely physical, but the result of the strategic behavior of organisations to deal with their own knowledge trade-off. Geography matters for the micro-motives of organisations for shaping knowledge relations. Indeed, geographical proximity between organisations involved in a partnership has ambivalent effects on their respective innovation capabilities (Boschma 2005). What these effects are will depend on at least two related criteria: the phases of the knowledge value chain, and the gap between their absorptive capabilities (Nooteboom 2000). Geographical proximity will be more appropriate between partners when they have to favor mutual understanding, and when their core capabilities are sufficiently distant to avoid the risks of unintended knowledge spillovers. Conversely, when partners share close capabilities and compete in few differentiated markets but find opportunities for cooperation (in standards setting for instance), the risk of unintended spillovers is high and geographical distance or temporary proximity are more compatible than proximity.

For a particular knowledge process in a particular technological field, a collaboration network will be defined as the set of organisations that are involved in the field and the set of knowledge ties between them. From this relational matrix and considering the location of organisations, structural properties of the network will be good markers of the channels through which knowledge flows and geography structures itself. The level of connectivity is a good marker to understand the co-existence of arms-length and network relations in a technological field (Uzzi 1997), in particular when compatibility and standardization matter (Cowan et al. 2004). Moreover, a knowledge network can be characterized by a heterogeneous level of relations for each organization, giving rise to particular structures, such as the regular core/periphery structure observed in many situations (Borgatti and Everett 1999). A network exhibits a core/periphery structure when a highly cohesive structure of knowledge interactions between organisations co-exists with organisations that are poorly connected between themselves and with the core. Such a structure shows that knowledge relations are not randomly distributed within a network and can be interpreted as a particular stage of its dynamics. The geography of a knowledge network will reflect these properties. Since Porter's research (Porter 1998), clusters have been seen as efficient structures that favor innovation and growth. Nevertheless, thinking about innovation by focusing only on geographical clusters is a narrow view of innovations occurring in most technological fields. If clusters exist, they are generally embedded in larger geographical structures, and connected through global pipelines (Bathelt et al. 2004; Trippl et al. 2009).

13.3 The Structural and Geographical Properties of the European GNSS Collaboration Network: Two Propositions

The structural properties of knowledge networks has been increasingly investigated in the last couple of years, theoretically (Ter Wal and Boschma 2009; Boschma and Frenken 2010) as well as empirically (Owen-Smith and Powell 2004; Autant-Bernard et al. 2007; Scherngell and Barber 2011; Vicente et al. 2011; Balland 2012; Broekel and Graf 2012). These studies concern different industrial sectors and technological fields, different geographical areas, and different sources of relational data. Here we discusse testable propositions that link cognitive and geographical dimensions of knowledge networks to their purely structural dimensions, focusing on the particular case of a technological field in which standardization and the emergence of a dominant design influences the structuring of R&D collaboration networks. GNSS is a standard term for systems that provide positioning and navigation solutions. These technologies were originally developed in the aerospace and defense industries. But nowadays, they find complementarities and integration opportunities in many other socio-economic contexts concerns by mobility. The diffusion of GNSS related innovations depends on a high level of interoperability and compatibility, as well as a growing number of applications for consumers. That is why innovations in the field are driven by public incentives as a strategic challenge for policy makers to set a European standard of navigation and positioning through the Egnos and Galileo programs.

13.3.1 Structural Properties of Networks and Technological Standard Diffusion

The structural organization of the GNSS technological field will depend on the interplay between the phases of the knowledge value chain (Cooke 2006), the degree of maturity of the field regarding the market conditions (Audretsch and Feldman 1996), as well as its degree of relatedness regarding the interoperability and compatibility constraints of technological standards diffusion (Vicente et al. 2011; Broekel and Graf 2012). Indeed, GNSS are considered general purpose technologies for which the willingness of consumers to pay and adopt depends on the weight of network externalities on the demand side, and thus requires a high level of interoperability between competing suppliers (Katz and Shapiro 1994). Firstly, the diffusion of GNSS will depend on the ability of the suppliers of the field to interact in order to pool together their knowledge and existing technologies around a common standard. Secondly, general purpose technologies such as GNSS cross different sectors and markets so that their diffusion depends on the variety of applications and new markets they water. Transport, telecommunications, software, safety, tourism, environmental observations, among others, are sectors concerned by GNSS-based innovations, and require a high level of knowledge integration between separated and sometimes cognitively distant knowledge in order to propose viable integrated systems to consumers. One can expect that the maturity of the industry goes with a high level of density and connectedness in the collaboration network, with a high level of closure and triangulation that favors the mutual understanding between partners and prevents opportunistic behaviors (Coleman 1988; Cowan et al. 2004).

Nevertheless, literature shows that highly cohesive structures of knowledge interactions produce conformism and display risks of lock-in (Ahuja et al. 2009). Redundant ties limit access to new information and fresh ideas (Burt 1992), and can sclerose the technological field as a whole. Technological fields characterized by a high level of closure can enter into a phase of inhibition, which is typical of the decline phase of the product life cycle. Technological fields will exhibit a long term viability and development, when, in parallel to the structuring of the core, a less cohesive but not disconnected pool of explorative knowledge remains at the periphery of the collaboration network. In the exploration phase, technologies are beta tests and compatibility constraints are not as critical as in the integration and exploitation phases. But this pool of fresh and news ideas should be connected to the core of knowledge interactions, in order to be turned into tradable innovations in the future. Under such structural conditions, the technological field develops an endogenous capability to grow through its periphery, in particular with the strategic and creative role played by the organisations that connect the core to the periphery (Cattani and Ferriani 2008). Between disconnected structures of knowledge interactions that typify the very early stage of a technological field, and the highly ossified and dense structures of interactions that could typify a lock-in process, an expected core/periphery hierarchy appears as a marker of the increasing maturity of the field.

13.3.2 Geographical Properties of Networks and Technological Standard Diffusion

Previous research has already demonstrated that industry life cycles are sensitive to geographical changes due to the increasing codification of knowledge along the cycle of a product (Audretsch and Feldman 1996). The clustering of innovative activities corresponds to the early stage of a product, while dispersion occurs when an industry reaches a high level of maturity. If these results have been abundantly evidenced, they failed to investigate the interaction structures that shape these geographical changes. From the very early phase of emerging ideas to the phase from which these ideas are turned into mass market products, structural as well as geographical properties of knowledge networks evolve.

For instance, Owen-Smith and Powell (2004) highlight structural and geographical patterns along the growing maturity of the biotech sector in Boston. They show that, at the early stage of the cluster, the cohesiveness of the local relational structure rested mainly on the active participation of public research organisations that connect disconnected private organisations in a very open structure of fundamental knowledge dissemination. At the same time, in a nested analysis of geographical scales, they compare the structural and cognitive properties of the local network to the ones of the network extended to other organisations in any locations that have a tie with local ones. They show that clustered relations depend on the dominance of an academic and open institutional regime, while pipelines relationships in which private and big firms are involved remain focused on a market regime in which knowledge appropriateness prevails. Ter Wal (2011) proposes a network-based empirical study in the same knowledge field in the case of the German co-inventors network and observes a similar pattern of the evolution. From the exploration phase in the 1980s to the exploitation phase in the 1990s, he observes a shift in the network strategies of biotech organisations. While the network grew initially along geographical proximity, the increase of knowledge codification along the maturity process has led companies to use global networks as a resource of triadic closure, which favor trust and knowledge appropriateness.

The geography of R&D collaboration networks is thus dependent on the attributes of the organisations and the knowledge value chain of innovations. If clusters remain crucial in the explorative knowledge phase through their ability to connect separated knowledge, they cannot be self-sufficient since diffusion and commercialization require an enlargement of networks in space. Such a geographical structure corresponds to a particular stage of the growing maturity of the field. Research-based organisations are still active in connecting knowledge assets in order to develop new ideas in clusters. At the same time, the incumbents and engineering companies develop pipelines, in parallel to their cluster embeddedness, in order to coordinate the definition of future technological standards and integrate knowledge stemming from other sectors in order to define these standards. Then he clusters/pipelines structure of knowledge interactions in a particular technological field is typical of the overlap between the phases of its knowledge value chain.

13.4 Data

As a relational data source, we use joint R&D projects funded by the Framework Programmes (FP) for research and technological development of the European Union. As such, we follow recent empirical studies emphasizing the advantage of this kind of relational data in economic geography (Autant-Bernard et al. 2007; Breschi et al. 2009; Scherngell and Barber 2011; Balland 2012). For the purpose of this paper, we exploited the GNSS Supervisory Authority[1] (GSA) database on joint GNSS R&D projects funded by the 5th and 6th FP from 2002 to 2007.

This primary database is mainly used to deduce two adjency matrixes: the network of projects and the network of organisations that will be analysed in the empirical section. To construct the network of projects, it is assumed that two projects are linked if at least one organization participates in these two projects. To construct the network of organisations, we have converted the primary 2-mode matrix into a 1-mode square matrix of collaborations between all the organisations. We assume that each project is fully connected (forming a clique), so that two organisations are linked if they participate to the same project. Descriptive statistics on the network of projects and the network of organisations are presented in Table 13.1. They show that both the network of projects (0,181) and the network of organisations display a relatively high density (0,055) and a high connectivity. Considering the network of projects in particular, we identify a principal component of 66 projects, meaning that only 6 projects are isolated during the period of study.

The degree centrality distribution exhibits an asymmetrical shape, indicating that only a few nodes have a high probability of having large number of relations (Fig. 13.1). This statistical signature suggests some interesting traits about the industrial structure of the GNSS sector, related to the setting and control of technological standards. Vertical firms and transnational corporations as well as spatial agencies are often representative organisations of this type of market.

[1] GSA is the European GNSS Agency, in charge of public interests related to GNSS programmes in Europe.

Table 13.1 Structural characteristics

Statistics	Network of projects	Network of organisations
Nb of nodes	72	360
Nb of links (valued)	1,512	7,842
Nb of links (dichotomized)	914	7,144
Density	0.181	0.055
Main component	66	339

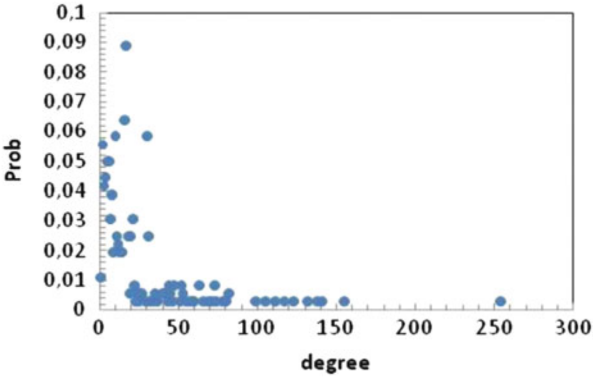


Fig. 13.1 Degree centrality distribution among the 360 organisations

13.5 Methodology

In this section, we describe the method we used to capture the nature of knowledge, in order to proceed to the social network analysis of both the network of organisations and the network of projects. The exploration-integration-exploitation taxonomy is discussed as well as the robustness of the final classification of projects we obtain. In addition, we explain the methodology for the empirical identification of clusters and pipelines in Europe.

13.5.1 Exploration – Integration – Exploitation: A Taxonomy

Joint R&D collaborative projects refer to a large variety of knowledge processes, ranging from exploration (fundamental research) to exploitation (applied focus) to follow the distinction proposed by March (1991). In the context of the GNSS industry, we also consider the integration category, for projects combining different existing technologies because this kind of project is concerned with specific standard and compatibility issues.

To proceed to the classification of the different projects into these three categories, we developed an approach making use of three criteria. First we analyzed the main goal of the project, as expressed in the title of the project, or in the abstract. It generally already gives a clear overview of whether projects are oriented from general concern for GNSS to very specific applications. Second, we used a criteria based on the redundancy of specific related key words (Table 13.2) in the abstract and in other available project documents (work-package reports for instance). By consequence, we classified in “exploration” a set of projects that do not develop direct applications, but aim at improving general knowledge for navigation and positioning. This consists of knowledge production far from clear market opportunities, even if prototypes or beta tests can sometimes result from fundamental research and models. For instance, projects that focus on research for accuracy and reliability of Galileo/Egnos signals, synchronization or calibration of atomic clocks can be considered as belonging to this early phase. On the other hand, we classified as “exploitation” the projects proposing to develop well defined GNSS applications, for instance the development of applications specifically required for transport regulation, air fleet management or emergency services. Finally, we found relevant to distinguish a third category: “integration”, for projects proposing technical integration of two technologies. For instance, in the database, most of the integrative projects are dedicated to the convergence and interoperability between GNSS, telecommunication and computer industries. The integration of two technologies requires additional R&D in order to ensure the compatibility between them.

13.5.2 Robustness of the Classification

We analyzed which type of organization is involved in which kind of project. Broekel and Graf (2012) directly use this kind of approach to distinguish between projects dedicated to basic and applied research, arguing that public research organisations and universities are more likely to be involved in the former, while firms are more likely to be involved in the latter. Following this reasonable assumption, we distinguish among research, engineering and market-related types of organization. We considered that public research organisations and universities belong to the “research” category. Firms specialized in satellite or telecommunications infrastructure, hardware or software, belong to the “engineering” category. “Market-related” category is an important residual one for the GNSS industry, involving final users, designers, associations and business consultants (Vicente et al. 2011). A large proportion of organisations developing engineering knowledge are found (192), with a balanced distribution of organisations developing research (84) and market-related (84) knowledge. This straightforward typology of knowledge bases of organisations allows us to control for our projects’ classification by combining the distribution of the knowledge types of the organisations with the knowledge nature of the projects. Each project displays a number of knowledge bases equal to its number of partners. We studied the distribution of the knowledge bases in the different projects, according to their knowledge phases (Table 13.3).

Table 13.2 Knowledge phase of the projects

	Exploration	Integration	Exploitation
Main goal	New knowledge for future applications	Combine pre-existing technologies	Develop GNSS-based applications and services
Key words	Concepts/theory	Technological standard	Market
	Research	Interoperability	Use
	Investigation	Combination	Applications
	Simulations	Satellite + ICT	Design
	Mathematical model	PDA	Development
	Study	Wireless	Services

Table 13.3 Types of organisations and cognitive nature of collaborations

	Exploration	Integration	Exploitation	Total
Research				
	62	37	25	124
(%)	52,5 %	15.9 %	9.2 %	20 %
Engineering				
(Nb of organisations)	46	163	169	378
(%)	39 %	70.3 %	62.4 %	60.8 %
Market-related				
(Nb of organisations)	10	32	77	119
(%)	8.5 %	13.8 %	28.4 %	19.2 %
Total				
(Nb of organisations)	118	232	271	621
(%)	100 %	100 %	100 %	100 %

This test confirms the robustness of our classification, as research organisations are more involved in exploration, engineering firms in integration, and market-related actors in exploitation.

13.5.3 Identification of Clusters and Pipelines

The GSA and FP databases provide systematic information on the country of the organisations and the name of a contact person, but information concerning postal addresses of organisations is not always indicated. However, the small size of the network allowed us to find missing postal addresses of organisations on their web sites, work packages of the projects or specialized GNSS websites. When a doubt still remained, especially for multi-establishment firms, more thorough research was undertaken in order to find the establishment of the engineers involved in the work packages we were considering. At the end, less than 8 % of the postal addresses are missing. On this base, we proposed a method to identify clusters and pipelines from the global network of organisations. Starting from the square matrix of organisations (360 $\times$ 360), we aggregated all the organisations belonging

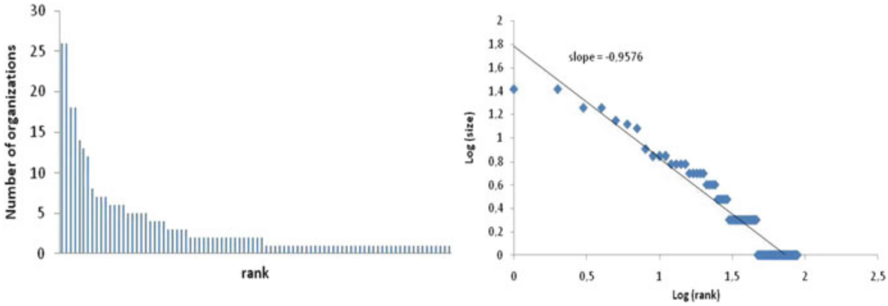


Fig. 13.2 Distribution of organisations among 88 NUTS II European regions

to the same region, taking NUTS2 regions as the spatial unit of analysis (Autant-Bernard et al. 2007; Scherngell and Barber 2011). Then we obtained a new 1-mode matrix of relations between regions, with the diagonal indicating the number of relations within the region.

Figure 13.2 represents the distribution of the number of organisations of the 88 NUTS2 European regions in which at least one organization is involved in the GNSS collaboration network. If we plot the regions against their rank with a log-log scale, it appears that this distribution follows a power law which is quite similar to Zipf law with a slope of −0.9576 obtained with a least square estimation. It is interesting to note the non-monotonic shape of the plot for the first seven values. Conformably to a Zipf like relation, it appears that only very few regions (7/88) concentrate a high number of organisations (more than 10) and a relational density higher than the average density of the network as a whole (see below). We considered that the main GNSS clusters are located in these seven regions. Then we drew a relational matrix for each of these clusters (i.e. we removed all organisations outside of the clusters) in order to study their cognitive structure. Pipelines were studied according to the block matrix of relations between regions.

13.6 Main Findings

This section presents the main empirical results concerning the influence of the nature of knowledge on structural and geographical properties of the GNSS technological field. Both the network of organisations and the network of projects are analyzed in a complementary way to provide empirical evidence for the proposition previously discussed.

13.6.1 Structural Organization of the Collaboration Network: A Core/Periphery Structure

We study the connectivity of the different R&D projects according to their knowledge features. We use the core/periphery model developed for social network analysis by Borgatti and Everett (1999). The core/periphery partition is obtained by using a genetic algorithm (Goldberg 1989). It maximizes the correlation between the observed core/periphery partition matrix and an ideal core/periphery pattern matrix where only core nodes are fully connected, while all peripheral nodes are isolated. Applying this model to the network of projects, we empirically identify a core formed by a group of densely connected projects, while another group of more loosely connected projects constitutes the periphery (Fig. 13.3). Table 13.4 presents the results of the model. Projects in the exploration phase are mostly peripheral, since only 4.4 % of the projects that are in the exploration phase are in the core. In contrast, 32 % of integrative projects and 41.7 % of exploitative projects belong to the core. The closer projects are to the market, the more they are interconnected. On the contrary, the upstream phase of knowledge value chain remains “located” at the periphery.

This result can be strengthened by an econometrical test in order to control for the size of the projects. Recall that we have shown above that organisations are not randomly distributed along the knowledge phases (exploration, integration, exploitation). Thus, we perform an econometrical test in order to estimate whether the knowledge profile of the partners (research, engineering, or market-related) influences the probability of the project belonging to the core of the network, with the size of the project as a control variable. To that end, for each of the 72 projects we distinguish the respective level of organisations belonging to research, engineering, and market-related categories. Then, we use a continuous variable range from 1 to 10 regarding the level of presence of each knowledge base.[2] For instance, a project of size 19 with 2 “research” organisations, 16 “engineering” organisations and 1 “market-related” organization is coded (2, 9, 1). This means that respectively 10.53 %, 84.21 %, 5.26 % of organisations are research, engineering, and market-related ones. We define $Y_{i \in \{1,72\}}$, as a binary variable taking the value 1 if the project i belongs to the core and the value 0 otherwise. The probability of belonging to the core is assumed to be related to the size of the project and the knowledge profile of the partners. The relationship is specified as:

$$Pr[Y_i = 1|X] = \Phi\left(\beta_0 + \beta_1 size + \beta_2 size^2 + \beta_3 research + \beta_4 engineering + \beta_5 market\right),$$

With $\Phi(.)$ representing the cumulative normal distribution function and X is the vector of regressors. We also estimate marginal effect which is the slope of the probability curve to each regressor X to $Pr[Y_i = 1|X]$, holding other variables

[2] For each project we code 1 if the project exhibits between 0 % and 10 % of organisations with a knowledge profile, 2 if the project exhibits between 10 % and 20 % . . . to 10 if the project exhibits between 90 % and 100 %.

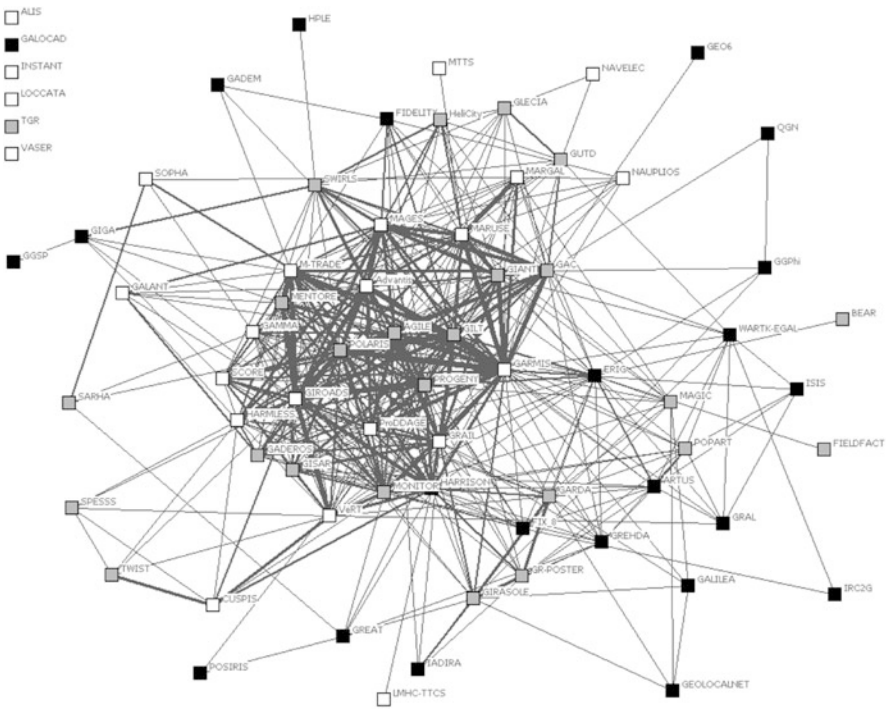


Fig. 13.3 Core & Periphery structure and nature of knowledge (*Black squares* represent projects dedicated to exploration, *grey squares* to integration, and *white squares* to exploitation. The *line* strength represents the number of organisations that tie projects, from 1 to 5 ties)

Table 13.4 Core & Periphery

	Core	Periphery	Total
Exploration			
Nb of projects	1	22	23
%	4.4 %	95.6 %	100 %
Integration			
Nb of projects	8	17	25
%	32 %	68 %	100 %
Exploitation			
Nb of projects	10	14	24
%	41.7 %	58.3 %	100 %
Total			
Nb of projects	19	53	72
%	26.4 %	74.6 %	100 %

Table 13.5 Probit estimation and marginal effect

Explained variable = belonging to the core	Probit estimation	Marginal effect
Size	0.925***	.0044***
	(0.204)	
Size^2	−0.019***	−.0000908***
	(0.004)	
Research	0.713	.003393
	(0.725)	
Engineering	1.604*	.0076339*
	(0.758)	
Market-related	1.206*	.0057391*
	(0.620)	
Constant	−23.962**	
	(9.497)	
Number of observations	72	
Log pseudolikelihood	−9.889	
Pseudo R2	0.7620	

Note: ***, **, * mean significant at the level of 1 %, 5 %, and 10 % respectively. Robust standard errors in parenthesis.

constant.[3] The following table displays the result of a probit estimation,[4] as well as the marginal effect of each variable (Table 13.5).

As we suspected, the probability of a project belonging to the core of the network is significantly influenced by engineering and market-related knowledge bases. Conversely, increasing the level of the research component has no effect on the probability of belonging to the core of the network. The marginal effect of the research component has no impact on the probability to belong to the core of the network. It also means that if a collaborative project has to belong to the core for market purpose or standardization consideration, increasing the level of the research base within the project has no effect on the probability of belonging to the core. The engineering component is the more influential determinant: a marginal positive variation of this knowledge base increases the probability of belonging to the core by 0.7 %. Finally, an interesting result appears regarding the size of the project. Increasing the size of the project has a positive effect on the probability of belonging to the core of the network but at a decreasing rate, which means the existence of a threshold above which the marginal actors negatively influence the probability of belonging to the core. As previously mentioned, one plausible explanation relies on the limited capabilities of various partners to efficiently manage coordination costs. This hypothesis is sustained in network literature on strategic networks stability (Jackson and Wolinsky 1996).

[3] Detailed about the econometric specification can be found in Cameron and Trivedi (2005).

[4] We control for Heteroscedasticity with White correction.

Fig. 13.4 GNSS clusters and pipelines in Europe

13.6.2 Geographical Organization: A Clusters/Pipelines Structure

The second set of results concern the way the features of knowledge influence the geographical structuring of the technological field. As previously said, clusters are identified on the basis of the number of organisations in the region that are involved in GNSS projects, but also according to the number of relations within the cluster. This methodology allows us to identify the main GNSS clusters and the pipelines between them (Fig. 13.4).

Table 13.6 presents descriptive statistics concerning the seven main GNSS clusters. Considering the number of relations, the biggest cluster is located in the Community of Madrid (132 ties within the cluster), the second one in the Lazio Region (74) and the third one in the Midi-Pyrenees Region (52). We can see that these three clusters include the three main organisations (according to their degree centrality): Thales Alenia Space (Toulouse), Telespazio (Roma) and GMV (Madrid).

In order to provide information about the cognitive structure of the GNSS clusters, each cluster's relational matrix has been divided into three matrixes (nodes are still organisations), according to the nature of relations: exploration, integration and exploitation. Table 13.7 shows how the nature of knowledge influences the geographical organization of the GNSS technological field.

Indeed, 48 % of the relations within the clusters belong to the exploration phase, 30 % to the integration phase and only 22 % to the exploitation phase. This result is

Table 13.6 Clusters and pipelines interaction structure

Clusters	Community of Madrid	Lombardy Region	Upper Bavaria	Midi-Pyrenees Region	Lazio Region	Inner London	Ile de France Region
Main organization	GMV	PRS	Astrium	TAS	Telespazio	Logica	FDC
Nb of organisations	26	13	12	18	18	14	26
Internal degree[a] (dichotomized)	132	20	18	52	74	14	38
Density (dichotomized)	0.203	0.128	0.136	0.169	0.241	0.076	0.058
Exploration	86	2	6	32	24	10	18
Integration	32	6	12	14	28	2	22
Exploitation	34	14	0	6	28	2	0
Internal degree (valued)	152	22	18	52	80	14	40
Pipelines							
Community of Madrid	–	22	34	74	57	37	79
Lombardy Region	22	–	8	13	47	5	11
Upper Bavaria	34	8	–	27	23	14	20
Midi-Pyrenees Region	74	13	27	–	40	30	57
Lazio Region	57	47	23	40	–	11	28
Inner London	37	5	14	30	11	–	25
Ile de France Region	79	11	20	57	28	25	–
External degree[b]	303	106	126	241	206	122	220
Cluster openness[c]	1.99	4.81	7	4.63	2.57	8.71	5.5

[a]Internal degree refers to the number of relations within the cluster
[b]External degree refers to the number of relations across the cluster, i.e. within the pipelines
[c]Cluster openness = external degree/internal degree

Table 13.7 Nature of knowledge flows in clusters and pipelines

	Exploration	Integration	Exploitation	Total
Within the clusters				
Nb of links	178	116	84	378
%	47 %	31 %	22 %	100 %
Within the pipelines				
Nb of links	462	588	274	1,324
%	35 %	44.5 %	20.5 %	100 %
Clusters/others				
Nb of links	1,482	1,610	890	3,982
%	37 %	40.5 %	22.5 %	100 %
Others/others				
Nb of links	210	376	478	1,064
%	20 %	35 %	45 %	100 %

in line with the literature, according to which geographical proximity is more important in the exploration phase (Audretsch and Feldman 1996). Similarly, the pipeline relational matrix has been divided into three matrixes (the nodes are still the seven clusters), according to the nature of relations: exploration, integration and exploitation. Table 13.8 reveals a radically different distribution than the one found for local knowledge relations. Indeed, now 35 % of the relations across the clusters belong to the exploration phase, but 44.5 % to the integration phase and only 20.5 % to the exploitation phase. This result shows that organisations are more likely to collaborate with others located in another dominant cluster when collaborating on a project in the integration phase. Thus, we have shown that the phases of knowledge, i.e. exploration, integration or exploitation, are not randomly developed in clusters and pipelines, but that exploration tends to require more geographical proximity.

13.7 Discussion: How Do Clusters/Pipelines and Core/Periphery Structures Work Together in R&D Collaboration Networks?

Firstly, the study of connectivity between projects suggests that organisations that are not directly tied in a project can be tied through intermediaries that connect separated projects, so that knowledge can potentially flow into the network. If arms' length relations exist, knowledge diffusion and exchange seem to prevail in a cohesive structure of relations. This means that most of the organisations are aware that GNSS are general-purpose technologies that require a high level of interoperability and compatibility between applications. Such a result is typical of

Table 13.8 Cognitive/geographical/structural properties and the phases of the knowledge value chain

	Knowledge exploration	Knowledge integration	Knowledge exploitation
Cognitive properties	*Research and fundamental knowledge*	*Engineering knowledge*	*Market-related knowledge*
Geographical properties	*Highly clustered in a couple of places*	*Pipelines, cluster relatedness*	*Dispersed and covering the European area*
Structural properties	*Periphery*	*Core and periphery*	*Core*

the "industry of networks", for which development and diffusion require standardization. This relatedness is also the result of the European Commission strategy that makes sure that research in the field rests on the setting of standards, in order that innovations turn into mass-market technologies. The overall connectivity of the GNSS network exhibits an interesting structural property of core/periphery, meaning that beyond the average level of connectivity between collaborative projects, some of them are highly interconnected while some others remain poorly connected. On one hand, the development of the market will be all the more extensive if organisations exchange knowledge in order to set and stabilize the standard. Nevertheless, a full cohesive structure can engender some risks of lock-in. That is why, on the other hand, exploration activities enter the network gradually through the periphery, in order to maintain research and upstream technological solutions that can diffuse to the core when market opportunities occur.

Secondly, it is noteworthy that the main geographical clusters of the GNSS network are typified by a high level of explorative relations and a decreasing share of relations from exploration to exploitation (Table 13.7). This is not really a surprising result since the literature shows that exploration phases compel a high level of fundamental and tacit knowledge that requires proximity between organisations and social network effects. If we turn to pipelines, Table 13.7 shows that pipelines gather a large part of collaborations in the integration phase. An efficient integration and combination process requires cooperation between complementary as well as competing companies located in different clusters in order to set up a technological standard as widely as possible. The "space alliance" being composed by a couple of clusters in Europe (Fig. 13.4), the existence of these pipelines in the engineering process confirms the usefulness of the Galileo project. This project intends to organize the viability of the technological field by creating incentives for cooperation, in order to guarantee the diffusion of GNSS-based applications. Finally, knowledge relations in the exploitation phase are poorly represented in the main clusters as well as in pipelines. A large share of exploitation relations involves organisations that are dispersed in Europe. This result is not a surprise since the main purpose of collaborations in this phase concerns market tradability and diffusion of technological applications. These dispersed networks are all the more necessary given that GNSS diffusion, as well as ICT demand, is influenced by network externalities and thus by a wide geographical availability of applications.

Finally, considering the combination of the structural and geographical dimensions, new findings in economic geography and knowledge economics emerge. Table 13.8 summarizes these findings, crossing the knowledge phases with the cognitive, structural and geographical statistics of the GNSS network.

The most noteworthy result is the negative linear relationship between the geographical and structural concentration of knowledge interactions. This means that the more projects are embedded in a highly cohesive structure, the less knowledge relations are clustered in particular locations. The fact that geographically clustered relations are "located" in the periphery of the network of projects does not mean that clusters host organisations that are poorly connected among themselves. Recall that Table 13.6 showed that the seven main clusters display an internal density higher than the average density of the network as a whole. On the contrary, clusters are highly cohesive sub-structures of knowledge relations focused mainly on explorative projects that are poorly connected to the core of projects of the European network. At the other extremity, the core of collaborative projects hosts organisations that are scattered across the European area. Between these two extremes, an intermediate level of geographical dispersion corresponds to the interconnection between clusters that supports the integration knowledge processes.

This negative linear relationship can be explained by the industrial and spatial organization that supports the viability of the GNSS technological field. If we suppose the GNSS network in the period under investigation to be in a particular stage of its endogenous dynamics, its core/periphery and cluster/pipeline structure will reflect its particular stage of maturity. If clusters have been considered in the literature as efficient structures of knowledge production, their existence and their high performance are not sufficient conditions of high performance in the technological field as a whole. To reach maturity, a technological field needs to be supported by a high level of spatial diffusion supported itself by the existence of norms, compatibility and interoperability. The existence of pipelines and the spatially dispersed core of the network is thus the illustration that the GNSS technological field has reached a certain level of maturity during the period under study. Nevertheless, an excess of cohesion in the network can be interpreted as a lock-in condition that excessively scleroses the knowledge dynamics at work within the network. That is why, as previously said, the periphery of the network is a condition of its viability, because it can introduce fresh ideas and new knowledge in order to strengthen and extend the increasing part of the curve of the technological life cycle, part in which clusters play a critical role.

13.8 Conclusion

Our results highlight how knowledge spills over geography and relational structures, and how a particular technological field structures itself along its knowledge value chain. The salient outcome is the negative linear relationship found between geographical cluster/pipeline and structural core/periphery structures in the

European GNSS technological field. We have shown that clusters are critical loci for exploration processes in the upstream phase of the knowledge value chain and contribute to the growth of the technological field. But clusters, in spite of the focus they constitute for innovation policies, do not contribute alone to the market success of technologies. At the periphery of the knowledge network, clusters play a critical role by preserving a pool of new and upcoming exploitable knowledge. But the new ideas in a technological field will be turned into mass market products if, in the downstream knowledge phase of integration and exploitation, tradable goods and technologies remain on a high level of spatial diffusion and technological standardization. So the viability of the technological field will depend on the existence of a cohesive structure of relations in the core of the network of knowledge projects that involve dispersed and distant organisations.

In terms of policy perspectives, our findings suggest that networks and geography matter for innovation. Policy makers have to deal with these two dimensions jointly. Indeed, on the one side, nations have progressively targeted their policies from an industrial policy focus, generally governed at the national level, to a more decentralized and regional emphasis, with the development of clusters policies. Such a move towards the increasing role of regions in knowledge-based economies is consistent with the necessity to support leading places in technological domains. On the other side, the creation of the European Research Area has certainly participated to a better dissemination of knowledge in Europe and then an increasing capacity to integrate separated pieces of knowledge to foster innovation. But our findings suggest that these two sides need to be strongly related and more coordinated at the European level. If regional or national clusters policies have definitely increased the capacity of regions to explore new technological domains, the chance to transform them into future dominant designs depends on the ability of clusters to be connected to largest networks (Frenken et al. 2009).

References

Ahuja G, Polodiro F, Mitchell W (2009) Structural homophily or social asymmetry? The formation of alliances by poorly embedded firms. Strateg Manage J 30:941–958

Antonelli C (2006) The business governance of localized knowledge: an information economics approach for the economics of knowledge. Ind Innov 13(3):227–261

Audretsch D, Feldman MP (1996) Innovative clusters and the industry life cycle. Rev Ind Organ 11:253–273

Autant-Bernard C, Billand P, Frachisse D, Massard N (2007) Social distance versus spatial distance in R&D cooperation: empirical evidence from European collaboration choices in micro and nanotechnologies. Pap Reg Sci 86(3):495–520

Balland PA (2012) Proximity and the evolution of collaboration networks: evidence from research and develoment projects within the GNSS industry. Reg Stud 46(6):741–756

Bathelt H, Malmberg A, Maskell P (2004) Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. Prog Hum Geogr 28(1):31–56

Borgatti SP, Everett MG (1999) Models of core/periphery structures. Soc Netw 21:375–395

Boschma RA (2005) Proximity and innovation: a critical assessment. Reg Stud 39(1):61–74

Boschma RA, Frenken K (2010) The spatial evolution of innovation networks. A proximity perspective. In: Boschma RA, Martin R (eds) Handbook on evolutionary economic geography. Edward Elgar, Cheltenham, pp 120–135

Breschi S, Lissoni F (2001) Knowledge spillovers and local innovation systems: a critical survey. Ind Corp Change 10(4):975–1005

Breschi S, Cassi L, Malerba F, Vonortas N (2009) Networked research: European policy intervention for information and communication technologies. Technol Anal Strateg Manage 21:833–857

Broekel T, Graf H (2012) Public research intensity and the structure of German R&D networks: a comparison of 10 technologies. Econ Innov New Technol 21(4):345–372

Burt R (1992) Structural holes. Harvard University Press, Cambridge, MA

Cameron C, Trivedi P (2005) Microeconometrics. Cambridge University Press, Cambridge, NY

Cattani G, Ferriani S (2008) A core/periphery perspective on individual creative performance: social networks and cinematic achievements in the Hollywood film industry. Organ Sci 6:824–844

Coleman JS (1988) Social capital in the creation of human capital. Am J Sociol 94:95–120

Cooke P (2006) Rational drug design, the knowledge value chain and biosciences megacentres. Camb J Econ 29:325–341

Cowan R, David PA, Foray D (2000) The explicit economics of knowledge codification and tacitness. Ind Corp Change 9:211–253

Cowan R, Jonard N, Özman M (2004) Knowledge dynamics in a network industry. Technol Forecast Soc Change 71:469–484

Feldman M (1999) The new economics of innovation, spillovers and agglomeration: a review of empirical studies. Econ Innov New Technol 8(1):5–25

Freeman C (1991) Networks of innovators: a synthesis of research issues. Res Policy 20:499–514

Frenken K, Hoekman J, Kok S, Ponds R, Van Oort F, Van Vliet J (2009) Death of distance in science? A gravity approach to research collaboration. In: Pyka A, Scharnhorst A (eds) Innovation networks – new approaches in modelling and analyzing. Springer, Berlin, pp 43–57

Goldberg DE (1989) Genetic algorithms. Addison Wesley, New York

Hagedoorn J (2002) Inter-firm R&D partnerships: an overview of major trends and patterns since 1960. Res Policy 31(4):477–492

Jackson MO, Wolinsky A (1996) A strategic model of social and economic networks. J Econ Theory 71:44–74

Katz ML, Shapiro C (1994) Systems competition and network effects. J Econ Perspect 8:93–115

March JG (1991) Exploration and exploitation in organizational learning. Organ Sci 2:71–87

Nooteboom B (2000) Learning by interaction: absorptive capacity, cognitive distance and governance. J Manage Gov 4(1–2):69–92

Owen-Smith J, Powell WW (2004) Knowledge networks as channels and conduits: the effects of spillovers in the Boston biotechnology community. Organ Sci 15(1):5–21

Porter ME (1998) On competition. Harvard Business School, Boston

Scherngell T, Barber M (2011) Distinct spatial characteristics of industrial and public research collaborations: evidence from the 5th EU Framework Programme. Ann Reg Sci 46:247–266

Ter Wal A (2011) The spatial dynamics of the inventor network in German biotechnology: geographical proximity versus triadic closure. Papers in evolutionary economic geography 11–02, Section of Economic Geography, Utrecht University

Ter Wal A, Boschma R (2009) Applying social network analysis in economic geography: framing some key analytic issues. Ann Reg Sci 43(3):739–756

Trippl M, Tödtling F, Lengauer L (2009) Knowledge sourcing beyond buzz and pipelines: evidence from the Vienna software cluster. Econ Geogr 85(4):443–462

Uzzi B (1997) Towards a network perspective on organizational decline. Int J Sociol Soc Policy 17:111–155

Vicente J, Balland PA, Brossard O (2011) Getting into networks and clusters: evidence from the Midi-Pyrenean GNSS collaboration network. Reg Stud 45(8):1059–1078

Part IV
Impact of R&D Networks and Policy Implications

Chapter 14
Proximity and Stratification in European Scientific Research Collaboration Networks: A Policy Perspective

Jarno Hoekman and Koen Frenken

Abstract In this chapter we introduce a framework to understand the geography of scientific research collaboration with an emphasis on empirical studies that evaluate the policy efforts to create a 'European Research Area' (ERA). We argue that the geography of scientific research collaboration follows a logic of proximity that provides researchers with solutions to the problem of coordination, and a logic of stratification that provides researchers with differential means to engage in collaboration. The policy efforts to create ERA can then be understood as strategic policy interventions at the European level that affect the form and nature of both structuring principles. More specifically, the aim of reducing '*fragmentation of research activities*, *programmes and policies*' affects the importance of several forms of proximity vis-à-vis each other, while the promotion of '*research excellence*' results in new forms of network stratification at multiple spatial scales. We provide an overview of recent empirical findings to illustrate these claims, and discuss potential implications for future ERA policies.

14.1 Introduction

Probably the largest transnational policy effort affecting the current geography of scientific research collaboration is the effort by the European Commission (EC) to create a European Research Area (ERA). The objective of ERA policy is to overcome "*fragmentation of research activities*, *programmes and policies across*

J. Hoekman (✉)
Department of Pharmaceutical Sciences & Department of Innovation Sciences, Utrecht University, Utrecht, The Netherlands
e-mail: j.hoekman@uu.nl

K. Frenken
School of Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: k.frenken@tue.nl

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_14,

Europe" (Commission 2007, p. 2) by removing "*barriers to the free flow of knowledge*" (European Council 2008, p. 5). This aim is pursued through direct funding of collaborative research projects, mobility schemes and streamlining of research policies. The Framework Programmes (FPs) of the European Commission (EC) constitute one of the main instruments to realize the ERA vision. They are specifically designed to pool resources and promote international scientific collaboration between EU member states by enabling and intensifying interactions among researchers. Ever since its inception, research budgets for the FPs have been on the rise and the budget for Horizon 2020 indicates a substantial increase over previous FPs (Commission 2013).

Despite the substantial resources supporting ERA policy, clarity is still lacking about what the ERA vision entails and about the rate of progress in moving towards this vision. Reading from EC policy documents, ERA is conceived as "*an 'internal market' in research, an area of free movement of knowledge, researchers and technology*" (Commission 2002, p. 4). From a geographical perspective, this vision is expressed in policy efforts to reduce the significance of spatial barriers that hinder European-wide research collaboration such as those following from regional and national boundaries. Yet, the intended geographical effects of these efforts are uncertain and the abstract nature of the vision masks the fact that there are trade-offs between more specific objectives defined under the heading of ERA policy. For instance, there has been much concern that competitive research policies compromise the cohesion objective of the European Union (Sharp 1998; Begg 2010) as those policies are not intended to intervene in the European scientific and technological landscape at large, but to bundle resources with the purpose of supporting collaborative efforts between '*excellent*' researchers in a few strategic scientific fields.

Against this background, the goal of this chapter is *first* to introduce a conceptual framework that can be used to understand the geography of scientific research collaboration and *second* to review recent empirical studies on the structuring principles of this framework in the context of ERA. Our conceptual framework starts from the observation that the geographical structure of scientific collaboration networks can be understood from the joint outcome of a logic of proximity and a logic of stratification (Hoekman et al. 2009). In the review we specifically focus on empirical studies that have addressed research collaboration in the scientific domain using (co-)publication data.

The remainder of this chapter is structured as follows. In the next section we provide a theoretical introduction to the geography of research collaboration. We subsequently pay attention to proximity and stratification as two organizing principles of research collaboration and show how they are affected by the European policies to create ERA. The implications of our findings are discussed in the concluding section.

14.2 Geography of Research Collaboration

The geography of research collaboration deals with the question how space structures collaborations between researchers, and how aggregates of such collaborations constitute spatial networks between locations that can be studied using spatial scientometric tools (Frenken et al. 2009). In this framework, research collaborations are structured according to a logic of proximity that provides solutions to the problem of coordination, and a logic of stratification that provides differential means to engage in collaboration (Hoekman et al. 2009). These logics are not stable over time but change as a consequence of globalisation. Globalisation follows from a process of time-space compression (Harvey 1990), which is made possible by advancements in transportation and ICTs. At the same time, globalisation is governed by institutional harmonisation at the transnational level as envisaged for instance by transnational institutions such as the European Commission. These institutions may directly affect the geography of collaboration through funding of international collaborative research, but also indirectly through the alignment of research agenda's and infrastructures between territories.

The understanding of space needs to be explicated in this context as contemporary geographers have provided multiple conceptions of place and space which refer to material as well as to perceptual dimensions (e.g. Lefebvre 1991; Massey 2004). We start from the physical location of individuals on the Euclidean surface and their media of communication and movement. Given these general elements, space can be defined as a fundamental material dimension that provides settings of interaction as a time-sharing activity between individuals (Hägerstrand 1970; Giddens 1984; Harvey 1990). This materiality can be both conceived in terms of places where researchers are co-present, as well as in terms of flows which allow for time-sharing activities at a distance (Castells 1996).

One can argue that research collaborations always involve some form of time-sharing activity between individuals, although it has been notoriously difficult to provide more exact definitions (Katz and Martin 1997). Traditionally, these settings of interactions follow from moments of physical co-presence when researchers meet at certain locations and interact with one another face-to-face. The complex nature of scientific activities makes this form of interaction essential as some aspects of knowledge are tacit, implying that they "*cannot be put into words*" (Polanyi 1958, p. 4) or "*cannot be or – have not been – set out or passed on in formulae, diagrams or verbal descriptions and instructions for actions*" (Collins 2001, p. 72). Acquisition of this knowledge therefore necessitates 'enculturation' between researchers ranging from short-time visits to institutionalization in 'master-apprentice relations' (Collins 1985). Furthermore, moments of co-presence facilitate the establishment of trust through the sensory effect that individuals have on one another when they are co-present (Simmel 1997; Urry 2000) which is essential to establish the credibility of research findings (Shapin 1995).

The materiality of space and the indivisibility of the body set limits on the co-presence of individuals in these settings of interaction. Individuals can only be at one location at the same time and movement in space involves movement in time (Hägerstrand 1970). Research collaborations that rely on moments of co-presence are thus structured by the location of scientists vis-à-vis each other and their means of mobility when they intend to meet. It follows that collaborating scientists may need to co-locate depending on the necessary frequency of co-presence and the advancement of media of mobility. More specifically, when either the necessary frequency of co-presence is high or the means of mobility are low, it becomes a necessary condition that researchers work in close physical proximity on a permanent base. In time-space geography this condition is visualised using time-space prisms that show the absolute boundaries of individual movement in space given that he/she needs to 'bundle' with other individuals at a particular moment in the future (Hägerstrand 1970).

Technological advancements provide the possibility to relax the necessary overlap between co-presence and co-location (Torre and Rallet 2005). Transportation technology, the development of related material infrastructure and a relative decline in the costs of mobility have rendered a 'shrinking of distance' (Janelle 1969) in terms of the time and money needed to travel from one location to the other. As a result individuals can travel longer distances than in the past without necessarily travelling longer. This process extends the spatial range that a researcher can cover given that s/he wants to return to his/her permanent location within a particular time-frame. Spatial range is not a simple function of the kilometric distance between the permanent location of researchers because the material infrastructure that supports differential means of mobility (e.g. highways, airports) is unequally distributed in space. Moreover, the actual perception of distance is a subjective matter and differs between individuals according to their mental maps (Milgram and Jodelet 1976).

Information technologies also make physical proximity on a permanent base less of a necessity, since researchers can interact through the material infrastructure that supports flows of communication between distant locations (Torre and Rallet 2005). In this context, Castells (1996) notes that these technologies create new spaces of their own which are not constituted by traditional settings of interactions based on co-presence, but are materialised in 'circuits of electronic exchange' that support time-sharing practice without physical proximity. As such, space can no longer be conceptualised based on physical proximity alone but its materiality should also be conceived in terms of flows and their particular spatial forms.

However, despite technological advancements, the substitution of communication technologies for moments of co-presence is limited. Olsen and Olsen (2000) question in this respect whether this substitution process can ever be perfect as modern media hinder the unique establishment of common reference frames and mutual understanding through amongst others rapid feedback, pointing and referring to objects in real space (i.e. acquiring ostensive knowledge), subtle communication, informal interaction before and after 'meetings' and a shared local context. Thus, a main tenet of the geography of research collaboration holds that despite technological advancement, the friction of distance still exerts gravitational force on collaborative knowledge production.

14.3 Logic of Proximity

Physical proximity between researchers provide solutions to the problem of coordination in actual collaborative practices which is a main concern surrounding the uncertain activity of knowledge production. Coordination involves the creation of alignments between researchers by integrating different pieces of a research project in order to accomplish collective tasks (Cummings and Kiesler 2007). As argued above, moments of co-presence are essential to create such alignment.

However, the exact intensity and duration of moments of co-presence that is necessary for successful coordination is conditioned by proximities other than physical proximity, which may already exist between collaborating researchers (Boschma 2005). Already established proximities mediate the success of coordination given a certain amount of co-presence. For example, researchers that already collaborated in the past created social and cognitive proximity which facilitates future collaborations. They will be more effective in communicating by means of ICTs because trust and common references frames have already been established (Amin and Roberts 2008). Hence, the need for co-presence is expected to decline over time in repeated collaborations. It has also been shown that there is less need for co-presence in research collaborations between universities than in university-industry research collaborations as in the former institutional proximity is already established at different locations, whereas in the latter it is not (Ponds et al. 2007).

Physical proximity between researchers is in itself neither a necessary nor a sufficient condition for successful research coordination, although it facilitates the establishment of other forms of proximity *via* moments of co-presence (Boschma 2005). As a result, most forms of proximity are geographically localized as they are established through recurrent moments of co-presence between researchers. For instance, socio-cognitive proximities established on the basis of previous moments of co-presence are often sustained in localised networks. Breschi and Lissoni (2009) show for instance that researchers' embeddedness in social networks decays with geographical distance. Storper and Venables (2004, p. 367) consider cities a main stage where socio-cognitive proximities are sustained by making reference to their importance for "*getting into loops which are associated with collocation*". Others have in this context pointed towards "*being there*" (Gertler 2003) and "*buzz*" (Bathelt et al. 2004) as organising principles for socio-cognitive systems that are bounded in space (see also Howells 2002).

Moreover, institutional proximities as defined by common 'rules of the game' (North 1990) that are enforced in particular locations are almost by definition geographically localised. In science, institutional proximity has historically been created at the national level by building national education systems and creating specific technological capabilities in order to stimulate economic growth (Lundvall 1988; Crawford et al. 1993). As a result there are many key institutional settings with national significance in research funding schemes, research infrastructures, research assessments, education systems, intellectual property regimes and labour markets, amongst others (Crescenzi et al. 2007). In addition to these national institutions, the emphasis on regional competiveness as an important policy goal has also contributed to a plethora of regional institutions that create institutional proximity between researchers at the sub-national level (Bristow 2005).

14.3.1 Proximity and ERA

European policy interventions to create ERA affect the spatiality of proximity dimensions in multiple ways. *First*, direct funding of transnational research projects and mobility schemes is expected to facilitate moments of co-presence between European researchers that are often not located in close physical proximity to each other. The main policy instruments to achieve this goal are FP projects with a temporal character, but also long-term collaboration networks such as the Virtual Knowledge and Innovation Communities that have recently been created under the heading of the European Institute of Innovation and Technology's (EIT) are important in this respect. The European character of these efforts follows from formal allocation criteria of funding that require the inclusion of researchers from multiple European member states. Given the pervasive geographical localization of research collaboration, these collaborative projects are unlikely to emerge in similar structure without strategic policy intervention. Hence, funding of collaborative research projects is instrumental in the creation of ERA through the establishment of novel socio-cognitive proximities between physically distant researchers. In doing so the EC aims to remove spatial barriers – especially national borders – that hamper collaboration between different nation-states within Europe.

Second, ERA policy is also expected to be instrumental in aligning regional and national institutions in which new forms of research collaboration may eventually become embedded. Initiatives to achieve this goal include ERA-NET that aims to counteract the fragmentation of national research policies and funding schemes between separate member states by networking and streamlining activities; ERA-WATCH that benchmarks information on the research policies and research systems of member state; ESFRI that coordinates investments in pan-European research infrastructures; and the Joint Programming Initiatives in which member states reach agreements on Strategic Research Agendas to address major societal challenges. In these programmes institutional alignment is realised through an Open Method of Coordination (OMC) which is characterised by soft regulations such as guidelines, indicators, benchmarking and learning through best practice. There are no official sanctions in OMC as it is believed that the method's effectiveness is ensured through a form of peer pressure and a process of 'naming and shaming'. As such this instrument functions as a catalyst for harmonisation between national policies.

The rate of progress towards the creation of ERA in science has been assessed by monitoring the evolution of spatial collaboration networks constructed from publication as well as from FP project data. The empirical results presented in these studies demonstrate that the incidence of cross-border research collaboration in Europe is increasing over time which goes at the expense of scientific research collaborations within sub-national regions and nation-states. More specifically, co-publication activities in Europe show a gradual tendency towards European integration judged from the observation that the importance of territorial borders reduces over time (Mattson et al. 2008; Hoekman et al. 2010; Chessa et al. 2013). There is evidence that (part of) this process of integration results from funding provided through the FPs. Hoekman et al. (2013) show for instance that the number of co-publications between international European regions is positively affected by joint participation of these regions in FP projects, even after controlling for prior co-publication activity.

Although these empirical results indeed suggest that ERA policy reduces fragmentation of scientific research activities across Europe, the findings have been qualified in a number of different ways. *First*, although there seems to be a tendency towards European integration, Europe's scientific landscape continues to consist mainly of a collection of regional and national research systems. This finding has been observed in studies using gravity models indicating that in the European context regional, national and language borders continue to have a large and independent negative effect on co-publication activity (Maggioni and Uberti 2009; Hoekman et al. 2009).

Second, although we observe that there is an increase in cross-border collaboration in the last decade, there is no evidence that the influence of physical proximity on structuring research collaborations is decreasing over time. This result may be surprising as we would expect internationalisation to go hand in hand with a decreasing effect of distance. However, although ERA policy is effective in reducing the importance of national borders, researchers continue to orient themselves mainly towards physically proximate, but possibly cross-territorial, partners. This observation particularly holds for the new member states of the European Union, which are rapidly catching up in scientific activity (Hoekman et al. 2010).

Third, when using co-publications to compare the growth of international scientific research collaboration between EU member states and non-EU OECD members, Chessa et al. (2013) observes that ever since 2003, international research collaboration within the EU is not growing at a more rapid pace than international collaboration between other OECD members. This result raises doubt on the extent to which ERA policy is effective in stimulating cross-border research collaboration and suggest that (part of) internationalization of science should be rather explained by a more general process of time-space compression following from mobility and ICT advancements.

14.4 Logic of Stratification

Science is a stratified institution as evidenced by the observation that "*power and resources are concentrated in the hands of a relatively small minority*" (Cole and Cole 1972, p. 368). Expressed in quantitative terms, the productivity of scientists follows a rank-size distribution where there are only a few scientists with very high productivity and many with low productivity (Price 1963; Stephan 2012). According to the sociology of science the unequal distribution of productivity reflects itself in the reward system that gives credit where credit is due, therefore effectively providing productive researchers with more recognition (Merton 1973). Recognition is not an isolated property based on past achievement alone. Rather, it is part of a cumulative cycle of conversion that conditions "scientist's abilities actually to do science" (Latour and Woolgar 1986, p. 198). Within this cycle, recognition can be transformed in instrumental assets such as money, equipment and data. Researchers 'invest' in these assets to produce new scientific knowledge with the intention of 'earning back' recognition with 'interest' after a complete cycle (Hessels 2010). Positive feedback mechanisms exist in the system and they may increase the interest rate to investments based on already established reputations of researchers (Merton 1973; Stephan 2012). Such positive feedback mechanisms are for instance observed when studying the attribution of reward as visible in scientific citations (Peterson et al. 2010)

Research collaboration is also a way to gain and sustain recognition as collaboration provides access to resources such as research infrastructure, information and training. Moreover, collaboration creates networks through which scientific knowledge and researchers' own reputation diffuses (Beaver and Rosen 1978). In doing so, the embeddedness of researchers in networks is a medium to mobilise 'allies' and to convince peers about the significance of research results (Latour 1987).

The structure of research collaboration that follows from this reward system can be considered an emergent, self-organising system insofar the selection of research partners is based upon choices made by researchers themselves, irrespective of their locations (Wagner and Leydesdorff 2005). However, these 'footloose' choices can only be made when researchers have the resources to organise the settings of interaction that are necessary for successful coordination of research collaboration in the first place. In this respect, the unequal distribution of rewards and its reinforcement through positive feedback mechanisms makes some researchers more footloose than others because it provides researchers with differential means to access mobility technology and ICTs. It follows that physical distance becomes relatively less of a concern for researchers with higher reputation as they have the resources to organise moments of co-presence on a temporal base. Reputable researchers are also more attractive collaborators and as a result other researchers have a higher preference to be co-present with them (both for training and collaboration).

Given this logic of stratification, the structure of scientific collaboration networks follows a 'preferential attachment' process which is especially observed for early-career researchers and young talent (e.g. Ph.D.s, post-docs). The minority of researchers with a high reputation are in this case like 'magnets' for the (yet) less reputable ones which makes it relatively easy for the former to hire new personnel, potentially over large distances. Mahroum (2000, p. 372 and p. 376) notes in this respect that "*mobility is a premier agent of scientific expansion [where] highly talented scientists flow to scientific institutions that are reputed for their excellence*". In this collaboration structure, reputable researchers have the means to collaborate over large distances and they can also organize the conditions to efficiently collaborate with peers in close vicinity. In contrast, it is expected that less reputable researchers will collaborate in closer proximity to their permanent location and that they choose more often to move on a permanent base to another location where they can be co-present with more reputable researchers.

14.4.1 ERA and Stratification

ERA policy on the stratification of scientific research collaboration networks starts from the observation that current research activities are already unevenly spatially distributed in Europe, even more so than economic activity (see Frenken et al. 2007; Matthiessen et al. 2010). Figures 14.1 and 14.2 show for instance that scientific publication output as well as scientific publication output per capita is concentrated in a group of 'core' regions located in a Western European axis stretching south-east from London towards Rome, in Scandinavian regions and in some large city-regions located in other parts of Europe (e.g. Berlin, Budapest, Glasgow/Edinburgh, Madrid, Vienna). This spatial pattern of knowledge activities was already present before the initiation of ERA policy (Moreno et al. 2005; Crescenzi et al. 2007).

A key question of ERA policy is whether it should further support those agglomerative tendencies by strengthening collaborations network between these agglomerations or whether it should allocate funding to peripheral actors as to provide these actors with opportunities to connect to the already established core of knowledge producing actors. The trade-off is reminiscent of the more general tension between EC's cohesion policy and research policy which both constitute significant shares of the European budget. In this respect, place-based cohesion policy provides resources to Europe's poorest regions. A major goal of these efforts is to strengthen the scientific capabilities of these cohesion regions through various instruments focused on scientific research infrastructures, network development and knowledge transfer (Musyck and Reid 2007; Begg 2010). In doing so, cohesion policy intends to support structural conditions that facilitate participation in ERA.

However, the actual participation rates in ERA is dependent on the allocation of research funds which is determined by EC's research policy rather than cohesion policy. As a result there have been worries that over the subsequent FPs the

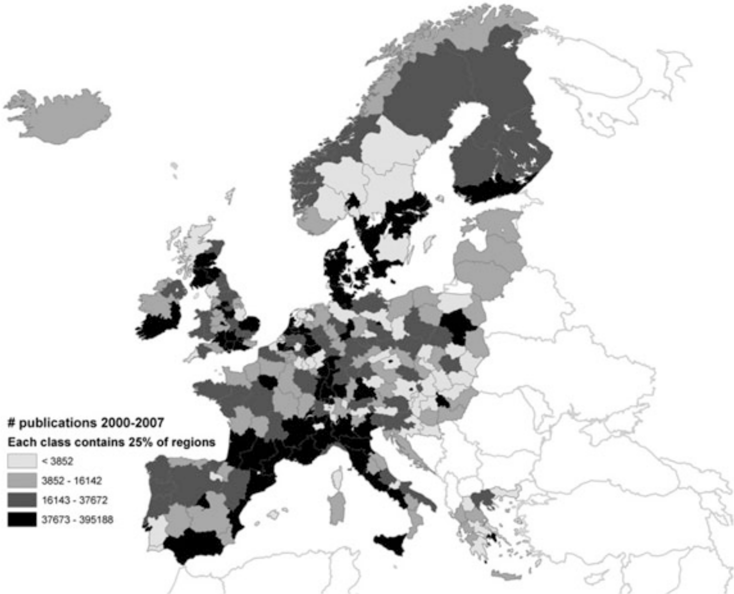


Fig. 14.1 Total number of publications in 2000–2007

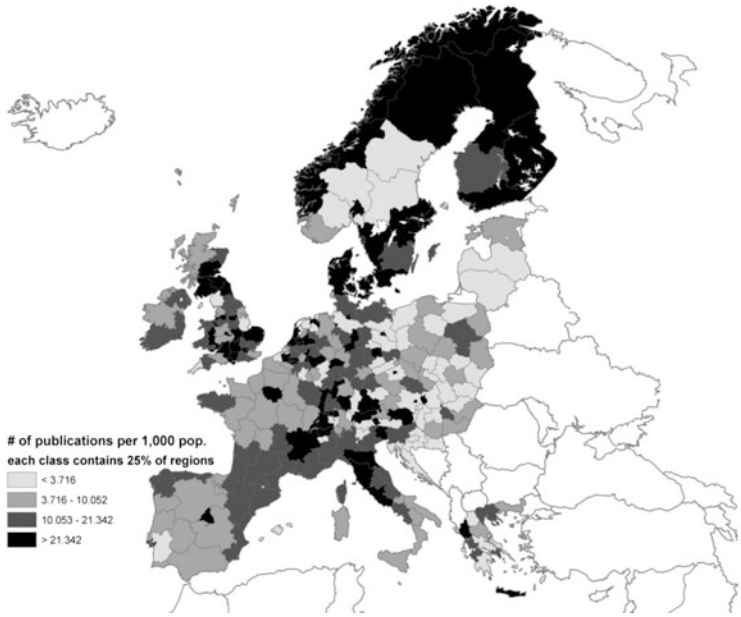


Fig. 14.2 Total number of publications per capita in 2000–2007

rationales and goals of the FPs have changed to such an extent that the cohesion objective no longer plays a role in the selection of FP projects that are being funded (Sharp 1998; Breschi and Malerba 2009). Rather ERA policy is increasingly focused on stimulating '*virtual centres of excellence*' (Commission 2007, p. 15) that strive to maximize the research potential of the European territory as a whole. It follows that over the successive FPs, funding has become increasingly based on criteria of research quality (i.e. scientific excellence), socio-economic relevance (i.e. tackling societal challenges and innovation potential) and critical mass, rather than on a redistribution criterion.

Turning to the empirical evidence on allocation of funding, Sharp (1998) found that funding in FP3 and FP4 favoured core regions only in absolute terms which was expected given the sheer number of researchers in these regions. Yet, after controlling for size peripheral countries managed to acquire more funding relative to their total research capacities. This finding was in line with the redistribution objective of the FPs at that time which treated proposals that included researchers from less developed regions as more favourable. In a more recent analysis, Hoekman et al. (2013) did not find the same result for FP5 and FP6; instead, they even observed that allocation of FP funds marginally increases with prior co-publication activity. Compared to earlier FP funding this finding can be interpreted as a move towards excellence, although the observed effect of 'excellence' funding remains is limited as of yet.

Part of the success of policy efforts to create more cohesive collaboration structures also depends on the extent to which funding is allocated to already established performers in terms of scientific collaboration networks. A number of empirical studies show in this respect that – similar to existing scientific research collaboration structures – the number of links between organisations in FP projects tends to decay with geographical distance and language barriers (Scherngell and Barber 2009, 2011; Maggioni and Uberti 2009), although these effects become less important over the successive FPs (Scherngell and Lata 2012). Importantly concerning stratification, it seems difficult for unconnected actors to acquire a central position in the FP funding networks. Breschi and Cusmano (2004), Autant-Bernard et al. (2007) and Wanzenböck et al. (2012) analyse the social network structures among FP participants and find that the funded collaboration networks are dominated by a small '*oligarchic core*' (Breschi and Cusmano 2004, p. 748) of research actors, whose central network positions in the programme have only strengthened over the successive funding rounds. This implies that participants are much more likely to acquire FP funding when they were already participating in previous FPs (Paier and Scherngell 2011), and that peripheral participants experience difficulties to enter the FP networks.

Turning to the effect of FP funding on the geography of scientific research collaboration, Hoekman et al. (2013) found that the effect of the FPs on raising co-publication output decreases when funding is allocated to regional pairs with already established scientific collaboration networks. This suggest that the FPs are more effective in establishing ties between poorly connected regions than in further strengthening existing ties between core regions. They conclude on the base of these findings that the effect of funding on raising co-publication output seems strongest in poorly performing regions, despite the fact that more resources flow to well performing regions.

14.5 Conclusion

Scientific research collaboration across territories is believed to be beneficial for the production and diffusion of scientific knowledge. However, long distance collaboration is still significantly hampered by the dominance of localised interactions within national and regional systems and by agglomeration dynamics that put a prime on face-to-face contact. Against this background, the efforts of the European Commission to create a European Research Area (ERA) is a significant attempt to overcome fragmentation and to increase excellence in the European scientific research system.

This chapter introduced proximity and stratification as two organizing principles that can be used to understand the geographical structure of European scientific collaboration networks. Concerning proximity we noted that despite efforts to integrate scientific research activities across borders, Europe remains a loosely connected group of national and regional science systems. With respect to stratification we concluded that there exists a tendency, even if small, towards excellence in funding, but that the effect of this policy in terms of raising cross-border scientific research collaborations remains questionable.

Given the substantial resources that have been spent on realizing ERA since its inception in 2000, it can be argued that this empirical reality is contrary to expectations. A more detailed assessment of the reasons for the observed lack of geographical effect of ERA policy seems therefore warranted. One straightforward explanation may be that despite substantial funding, European research budgets remain a minor funding source when compared to with national and regional research budgets. If this is indeed the case we may expect more from the Horizon 2020 programme that shows an increase in funding over previous programmes and stresses the significance of excellence as evidenced amongst others by the expansion of the European Research Council.

Another reason for the persistence of physical proximity in scientific research collaboration may be that the nature of the problems being studied in scientific collaborations have become increasingly complex over time, necessitating equally frequent moments of co-presence, despite advancements in ICTs and mobility. Such an increase in complexity may follow from the internal dynamic of science where researchers create new forms of 'complementarities' between specialised fields of knowledge and heterogeneous groups of organisations (Bonaccorsi 2008). They may also be driven by external pressures of governments and society to come up with solutions to 'grand challenges' (Gibbons et al. 1994; Nowotny et al. 2001). The identification of such complex thematic goals in recent ERA policy documents is illustrative of this phenomenon.

In light of our findings, it seems essential to monitor and evaluate the geographical effects of ERA policy efforts in future studies. It is encouraging that evidence based evaluation is slowly becoming an important pillar of EC's research policy. At the same time, it should be realized that impact assessments based on bibliometric data provide only a partial window on the state and dynamics of ERA. Despite the validity of these indicators in the scientific domain, our conclusions should be interpreted with caution when extrapolating to other contexts. Yet, when treated with appropriate caution such data can provide useful comparative empirical evidence that is both compelling and politically informative.

References

Amin A, Roberts J (2008) Knowing in action: beyond communities of practice. Res Policy 37:353–369

Autant-Bernard C, Billand P, Frachisse P, Massard N (2007) Social distance versus spatial distance in R&D cooperation: empirical evidence from European collaboration choices in micro and nanotechnologies. Pap Reg Sci 86:495–519

Bathelt H, Malmberg A, Maskell P (2004) Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. Prog Hum Geogr 28:31–56

Beaver DB, Rosen R (1978) Studies in scientific collaboration. Part 1: the professional origins of scientific co-authorship. Scientometr 1:65–84

Begg I (2010) Cohesion or confusion: a policy searching for objectives. J Eur Integr 32:77–96

Bonaccorsi A (2008) Search regimes and the industrial dynamics of science. Minerva 46:285–315

Boschma R (2005) Proximity and innovation: a critical assessment. Reg Stud 39:61–74

Breschi S, Cusmano L (2004) Unveiling the texture of a European Research Area: emergence of oligarchic networks under EU Framework Programmes. Int J Technol Manage 27:747–772

Breschi S, Lissoni F (2009) Mobility of skilled workers and co-invention networks: an anatomy of localised knowledge flows. J Econ Geogr 9:439–468

Breschi S, Malerba F (2009) ERA and the role of networks. In: de Langhe H, Muldur U, Soete L (eds) European science and technology policy: towards integration or fragmentation? Edward Elgar, Cheltenham, pp 160–174

Bristow G (2005) Everyone's a 'winner': problematising the discourse of regional competitiveness. J Econ Geogr 5:285–304

Castells M (1996) The rise of the network society, vol 1, The information age: economy, society and culture. Blackwell, Oxford

Chessa A, Morescalchi A, Pammolli O, Penner O, Petersen AM, Riccaboni M (2013) Is Europe evolving towards an integrated research area. Science 339:650–651

Cole S, Cole R (1972) The Ortega hypothesis. Science 178:368–375

Collins HM (1985) Changing order: replication and induction in scientific practice. The University of Chicago Press, Chicago

Collins HM (2001) Tacit knowledge, trust and the Q of Sapphire. Soc Stud Sci 31:71–85

Commission of the European Communities (2002) The European Research Area. Providing new momentum. COM (2002) 565 final, Brussels, 16 Oct 2002

Commission of the European Communities (2007) Green paper 'The European Research Area: New Perspectives'. {SEC(2007) 412}, COM(2007)161 final, Brussels, 4 Apr 2007

Commission of the European Communities (2013) One trillion euro to invest in Europe's future – the EU's budget framework 2014–2020. COM (2013) IP/13/1096, Brussels, 19 Nov 2013

Crawford E, Shin T, Sörlin S (1993) The nationalization and denationalization of the sciences: an introductory essay. In: Crawford E, Shin T, Sörlin S (eds) Denationalizing science: the contexts of international scientific practice. Kluwer Academic Publishers, Dordrecht, pp 1–40

Crescenzi R, Rodriguez-Pose A, Storper M (2007) The territorial dynamics of innovation: a Europe-United States comprative analysis. J Econ Geogr 7:673–709

Cummings JN, Kiesler S (2007) Coordination costs and project outcomes in multi-university collaborations. Res Policy 36:1620–1634

de Price DS (1963) Little science, big science. Columbia University Press, New York/London

European Council (2008) Presidency conclusions, Brussels European Council, 13–14 Mar 2008

Frenken K, Hoekman J, Van Oort F (2007) Towards a European Research Area. RPB/NAi Publishers, The Hague/Rotterdam

Frenken K, Hardeman S, Hoekman J (2009) Spatial scientometrics: towards a cumulative research program. J Informetr 3:222–232

Gertler MS (2003) Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). J Econ Geogr 3:75–99

Gibbons M, Limoges C, Nowotny H, Schwartzman S, Scott P, Trow M (1994) The new production of knowledge: the dynamics of science and research in contemporary societies. Sage, London

Giddens A (1984) The constitution of society. Polity Press, Cambridge, UK

Hägerstrand T (1970) What about people in regional science. Pap Reg Sci 24:6–21

Harvey D (1990) The condition of postmodernity. Basil Blackwell, Cambridge, MA/Oxford, UK

Hessels L (2010) Science and the struggle for relevance. Ph.D. thesis, Universiteit Utrecht, Utrecht

Hoekman J, Frenken K, Van Oort F (2009) The geography of collaborative knowledge production in Europe. Ann Reg Sci 43:721–738

Hoekman J, Frenken K, Tijssen R (2010) Research collaboration at a distance: changing spatial patterns of research collaboration in Europe. Res Policy 39:662–673

Hoekman J, Scherngell T, Frenken K, Tijssen R (2013) Acquisition of European research funds and its effect on international scientific collaboration. J Econ Geogr 13:23–52

Howells JM (2002) Tacit knowledge, innovation and economic geography. Urban Stud 39:871–884

Janelle DG (1969) Spatial organization: a model and concept. Ann Association Am Geogr 59:348–364

Katz JS, Martin BR (1997) What is research collaboration? Res Policy 26:1–18

Latour B (1987) Science in action: how to follow scientists and engineers through society. Harvard University Press, Cambridge

Latour B, Woolgar S (1986) Laboratory life: the construction of scientific facts. Princeton University Press, Princeton

Lefebvre H (1991) The production of space. Basil Blackwell, Oxford

Lundvall BA (1988) Innovation as an intetractive process: from user-produce interaction to the national system of innovation. In: Dosi G, Freeman C, Nelson RR, Silverberg G, Soete L (eds) Technical change and economic theory. Pinter, London, pp 349–369

Maggioni MA, Uberti TE (2009) Knowledge networks across Europe: which distance matters? Ann Reg Sci 43:691–720

Mahroum S (2000) Scientific mobility: an agent of scientific expansion and institutional empowerment. Sci Commun 21:367–378

Massey D (2004) For space. Blackwell, London

Matthiessen CW, Schwarz AW, Find S (2010) World cities of scientific knowledge: systems, networks and potential dynamics. An analysis based on bibliometric indicators. Urban Stud 47:1879–1897

Mattson P, Laget P, Nillson A, Sundberg CJ (2008) Intra-EU vs. extra-EU scientific co-publication patterns in EU. Scientometrics 75:555–574

Merton R (1973) The sociology of science: theoretical and empirical investigations. The University of Chicago Press, Chicago

Milgram S, Jodelet D (1976) Psychological maps of Paris. In: Proshansky HM, Ittelson WH, Rivlin LG (eds) Environmental psychology, 2nd edn. Holt, Rinehart and Winston, New York, pp 104–124

Moreno R, Paci R, Usai S (2005) Geographical and sector clusters of innovation in Europe. Ann Reg Sci 39:715–739

Musyck B, Reid A (2007) Innovation and regional development: do European structural funds make a difference? Eur Plann Stud 15:961–983

North D (1990) Institutions, institutional change, and economic performance. Cambridge University Press, Cambridge, MA

Nowotny H, Scott P, Gibbons M (2001) Re-thinking science. Knowledge and the public in an age of uncertainty. Reino Unido Polity Press, Cambridge, MA

Olsen GM, Olsen JS (2000) Distance matters. Human–computer. Interaction 15:139–178

Paier M, Scherngell T (2011) Determinants of collaboration in European R&D networks: empirical evidence from a discrete choice model. Ind Innov 18:89–104

Peterson GJ, Presse S, Dill KA (2010) Non-universal power law scaling in the probability distribution of scientific citations. Proc Natl Acad Sci 107:16023–16027

Polanyi M (1958) Personal knowledge. Routledge/Kegan Paul, London

Ponds R, van Oort FG, Frenken K (2007) The geographical and institutional proximity of research collaboration. Pap Reg Sci 86:423–443

Scherngell T, Barber M (2009) Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU Framework Programme. Pap Reg Sci 88:531–546

Scherngell T, Barber M (2011) Distinct spatial characteristics of industrial and public research collaborations: evidence from the 5th EU Framework Programme. Ann Reg Sci 46:247–266

Scherngell T, Lata R (2012) Towards an integrated European Research Area? Findings from eigenvector spatially filtered interaction models using European Framework Programme data. Pap Reg Sci. doi:10.1111/j.1435-5957.2012.00419.x

Shapin S (1995) Cordelia's love: credibility and the social studies of science. Perspect Sci 3:255–275

Sharp M (1998) Competitiveness and cohesion. Are the two compatible? Res Policy 27:569–588

Simmel G (1997) Sociology of senses. In: Frisby D, Featherstone M (eds) Simmel on culture. Sage, London, pp 109–120

Stephan P (2012) How economics shapes science. United States, Cambridge, MA

Storper M, Venables T (2004) Buzz: face-to-face contact and the urban economy. J Econ Geogr 4:351–370

Torre A, Rallet A (2005) Proximity and localization. Reg Stud 39:47–51

Urry J (2000) Sociology beyond societies: mobilities for the twenty-first century. Routledge, London

Wagner C, Leydesdorff L (2005) Network structure, self-organisation and the growth of international collaboration in science. Res Policy 34:1608–1618

Wanzenböck I, Scherngell T, Lata R (2012) Embeddedness of European regions in EU funded R&D networks: a spatial econometric perspective. Paper presented at the first Eurolio European Seminar on Geography of Innovation, 26–28 Jan 2012, Congress Center, Saint-Etienne

Chapter 15
The Embeddedness of Regions in R&D Collaboration Networks of the EU Framework Programmes

Iris Wanzenböck and Barbara Heller-Schuh

Abstract This article focuses on the embeddedness of European regions in networks of R&D collaborations of the EU Framework Programmes. Network embeddedness is defined in terms of network analytic centrality measures calculated for organisations and aggregated to the regional level. The objective is to estimate how region-internal and region-external characteristics affect a region's positioning in the thematic networks of Information and Communication Technologies, Sustainable Development and Life Sciences. In our modelling approach, we employ panel spatial Durbin error models, linking a region's centrality in the network to knowledge production and general economic characteristics of regions, and their neighbours, respectively. We found evidence that financial R&D resources, human capital and the level of socio-economic development are important general determinants of a region's network positioning. By linking European R&D networks with regional innovativeness, the study provides important implications for setting priorities in a regional innovation policy context.

15.1 Introduction

In the recent past, Research, Innovation and Technology (RTI) policies have emphasised supporting interactions and networks between organisations of the innovation system. At the EU level, the key policy instruments are the EU Framework Programmes (FPs) that support pre-competitive R&D projects, creating a pan-European network of organisations performing joint R&D. This policy focus is based on theoretical and empirical literature on the economics of innovation that emphasises two arguments in this respect: *First*, interactions, research collaborations and networks of actors are crucial for successful innovation (Powell and

I. Wanzenböck (✉) • B. Heller-Schuh
Innovation Systems Department, AIT Austrian Institute of Technology, Donau-City-Str. 1, Vienna 1220, Austria
e-mail: iris.wanzenboeck@ait.ac.at

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_15,

Grodal 2005; Fischer 2001), and, *second*, innovation and knowledge diffusion are the key vehicles for sustainable economic competitiveness (Romer 1990).

Furthermore, the close relationship between organisational innovativeness and the regional environment in which organisations are locally embedded has been emphasised in Regional Science and Economic Geography (Asheim and Gertler 2005). Recent literature streams in these fields lay particular emphasis on long-distance R&D alliances, highlighting their role for localized knowledge production structures (Asheim et al. 2011; Boschma and Frenken 2010). This is also central to the idea of this study, assuming that participation of organisations in inter-regional R&D networks enhances not only the organisations own innovation capability, but also indirectly – due to the presence of geographically localised knowledge spillovers (Breschi and Lissoni 2009; Karlsson and Manduchi 2001) – spurs innovativeness of the entire regional innovation system (Cooke et al. 1997).

The focus of the study is on regional characteristics that affect embeddedness of European regions in R&D networks as captured by the participation in joint R&D projects. Network embeddedness is defined in terms of centrality as applied in the Social Network Analysis (SNA) literature calculated for organisations and aggregated to the regional level. It is assumed that vertices showing a more central network position more likely benefit from network advantages in terms of preferential information and knowledge access within the network (Borgatti 2005; Wasserman and Faust 1994). Since such a privileged network position of organisations may be beneficial for the entire region, it is a crucial task for regional policy to provide framework conditions that stimulate participation intensity of organisations in inter-regional R&D networks. Thus, we aim to identify region-specific characteristics that influence a region's embeddedness in European R&D networks, such as knowledge production capacities, technology-related conditions, agglomeration effects and economic structure. Further, we consider the influence of the characteristics of neighbouring regions referred to as inter-regional spatial spillovers (Fischer et al. 2009).

In this study we apply an innovative approach, representing the network structure at the organisational level by the means of a bipartite graph, while at the same time keep the regional focus that is our relevant unit of analysis in our panel spatial Durbin error modelling (SDEM) approach (Le Sage and Pace 2009; Elhorst 2003). We construct the thematic R&D collaboration networks for 241 NUTS-2 regions of the EU-25 countries for the years 1998–2006 using data on joint FP projects in the thematic fields of Information and Communication Technologies, Life Sciences and Sustainable Development. Individual analysis of thematic R&D networks allows us to explicitly account for peculiarities in the way of performing R&D.

The remainder of the study is organised as follows. Section 15.2 sets forth the theoretical background and derives the main hypotheses for the empirical analysis. Section 15.3 clarifies the notion of network embeddedness, and outlines our measurement approach to empirically observe the network positioning of a region from a SNA perspective. Section 15.4 formalises the panel version of the SDEM approach and introduces our set of regional characteristics, before Sect. 15.5 presents the estimation results. Section 15.6 concludes with a summary of the main results and some policy implications.

15.2 A Regional Perspective on R&D Network Embeddedness

Today it is widely agreed that joint R&D activities, networks and collaborations are conducive to knowledge production and successful innovation (Powell and Grodal 2005). Firms, universities and research organisations increasingly search for external knowledge sources in order to keep pace in the global competition on ideas, technological developments and innovative products. In this regard, R&D networks have been increasingly recognized as efficient approach to access external, often spatially distant, knowledge in a rapid and targeted way (Hoekman et al. 2009).

The spatial structure of R&D networks has gained particular interest in recent scholarly research on the geography of innovation (see, for instance, Scherngell and Lata 2013; Autant-Bernard et al. 2007). In essence, this is based on observations that R&D capabilities tend to be spatially concentrated in a certain regional environment (Asheim et al. 2011), and that locally embedded actors frequently search for knowledge sources located in more distant geographical spaces to refine their own knowledge base with very specific knowledge components (Scherngell and Barber 2009; Maggioni et al. 2007). To access new knowledge that is not available within spatial proximity, innovative actors more and more rely on longer-distance, cross-regional R&D collaborations, for instance in the form of joint R&D projects, or joint assignment of patents or co-publications. Often there are specific key players – universities, research organisations, large knowledge-intensive or small highly specialised firms – that are the important driving forces in establishing and maintaining such inter-regional links of knowledge transmission (Morrison 2008).

One basic theoretical assumption in this context is that these key players act as levers for knowledge diffusion to the regional environment. Knowledge gained by such inter-regional network channels may be injected to intra-regional knowledge diffusion channels, and, by this, enhance the general knowledge transmission dynamics within the regional system (Giuliani 2007; Bathelt et al. 2004). In this sense, it is assumed that the innovativeness of a region depends not only on internal conditions for knowledge production and diffusion, but also on the ability of its actors to identify and quickly access region-external knowledge sources, and by this, on their ability to participate in inter-regional R&D networks.

Following network theoretical considerations, not only being part in a network but the strategic positioning is essential to reap full benefits of networking connections, in Social Network Analysis (SNA) referred to as centrality or prestige of an actor (Wasserman and Faust 1994). For R&D networks, this implies that organisations involved in several collaborative arrangements are well interlinked to other organisations, show short pathways to diverse sets of nodes, and therefore, take up a central position within the whole knowledge network (Borgatti 2005). They act as hubs or gatekeepers for knowledge diffusion, spreading knowledge throughout several actors in the entire network. In this sense, not only the organisations own knowledge bases, but also the knowledge to which the respective organisation has direct or indirect access through its network links determines its prestige in the network (Powell et al. 2005).

However, the way in which R&D network links are established depends on the technological regimes and prevailing research strategies in a specific field (Gilsing et al. 2008) and differs due to the geographical, technological or institutional background of the collaborating actors (Paier and Scherngell 2011; Boschma and Frenken 2010; Scherngell and Barber 2011; Ponds et al. 2007). Beyond the dyadic level, studies at the network level reveal remarkable variations in the structural properties of R&D networks across research areas and thematic priorities (see, for instance, Barber et al. 2011; Heller-Schuh et al. 2011). We therefore assume that specific characteristics of research processes in different thematic fields influence the composition of the network structure, and in turn, the positioning of actors in the network.

In this study we focus on the centrality of actors across regions in distinct thematic networks of R&D cooperation constituted under the heading of the EU FPs. The FPs create a pan-European network of actors performing joint R&D by supporting pre-competitive R&D projects (Breschi and Malerba 2009). A central embeddedness of these actors in FP networks may ease the establishment of contacts to strategic important region-external knowledge sources that might – given the theoretical considerations above – stimulate region-internal dynamics and knowledge diffusion. From a Regional Science perspective, the question comes to mind which local and global conditions drive the centrality of organisations located in one region, and, by this, the overall regional visibility in FP networks of a specific thematic field. Basically, a central position in R&D networks may be determined by the overall characteristics of the regional innovation environment, particularly reflected in the interrelation of region-specific knowledge production capabilities, as well as economic and institutional conditions (Broekel and Brenner 2011; Fritsch and Slavtchev 2011; Rodriguez-Pose and Crescenzi 2008). In addition to region-internal factors that may drive a region's centrality in the thematic networks of R&D cooperation in Europe, we apply in our study a spatial perspective explicitly taking account of characteristics in neighbouring regions (Fischer et al. 2009; Fischer and Varga 2003). The following hypotheses are to be tested:

i. We assume that *knowledge production capacities* in terms of financial and human resources as well as a strong technological knowledge base are decisive for reaching a certain network centrality. They enable to explore, exploit and transfer knowledge from external sources, irrespective of the thematic field. We further assume that specialised regions that supply distinctive technologies have strategic advantages in thematic R&D networks, but the importance may vary depending on the overall technological orientation of the specific thematic field.
ii. Concerning the effect of *general economic conditions and agglomeration*, we assume that regions with a higher level of socio-economic development are more centrally embedded in FP networks. This might specifically apply to urban regions that host important scientific organisations and multinational firms that benefit from domestic industrial agglomeration effects and at the same time more intensively exploit external knowledge sources to keep path in global competition.

iii. Finally, we assume that a region's embeddedness in the European FP networks is affected by indirect effects in the form of *spatial spillovers from neighbouring regions*. Thus, we hypothesise that geographical space in terms of regional interaction effects does matter in that research-related and general economic conditions in neighbouring regions impact a region's network position.

15.3 Measuring Network Embeddedness of Regions

Social Network Analysis (SNA) provides a rich analytical toolset to observe network structures and to characterise the role of specific actors in the network, such as their centrality. In our study, network embeddedness refers to the centrality of *organisations* – located in a specific region – in the European R&D collaboration networks from a SNA perspective.

In our measurement approach, we account for the role of specific organisations in the network, explicitly making use of the *organisational* level to identify a region's network embeddedness in the European network of R&D cooperation. The centrality of a region $i = 1, \ldots, n$ may then be viewed as the sum of all centralities of organisations participating in the FPs that are located in that region i. In the following, we formalise this approach by defining our network and the centrality measures at the organisational level, and demonstrate how we come back to the regional level of analysis in the context of our main research question.

To observe our networks of R&D collaboration in Europe, we draw on data from the EUPRO database which provides comprehensive information on funded research projects of the EU FPs and all participating organisations. For the study at hand, we use data on projects running between 1998 and 2006 in three thematic fields that are *Information and Communication Technologies (ICT)*, *Life Sciences (LS)* and *Sustainable Development (SD)*.[1]

Our network is formalised by representing FP project collaborations at the organisational level as a bipartite graph $G(V = V_1 + V_2, E)$, letting V_1 be a set of vertices representing $u, v = 1, \ldots, m$ organisations participating in the FPs in year t, and V_2 be a set of $l = 1, \ldots, L$ vertices representing FP projects funded in the same

[1] In our definition of the distinct thematic areas of the FPs we basically follow the study of Hoekman et al. (2012). Our thematic priorities consist of the following programme lines in the distinct FP: FP4 programmes ENV2C, MAST3, JOULE and THERMIE, FP5-EESD, FP6-SUSTDEV for Sustainable Development; FP4-BIOTECH2, FP4-BIOMED2 and FP4-FAIR, FP5-Quality of Life, FP6-Food, FP6-LIFESCIHEALTH for Life Sciences; FP4-ACTS, FP4-ESPRIT4 and FP4-TELEMATICS 2C, FP5-IST, FP6-IST for the thematic priority ICT (Rietschel et al. 2009). The thematic areas we include make up 72.5 % of total funding in FP5, and 63.3 % of total funding in FP6 (Hoekman et al. 2012). Details on the network structure of the thematic FP networks are given in the Appendix.

year t, with an edge between two vertices if – and only if – one vertex is a project and the other is an organisation that takes part in the project, giving rise to the set of edges E. G is said to be *bipartite* when there are no edges between pairs of elements within V_1 or V_2.

Note that one might simply define edges between organisations when the organisations are separated by a path of length two in the bipartite graph to obtain a one-mode organisational representation. The topology of our bipartite graph may be encoded in the $m \times m$ adjacency matrix $\boldsymbol{A}$ for a given year t by

$$\mathbf{A}_t(u,v) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix} \quad u,v = 1,\ldots,m \tag{15.1}$$

where the element a_{uv} contains the collaboration intensity as measured in terms of joint FP projects between organisations u and v. Furthermore, the number of edges incident on a vertex u is called the degree k_u.

In our study, network embeddedness of organisation u – and in a second step of region i (see below) – is captured by two distinct centrality measures, namely *betweenness* and *eigenvector centrality*.[2] The betweenness concept captures the centrality of a node in terms of its position for controlling the flow of information within the network (Freeman 1979). Thus, central organisations benefit from gaining access to various knowledge sources, and, at the same time, take up – independent of their degree – a significant position in influencing the transfer of knowledge within the whole network. In other words, they act as 'gatekeepers' by exerting control over the knowledge flowing through them. Mathematically, betweenness centrality g_{ut} of organisation u for a given year t measures how often organisation u is situated between other not directly interlinked organisations in time period t, as defined by

$$g_{ut} = \sum_{\substack{v=1 \\ v<q}}^{m} d_{vqt}(u)/d_{vqt} \tag{15.2}$$

where $d_{vqt}(u)$ is the shortest path[3] between organisations v and q going through organisation u at time t, for $u \neq v \neq q$.

[2] Further centrality measures commonly used in SNA are degree and closeness centrality. Degree centrality focuses on connections directly attached to a vertex; closeness centrality indicates how close a distinct vertex is to all other vertices in the network (Faust 1997).

[3] A path is the alternating sequence of vertices and links, beginning and ending with a vertex, so that the shortest path or geodesic distance d_{uvt} between two organisations u and v in time period t. is defined as the number of vertices to be passed in the shortest possible path from one vertex to another (see Wasserman and Faust 1994 for further details).

We additionally investigate the eigenvector centrality by examining all organisations in parallel and assigning centrality weights that correspond to the average degree of all linked organisations (Bonacich 1987). The concept lays emphasis on the importance of direct linkages of a vertex in the network, but additionally takes the degree of all other connected vertices into account. Eigenvector centrality r_{ut} of organisation u at time t is defined to be proportional to the sum of degrees of organisations to which it is connected[4]:

$$r_{ut} = \frac{1}{\lambda} \sum_{v=1}^{m} a_{uvt} k_{vt} \tag{15.3}$$

where λ is the largest eigenvalue of $\mathbf{A}_t$.[5]

Since we are interested in a region's centrality in the European networks of R&D collaboration, and how different regional characteristics affect this centrality, we aggregate our organisational centralities to the regional level. For this reason, we sum up all normalised centralities of organisations that are located in region $i = 1, \ldots, n$, so that the region-specific centrality is to be derived by

$$y_{it}^{(b)} = \sum_{u=1}^{m} g_{iut} \tag{15.4}$$

$$y_{it}^{(ei)} = \sum_{u=1}^{m} r_{iut} \tag{15.5}$$

where $y_{it}^{(b)}$ and $y_{it}^{(ei)}$ is the betweenness and eigenvector centrality of region i at time t, respectively. By this, we are presenting an innovative approach, using SNA measures to characterise regional embeddedness in networks, but taking the network structure at the organisational level as starting point.

15.4 Modelling Regional Network Embeddedness

We seek to measure how different region-internal and region-external characteristics affect a region's embeddedness in thematic FP networks in Europe as measured by betweenness centrality $y_{it}^{(b)}$ or eigenvector centrality $y_{it}^{(ei)}$ defined by Eqs. 15.4

[4] For practical purposes, we draw on the adjacency matrix $\mathbf{A}_t$ defined by Eq. 15.1 instead of the bipartite graph in our formal description.

[5] A common notation used in this context is the eigenvector equation as given by $\lambda\,\mathbf{x} = \mathbf{A}\,\mathbf{x}$, where $\mathbf{x}$ is a vector of centralities $\mathbf{x} = (x_1, x_2, \ldots.)$ denoting the eigenvector of the adjacency matrix $\mathbf{A}$ with eigenvalue λ (Bonacich 1987).

and 15.5, respectively.[6] Note that we calculate these centralities separately for each thematic priority but refrain from introducing an additional sector index for purposes of readability. Thus, for each thematic network, the situation we are considering is one of observations y_{it} (i, $j = 1, \ldots, n = 241$; $t = 1, \ldots, T = 9$) on stochastic variables, say Y_{it}, corresponding to the centrality in the FP network of region i at time t. We assume – based on our theoretical consideration in Sect. 15.2 – an outcome of y_{it} to be determined by the 1-by-Q row vector ($q = 1, \ldots, Q$) of variables $\boldsymbol{c}$ accounting for the *knowledge production capacity* of a region, and by the 1-by-S row vector ($s = 1, \ldots, S$) of variables $\boldsymbol{z}$ accounting for the *regional economic conditions and agglomeration effects*. We specify a space-time model that fits our dataset as follows:

$$\begin{aligned} y_{it} = \alpha + \boldsymbol{c}_{it}\beta^{(c)} + \sum_{j=1}^{n} w_{ij}\boldsymbol{c}_{jt}\boldsymbol{\gamma}^{(c)} + \boldsymbol{z}_{it}\beta^{(z)} \\ + \sum_{j=1}^{n} w_{ij}\boldsymbol{z}_{jt}\boldsymbol{\gamma}^{(z)} + \sum_{t=1}^{T} \boldsymbol{b}_t\tau_t + \mu_i + u_{it} \end{aligned} \tag{15.6}$$

with

$$u_{it} = \rho \sum_{j=1}^{n} w_{ij}u_{it} + \varepsilon_{it} \tag{15.7}$$

where α is a scalar parameter, $\boldsymbol{\beta}^{(c)}$ (Q-by-1) and $\boldsymbol{\beta}^{(z)}$ (S-by-1) are associated parameter vectors estimating the influence of the regional knowledge production capacity $\boldsymbol{c}_{it}$, and effects of the regional economic condition and agglomeration $\boldsymbol{z}_{it}$ for region i at time t. The parameter vector $\boldsymbol{\gamma}^{(c)}$ (Q-by-1) and $\boldsymbol{\gamma}^{(z)}$ (S-by-1) reflect spatially weighted exogenous interaction effects that are directly interpretable as local multipliers (Le Sage and Pace 2009). w_{ij} is an element of the non-stochastic, time-invariant n-by-n spatial weights matrix $\boldsymbol{W}$ describing the spatial arrangement of our set of n regions.[7] $\boldsymbol{b}_t$ is the n-by-1 vector controlling for time-specific idiosyncrasies, and τ_t being the associated scalar parameter, while μ_i denotes the region-specific effect accounting for omitted space-specific but time-invariant

[6] A descriptive analysis of our centrality measures as used in the spatial modelling approach are given in the Appendix.

[7] We define $w_{ij} = 1$ if $_i$ and $_j$ are spatial neighbours in the form that they are sharing a common border, and zero otherwise, with $w_{ii} = 0$. We use a row standardized version of $\boldsymbol{W}$ allowing interpretation of the spatial lags of the independent variables being the weighted average impact on region i by their neighbouring regions. For the SDEM, interpretation of both direct and indirect effects is directly associated with the parameter estimates as opposed to specifications that contain spatial lags of the dependent variable, such as the SDM (Le Sage and Pace 2009).

variables, following a random effects specification; ρ is the spatial autocorrelation coefficient and ε_{it} is the IID error term. Equation 15.6 with error specification Eq. 15.7 is referred to as the panel version of the SDEM that controls for a bias due to omitted spatially autocorrelated independent variables while at the same time allows to estimate spatial spillovers (see Le Sage and Pace 2009 for further details on properties of the SDEM). We use Maximum Likelihood estimation procedures to estimate the parameters (see Elhorst 2003 for details; Baltagi 2008).

In modelling a region's network embeddedness we consider two different types of independent variables. The following $q = 1, \ldots, Q = 4$ variables $\boldsymbol{c}$ account for regional *knowledge production capacity*:

i. $c_{it}^{(1)}$ captures total regional R&D expenditures (log of public and private R&D expenditures in % of GRP), used as a proxy for the level of financial knowledge production inputs.
ii. $c_{it}^{(2)}$ is the logarithmic share of population with tertiary education (corresponding to levels 5 and 6 of the ISCED 1997 classification system), measuring a region's endowment with human capital.
iii. $c_{it}^{(3)}$ is the region's R&D activities in high-tech sectors measured by the number of high-tech patents per million employees and used in logarithmic form.
iv. $c_{it}^{(4)}$ captures the degree of technological specialisation within region i, using an index of specialisation of a region's patent portfolio.[8]

   Then we include $s = 1, \ldots, S = 3$ variables $\boldsymbol{z}$ accounting for the *regional economic conditions* and *agglomeration effects*.

v. $z_{it}^{(1)}$ is the degree of industrial diversity within region i measured in terms of an industrial diversity index.[9]
vi. $z_{it}^{(2)}$ is the logarithmic form of the gross regional product (GRP) per capita, proxying the general socio-economic potential and development of a region.
vii. $z_{it}^{(3)}$ denotes the region's population density as measured by the number of inhabitants per square kilometre, used as proxy variable for the degree of urbanisation, and in this context, for agglomeration effects.

[8] The index is defined by $c_{it}^{(4)} = \frac{1}{2}\sum_{p}\left|s_{ip} - \overline{s}_{p}\right|$ where s_{ip} is the region's i share of patents in a specific IPC class p and $\overline{s}_p$ is the mean of IPC class p. Patents were taken into account at a three-digit level corresponding to the International Patent Classification (IPC).

[9] We include five different main economic sectors, namely agriculture, manufacturing, construction, private services and non-market service sector. The index of specialisation to account for industrial diversity is defined as $z_{it}^{(1)} = \frac{1}{2}\sum_{p}\left|o_{ip} - \overline{o}_{p}\right|$ where o_{ip} is the region's i share of gross value added in a specific sector p (indexed $p = 1, \ldots., 5$) and $\overline{o}_p$ is the mean of sector p for $n = 241$ regions.

Data for most independent variables have been drawn from the Eurostat regional database, while information on patents was taken from the European Patent Office (EPO) database. Our sample comprises full data for 241 European NUTS-2 regions over the period 1998–2006.

15.5 Estimation Results

This section discusses the Maximum Likelihood (ML) estimates for our network embeddedness models as specified by Eqs. 15.6 and 15.7 in terms of betweenness centrality (BC) and eigenvector centrality (EC), estimated separately for each thematic priority under consideration. Table 15.1 provides details on the estimation results of the SDEM. Asymptotic standard errors are given in brackets, and time controls as well as various model diagnostics and goodness-of-fit measures are given at the bottom.

The results throughout all thematic areas show that endowment with *knowledge production capacities* is a crucial explanation factor for a region's centrality in the European network of R&D collaboration. Thus, we can confirm our hypothesis *i*). Interestingly, the findings provide evidence that different aspects are decisive in distinct thematic areas: We observe significantly positive estimates for our R&D expenditures variable ($\beta_1^{(c)}$) in all thematic fields, indicating that the higher the financial resources devoted to R&D, the higher is the ability of a region and its domestic organisations to reach centrality in R&D networks. However, parameter estimates for our human capital variable ($\beta_2^{(c)}$) are more diversified for the different model specifications. We find that a highly educated pool of labour is essential for eigenvector centrality in all thematic areas, and particularly in LS. In terms of betweenness centrality, a positive influence of human capital can only be confirmed for the LS but not for ICT and SD FP networks.

Moreover, quality of a region's technological knowledge base measured in terms of high-tech patents ($\beta_3^{(c)}$) enhances betweenness and eigenvector centrality in all thematic networks, except eigenvector centrality in the LS FP network. It seems that distinct motives and types of R&D collaborations in a specific thematic field (science-driven and university-led research activities in LS vs. more industry-based and application-oriented R&D collaborations in ICT and SD) could explain the variation in our results for knowledge production specific factors, which become even more obvious in our betweenness and eigenvector centrality model specifications.

Table 15.1 ML estimation results for the panel spatial Durbin error model (SDEM)

	ICT				Life Sciences				Sustainable Development			
	BC		EC		BC		EC		BC		EC	
Knowledge production capacities												
R&D expenditures $[\beta_1^{(c)}]$	0.961^{***}	(0.228)	0.653^{***}	(0.235)	0.525^{***}	(0.197)	1.090^{***}	(0.209)	0.377^{*}	(0.214)	0.912^{***}	(0.205)
Human capital $[\beta_2^{(c)}]$	0.582	(0.636)	1.052^{*}	(0.637)	0.985^{*}	(0.548)	2.088^{***}	(0.556)	0.414	(0.593)	1.558^{***}	(0.554)
High-tech patents $[\beta_3^{(c)}]$	0.039^{**}	(0.016)	0.077^{***}	(0.017)	0.027^{**}	(0.014)	0.023	(0.016)	0.035^{**}	(0.015)	0.052^{***}	(0.015)
Tech. spec. $[\beta_4^{(c)}]$	-0.197^{**}	(0.077)	−0.080	(0.082)	−0.105	(0.065)	-0.047^{***}	(0.074)	−0.096	(0.072)	0.144^{**}	(0.071)
Economic conditions and agglomeration												
Industrial diversity $[\beta_1^{(z)}]$	0.126^{***}	(0.043)	0.030	(0.042)	0.143^{***}	(0.039)	0.115^{***}	(0.036)	0.113^{***}	(0.041)	0.024	(0.037)
GRP p.c. $[\beta_2^{(z)}]$	3.777^{***}	(0.645)	3.507^{***}	(0.655)	4.631^{***}	(0.603)	2.406^{***}	(0.571)	3.774^{***}	(0.624)	2.426^{***}	(0.595)
Pop. density $[\beta_3^{(z)}]$	0.001^{***}	(0.000)	0.001	(0.000)	0.001^{**}	(0.000)	0.000	(0.000)	0.001^{**}	(0.000)	0.000	(0.000)
Spatial spillovers												
w. R&D exp. $[\gamma_1^{(c)}]$	-1.271^{***}	(0.466)	−0.723	(0.479)	-0.976^{**}	(0.447)	-1.017^{***}	(0.418)	−0.389	(0.457)	-0.990^{**}	(0.441)
w. Human capital $[\gamma_2^{(c)}]$	−1.146	(0.865)	−1.272	(0.874)	−1.265	(0.798)	-1.615^{**}	(0.760)	-2.227^{***}	(0.831)	-3.069^{***}	(0.791)
w. High-tech pat. $[\gamma_3^{(c)}]$	0.014	(0.046)	0.043	(0.050)	0.004	(0.039)	0.053	(0.046)	−0.034	(0.044)	−0.026	(0.044)
w. Tech. spec. $[\gamma_4^{(c)}]$	0.123	(0.150)	0.210	(0.161)	0.471^{***}	(0.132)	0.202	(0.146)	0.196	(0.143)	−0.103	(0.144)
w. Ind. diversity $[\gamma_1^{(z)}]$	-0.153^{**}	(0.066)	-0.127^{*}	(0.066)	-0.214^{***}	(0.063)	-0.139^{**}	(0.056)	-0.208^{***}	(0.064)	-0.162^{***}	(0.060)
w. GRP p.c. $[\gamma_2^{(z)}]$	0.731	(0.620)	0.855	(0.626)	0.792	(0.574)	1.058^{*}	(0.544)	1.534^{**}	(0.596)	2.237^{***}	(0.567)
w. Pop. density $[\gamma_3^{(z)}]$	−0.001	(0.001)	−0.001	(0.001)	−0.001	(0.001)	−0.001	(0.001)	-0.002^{**}	(0.001)	−0.001	(0.001)

(continued)

Table 15.1 (continued)

	ICT		Life Sciences		Sustainable Development	
	BC	EC	BC	EC	BC	EC
SAR coefficient (ρ)	0.086*** (0.029)	0.169*** (0.029)	0.238*** (0.027)	0.203*** (0.028)	0.169*** (0.030)	0.268*** (0.028)
Constant (α)	−50.357*** (5.577)	−47.947*** (5.884)	−64.277*** (5.653)	−40.305*** (5.201)	−57.581*** (5.463)	−45.662*** (5.609)
Model fit and diagnostics						
Log Likelihood	−5477.16	−5605.80	−5096.96	−5427.79	−5323.25	−5288.79
LR Test for SEM	23.06***	19.03***	48.40***	29.03***	32.31***	36.19***
BSK-Test	29.57***	25.09***	27.47***	21.41***	27.04***	23.40***

Estimations with random region-specific effect. Estimates for time dummies are not presented in the table. BSK-test is the Baltagi-Song-Koh Test, a conditional LM-Test for the existence of spatial error correlation assuming possible random effects as outlined by Baltagi et al. (2007)
***significant at the 0.01 significance level, **significant at the 0.05 significance level, *significant at the 0.1 significance level

Regarding the effects of a region's technological specialisation ($\beta_4^{(c)}$), we do not find a consistent pattern in our different model versions that might broadly explain network embeddedness in different FP networks. The results show that technological specialisation negatively influences a region's network embeddedness in terms of betweenness centrality only in ICT, while it has no statistically significant effect in the other two thematic areas. In terms of eigenvector centrality, we observe that technological specialisation even impedes a region's strategic network position in the LS network, but fosters embeddedness in the SD network. Against our general assumption of positive technological specialisation effects (hypothesis *i*), we cannot identify a homogenous pattern across the thematic fields under consideration.

In addition, we account for *economic conditions and regional agglomeration effects* in modelling a region's network embeddedness. In this regard, we found evidence that industrial diversity ($\beta_1^{(z)}$) in a distinct region fosters betweenness centrality in all thematic networks; for eigenvector centrality such a positive effect is observed only in the LS network. Concerning the influence of a region's general socio-economic potential measured in terms of its GRP per capita ($\beta_2^{(z)}$), we observe the highest positive effects for all model specifications, indicating that the stage of socio-economic development of a region is one of the most crucial regional factors that fosters a central position of the respective region in the European network of R&D collaboration. These findings confirm our hypothesis (*ii*). However, we further found that urbanisation per se is not a sufficient driving force for R&D network embeddedness of a region; the estimate $\beta_3^{(z)}$ for population density is significant for betweenness centrality in the SD network, but exerts only a marginal negative influence.

Furthermore, we assume that *spatial spillovers from neighbouring regions* influence a region's network embeddedness (hypothesis *iii*). This is confirmed as the results show that high knowledge production capacities of neighbouring regions (as indicated by the estimates for $\gamma_1^{(c)}$ and $\gamma_2^{(c)}$) considerably decrease a region's own network embeddedness. This might be explained by the fact that financial or human R&D resources are not freely available within regions, but first, limited in availability, and second, rather mobile as far as spatially proximate regions are concerned. Thus, a drain of crucial knowledge production inputs to more research-intensive and productive regions is conceivable, especially due to centre-periphery structures and agglomeration tendencies across adjacent regions. In contrast, we can observe positive indirect effects induced by the general socio-economic potential of neighbouring regions, reflected in the spatially lagged parameter estimates for GRP per capita ($\gamma_2^{(z)}$). The effects are particularly strong in the SD network for both centrality specifications, while the stage of socio-economic development of adjacent regions seems to have no influence on embeddedness in the ICT network. Furthermore, the results show that the industrial structure in adjacent regions has adverse effects compared to their region-internal counterparts; a high degree of industrial diversity in adjacent regions ($\gamma_1^{(z)}$) negatively affects a region's embeddedness in all three thematic R&D networks.

15.6 Concluding Remarks

R&D collaborations have attracted a great deal of attention in the scientific literature of economics and geography of innovation. In spatial terms, they constitute valuable means allowing organisations direct access to knowledge that is widely dispersed in geographic space. Thus, they have often been analysed from a dyadic or micro-level perspective. Only recently, the network perspective has gained importance in the analysis of R&D collaborations, highlighting the possibility to tap knowledge that diffuses through the entire network via indirect allies. From this view, it is assumed that actors are embedded in a web of direct and indirect ties where knowledge is transmitted. Thus, strategic positioning in such R&D networks is of central importance to reap full benefits of the entire network structure.

This study focused on the embeddedness of European regions in the network of R&D collaborations constituted under the heading of the EU Framework Programmes (FPs). We analysed R&D networks for three distinct thematic FP programmes from a regional perspective. The aim was to identify regional factors that impact a region's network position, i.e. its centrality in the European R&D collaboration network. We took a Social Network Analysis (SNA) perspective, defining a region's embeddedness as the aggregate of the centralities of each organisation located in that region. We further differentiated between eigenvector centrality, i.e. importance of an organisation in terms of connectedness to central hubs in the network, and betweenness centrality, i.e. an organisation's ability to access and control a diverse set of knowledge flows in the network. A panel version of the spatial Durbin error model was estimated to relate regional network embeddedness to a set of independent variables encompassing the knowledge production capacity and the general economic structure of a region as well as spatial spillovers of neighbouring regions.

We explicitly account for individual network structures and characteristics of the thematic FP programmes by estimating individual model versions for ICT, Life Sciences and Sustainable Development. The different model specifications point to striking similarities across thematic programmes and centrality concepts. We have identified a set of the most conducive region-internal and region-external factors for a region's embeddedness that spans across thematic networks of R&D collaboration in Europe:

First, high regional knowledge production capacities are crucial for reaching high visibility in the thematic networks of R&D collaborations across Europe. Knowledge-intensive regions holding high amounts of financial R&D inputs and human capital more likely participate in R&D collaborations, and thus, more likely become a core player in the European R&D collaboration landscape. However, knowledge production capacities of neighbouring regions decrease a region's own network centrality, possibly due to the presence of centre-periphery and agglomeration tendencies across nearby regions that are induced by limited financial and human R&D recourses. *Second*, highly developed regions that encompass a diversified industrial structure more likely gain an advantageous position in the European networks of R&D collaboration. In this regard, a region's network embeddedness is additionally driven by the economic strength of nearby regions, while negative indirect effects are induced by industrial diversity in neighbouring regions.

The question of how to gain an advantageous position in the European network of R&D collaboration may be of particular relevance for regional innovation policy, particularly in light of the necessity to maintain or enhance innovativeness of regions and their locally embedded actors. In this context, our study provides general evidence that it is a mixture of specific R&D strengths and general economic factors that is important for a central network embeddedness of regions. Increasing regional R&D inputs such as financial investments and human capital is therefore essential for less central regions to enhance visibility of their organisations in inter-regional R&D networks. Likewise, more ambitious framework conditions for R&D might reduce a drain of highly educated people to neighbouring regions, which otherwise would have negative consequences for a region's R&D network position. However, we do not find broad evidence in this study that regional specialisation on just a few technological core fields has an influence on the strategic positioning in long-distance R&D collaboration networks.

Acknowledgements We are grateful to Thomas Scherngell (AIT) for providing valuable research assistance. We thank Michael Barber (AIT) and two anonymous referees for comments on earlier versions of this study.

Appendix

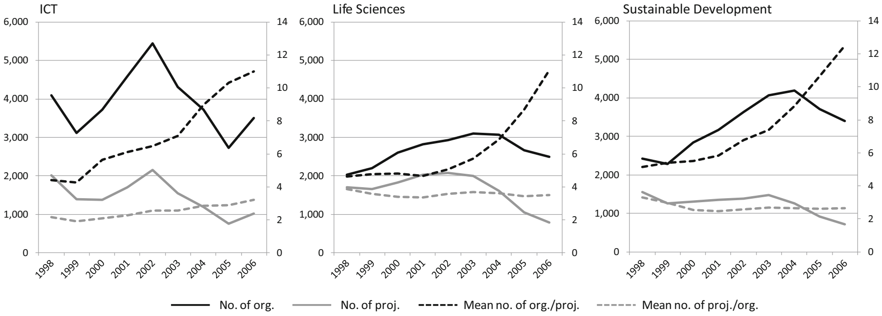


Fig. 15.1 Network characteristics by thematic field

Table 15.2 Descriptive statistics on regional centrality by thematic field

	Year	Mean	Median	Std.dev.	Skewness	Kurtosis
Information and communication technologies (ICT)						
Betweenness centrality	1998	0.008	0.000	0.024	6.112	49.728
	2006	0.007	0.001	0.024	8.120	75.972
Eigenvector centrality	1998	0.084	0.014	0.235	7.825	83.440
	2006	0.170	0.055	0.392	7.496	77.951
Life sciences (LS)						
Betweenness centrality	1998	0.007	0.001	0.017	5.102	30.620
	2006	0.006	0.001	0.018	6.759	58.872
Eigenvector centrality	1998	0.180	0.054	0.346	4.231	23.455
	2006	0.153	0.051	0.329	5.899	48.336
Sustainable development (SD)						
Betweenness centrality	1998	0.006	0.001	0.016	5.071	32.323
	2006	0.007	0.001	0.018	4.871	31.990
Eigenvector centrality	1998	0.127	0.034	0.245	3.902	20.540
	2006	0.298	0.102	0.559	5.212	39.997

Note: Regional centralities are considered as sum of organisations' centralities; betweenness and eigenvector centrality are normalised between zero and one at the organisational level

Table 15.3 Top-10 regions for betweenness and eigenvector centrality by thematic field (2006)

Eigenvector centrality

Region	ICT		Region	LS		Region	SD	
Île de France	4.713	(183)	Île de France	3.455	(131)	Île de France	5.684	(140)
Oberbayern	2.067	(69)	Inner London	2.118	(50)	Denmark	3.338	(111)
Madrid	1.218	(86)	Lombardia	1.375	(74)	Oberbayern	2.331	(69)
Attiki	1.212	(107)	Denmark	1.358	(94)	Lazio	2.244	(53)
Lombardia	1.100	(101)	Oberbayern	1.075	(44)	Madrid	1.639	(83)
Lazio	1.041	(80)	Stockholm	0.861	(37)	Attiki	1.577	(63)
Inner London	1.036	(77)	Cataluña	0.824	(54)	Etelä-Suomi	1.531	(53)
Köln	0.876	(60)	Zuid-Holland	0.816	(35)	Zuid-Holland	1.531	(81)
Stockholm	0.873	(36)	Madrid	0.779	(45)	Köln	1.356	(50)
Cataluña	0.792	(72)	Lazio	0.777	(46)	Inner London	1.336	(76)

Betweenness centrality

Region	ICT		Region	LS		Region	SD	
Oberbayern	0.249	(69)	Île de France	0.202	(131)	Île de France	0.165	(140)
Île de France	0.223	(183)	Inner London	0.108	(50)	Zuid-Holland	0.099	(81)
Madrid	0.078	(86)	Denmark	0.092	(94)	Attiki	0.082	(63)
Attiki	0.059	(107)	Gelderland	0.067	(20)	Denmark	0.063	(111)
Lombardia	0.049	(101)	Oberbayern	0.064	(44)	Lazio	0.061	(53)
Vlaams-Brabant	0.046	(23)	Stockholm	0.061	(37)	Stuttgart	0.058	(71)
Lazio	0.046	(80)	Madrid	0.056	(45)	Oberbayern	0.057	(69)
Etelä-Suomi	0.039	(65)	Lombardia	0.045	(74)	Lisboa	0.057	(41)
Stockholm	0.039	(36)	Zuid-Holland	0.042	(35)	Noord-Holland	0.054	(38)
Közép-Magyarország	0.035	(77)	Lazio	0.035	(46)	Madrid	0.048	(83)

Note: No. of participating organisations are given in brackets. Regional centralities are considered as sum of organisations' centralities; betweenness and eigenvector centrality are normalised between zero and one at the organisational level

References

Asheim B, Gertler S (2005) The geography of innovation: regional innovation systems. In: Fagerberg J, Mowery DC, Nelson RR (eds) The Oxford handbook of innovation. Oxford University Press, Oxford, pp 291–317

Asheim B, Lawton Smith H, Oughton C (2011) Regional innovation systems: theory, empirics and policy. Reg Stud 45(7):875–891

Autant-Bernard C, Billand P, Frachisse D, Massard N (2007) Social distance versus spatial distance in R&D cooperation: empirical evidence from European collaboration choices in micro and nanotechnologies. Pap Reg Sci 86:495–519

Baltagi B (2008) Econometric analysis of panel data, 4th edn. John Wiley & Sons, Chichester

Baltagi BH, Song SH, Jung B, Koh W (2007) Testing panel data regression models with spatial and serial error correlation. J Econom 140:5–51

Barber MJ, Fischer MM, Scherngell T (2011) The community structure of R&D cooperation in Europe: evidence from a social network perspective. Geogr Anal 43:415–432

Bathelt H, Malmberg A, Maskell P (2004) Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. Prog Hum Geogr 28:31–56

Bonacich P (1987) Power and centrality: a family of measures. Am J Sociol 92:1170–1182
Borgatti SP (2005) Centrality and network flow. Soc Netw 27:55–71
Boschma R, Frenken K (2010) The spatial evolution of innovation networks. A proximity perspective. In: Boschma R, Martin R (eds) The handbook of evolutionary economic geography. Edward Elgar, Cheltenham, pp 120–135
Breschi S, Lissoni F (2009) Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. J Econ Geogr 9:439–468
Breschi S, Malerba F (2009) ERA and the role of networks. In: Delanghe H, Muldur U, Soete L (eds) European science and technology policy. Towards integration or fragmentation? Edward Elgar, Cheltenham, pp 160–174
Broekel T, Brenner T (2011) Regional factors and innovativeness: an empirical analysis of four German industries. Ann Reg Sci 47:169–194
Cooke P, Gomez-Uranga M, Etxebarria G (1997) Regional innovation systems: institutional and organizational dimensions. Res Policy 26:475–491
Elhorst JP (2003) Specification and estimation of spatial panel data models. Int Reg Sci Rev 26(3): 244–268
Faust K (1997) Centrality in affiliation networks. Soc Netw 19:157–191
Fischer MM (2001) Innovation, knowledge creation and systems of innovation. Ann Reg Sci 35: 199–216
Fischer MM, Varga A (2003) Spatial knowledge spillovers and university research. Ann Reg Sci 37: 302–322
Fischer MM, Scherngell T, Reismann M (2009) Knowledge spillovers and total factor productivity. Evidence using a spatial panel data model. Geogr Anal 41:204–220
Freeman LC (1979) Centrality in social networks I: conceptual clarification. Soc Netw 1:215–239
Fritsch M, Slavtchev V (2011) Determinants of the efficiency of regional innovation systems. Reg Stud 45(7):905–918
Gilsing V, Nooteboom B, Vanhaverbekec W, Duystersd G, van den Oorda A (2008) Network embeddedness and the exploration of novel technologies: technological distance, betweenness centrality and density. Res Policy 37:1717–1731
Giuliani E (2007) The selective nature of knowledge networks within clusters. J Econ Geogr 7: 139–168
Heller-Schuh B, Barber M, Henriques L, Paier M, Pontikakis D, Scherngell T, Veltri GA, Weber M (2011) Analysis of networks in European Framework Programmes (1984–2006). JRC scientific technological reports JRC 63467, European Commission, Luxembourg
Hoekman J, Frenken K, van Oort F (2009) The geography of collaborative knowledge production in Europe. Ann Reg Sci 43:721–738
Hoekman J, Scherngell T, Frenken K, Tijssen R (2012) Acquisition of European research funds and its effect on international scientific collaboration. J Econ Geogr. doi:10.1093/jeg/lbs011
Karlsson CA, Manduchi A (2001) Knowledge spillovers in a spatial context – a critical review and assessment. In: Fischer MM, Fröhlich J (eds) Knowledge, complexity and innovation systems. Springer, Berlin/Heidelberg/New York, pp 101–123
Le Sage J, Pace RK (2009) Introduction to spatial econometrics. CRC Press, Boca Raton/London/ New York
Maggioni MA, Nosvelli M, Uberti TE (2007) Space versus networks in the geography of innovation: a European analysis. Pap Reg Sci 86:471–493
Morrison A (2008) Gatekeepers of knowledge within industrial districts: who they are, how they interact. Reg Stud 42:817–835
Paier M, Scherngell T (2011) Determinants of collaboration in European R&D networks: empirical evidence from a discrete choice model. Ind Innov 18:89–104
Ponds R, van Oort F, Frenken K (2007) The geographical and institutional proximity of research collaboration. Pap Reg Sci 86(3):423–443
Powell WW, Grodal S (2005) Networks of innovators. In: Fagerberg J, Mowery DC, Nelson RR (eds) The Oxford handbook of innovation. Oxford University Press, Oxford, pp 56–85

Powell WW, White D, Koput KW, Owen-Smith J (2005) Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences. Am J Soc 110:1132–1205

Rietschel ET, Arnold E, Čenys A, Dearing A, Feller I, Joussaume S, Kaloudis A, Lange L, Langer J, Ley V, Mustonen R, Pooley D and Stame N (2009) Evaluation of the six Framework Programmes for research and technological development 2002–2006. Report of the expert group to the European Commission, Berlin/Brussels

Rodriguez-Pose A, Crescenzi R (2008) R&D, spillovers, innovation systems and the genesis of regional growth in Europe. Reg Stud 42:51–67

Romer PM (1990) Endogenous technological change. J Pol Econ 98:71–102

Scherngell T, Barber M (2009) Spatial interaction modelling of cross-region R&D collaborations. Empirical evidence from the 5th EU Framework Programme. Pap Reg Sci 88:531–546

Scherngell T, Barber M (2011) Distinct spatial characteristics of industrial and public research collaborations: evidence from the 5th EU Framework Programme. Ann Reg Sci 46:247–266

Scherngell T, Lata R (2013) Towards an integrated European Research Area? Findings from Eigenvector spatially filtered spatial interaction models using European Framework Programme data. Pap Reg Sci 92:555–577

Wasserman S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, Cambridge

Chapter 16
Local Buzz Versus Global Pipelines and the Inventive Productivity of US Cities

Stefano Breschi and Camilla Lenzi

Abstract Drawing on recent research emphasizing the role played by social and collaboration networks in driving the spatial diffusion of scientific and technological knowledge, this chapter presents new evidence on the structural properties of knowledge networks in 331 US cities based on European Patent Office data for the period 1990–2004. Interestingly, and differently from previous studies, the chapter not only looks at cities' internal network topological structure, but also at the embeddedness of metropolitan inventors within the broader US-wide collaboration network. To this end, it proposes new indicators aimed to capture US cities' propensity to engage not only in local, but also in global knowledge exchanges. In particular, the chapter proposes a classification of US cities according to these dimensions and examines the evolution of metropolitan co-invention networks structural properties in a diachronic perspective. These trends are finally associated to cities' inventive and economic performance.

16.1 Introduction

The importance of knowledge spillovers and exchanges for innovative and economic performance can be hardly neglected and has been supported in the literature by mounting theoretical and empirical research at different levels of analysis since the early formulation of the well-known Marshallian externalities (Audretsch 1998; Sorenson 2003; Beaudry and Schiffauerova 2009). More recently, research has specifically focused on the mechanisms driving the diffusion of knowledge

S. Breschi
Department of Management and Technology & CRIOS, Università L. Bocconi, Via Sarfatti 25, 20136 Milano, Italy

C. Lenzi (✉)
Department of Architecture, Built Environment and Construction Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy
e-mail: camilla.lenzi@polimi.it

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_16,

and notably knowledge networks. In particular, a relatively broad consensus has emerged in the literature about the importance of social proximity within well-defined knowledge communities vis à vis pure geographical co-location to explain knowledge flows and their spatial reach (Singh 2005; Breschi and Lissoni 2009).

Although the literature in economic geography and urban economics has largely emphasized the benefits arising from local networking, recent research has started to question that closeness among actors is beneficial per se while it may also show negative consequences (Burt 2001, 2004; Uzzi 1997). For example, Boschma and Frenken (2010) claim that knowledge exchanges and repeated interactions among the same set of co-localized actors become less valuable over time as knowledge becomes redundant and opportunities for recombination of different but complementary pieces of knowledge are exhausted. Similarly, Fratesi and Senn (2009, p. 17) argue that an excess of inward exchanges that are not complemented by external ones may bring 'the risk of localism, which implies that a regional economy is unable to acquire and master external knowledge and is hence likely to be less innovative'. A disproportionate inward orientation may reduce the potential for knowledge exploration and recombination thus leading to decreases in creativity and losses of positions in the spatial ranking (Neal 2011). Localism may crystallize the existing knowledge basis by reducing heterogeneity and by neglecting alternative technological approaches and solutions, thus inevitably increasing the risks of lock-in. Therefore, the 'local buzz' effect, associated to the rapid diffusion and recombination of ideas and knowledge in clusters and urban settings, may decline and fail to sustain high levels of inventive performance (Bathelt et al. 2004; Storper and Venables 2004).

On the other hand, external linkages with distant regions and cities may provide some shelter to this 'trap'. External sources of new and non redundant knowledge, i.e. 'global pipelines', may inject into the local network new information about market opportunities and still unmet demand (Bresnahan et al. 2001), specialized skills and human capital (Gittelman 2007) and may give access to a larger repertoire of technological and organizational solutions (Owen-Smith and Powell 2004). The role of external links has been also highlighted in the literature on industrial districts, pointing to the key role of leading firms in clusters for the access to external knowledge and, possibly, its transfer to other firms nearby located (Giuliani and Bell 2005; Morrison 2008). Overall, embeddedness in broad knowledge networks, beyond local boundaries, may provide considerable resources and information advantages locally not available and may mitigate the risks of 'entropic death' and lock-in to an obsolete set of technologies.

Still, whereas an excessive inward orientation may reduce the creative drive, an excessive external exposure may lead to technological dependence and exhaustion of endogenous capabilities of autonomous innovation (Evangelista et al. 2002). In fact, a certain level of local knowledge base and absorptive capacity is needed to trans-code, absorb and diffuse the externally sourced knowledge. Hence, a balanced mix of internal and external sources of knowledge (i.e. local buzz and global pipelines) seems to be necessary to promote knowledge diffusion and creation at the local level. In other words, internal and external sources of knowledge are complementary and mutually reinforcing rather than substituting each other (Bathelt et al. 2004; Graf 2011).

On this ground, the present chapter explores the relevance of local buzz and global pipelines and their link to inventive and economic performance by presenting new evidence about co-invention networks of 331 US cities based on European Patent Office (EPO) data for the period 1990–2004. The focus on urban settings particularly suits the study of the link between co-invention network structure and new knowledge creation, as creativity and invention in the US have ever been and still are a predominantly metropolitan phenomenon (Carlino et al. 2007; Feller 1971, 1973; Lamoreaux and Sokoloff 2000; Pred 1966, 1973).

To this end, the remainder of the chapter is organized as follows. The next section explains how network indicators can be successfully used to capture the intensity of local buzz and global pipelines in a city and proposes a taxonomy of US cities according to their propensity to engage in local or global relationships or both. Section 16.3 presents the data set used to draw the co-invention network and to develop the proposed indicators. Section 16.4 comments on how co-invention structural properties link to a city inventive and economic performances. Lastly, Sect. 16.5 concludes by summarizing the main results and by advancing some suggestions for future research.

16.2 The Data Set and the Co-Invention Network's Construction

The use of patent data as relational data enables to map and to study the socio-professional networks in which inventors are embedded through the tools of social network analysis and graph theory (Breschi and Lissoni 2004; Singh 2005; Ter Wal and Boschma 2009). In such a framework, the nodes of the network are inventors and the edges of the network link co-inventors listed on the same patent document. In other words, a pair of inventors is connected if they are designated as inventors in one or more patent documents. Whereas co-invention links certainly capture a subset of all relevant knowledge exchanges and links among individuals within and across cities, these are neither unintentional nor unchecked. The network of inventors is in fact the most immediate and influential social environment from which ideas and information can be drawn, at least for the technical contents of their patents (see Breschi and Lissoni (2004) for additional details on this issue).

Despite the limitations of patent data widely discussed in the literature, patent data present, especially for spatial analyses, two specific advantages as they provide information on the address of each individual inventor who contributed to produce the invention and the list of individual inventors that have produced it. From the latter, the whole set of co-invention ties linking individuals can be derived. On the basis of inventors' addresses, it is then possible to distinguish between (co-invention) ties linking inventors located in the same city, and co-invention ties linking individuals located in different cities.[1]

To build the co-invention network within and across US cities, all patent applications made by US organizations at the European Patent Office (EPO) in 1990–2004 were extracted using the CRIOS-PATSTAT database.[2] Next, the names and addresses of the inventors listed on each patent were collected and harmonized so to minimize spelling errors. The correct identification of individual inventors (i.e., nodes) is crucial to ensure the validity of results derived from social network analysis techniques and requires cleaning and standardization procedures to code names and addresses. To this purpose, we have implemented the Massacrator routine. This is a SQL-based algorithm that compares all inventors with the same name and surname, but different addresses, in pairs. The routine compares information on each inventor in the pair, namely biographical information, the technological contents (i.e. the IPC code) and applicant of each inventor's patents, citation relationships and co-invention ties between the inventors. The routine computes a cumulative 'similarity score' for each pair of inventors with the same name and surname, but different address, according to the similarities in the criteria listed above. The greater the similarity in the information set, the greater the score and the greater the probability that the two inventors are actually the same person. The threshold value of the score for accepting identity between two individuals has been set to a relatively high value to ensure a rather conservative approach in coupling individuals. In other words, the routine has been tuned in a way to minimize false positives (i.e. the probability to identify two inventors as the same person, when they are not, is minimized), although this comes with the cost of having some false negatives (i.e. the data set can still include some cases in which the same person is listed as two different inventors). Lissoni et al. (2006) provide fuller details on the Massacrator routine.

The reported addresses were used to assign each US inventor to one of the 370 US Metropolitan Statistical Areas (MSAs), using the definition files available on the U.S. Census Bureau website.[3] The US Office of Management and Budget defines MSAs as urban core areas of at least 50,000 people, plus adjacent counties that have a high degree of social and economic integration with the core, as measured by commuting ties. The MSAs of Hawaii, Puerto Rico, and Alaska,

[1] Our definition of internal (i.e. within cities) vs. external (across cities) ties is based on inventors' address as available from patent data, regardless they work in the same area or not.

[2] The reference year is the priority year, i.e., the first date at which the patent was applied for anywhere in the world, as it is the closest to the actual time of the invention.

[3] MSAs are defined according to the June 2003 definition of MSAs (http://www.census.gov/population/www/metroareas/metrodef.html). In absence of information on the MSA, information on the state and the zip code were used to assign an inventor to the corresponding MSA using ZIPList5, a commercial database listing every active ZIP code currently defined by the U.S. Postal Service (http://www.zipinfo.com/products/z5cbsa/z5cbsa.htm). For each ZIP code the database identifies the MSA in which the ZIP code lies.

along with 30 other MSAs were excluded from this analysis as there were no patents filed in the MSA in the sample period or because the data used to control for other economic characteristics of the city were missing or incomplete. The final data set includes information on 331 MSAs, 378,167 patents, and 418,228 inventors which amount to, respectively, 96.7 % of all EPO patent applications made by US organizations, and 94.5 % of all US inventors, in 1990–2004.

One of the main criticisms to the use of patent data relates to their uneven quality. In this case, by using patent applications of US organizations at EPO, patents of low quality or with low commercial value, which are not worth extending to Europe (through a costly procedure such as the EPO one), are dropped from the analysis thus mitigating this risk. Several scholars have argued that, precisely in the period under consideration, the average quality of patents issued by the USPTO has declined due to a series of factors, such as the expansion of patentable subject matters and the increasing use of patents for strategic purposes, which have generated a dramatic surge in the number of patent applications and an increased patent office backlogs. As a consequence of these trends, patents of insufficient quality or with inadequate search of prior-art have been increasingly issued (Hall et al. 2004; Jaffe and Lerner 2004).

Consistently with existing studies in the management and economic geography literature (Fleming et al. 2007; Lobo and Strumsky 2008), the co-invention network is considered as a binary network despite, from a technical point of view, it is the one-mode projection of an affiliation (or two-mode) network and as such, it is a valued network, where the line value indicates the number of patents (i.e. the events) in which two inventors have been part of the same team.

Finally, to account for the possibility of ties' decay, the co-invention network at time t includes only the co-invention ties that have been formed between (t-1) and (t-5); each year it is updated by dropping older than 5-year ties and adding up new ones. In fact, the effectiveness with which a co-invention tie transmits knowledge between inventors is likely to decay with the age of the link, especially for co-invention tie established long ago and never renewed. This time window is consistent with other studies (Fleming et al. 2007; Schilling and Phelps 2007; Lobo and Strumsky 2008) and adopting different time windows did not substantively change the results.

16.3 The Measurement of Local Buzz and Global Pipelines Through Network Indicators and a Classification of US Cities

The measurement of structural properties of co-invention networks within cities and regions (i.e. the intensity of the local buzz) has mostly relied upon indicators aimed at capturing the pervasiveness of knowledge exchanges and scientific collaborations of local actors.[4] In fact, in networks where members are closely connected knowledge and information tend to diffuse more rapidly, and with less noise, than in networks where members are connected by longer chains of ties. As a consequence, new information or ideas generated within the network may rapidly reach (or spill over to) all other members of the network and be recombined with their own knowledge. Accordingly, indicators like the density of the urban/regional network, the average distance and the share of actors belonging to the largest component[5] have been extensively used to account for the connectivity of local networks (see among the many others Fleming et al. 2007 and Lobo and Strumsky 2008).

However, such measures look problematic in the case of sparse and fragmented networks as co-invention networks, where density is low and actors are distributed in small unconnected components. Also, these indicators are dimensionless whereas the amount of knowledge circulating in a city is likely to depend not only on how proximate are the actors in a network but also on how many other nodes in the network a given node is able to connect (either directly or indirectly). In other words, knowledge flows depend both on the *size* of the network and its *connectivity*. In fact, given two networks with the same average distance among nodes, the amount of knowledge flowing within a city will be higher, the larger the number of nodes in the network.

Interestingly, the distance-weighted internal reach indicator proposed by Borgatti (2006) enables to take into account both dimensions. Formally, for any individual, this is defined as the sum of the reciprocal distances to all other inventors k s/he can reach in the co-invention network within the city, as summarised in Eq. 16.1 below:

$$dwr_j = \sum \begin{matrix} n \\ k = 1 \\ j \neq k \end{matrix} \frac{1}{d_{jk}} \tag{16.1}$$

where d_{jk} is the geodesic distance (i.e., the shortest path) that separates inventor j from inventor k in the co-invention network internal to a city.[6] Taking the overall city network, the *average distance-weighted internal reach* is this measure averaged across all nodes (i.e. inventors) in the network:

[4] As in Fleming et al. (2007) and Lobo and Strumsky (2008), the metropolitan (also termed as internal or local) co-invention network is composed of the subset of nodes located in a given city and the ties among them; its structural properties therefore determines a city's network structure. Differently, the external network is composed of the links connecting nodes residing in different cities.

[5] This is the largest group of connected nodes in a network; more formally, it is the largest sub-graph that contains the largest number of nodes.

[6] The reciprocal of an infinite distance, i.e., when two inventors in the network are not *reachable*, is set at 0.

$$Internal\ reach_{MSA_i} = \frac{\sum_{j=1}^{n} \sum_{\substack{k=1 \\ j \neq k}}^{n} \frac{1}{d_{jk}}}{n_i} \tag{16.2}$$

This index varies between 0 and n_i; it is equal to 0 when every inventor in a city is an isolate, i.e., when s/he does not collaborate with any other inventor in the city and it takes a value equal to n_i when every inventor is directly connected to every other inventor without intermediaries (s/he directly collaborates with every other inventor in the city).

This index suits very well the case of sparse and fragmented networks as co-invention networks. First, it allows considering all inventors in a city whereas indicators like the size of the largest component or the average distance focus on one component only that may capture only a low fraction of all inventors (the average share of inventors in the largest component in the period 1990–2004 is only 17.5 %).

Second, it not only provides information on how well connected are actors within a city and therefore on how smoothly knowledge flows but also on the scale of such effects. In fact, it is normalized by n_i and, therefore, it captures a network's *size* and *connectivity* simultaneously, as both aspects matter in explaining the local buzz effect. In cities where *internal reach* is high, connectivity is high (i.e., inventors are closer and less fragmented into multiple disconnected components), and the number of inventors is larger. Knowledge, therefore, is likely to flow faster, with less noise and to benefit a larger number of individuals.

Third, it considers both first-order ties (i.e., inventors directly connected because of joint patents) but also second-order and higher-order co-invention links, and thus the overall scope of a city's ties. The intensity of the local buzz effect does not depend only on the number of direct ties but also on the number of indirect ties, though the value and impact of the knowledge received may be subject to distance decay effects between sender and receiver (Ahuja 2000).

The measurement of structural properties of co-invention networks among cities and regions (i.e. the intensity of global pipelines) has mostly relied upon indicators aimed at capturing the degree of openness of local networks, for example through the number of external co-inventors to the city (Fleming et al. 2007; Lobo and Strumsky 2008) or centrality indexes (Giuliani and Bell 2005). However, for the reasons discussed above, such indexes may be quite unsatisfactory in the case of sparse and fragmented networks. Therefore, an adapted version of the distance-weighted reach indicator proposed by Borgatti (2006) is used to capture the extent to which a city's inventors forge external ties with all other inventors located in all other cities. Formally, this index is defined as follows:

$$External\ reach_{MSA_i} = \frac{\sum_{i=1}^{n_i} \sum_{h=1}^{n_h} \frac{1}{d_{ih}}}{n_i} \tag{16.3}$$

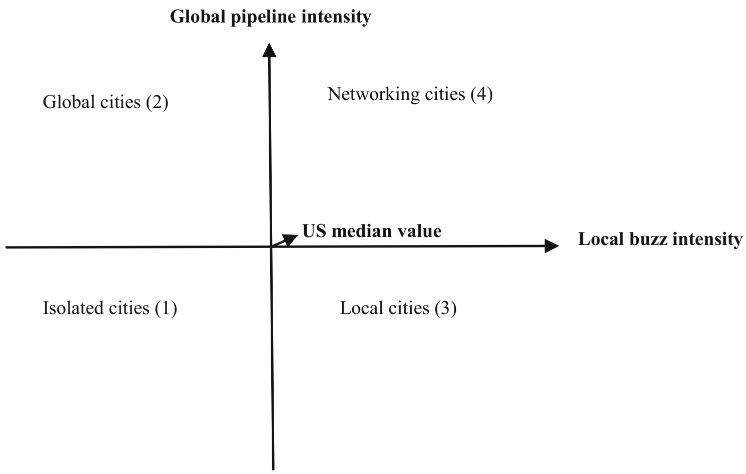


Fig. 16.1 A proposed classification of US cities

where n_i denotes the number of inventors located in city i and n_h denotes the number of inventors located in other cities (i.e., not located in city i), and d_{ih} denotes as before the geodesic distance (i.e., shortest path) in the global co-invention network between inventor i and inventor h. The index takes a minimum value of zero (i.e., all inventors in city i are not connected to any external inventor). In the (theoretical) case in which every inventor in city i directly collaborates to every other inventor in every other city, the index takes value (n_h). Similarly to internal reach, higher values of the external reach index imply that a city has faster access to a larger pool of external knowledge and resources.

On the basis of these indicators, it is possible to classify US cities in terms of their intensity of internal and external links, i.e. the intensity of the local buzz and global pipelines effects. A simple and somewhat sketchy taxonomy can help to specify the different propensity across cities towards inward and outward connections and modes of integrating them. In particular, cities can be classified according to their positioning in terms of internal reach and external reach with respect to the US median value, as displayed in Fig. 16.1.[7]

Cities can show higher values of both indicators with respect to the US median; accordingly, these cities can be named as networking. Moreover, there may be cities that are predominantly inward oriented (i.e. internal reach is greater than the US median but external reach is lower than the US median); they may be termed as local cities. By contrast, there may be cities that are predominantly outward oriented (i.e. external reach is greater that the US median but internal reach is

[7] The choice of the median value for the creation of the classification is preferable to the use of average values as both internal and external reach show a very skewed distribution as summary statistics in Table 16.1 show.

lower than the US median); these cities can be therefore labeled as global cities. Lastly, there may be cities with a very limited networking attitude as both internal and external reach show values below the US median. These cities can be therefore defined as isolated cities.

This simple classification can provide a useful framework in which to read the heterogeneity across cities in their propensity to forge internal and external links and how this links to inventive and economic performance, i.e. what is the relative importance of local buzz and global pipelines for innovation and competitiveness.

16.4 The Relevance of Local Buzz and Global Pipelines for US Cities Inventive and Economic Performance

The wide temporal coverage of the present data set interestingly enables to monitor the evolution over time of the two central variables, internal reach and external reach. Table 16.1 reports their average values for three non-overlapping sub-periods of equal length of the 1990–2004 time span.

Besides the large spatial heterogeneity of the two indicators, attested by the large variance and difference between minimum and maximum values, it is interesting to notice the sharp increase of both variables in the 15 years under observation. This is certainly consistent with the well documented patent explosion during the 1990s and especially in the US. Importantly, the average value of internal reach more than doubled whereas the external reach indicator grew more than 30 times.[8]

This increasing trend in the number of external relationships is also visible in Fig. 16.2 that plots the average ratio of external to internal ties for the 331 cities analyzed in the period 1990–2004. External ties are those linking an inventor in city i to an inventor located in city j. Internal ties, instead, link two inventors located in the same city. This ratio rose steadily from around 2.5 to around 4.3: links across cities have been increasing much faster than links within cities. The outward-looking propensity of agents within cities have spurred the search for knowledge outside the boundaries of the area in which they were located. However, this process was quite spatially unbalanced as confirmed by the coefficient of variation with value greater than one: cities were quite heterogeneous in their ability to forge external ties.

By applying the classification proposed in Sect. 16.3, US cities can be grouped on the basis of their propensity to engage into local and external relationships (i.e. local buzz vs. global pipelines) and can be divided accordingly in four clusters. To this end, we computed for each city the average value of the internal reach and external reach indicators in the three sub-periods 1990–1994, 1995–1999, and

[8] Because both internal and external reach are scale variant, we cannot exclude that this result can be influenced by an increase of the external network greater than the increase of the average internal network.

Table 16.1 Summary statistics for internal reach and external reach, 1990–2004

	Internal reach			External reach		
	1990–1994	1995–1999	2000–2004	1990–1994	1995–1999	2000–2004
Mean	2.335	3.803	5.409	15.227	125.114	512.4285
Standard deviation	4.646	11.222	15.989	26.718	138.674	476.918
Median	1.298	1.536	1.818	6.386	80.536	383.314
Minimum	0	0	0	0	0.253	0.452
Maximum	61.809	122.848	173.369	202.432	919.602	2,726.617
Observations	331	331	331	331	331	331

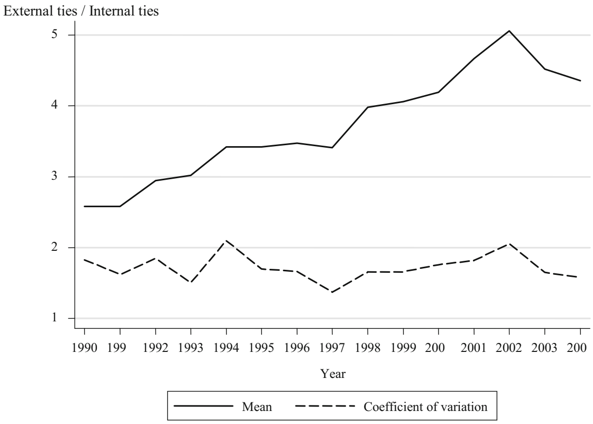


Fig. 16.2 Average ratio of external to internal ties 1990–2004, (331 MSAs) Source: Breschi and Lenzi 2011

2000–2004. About 30 % of US cities fall in the isolated group. Perhaps these cities show a limited inventive activity that hinder opportunities for collaboration both at the local and global levels. The two groups of local cities on the one hand and of global cities on the other have comparable size (almost 20 % of US cities each), attesting that a non negligible number of cities either exploit only local knowledge exchanges or, alternatively, rely on external knowledge sources. Lastly, networking cities are also quite a large group (30 %), suggesting that internal and external ties could be complementary and mutually reinforcing rather than substitute each other (Table 16.2).

Interestingly enough, the relative size of the four groups tends to be rather persistent in the three sub-periods, although there is some churning and erraticism in cities' classification as several of them change category over time as reported in the transition matrixes below (Tables 16.3, 16.4, and 16.5).

Importantly, Table 16.6 reports the average values of the internal reach and external reach indicators and the significance level of the ANOVA test for the four groups of cities in the three sub-periods 1990–1994, 1995–1999 and 2000–2004.

Table 16.2 Number of cities by group, 1990–2004

	1990–1994	1995–1999	2000–2004
Isolated cities (1)	104	105	101
Global cities (2)	62	60	64
Local cities (3)	62	61	64
Networking cities (4)	103	105	102

Table 16.3 Transition matrix, 1990–1999

		1995–1999				
		Isolated cities (1)	Global cities (2)	Local cities (3)	Networking cities (4)	Total
1990–1994	Isolated cities (1)	66	23	12	3	104
	Global cities (2)	20	29	1	12	62
	Local cities (3)	15	3	31	13	62
	Networking cities (4)	4	5	17	77	103
	Total	105	60	61	105	331

Table 16.4 Transition matrix, 1995–2004

		2000–2004				
		Isolated cities (1)	Global cities (2)	Local cities (3)	Networking cities (4)	Total
1995–1999	Isolated cities (1)	73	19	10	3	105
	Global cities (2)	12	38	3	7	60
	Local cities (3)	10	2	38	11	61
	Networking cities (4)	6	5	13	81	105
	Total	101	64	64	102	331

Table 16.5 Transition matrix, 1990–2004

		2000–2004				
		Isolated cities (1)	Global cities (2)	Local cities (3)	Networking cities (4)	Total
1990–1994	Isolated cities (1)	62	24	15	3	104
	Global cities (2)	17	29	4	12	62
	Local cities (3)	15	5	26	16	62
	Networking cities (4)	7	6	19	71	103
	Total	101	64	64	102	331

The highly significant differences in the average values of both indexes across the four groups indicate that the proposed classification of US cities does capture quite heterogeneous behaviours in metropolitan networking intensity and attitudes as well as, possibly, in cities' inventive activities.

Table 16.6 Internal reach and external reach average values and ANOVA test statistical significance (p-value), by group of cities and periods

	Internal reach			External reach		
	1990–1994	1995–1999	2000–2004	1990–1994	1995–1999	2000–2004
Isolated cities (1)	0.476	0.683	0.815	1.776	24.404	170.689
Global cities (2)	0.67	0.677	0.739	19.214	195.388	712.993
Local cities (3)	2.778	3.169	3.513	2.699	38.1	178.064
Networking cities (4)	4.947	9.079	14.077	33.952	236.218	934.771
ANOVA significance	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$

In order to explore this possibility, Tables 16.7, 16.8, and 16.9 investigate whether the increasing outward-looking propensity of cities is associated to a larger inventive performance in the three sub-periods 1990–1994, 1995–1999 and 2000–2004. In particular, Tables 16.7, 16.8, and 16.9 report the average values in the four groups of cities, in the US and the ANOVA significance test for different measures of inventive output computed in the respective period under examination.

The first indicator considered is the total number of patents weighted by the number of jobs.[9] The second indicator is the number of internal patents (weighted by the number of jobs) applied for only by inventors located in the focal MSA (i.e. without inventors external to the city) and it provides a measure of a city's endogenous inventive capabilities. The third indicator is the number of mixed patents (weighted by the number of jobs) that comprise at least two internal inventors and one external inventor, i.e. they are the outcome of both within and across cities collaboration. The fourth indicator is the number of external patents (weighted by the number of jobs) that include only one internal inventor and at least one external inventor, i.e. they are the outcome of cross-cities collaboration. To account for the uneven quality and technological impact of patents, also the number of patents weighted by the number of citations received in the first 5 years after application (self-citations excluded) is considered. Originally, also two indicators of technological recombination capabilities are considered. These are derived by exploiting information on all technological classes listed in each patent document. Firstly, for each city, we computed all pairs of technological classes associated to each internal patent. From this, we counted the number of new pairs of technological classes in the city that are new to the US (i.e. radical creativity) and that are new only to the city itself (i.e. incremental creativity). As both indicators tend to increase with the number of patents developed, they have been both weighted by the

[9] Source: US Bureau of Economic Analysis (BEA) Regional Economic Accounts (http://www.bea.gov/regional/reis/).

Table 16.7 Mean values by group of cities and in US and ANOVA test statistical significance (p-value), 1990–1994

1990–1994	Isolated cities (1)	Global cities (2)	Local cities (3)	Networking cities (4)	US average	ANOVA significance
Total patents	0.063	0.135	0.148	0.339	0.178	$p < 0.01$
Internal patents	0.034	0.05	0.096	0.186	0.096	$p < 0.01$
Mixed patents	0.003	0.011	0.015	0.041	0.018	$p < 0.01$
External patents	0.020	0.056	0.029	0.085	0.049	$p < 0.01$
High quality patents	0.015	0.032	0.086	0.169	0.079	$p < 0.01$
Radical creativity	8.348	10.188	12.25	12.994	10.869	$p < 0.10$
Incremental creativity	18.569	27.78	24.936	31.757	25.403	$p < 0.05$
Population density	141.183	223.389	218.429	407.613	253.957	$p < 0.01$
PCPI	12,440.1	13,323.7	13,434.3	14,673.4	13,486.8	$p < 0.01$

Table 16.8 Mean values by group of cities and in US and ANOVA test statistical significance (p-value), 1995–1999

1995–1999	Isolated cities (1)	Global cities (2)	Local cities (3)	Networking cities (4)	US average	ANOVA significance
Total patents	0.094	0.168	0.191	0.447	0.237	$p < 0.01$
Internal patents	0.042	0.056	0.109	0.221	0.113	$p < 0.01$
Mixed patents	0.007	0.014	0.021	0.063	0.029	$p < 0.01$
External patents	0.037	0.074	0.047	0.125	0.073	$p < 0.01$
High quality patents	0.011	0.024	0.073	0.154	0.07	$p < 0.01$
Radical creativity	5.886	8.121	6.885	8.957	7.449	$p < 0.01$
Incremental creativity	16.748	25.337	18.803	26.758	21.859	$p < 0.01$
Population density	166.744	184.44	254.1	424.985	267.97	$p < 0.01$
PCPI	13,433.3	13,621.4	14,653.7	15,768.	14,433.	$p < 0.01$

Table 16.9 Mean values by group of cities and in US and ANOVA test statistical significance (p-value), 2000–2004

2000–2004	Isolated cities (1)	Global cities (2)	Local cities (3)	Networking cities (4)	US average	ANOVA significance
Total patents	0.107	0.187	0.239	0.552	0.285	$p < 0.01$
Internal patents	0.044	0.056	0.122	0.264	0.129	$p < 0.01$
Mixed patents	0.009	0.015	0.031	0.086	0.038	$p < 0.01$
External patents	0.043	0.091	0.067	0.154	0.091	$p < 0.01$
High quality patents	0.004	0.006	0.016	0.049	0.021	$p < 0.01$
Radical creativity	3.900	4.649	4.179	5.534	4.602	$p < 0.01$
Incremental creativity	12.574	18.768	13.967	20.255	16.401	$p < 0.01$
Population density	188.14	169.634	252.976	460.121	280.911	$p < 0.01$
PCPI	14,826.8	14,907.7	15,714.9	17,634.6	15,879.4	$p < 0.01$

total number of patents in a city. Finally, an indicator of population density and wealth (PCPI, i.e. deflated per capita personal income) are added.[10]

In terms of total patents, Tables 16.7, 16.8, and 16.9 indicate a ranking from isolated cities up to networking cities. Local cities look somewhat more inventive than global cities; however, their similar performance is driven by two opposite networking mechanisms (i.e. internal vs. external patenting), consistently with the proposed classification of cities. This is consistent with the view that an excessive external exposure does not necessarily lead to higher innovative performances and some local absorptive capacity (i.e. local buzz) is needed to effectively exploit externally sourced knowledge. The figures on patents weighted by the number of citations received confirm further this result, being local cities able to produce consistently more valuable knowledge.

Interestingly, however, the two indicators of creativity and technological recombination potentials indicate that external ties are associated to higher local creativity (especially in incremental inventions), possibly by injecting non redundant knowledge at the local level and by opening opportunities for technological exploration of new fields.

Importantly, results suggest the existence of complementary effects in both productivity and creativity. In fact, networking cities show superior performances

[10] Source: http://www.bea.gov/iTable/iTable.cfm?reqid=70&step=1&isuri=1&acrdn=5#reqid=70&step=26&isuri=1&7023=7&7024=Non-Industry&7001=720&7090=70&7029=20&7031=5&7025=5&7022=20 (for regional per capita income data) and ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt (for CPI data).

according to any indicator with respect to all the other groups of cities. The combination of local buzz and global pipelines seems therefore key to achieve higher inventive outputs whereas the presence of one of the two effects only does not seem sufficient. Neither a vibrant local buzz nor intense global pipelines per se are associated to greater performances.

As a final observation, it is also worth remarking that the differences detected in inventive and creative performances are matched by the differences in terms of population density and per capita wealth, indicating that knowledge networks (either internal or external to a city) develop more easily in densely populated and wealthier settings, as highlighted in the literature on agglomeration economies (Duranton and Puga 2001; Carlino et al. 2007).

16.5 Conclusions

This chapter has presented new evidence on the topological properties of knowledge networks in 331 US cities on the basis of European Patent Office data for the period 1990–2004. In particular, the chapter has examined simultaneously the structure of the co-invention network within a city and the embeddedness of metropolitan inventors within the broader US-wide collaboration network.

The indicators developed in this chapter allow to capture the propensity of US cities to engage not only in local networking (i.e. the intensity of the local buzz) but, more importantly, to entertain knowledge exchanges with actors located in other places (i.e. the intensity of global pipelines). US cities have been classified according to these dimensions and four main groups have been identified according to the intensity of local and external connections with respect to the US median values.

Interesting results came out from this analysis. Firstly, the data suggest that internal networking is crucial for external knowledge acquisition, absorption, recombination and socialization at the local level, as networking cities exhibit the highest performances in any respect. On the one hand, relying mostly on external sources of knowledge, as for global cities, may open to risks of technological dependence, to vulnerability to the interruption of knowledge flows through relied and trustworthy channels and, overall, to lower inventive performance in terms of better quality patents. On the other hand, the local buzz effect, prevailing in local cities, may be of moderate impact if it is not nourished, enriched and complemented by external linkages that can inject fresh and non redundant knowledge in the metropolitan network. Local buzz and global pipelines seem therefore truly complementary and bring super-linear effects, as the performance of networking cities demonstrate. However, our results are based on a descriptive analysis comparing different groups of cities according to their network structure with respect to their performance in the same period, which prevents to advance causal claims on the relationship between network structure and inventive performance. We hope to extend our future research in this direction and to explore the interplay between local buzz and global pipelines and their causal effects on cities inventive productivity in a dynamic perspective.

Secondly, external connections seem to be a vital mechanism to enhance local creativity potential. Technological recombination and exploration critically impinges on the capacity to scan external (and perhaps cognitively distant) fields and bring in-house fresh and not redundant knowledge. External ties could be strategically designed and exploited in order to enhance and improve knowledge transfer and acquisition as well as to tap into specific technological domains and to learn and improve upon them. Those nodes performing an interfacing function between local and external knowledge systems, frequently termed in the literature as gatekeepers (Giuliani and Bell 2005; Graf 2011), could be crucial in this regard. It would be interesting to know more about the attributes of these actors and their impact on a city rate of innovation.

Some cautionary notes about the interpretation of our findings should be finally mentioned. Firstly, the use of patent data for analysis of innovation process and performance is in fact subject to a number of limitations, and possibly distortions, as frequently highlighted in the literature. Secondly, whereas the geographical roots and nature of knowledge networks can be hardly overlooked as new ties are primarily forged locally and ties tend to persist also after re-location into different areas, social relationships and knowledge linkages may develop over and across different geographical units, thus ultimately being a-spatial. The indicators and analysis proposed in this chapter are a preliminary effort to take into account the spatial dimension of knowledge networks and their impact on knowledge creation.

References

Ahuja G (2000) Collaboration networks, structural holes, and innovation: a longitudinal study. Adm Sci Q 45(3):425–455

Audretsch DB (1998) Agglomeration and the location of innovative activity. Oxford Rev Econ Policy 14(2):18–29

Bathelt H, Malmberg A, Maskell P (2004) Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. Prog Hum Geogr 28(1):31–56

Beaudry C, Schiffauerova A (2009) Who's right, Marshall or Jacobs? The localization versus urbanization debate. Res Policy 38(2):318–337

Borgatti SP (2006) Identifying sets of key players in a social network. Comput Math Organ Theor 12(1):21–34

Boschma R, Frenken K (2010) The spatial evolution of innovation networks. A proximity perspective. In: Boschma R, Martin R (eds) The handbook of evolutionary economic geography. Edward Elgar, Cheltenham, pp 120–135

Breschi S, Lenzi C (2011) Net city. How co-invention networks shape the inventive productivity of US cities, presented at the 51st ERSA Congress, University of Barcelona, Barcelona

Breschi S, Lissoni F (2004) Knowledge networks from patent data: methodological issues and research targets. In: Glanzel W, Moed H, Schmoch U (eds) Handbook of quantitative S&T research: the use of publication and patent statistics in studies of S&T systems. Springer, Berlin, pp 613–643

Breschi S, Lissoni F (2009) Mobility of skilled workers and co-invention net-works: an anatomy of localized knowledge flows. J Econ Geogr 9(4):439–468

Bresnahan T, Gambardella A, Saxenian A (2001) 'Old economy' inputs for 'new economy' outcomes: cluster formation in the new Silicon Valleys. Ind Corp Change 10(4):835–860

Burt R (2001) Structural holes versus network closure as social capital. In: Lin N, Cook KS, Burt RS (eds) Social capital: theory and research. Aldine de Gruyter, New York, pp 31–56
Burt R (2004) Structural holes and good ideas. Am J Sociol 110(2):349–399
Carlino GA, Chatterjee S, Hunt RM (2007) Urban density and the rate of invention. J Urban Econ 61(3):389–419
Duranton G, Puga D (2001) Nursery cities. Am Econ Rev 91:1454–1475
Evangelista R, Iammarino S, Mastrostefano V, Silvani A (2002) Looking for regional systems of innovation: evidence from the Italian Innovation Survey. Reg Stud 36(2):173
Feller I (1971) The urban location of United States invention, 1860–1910. Explor Econ Hist 8 (3):285–303
Feller I (1973) Determinants of the composition of urban inventions. Econ Geogr 49(1):47–58
Fleming L, King C, Juda AI (2007) Small worlds and regional innovation. Org Sci 18(6):938–954
Fratesi U, Senn L (2009) Regional growth, connections and economic modelling: an introduction. In: Fratesi U, Senn L (eds) Growth and innovation of competitive regions: the role of internal and external connections. Springer, Berlin, pp 3–28
Gittelman M (2007) Does geography matter for science-based firms? Epistemic communities and the geography of research and patenting in biotechnology. Org Sci 18(4):724–741
Giuliani E, Bell M (2005) The micro-determinants of meso-level learning and innovation: evidence from a Chilean wine cluster. Res Policy 34(1):47–68
Graf H (2011) Gatekeepers in regional networks of innovators. Camb J Econ 35(1):173–198
Hall BH, Graham S, Harhoff D, Mowery DC (2004) Prospects for improving U.S. patent quality via postgrant opposition. Innov Policy Econ 4:115–143
Jaffe AB, Lerner J (2004) Innovation and its discontents: how our broken patent system is endangering innovation and progress, and what to do about it. Princeton University Press, Princeton
Lamoreaux NR, Sokoloff KL (2000) The geography of invention in the American Glass Industry, 1870–1925. J Econ Hist 60(3):700–729
Lissoni F, Sanditov B, Tarasconi G (2006) The Keins database on academic inventors: methodology and contents. Cespri working papers 181. Available at http://www.francescolissoni.com/rp_g000004.pdf
Lobo J, Strumsky D (2008) Metropolitan patenting, inventor agglomeration and social networks: a tale of two effects. J Urban Econ 63(3):871–884
Morrison A (2008) Gatekeepers of knowledge within industrial districts: who they are, how they interact. Reg Stud 42(6):817–835
Neal ZP (2011) From central places to network bases: a transition in the U.S. Urban Hierarchy, 491 1900–2000. City Commun 10(1):49–75
Owen-Smith J, Powell WW (2004) Knowledge networks as channels and conduits: the effects of spillovers in the Boston Biotechnology Community. Org Sci 15(1):5–21
Pred AR (1966) The spatial dynamics of U.S. urban-industrial growth, 1800–1914; interpretative and theoretical essays. MIT Press, Cambridge, MA
Pred AR (1973) Urban growth and the circulation of information: the United States System of Cities, 1790–1840. Harvard University Press, Cambridge, MA
Schilling MA, Phelps CC (2007) Inter-firm collaboration networks: the impact of large-scale network structure on firm innovation. Manage Sci 53(7):1113–1126
Singh J (2005) Collaborative networks as determinants of knowledge diffusion patterns. Manage Sci 51(5):756–770
Sorenson O (2003) Social networks and industrial geography. J Evol Econ 13(5):513–527
Storper M, Venables AJ (2004) Buzz: face-to-face contact and the urban economy. J Econ Geogr 4 (4):351–370
Ter Wal A, Boschma R (2009) Applying social network analysis in economic geography: framing some key analytic issues. Ann Reg Sci 43(3):739–756
Uzzi B (1997) Social structure and competition in interfirm networks: the paradox of embeddedness. Adm Sci Q 42(1):35–67

Chapter 17
Research Collaboration and Regional Knowledge Production in Europe

Slavomir Hidas, Martyna Wolska, Manfred M. Fischer, and Thomas Scherngell

Abstract The focus of this study is on regional knowledge production in Europe, with special emphasis on the interplay between intra- and inter-regional research collaboration. The objective is to identify and measure effects of research collaboration on knowledge production at the level of European regions. We use a panel version of the spatial Durbin model (SDM) for empirical testing. The European coverage is achieved using 228 NUTS-2 regions covering all pre-2007 EU member states except Cyprus, Greece and Malta. The dependent variable, regional knowledge production, is measured in terms of fractional patent counts at the regional level in the time period 2000–2008, using patents applied at the European Patent Office (EPO). The independent variables include an agglomeration variable, reflecting intra-regional research collaboration, measured in terms of employment in knowledge intensive sectors, and a network variable, reflecting extra-regional research collaboration, measured in terms of a region's collaboration activities in the EU Framework programmes (FPs), weighted by R&D expenditures in network partner regions. We implement a panel version of the standard SDM that controls for spatial autocorrelation as well as individual heterogeneity across regions, and allows for the estimation of spatial spillovers from neighbouring regions. The estimation results confirm the prevalence of agglomeration effects for regional knowledge production, and, by this, the importance of co-location of R&D actors.

S. Hidas • M. Wolska
Institute for Economic Geography and GIScience, Vienna University of Economics and Business (WU Wien), Vienna, Austria

Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria

M.M. Fischer
Institute for Economic Geography and GIScience, Vienna University of Economics and Business (WU Wien), Vienna, Austria

T. Scherngell (✉)
Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria
e-mail: thomas.scherngell@ait.ac.at

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_17,

Furthermore, the study provides evidence that inter-regional R&D collaborations in the FPs significantly contribute to regional knowledge production.

17.1 Introduction

Today it is widely recognised that interactions and research collaborations among organisations are essential elements of knowledge production processes (see, for instance, Powell and Grodal 2005). Organisations must collaborate more actively and more purposefully with each other in order to cope with converging technologies, and increasing market pressures due to changing patterns of demand in a globalising world (see, for instance, Fischer 2001). In particular, firms have expanded their knowledge bases into a wider range of technologies (Granstand 1998), requiring more diverse knowledge, so firms must learn how to integrate new knowledge into existing products or production processes (Cowan 2004). It may be difficult for a firm to develop this knowledge alone or acquire it via the market. Thus, firms aim to form co-operative arrangements with other firms, universities or research organisations that already have this knowledge to get earlier access to it.

In the recent past, organisations seem to have expanded the spatial range of their collaboration activities, referred to as local buzz vs. global pipelines or the local–global duality in the process of knowledge creation (see, for instance, Bathelt et al. 2004). On the one hand, as a consequence of the globalisation process, knowledge production becomes increasingly interconnected and internationalised. The network of interactions between R&D actors rises considerably. On the other hand, R&D activities remain bounded within a relatively narrow geographic area. Taking regions – defined as subnational spatial units – as essential sites of knowledge creation (see, for instance, Lagendijk 2001), this local–global duality is reflected by the co-existence of, on the one hand, the co-location of actors producing knowledge inducing geographically localised, mostly intra-regional knowledge spillovers (see, for instance, Fischer et al. 2006), and, on the other hand, of global, more far-reaching research collaborations taping specific pieces of region-external knowledge (see, for instance, Varga et al. 2010).

In a policy context, it is notable that regional, national and supranational Science, Technology and Innovation (STI) policies as well as regional innovation policies have shifted attention to supporting research collaborations between various organisations, in particular among firms and universities (see Caloghirou et al. 2002, among others).[1] Policy makers have to balance between two types of

[1] This policy focus has been mainly triggered by various considerations in theoretical and empirical literature of Economics of Innovation, Economic Geography, Regional Science and Management Science (see Fagerberg and Verspagen 2009 for an overview). In particular, two arguments are essential in this respect: First, innovation, knowledge creation and the diffusion of new knowledge are the key vehicles for sustained economic growth of firms, industries or regions, and, thus, are essential for achieving sustained competitive advantage in the economy (see, for example, Romer 1990). Second, as mentioned above, interactions, research collaborations and networks of actors are crucial for successful innovation (see, for instance, Fischer 2001).

policies: on the one hand, policy that leads to economies of scale in knowledge production by supporting further regional specialisation, and, on the other hand, policy that promotes cross-regional R&D collaboration and accelerates inter-regional knowledge diffusion particularly to regions where given knowledge is not available (Pontifakis et al. 2009). While regional and national policy programmes mainly address collaborative knowledge production within one region or country, at the supranational level, such as the EU, more far-reaching, large-distance collaboration is encouraged. The prime examples at the European level are the European Framework Programmes (FPs) for Research and Technological Development (RTD). They support pre-competitive R&D projects, creating a pan-European network of actors performing joint project-based R&D (see, for instance, Scherngell and Lata 2013).

Up to now, there is only little empirical evidence on the local–global duality in knowledge creation at the regional level. In this study we take a regional perspective to address this question drawing on novel data sets providing information on project based networking activities in the FPs. The objective is to identify and measure effects of intra- and interregional research collaboration on knowledge production at the level of European regions. We use a panel version of the spatial Durbin model (SDM) for empirical testing. The European coverage is achieved using 228 NUTS-2 regions covering all pre-2007 EU member states except Cyprus, Greece and Malta. The dependent variable, regional knowledge production, is measured in terms of fractional patent counts at the regional level in the time period 2000–2008, using patents applied for at the European Patent Office (EPO). The independent variables include an agglomeration variable, reflecting intra-regional research collaboration, measured in terms of employment in knowledge intensive sectors, and a network variable, reflecting extra-regional research collaboration, measured in terms of a regions' collaboration activities in the EU Framework programmes (FPs), weighted by R&D expenditures in network partner regions. By this, we are able to estimate the distinct effects of network participation and agglomeration on regional knowledge production. In estimating the effects, we implement a panel version of the standard SDM that controls for spatial autocorrelation as well as individual heterogeneity across regions. The specification incorporates a spatial lag of the dependent variable as well as spatial lags of the independent variables. This allows for the estimation of spatial spillovers of agglomeration and network effects from neighbouring regions by calculating scalar summary measures of impacts.

The paper is organised as follows. Section 17.2 sheds some light on the theoretical background for the study, focusing on regional knowledge production and the importance of extra-regional research collaboration for gaining access to external knowledge sources. Section 17.3 outlines the econometric framework, specifying the empirical model in form of a panel version of the SDM relationship to be estimated. Section 17.4 comprises a detailed description of the empirical setting, presenting the data and the dependent and independent variables as well as some descriptive statistics. Section 17.5 presents the estimation results and their interpretation, before Sect. 17.6 concludes with a summary of the main results and an outlook for future research.

17.2 Theoretical Background

The importance of research collaborations for generating new knowledge[2] is nowadays widely accepted (see, for instance, Powell and Grodal 2005). The motives and drivers for organisations to engage in R&D collaborations with firms, research organisations and universities are manifold; one of the most striking arguments is the increasing complexity of innovation processes, most notably in the context of converging and rapidly developing technologies (see, for instance, Pavitt 2005). Consequently, the absorption and integration of new knowledge from various sources as well as a permanent search for novel combination opportunities of complementary knowledge bases is the key to sustainable innovative capability.

As noted by Granstand (1998), innovating organisations have expanded their knowledge bases into a wider range of technologies which requires the integration of a more diverse set of external knowledge pieces. Collaborative arrangements create incentives for interactive organisational learning which leads to faster knowledge diffusion and stimulates the creation of new knowledge or the combination of pieces of existing knowledge in a new way. Such collaborations are particularly useful in an environment characterised by uncertainty and complexity such as knowledge production processes. Collaborating reduces the degree of uncertainty and provides faster access to different kinds of knowledge, in particular tacit knowledge (see, for example, Kogut 1988).[3]

[2] Knowledge can be seen as a process, embedded in employees and firms' routines (Fischer and Froehlich 2001). For the purpose of this study, it is useful to distinguish between two types of knowledge – tacit and codified (see, among others Polanyi 1967; Nonaka and Takeuchi 1995; Fischer 2001). Tacit knowledge is embodied in a person and can be obtained by experience. It requires rather interpersonal contact to diffuse, and, thus, is conditional on geographical proximity (Fischer 2001; Varga et al. 2010). On the contrary, codified (explicit) knowledge is stated in an explicit form, can be stored and transmitted easily over long distances almost frictionless (Bathelt et al. 2004).

[3] Incentives to cooperate and advantages arising from R&D collaborations may also be identified using other theoretical arguments (Hagedoorn et al. 2000; Caloghirou et al. 2003). From the perspective of transaction costs, firms and organisations entering into collaborative arrangements can avoid high costs of internalising R&D activities. Industrial organisation theory argues that R&D collaborations are suitable strategies to capture external knowledge. In addition, the managerial perspective highlights an ability of a firm to learn from cooperation, thereby adopting new skills and abilities, and, thus, improving its own competitive position after all. Both, managerial and industrial organisation views, implicitly include further advantages arising from R&D collaborations, such as R&D costs sharing, economies of scale and scope, risk pooling or access to complementary resources. Close interactions build trust and reduce the uncertainty, and, thus, the complexity of production.

The fundamental importance of research collaborations for knowledge production is also reflected in the in the various systems of innovation concepts (see Lundvall 1992 among many others). In this conception the sources of new knowledge are often established between firms, universities, suppliers and customers. In the concept of the *regional innovation system*, it is further assumed that innovating actors are embedded in a regional innovation system – where the region is defined as a subnational spatial unit – benefiting from spatial proximity to other actors (see Asheim and Gertler 2005). Spatial proximity is considered to be of particular importance since knowledge is in part tacit; Krugman (1991) argues that knowledge flows are restricted with geographical boundaries due to cost of (especially tacit) knowledge transmission, which in contrast to costs for the transmission of information, rises with geographical distance. The distance decay pattern of knowledge externalities has been confirmed in various empirical studies, beginning with the pioneering study of Jaffe et al. (1993), followed by Autant-Bernard (2001), Maurseth and Verspagen (2002), Fischer und Varga (2003), Fischer et al. (2006), LeSage et al. (2007) or Fischer et al. (2009). Audretsch and Feldman (1996) provide evidence that in industries, for which knowledge diffusion is particularly important, innovative activity tends to be more spatially concentrated. It implies that knowledge flows are encouraged by spatial proximity of different R&D actors including firms, public and private research institutes, universities etc. Such organisations are taking advantage of their co-location. These gains are also referred to as agglomeration economies or external economies of scale[4] (Rosenthal and Strange 2004).

However, key players of the regional innovation systems, such as universities and large knowledge-intensive firms, do not only benefit from the local knowledge base, but increasingly are compelled to search for knowledge sources that are geographically located further away in order to keep pace in the global innovation competition (see, for example, Maggioni et al. 2007; Scherngell and Barber 2009, 2011; Wanzenböck et al. 2012). Such region-external knowledge sources are tapped via region-external research collaboration activities – for instance in the form of joint R&D projects, joint assignment of patents or joint conduction of scientific publications – and/or labour mobility. These knowledge sources may be explicitly valuable for such organisations to gain contact with less familiar pieces of knowledge that may be important for their long-term development (see Maskell et al. 2006).

In a policy context, the importance of research collaboration has also been affirmed by the common vision of the EU to develop the European Research Area, intended to integrate national science, technology and innovation (STI)

[4] Two types of agglomeration economies may be specified. Localisation economies (called also intra-industry externalities) emerge from the spatial concentration of economic activity within one single industry, hence from the scale of the industrial specialisation (Marshall 1920; Arrow 1962; Romer 1986). Urbanisation economies (intra-industry externalities) arise from the industrial diversification, region-size (Jacobs 1969).

policies (European Commission 2000), and to support international research collaboration across Europe. The main instrument to reach this goal are the EU Framework Programmes (FPs) for Research and Technological Development (RTD) that are funding programmes created to support and stimulate R&D projects[5] between European organisations in order to boost technological competitiveness, on the one hand, while to ensure cohesion, on the other hand. By this, the FPs provide a significant channel for organisations to tap region-external knowledge sources, and may represent an example of geographically dispersed R&D collaborations[6]. Furthermore, increasing inter-regional connectedness that may be viewed as an alternative explanation of regional knowledge production in addition to conventional agglomeration economies, may provide regions with rather weak agglomeration characteristics an opportunity to be highly productive in case of being well inter-linked to inter-regional R&D collaboration networks (Varga et al. 2010). The focus of this study is to test the interdependencies between region-internal research collaboration – proxied by regional agglomeration effects – and region external research collaboration – proxied by regional participation in the FPs. By this, the study contributes to the literature on the local–global duality of knowledge production processes from a regional perspective.

17.3 The Empirical Model

In order to estimate the relationship between regional knowledge production and region-internal and region-external research collaboration, we use a panel version of the Spatial Durbin Model (SDM) as introduced by Elhorst (2003). This is an appropriate way to deal with the problem of spatial autocorrelation, and to estimate the influence of spatial spillover effects. The panel version of the standard SDM model controls not only for spatial autocorrelation but also for individual heterogeneity across regions (see LeSage and Fischer 2012). Denoting our set of regions

[5] See, for instance, Breschi and Cusmano (2002) for a preliminary view on the emergent pan-European network of firms, public research organisations, universities, consultants and government institutions jointly collaborating on projects across different research areas.

[6] The geography of R&D collaborations within the FPs has attracted increasing attention in the recent past. The study of Constantelou et al. (2004) confirms significant collaborative activity among clusters of neighboring countries. Autant-Bernard et al. (2007) find that relational distance by means of the firms' position within a network matters more than their geographical location. Maggioni et al. (2007) suggest that a region's knowledge production is mainly influenced, besides by regions that are located close in geographical space, also by regions that are close in relational space. The study of Schnerngell and Barber (2009) provides evidence that geographical factors matters for interregional collaboration intensities, whereas the effect of technological proximity prevails. Schnerngell and Barber (2011) further show that geographical factors are less significant for public research networks in comparison with the greater impact of geography on patterns of industrial R&D collaboration networks.

by $i = 1, \ldots, N$ and our time periods by $t = 1, \ldots, T$, the empirical model to estimate the relationship between research collaboration and regional knowledge production is given by

$$y_t = \rho W y_t + \delta_1 a_t + \delta_2 W a_t + \gamma_1 k_{t-2} + \gamma_2 W k_{t-2} + \eta_t \tag{17.1}$$

with

$$\eta_t = \mu + \varepsilon_t \tag{17.2}$$

where y_t is the N–by–1 vector of observations on regional knowledge production in N regions at time t. a_t denotes the N–by–1 vector of observations on the agglomeration variable at time t, capturing intra-regional research collaborations, while k_{t-2} is the N–by–1 vector reflecting the observations on the network variable at time $t-2$,[7] measuring inter-regional research collaboration activities. δ_1, γ_1 are scalar parameters to be estimated.

W is the N-by-N matrix of spatial weights reflecting the spatial configuration of the regions with elements

$$w_{ij} = \begin{cases} 1 & if\ s_{ij}^{(1)} \leq s_{ik_{(i)}}^{(1)} \\ 0 \text{ otherwise} \end{cases} \tag{17.3}$$

where $s_{ij}^{(1)}$ measures the geographical distance between two given regions i and j. k_i indicates the k nearest neighbouring of region i. Following previous empirical research, we set $k = 5$ (see, among others, LeSage and Pace 2008; Scherngell and Lata 2013).

As a consequence, Wy_t denotes the N–by–1 vector representing the spatial lag of regional knowledge production in k nearest neighbours at time t. Its coefficient ρ measures the strength of spatial dependence. Similarly, N–by–1 vectors Wa_t and Wk_{t-2} denote the average of observations on the agglomeration and the network variable in k nearest neighbours at time t and $t-2$, respectively. δ_2, γ_2 are the associated scalar parameters to be estimated.

$\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$ is the N–by–1 vector of disturbances for time period t which is independently and identically distributed with zero mean and variance σ^2_ε. $\mu = (\mu_1, \ldots, \mu_N)'$ is the N–by–1 vector representing random spatial specific effects, i.e. μ is treated as a random element and is assumed to be independently and

[7] Following previous empirical studies (Furman et al. 2002; Varga et al. 2010), we decide to impose a lag of 2 years on the network variable, as it takes some time between the inputs translate into measurable outputs. In case of the agglomeration variable, time lag is not necessary, as the variable varies only slightly over the analysed period.

identically distributed with zero mean and variance σ^2_μ.[8] Since space-specific time-invariant effects are likely to have an impact on the dependent variable, their omission could lead to a biased and inconsistent estimation result (Elhorst 2010b).

Inclusion of lags of both dependent and independent variables allows to account for spatially autocorrelated omitted variables that are likely to be correlated with the included explanatory variables (LeSage and Pace 2009). Furthermore, the SDM model specification offers great analytical opportunities. Having the unconstrained SDM model as an initial model enables us to follow the general-to-simple model selection rule by testing whether the model can be simplified (Fischer and Wang 2011). The SDM model nests a spatial lag (SAR) and a spatial error (SEM) model as special cases. Even when one of these models is the true data generating process, the SDM model provides unbiased estimation results (LeSage and Pace 2009). When $\delta_2 = \gamma_2 = 0$, the model is reduced to the SAR model and comprises only a spatial lag of the dependent variable. By setting $[\delta_2\ \gamma_2] + \rho\ [\delta_1\ \gamma_1] = 0$, the SDM model is simplified to the SEM model.[9] However, this restriction is only correct when there are no omitted variables correlated with the included explanatory variables (LeSage and Fischer 2008). In order to find an appropriate model specification, a likelihood ratio test is carried out. The double difference between the values of log-likelihood function for the SDM model and a model with a restriction is chi-squared distributed with a number of degrees of freedom reflecting the number of imposed restrictions.

An additional specific advantage of using the SDM in the context of our research focus is the possibility to measure the scale of intra- and inter-regional spillovers, or so called direct and indirect effects (LeSage and Pace 2008).[10] Besides direct impacts of a change of independent variables a_i and k_i on knowledge production measured by means of patents y of their respective region i (direct effects), we can additionally observe the effect of changes of these variables in other regions j on region i (indirect effects). Such partial derivatives represent possible spillover impacts from all other regions $N-1$. Since we consider changes in each

[8] The model with random effects is more appropriate in case of our sample data, because variables that do not change or change only slightly over time periods cannot be estimated using the model with fixed effects, since they are eliminated in estimation process (Elhorst 2010b). Such a variable in our model is the agglomeration variable and its spatial lag. Moreover, the model with fixed effects can be estimated consistently only when the time domain T is sufficiently large (Elhorst 2010b). As our sample comprises a relatively small number of time periods $T = 9$ as compared to the number of cross sectional observations $N = 228$, the model with random effects is more suitable.

[9] There is also a possibility to derive a model where only independent variables exhibit spatial dependence and observations of the dependent variable are assumed to be spatially independent ($\rho = 0$). Finally, the restrictions $\rho = 0$ and $\delta_2 = \gamma_2 = 0$ would result in a standard OLS model (Fischer and Wang 2011).

[10] Taking only the parameter estimates δ_1 and γ_1 for the agglomeration and network variables into account would be an incorrect interpretation of the model, since they do not include the effect of so called feedback loops that arise as a result of impacts passing through neighboring regions and back to the regions themselves (LeSage and Pace 2009).

$j = 1, \ldots, N-1$ region including changes in the own region, these results can be expressed by means of N-by-N matrices for both independent variables:

$$S_a(W) = \partial y/\partial a = (I_N - \rho W)^{-1}(I_N\delta_1 + W\delta_2) \tag{17.4}$$

$$S_k(W) = \partial y/\partial k = (I_N - \rho W)^{-1}(I_N\gamma_1 + W\gamma_2) \tag{17.5}$$

The matrices $S_a\,(W)$ and $S_k\,(W)$ of all partial derivatives are correct measures of local (direct) and spillover (indirect) impacts arising from changes in the independent variables a and k of each region i on the dependent variable y of the respective region and all other regions (LeSage and Pace 2009; Elhorst 2011). The off-diagonal elements represent cross-partial derivatives, which can be summarised into scalar measures representing indirect impacts using the average of either the row-sums or column-sums of the matrix elements excluding the diagonal. The average summary measure of direct effects is defined as the average of the sum of the own partial derivatives on the main diagonal of the matrices. The average total scalar measure is represented by the sum of direct and indirect effect averages (LeSage and Pace 2009).

17.4 Data and Variables

In this study, European coverage is achieved using $N = 228$ NUTS-2 regions (revision 2003) covering all pre-2007 EU member states except Cyprus, Greece and Malta. The choice of NUTS-2 regions is motivated by the fact that they have an appropriate size to catch sub-national characteristics (see, for instance, LeSage and Fischer 2012). The time domain comprises $T = 9$ time periods from 2000 to 2008.

To measure regional knowledge production, we use fractional counts of patent applications to the European Patent Office sorted by the by priority year (date of application) derived from Eurostat.[11] We use fractional counts, i.e. we count patents based on the number of inventors listed on a patent application, dividing the number of inventors by the number of different regions in which they are located. For a patent with three different inventors in three different regions we count 1/3 for each region so that the total sum of counts for one patent equals to 1 (Eurostat 2007).

As introduced in the previous section, our independent variables consist of the *agglomeration variable* and the *network variable*. We use employment in knowledge intensive sectors derived from Eurostat as a proxy for agglomeration effects (see, for instance, Varga et al. 2010). By knowledge intensive sectors, we

[11] As inventions have to be novel, non-trivial and commercially applicable in order to be protected by a patent, patents can be recognised as quantitative indicators of inventions. Nevertheless, the use of patents has some limitations. Not all inventions that could be patented are actually patented, because patenting is a voluntary strategic decision. Further, not all inventions are allowed to be patented, for example a program code (OECD 1994).

understand high- and medium-high-technology manufacturing, high-technology knowledge intensive services, knowledge intensive market services, financial services as well as the education and the health sector, as defined by Eurostat.

The *network variable* is measured in terms of the number of regional EU Framework programme (FPs) participations, weighted by R&D expenditures in partner regions. Thus, the measure is defined as a product of a N-by-N collaboration matrix (see Scherngell and Barber 2009 and 2011), and a N–by–1 vector of total regional R&D expenditures for each time period. The data on regional R&D expenditures come from Eurostat. For the construction of the collaboration matrix we use data from the EUPRO database that contains information on research collaborations of participating firms and organisations within the FPs. The time period 1998–2006 covers the fifth (1998–2002) and the sixth (2002–2006) FP. For each time period, the collaboration matrix contains the number of linkages in terms of joint project participations between all (i, j)-region pairs, given $i = 1, \ldots, N$ regions in the rows and $j = 1, \ldots, N$ regions in the columns. Since the network variable acts as a proxy for extra-regional research collaboration, we do not consider intraregional knowledge flows.

Table 17.1 presents some summary statistics on the three model variables. It can be seen that for the dependent variable, that is regional knowledge production, as well as for the agglomeration variable, we cannot observe a time trend concerning mean knowledge production – as captured by regional patenting – and mean degree of agglomeration – as captured by employment in knowledge intensive sectors. In contrast, for the network variable – captured by regional participation in the FPs weighted by R&D expenditures in network partner regions – we can observe a sharp increase in mean regional FP participation intensity between 2000 and 2008.

17.5 Estimation Results

In this section we present and discuss our empirical findings. All variables in the model are defined in log form. Table 17.2 presents the parameter estimates of the SDM model. Furthermore, we report model specification tests that confirm the choice of the SDM specification with random effects. Using a likelihood ratio test, we can reject the restriction of the model to the SAR model, which includes only a spatial lag of the dependent variable (284.26, $p = 0.000$). A Breusch-Pagan test statistic validates the significance of the random effects (4,721.03, $p = 0.000$). All parameter estimates of the independent variables in the SDM model specification are highly significant. However, these estimates cannot be interpreted as marginal effects of changes in the agglomeration and network variables on the knowledge production variable. As mentioned in Sect. 17.3, the parameters estimates differ from direct effect estimates that contain also feedback effects arising partly due to the coefficient of spatially lagged dependent variable, which we find highly statistically significant, and partly due to the highly significant coefficients of spatially lagged independent variables (Elhorst 2010a). It is also important to remark that

Table 17.1 Descriptive statistics

	2000	2001	2002	2003	2004	2005	2006	2007	2008
Regional knowledge production									
Mean	223.8	222.2	222.2	228.1	238.2	243.9	247.5	245.9	208.2
Std.	410.3	420.1	403.0	410.2	425.9	423.1	426.0	420.2	354.2
Min	0.3	0.3	0.5	0.2	0.3	0.2	0.4	0.9	0.5
Max	2,913.5	2,922.6	2,792.5	3,098.4	3,317.7	3,019.0	3,073.5	3,087.4	2,689.1
Agglomeration variable									
Mean	295.6	303.5	307.3	312.4	317.3	322.8	331.2	340.3	345.5
Std.	264.6	272.1	275.7	273.1	282.4	288.7	293.9	301.1	307.3
Min	2.0	2.9	2.7	2.7	2.9	2.3	2.2	5.0	2.6
Max	2,387.3	2,473.4	2,474.6	2,404.8	2,414.7	2,470.3	2,444.2	2,516.4	2,562.1
Network variable									
Mean	3,121.3	2,817.8	3,427.7	4,155.5	5,055.5	5,265.1	7,186.4	7,638.3	8,194.4
Std.	4,921.0	4,358.2	5,380.2	6,473.6	7,923.3	8,209.4	11,759.5	12,646.5	13,788.6
Min	0	0	0	0	1.2	0	0	0	0
Max	45,617.7	40,613.2	51,987.3	61,428.1	76,191.6	79,236.3	119,005.1	129,465.8	140,499.7

Table 17.2 Parameter estimates from the SDM with random effects (Nobs = 2052)

Variable	Coefficient	Standard error	p-value
Agglomeration variable (δ_1)	0.523	0.058	0.000
Network variable (γ_1)	0.075	0.011	0.000
Spatially lagged variables			
Knowledge production (ρ)	0.413	0.028	0.000
Agglomeration variable (δ_2)	−0.244	0.062	0.000
Network variable (γ_2)	0.059	0.015	0.000
Model specification tests			
Log Likelihood	−1,807.91		
LR test for spatial lag	284.26 ($p = 0.000$)		
BP test for random effects	4,721.03 ($p = 0.000$)		

The dependent and the independent variables are defined as given in the text. LR denotes the Likelihood Ratio test for the spatial lag specification, while BP denotes the Breusch-Pagan test for the random effects specification.

highly significant spatially lagged variables do not imply significant indirect effects of the respective variables (see Table 17.3). The spatially lagged variables indicate just impacts of nearest neighbouring regions as defined by the spatial weight matrix W.

Table 17.3 reports the average summary measures along with 95 % credible intervals indicating that the direct, indirect and total effects of the explanatory variables, except the indirect effect of the agglomeration variable, are different from zero based on the credible intervals. If we consider the average direct impacts, it is important to note that they are close to the SDM model coefficient estimates reported in Table 17.2. Differences between these two measures represent feedback effects that arise from induced effects in the neighbours of the neighbours of region i, successively in the neighbours of those neighbours, and continuing throughout the whole system, including some feedback effects to the region i itself.

The direct effect of the agglomeration variable that is highly significant appears to be 0.519. Since the coefficient estimate is equal to 0.523, the feedback effect of this variable amounts to −0.004 or 0.8 % of the direct effect. Similarly, the feedback effect of the network variable is the difference between the highly significant direct effect 0.082 and the parameter estimate 0.075, that is 0.007 or 8.5 % of the direct effect. Thus, the feedback effect turns out to be relatively small and negative for the agglomeration variable. The feedback effect of the network variable, although still relatively small, shows much stronger and positive impact than in the previous case.

Direct effect estimates show in both cases a positive impact, i.e. a change of the independent variable in region i on the knowledge production in that region. This impact is much higher in magnitude in case of the agglomeration variable (0.519). It confirms the importance of co-location of R&D actors. The direct impact of the network variable, i.e. a region's own collaboration activity with other regions, is lower as compared to the agglomeration variable (0.082). However, the results confirm the direct impact of research collaborations within the EU FPs on regional

Table 17.3 Average scalar summaries from the SDM

Variable	0.05 level	Mean	0.95 level
Agglomeration variable			
Direct impact	0.402	0.519* (0.055)	0.622
Indirect impact	−0.170	−0.047 (0.063)	0.079
Total impact	0.401	0.472* (0.035)	0.538
Network variable			
Direct impact	0.062	0.082* (0.010)	0.102
Indirect impact	0.108	0.147* (0.021)	0.188
Total impact	0.192	0.229* (0.035)	0.269

*significant at the 0.001 significance level; standard errors in brackets

knowledge production, when considering patents as an output of knowledge production, though the agglomeration characteristics of a region play a much more prominent role.

Indirect effects of the agglomeration variable are not significant suggesting that the employment in knowledge intensive sectors has only a local impact, in other words, it influences only its own region. On the contrary, the indirect impact estimate for the network variable indicates considerable average spillover effect to other regions (0.147). The indirect effect of a change in the network variable appears to be 1.8 times the magnitude of the direct effect of the same variable. Thus, this result suggests that regions with less developed R&D infrastructure may profit from collaborations with other regions. The total impacts of both independent variables on knowledge production are positive and highly significant (0.472 and 0.229). A 10 % increase in the agglomeration variable increases regional production by 4.72 %. Similarly, a 10 % increase in the network variable results in a 2.29 % increase in regional knowledge production.

17.6 Conclusions

Research collaborations are nowadays to be seen as one of the most essential elements for the knowledge production of firms, universities and research organisations. The focus of this study has been on regional knowledge production in Europe, devoting special emphasis to the question how research collaborations contribute to knowledge production processes from a regional perspective. We have employed a spatial Durbin model (SDM) relationship to test whether region-internal and region-external research collaboration contribute to regional knowledge production, using 228 NUTS-2 regions of Europe as our spatial framework, and accounting for spatial spillovers between our system of spatial units. Regional knowledge production has been proxied by using information on regional patenting for the years 2000–2008, while region-internal research collaboration has been measured by means of an agglomeration variable that is defined by the share of a region's employment in knowledge intensive sectors, and region-external research collaboration by regional participation in the EU Framework Programmes (FPs) that have been specifically designed to foster international research collaboration across Europe.

The study produces promising results in the context of the literature dealing with the local–global duality of knowledge production, also referred to as the local-buzz vs. global pipelines in the process of knowledge creation. The estimation results confirm the prevalence of agglomeration effects for regional knowledge production, and, by this, the importance of co-location of R&D actors. However, the most important outcome of the study is that it provides statistical evidence that inter-regional R&D collaborations in the FPs significantly contribute to regional knowledge production, i.e. knowledge flows via such global knowledge pipelines – often corresponding to large-distance collaborations of key players of the regional innovation system – significantly contribute to the overall regional knowledge production output in form of regional patents.

The results are also important in a policy perspective, as this study is one of the first few studies that provides systematic statistical evidence on the positive contribution of participation in the FPs to knowledge production across Europe, and that such FP collaborations may indeed induce knowledge flows between regions that are located further away, complementing intra-regional inputs to the knowledge production process. Further, the results imply that considerable benefits may arise from R&D collaborations for lagging regions.

Some ideas for a future research agenda come to mind. *First*, alternative measurements of research collaboration may be considered, in particular for extra-regional research collaborations, having in mind that research collaborations in the FPs constitute only a very small and specific subsample of total research collaborations. *Second*, other model specifications may be considered, for instance models for dynamic spatial panels (see Elhorst 2011), in order to be able to disclose and characterise dynamic effects in the relationship between regional knowledge production and intra-regional vs. extra-regional research collaboration.

Acknowledgements This work has been funded by the FWF Austrian Science Fund (Project Nr. P21450) and the Innovation Economic Talent Development Programme of the AIT Austrian Institute of Technology and the Vienna University of Economics and Business.

Appendix: List of Regions

In this study we use 228 NUTS-2 regions (revision 2003) covering all pre-2007 EU member states except Cyprus, Greece and Malta. In addition, the list does not include the Spanish North African territories of Ceuta y Melilla, the Portuguese non-continental territories Azores and Madeira, and the French Departments d'Outre-Mer Guadeloupe, Martinique, French Guayana and Reunion.

Austria: Burgenland, Kärnten, Niederösterreich, Oberösterreich, Salzburg, Steiermark, Tirol, Vorarlberg, Wien

Belgium: Prov. Antwerpen, Prov. Brabant-Wallon, Prov. Hainaut, Prov. Limburg (B), Prov. Liège, Prov. Luxembourg (B), Prov. Namur, Prov. Oost-Vlaanderen, Prov. Vlaams-Brabant, Prov. West-Vlaanderen, Région de Bruxelles-Capitale/Brussels Hoofdstedelijk Gewest

Czech Republic: Jihovýchod, Jihozápad, Moravskoslezsko, Praha, Severovýchod, Severozápad, Střední Morava, Střední Čechy

Denmark: Danmark

Estonia: Eesti

Finland: Åland, Etelä-Suomi, Itä-Suomi, Länsi-Suomi, Pohjois-Suomi

France: Alsace, Aquitaine, Auvergne, Basse-Normandie, Bourgogne, Bretagne, Centre, Champagne-Ardenne, Corse, Franche-Comté, Haute-Normandie, Île de France, Languedoc-Roussillon, Limousin, Lorraine, Midi-Pyrénées, Nord – Pas-de-Calais, Pays de la Loire, Picardie, Poitou-Charentes, Provence-Alpes-Côte d'Azur, Rhône-Alpes

Germany: Arnsberg, Berlin, Brandenburg, Braunschweig, Bremen, Chemnitz, Darmstadt, Dessau, Detmold, Dresden, Düsseldorf, Freiburg, Gießen, Halle, Hamburg, Hannover, Karlsruhe, Kassel, Koblenz, Köln, Leipzig, Lüneburg, Magdeburg, Mecklenburg-Vorpommern, Mittelfranken, Münster, Niederbayern, Oberbayern, Oberfranken, Oberpfalz, Rheinhessen-Pfalz, Saarland, Schleswig-Holstein, Schwaben, Stuttgart, Thüringen, Trier, Tübingen, Unterfranken, Weser-Ems

Hungary: Dél-Alföld, Dél-Dunántúl, Észak-Alföld, Észak-Magyarország, Közép-Dunántúl, Közép-Magyarország, Nyugat-Dunántúl

Ireland: Border, Midland and Western; Southern and Eastern

Italy: Abruzzo, Basilicata, Calabria, Campania, Emilia-Romagna, Friuli-Venezia Giulia, Lazio, Liguria, Lombardia, Marche, Molise, Piemonte, Puglia, Sardegna, Sicilia, Toscana, Trentino-Alto Adige, Umbria, Valle d'Aosta/Vallée d'Aoste, Veneto

Latvia: Latvija

Lithuania: Lietuva

Luxembourg: Luxembourg (Grand-Duché)

Netherlands: Drenthe, Flevoland, Friesland, Gelderland, Groningen, Limburg (NL), Noord-Brabant, Noord-Holland, Overijssel, Utrecht, Zeeland, Zuid-Holland

Poland: Dolnośląskie, Kujawsko-Pomorskie, Lubelskie, Lubuskie, Łódzkie, Mazowieckie, Małopolskie, Opolskie, Podkarpackie, Podlaskie, Pomorskie, Śląskie, Świętokrzyskie, Warmińsko-Mazurskie, Wielkopolskie, Zachodniopomorskie

Portugal: Alentejo, Algarve, Centro (P), Lisboa, Norte

Slovakia: Bratislavský kraj, Stredné Slovensko, Východné Slovensko, Západné Slovensko

Slovenia: Slovenija

Spain: Andalucía, Aragón, Cantabria, Castilla y León, Castilla-La Mancha, Cataluña, Comunidad Foral de Navarra, Comunidad Valenciana, Comunidad de Madrid, Extremadura, Galicia, Illes Balears, La Rioja, País Vasco, Principado de Asturias, Región de Murcia

Sweden: Mellersta Norrland, Norra Mellansverige, Småland med öarna, Stockholm, Sydsverige, Västsverige, Östra Mellansverige, Övre Norrland

United Kingdom: Bedfordshire & Hertfordshire; Berkshire, Buckinghamshire & Oxfordshire; Cheshire; Cornwall & Isles of Scilly; Cumbria; Derbyshire & Nottinghamshire; Devon; Dorset & Somerset; East Anglia; East Riding & North Lincolnshire; East Wales; Eastern Scotland; Essex; Gloucestershire, Wiltshire & North Somerset; Greater Manchester; Hampshire & Isle of Wight; Herefordshire, Worcestershire & Warkwickshire; Highlands and Islands; Inner London; Kent; Lancashire; Leicestershire, Rutland and Northamptonshire; Lincolnshire; Merseyside; North Eastern Scotland; North Yorkshire; Northern Ireland; Northumberland and Tyne and Wear; Outer London; Shropshire & Staffordshire; South Western Scotland; South Yorkshire; Surrey, East & West Sussex; Tees Valley & Durham; West Midlands; West Wales & The Valleys; West Yorkshire

References

Arrow KJ (1962) The economic implications of learning by doing. Rev Econ Stud 29:155–173

Asheim B, Gertler S (2005) The geography of innovation: regional innovation systems. In: Fagerberg J, Mowery DC, Nelson RR (eds) The Oxford handbook of innovation. Oxford University Press, Oxford, pp 291–317

Audretsch DB, Feldman MP (1996) R&D spillovers and the geography of innovation and production. Am Econ Rev 86:630–640

Autant-Bernard C (2001) The geography of knowledge spillovers and technological proximity. Econ Innov New Technol 10:237–254

Autant-Bernard C, Billand P, Frachisse D, Massard N (2007) Social distance versus spatial distance in R&D cooperation: empirical evidence from European collaboration choices in micro and nanotechnologies. Pap Reg Sci 86:495–519

Bathelt H, Malmberg A, Maskell P (2004) Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. Prog Hum Geogr 28:31–56

Breschi S, Cusmano L (2002) Unveiling the texture of a European Research Area: emergence of oligarchic networks under EU Framework Programmes. KITeS working papers 130, KITeS, Centre for Knowledge, Internationalization and Technology Studies, Universita Bocconi, Milano

Caloghirou Y, Vonortas N, Ioannide S (2002) Science and technology policies towards research joint ventures. Sci Public Policy 29:82–94

Caloghirou Y, Ioannides S, Vonortas NS (2003) Research joint ventures. J Econ Surv 17:541–570

Constantelou A, Tsakanikas A, Caloghirou Y (2004) Inter-country technological linkages in European Framework Programmes: a spur to European integration? Int J Technol Manage 27:773–790

Cowan R (2004) Network models of innovation and knowledge diffusion. Unu-MERIT working paper series, 2004–016

Elhorst JP (2011) Dynamic spatial panels: models, methods and inferences. J Geogr Syst 14:5–38

Elhorst JP (2010a) Matlab software for spatial panels. Presented at the IVth world conference of the Spatial Econometrics Association (SEA), 9–12 June 2010, Chicago

Elhorst JP (2010b) Spatial panel data models. In: Fischer MM, Getis A (eds) Handbook of applied spatial analysis. Springer, Berlin, pp 377–407

Elhorst JP (2003) Specification and estimation of spatial panel data models. Int Reg Sci Rev 26:244–268

European Commission (2000) Towards a European Research Area, COM 2000/6, Brussels

Eurostat (2007) Criteria used to count patents used in Eurostat's patent domain. http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/Annexes/pat_esms_an7.pdf

Fagerberg J, Verspagen B (2009) Innovation studies – the emerging structure of a new scientific field. TIK working paper on innovation studies no. 20090104, Centre for Technology, Innovation and Culture, Oslo

Fischer MM (2001) Innovation, knowledge creation and systems of innovation. Ann Reg Sci 35:199–216

Fischer MM, Fröhlich J (2001) Knowledge, complexity and innovation systems: prologue. In: Fischer MM, Fröhlich J (eds) Knowledge, complexity and innovation systems. Springer, Berlin/Heidelberg/New York, pp 101–124

Fischer MM, Varga A (2003) Spatial knowledge spillovers and university research. Ann Reg Sci 37:302–322

Fischer MM, Wang J (2011) Spatial data analysis: models, methods and techniques. Springer, Berlin/Heidelberg

Fischer MM, Scherngell T, Jansenberger E (2006) The geography of knowledge spillovers between high-technology firms in Europe. Evidence from a spatial interaction modelling perspective. Geogr Anal 38:288–309

Fischer MM, Scherngell T, Reismann M (2009) Knowledge spillovers and total factor productivity. Evidence using a spatial panel data model. Geogr Anal 41:204–220

Furman JL, Porter ME, Stern S (2002) The determinants of national innovative capacity. Res Policy 31:899–933

Granstand O (1998) Towards a theory of the technology-based firm. Res Policy 27:465–489

Hagedoorn J, Link AN, Vonortas NS (2000) Research partnerships. Res Policy 29:567–586

Jaffe A, Trajtenberg M, Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. Q J Econ 108:577–598

Jacobs J (1969) The economy of cities. Random House, New York

Kogut B (1988) Joint ventures: theoretical and empirical perspectives. Strateg Manage J 9:319–332

Krugman PR (1991) Geography and trade. MIT Press, Boston

Lagendijk A (2001) Scaling knowledge production: how significant is the region. In: Fischer MM, Fröhlich J (eds) Knowledge, complexity and innovation systems. Springer, Berlin/Heidelberg/New York, pp 79–100

LeSage JP, Pace RK (2009) Introduction to spatial econometrics. CRC Press, Boca Raton/London/New York

LeSage JP, Pace RK (2008) Spatial econometric modeling of origin–destination flows. J Reg Sci 48:941–967

LeSage JP, Fischer MM (2012) Estimates of the impact of static and dynamic knowledge spillovers on regional factor productivity. Int Reg Sci Rev 35:103–127

LeSage JP, Fischer MM (2008) Spatial growth regressions: model specification, estimation and interpretation. Spat Econ Anal 3:275–304

LeSage JP, Fischer MM, Scherngell T (2007) Knowledge spillovers across Europe. Evidence from a Poisson spatial interaction model with spatial effects. Pap Reg Sci 86:393–421
Lundvall BA (1992) National innovation systems. Pinter, London
Maggioni MA, Nosvelli M, Uberti TA (2007) Space versus networks in the geography of innovation: a European analysis. Pap Reg Sci 86:471–493
Marshall A (1920) Principles of economics. Macmillan, London
Maskell P, Bathelt H, Malmberg A (2006) Building global knowledge pipelines: the role of temporary clusters. Eur Plann Stud 14:997–1013
Maurseth PB, Verspagen B (2002) Knowledge spillovers in Europe: a patent citations analysis. Scand J Econ 104:531–545
Nonaka I, Takeuchi H (1995) The knowledge creating company. How Japanese companies create the dynamics of innovation. Oxford University Press, New York/Oxford
OECD (1994) The measurement of scientific and technological activities. Using Patent Data as Science and Technology Indicators-Patent Manual 1994. OECD, Paris
Pavitt K (2005) Innovation processes. In: Fagerberg J, Mowery DC, Nelson RR (eds) The Oxford handbook of innovation. Oxford University Press, Oxford, pp 86–114
Polanyi M (1967) The tacit dimension. Doubleday, New York
Pontikakis D, Kyriakou D, van Bavel R (2009) The question of R&D specialisation: perspectives and policy implications. JRC scientific and technical reports. http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=2339
Powell WW, Grodal S (2005) Networks of innovators. In: Fagerberg J, Mowery DC, Nelson RR (eds) The Oxford handbook of innovation. Oxford University Press, Oxford, pp 56–85
Romer PM (1986) Increasing returns and long run growth. J Politi Econ 94:1002–1037
Romer PM (1990) Endogenous technological change. J Politi Econ 98:71–102
Rosenthal S, Strange W (2004) Evidence on the nature and sources of agglomeration economies. In: Henderson V, Thisse J (eds) Handbook of regional and Urban economics, vol 4, pp 2119–2171
Scherngell T, Lata R (2013) Towards an integrated European Reseach Area? Findings from Eigenvector spatially filtered spatial interaction models using European Framework Programme data. Papers in Regional Science 92:555–577
Scherngell T, Barber M (2011) Distinct spatial characteristics of industrial and public research collaborations: evidence from the 5th EU Framework Programme. Ann Reg Sci 46:247–266
Scherngell T, Barber M (2009) Spatial interaction modelling of cross-region R&D collaborations. Empirical evidence from the 5th EU Framework Programme. Pap Reg Sci 88:531–546
Wanzenböck I, Scherngell T, Lata R (2012) Embeddedness of European regions in EU funded R&D networks: a spatial econometric perspective. Paper presented at the first Eurolio European Seminar on Geography of Innovation, Congress Center, Saint-Etienne, 26–28 January 2012
Varga A, Pontikakis D, Chorafakis G (2010) Agglomeration and interregional network effects on European R&D productivity. Working papers 2010/3, University of Pécs, Department of Economics and Regional Studies

Chapter 18
Policy Induced Innovation Networks: The Case of the German "Leading-Edge Cluster Competition"

Uwe Cantner, Holger Graf, and Susanne Hinzmann

Abstract The last decades saw a pronounced shift in innovation policy in Germany and many other countries towards increased funding of cooperative R&D. Over the last years, competitions between regional initiatives pushed this trend even further by adding a regional perspective, by increasing the scope of funding, and by fostering interaction between a large number of actors. In 2007 the German ministry for education and research (BMBF) started the Leading-Edge Cluster Competition (Spitzencluster-Wettbewerb) in which 15 clusters were selected in three waves (2008, 2010, 2012) and are funded for a 5-year period with up to 40 million Euro each. Our paper presents selected results regarding the influence of government funding on cooperation networks within four of the clusters that were successful in the first wave of the Leading-Edge Cluster Competition. More specifically, we analyse the extent of policy influence on the network of most important cooperation partners, its geographic reach, and the changes of network structure in general. Our empirical analysis is based on original data that was collected in 2011 with cluster actors (firms and public research) who received government funding. Our results indicate that the program was quite effective in initiating new cooperations between cluster actors and in intensifying existing linkages. The vast majority of the linkages which are influenced by the cluster program are between actors located in the cluster region. With respect to the influence of the cluster policy on network structure, we find an increase in network centralization. Small and medium

U. Cantner (✉)
Friedrich Schiller University Jena, Germany and University of Southern Denmark, Jena, Germany
e-mail: uwe.cantner@uni-jena.de

H. Graf
Friedrich Schiller University Jena, Germany and TU Bergakademie Freiberg, Jena, Germany
e-mail: holger.graf@uni-jena.de

S. Hinzmann
Friedrich Schiller University Jena, Jena, Germany
e-mail: susanne.hinzmann@uni-jena.de

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_18,

sized enterprises used the chance to connect with the local ‘stars’, but not as much among each other.

18.1 Introduction

The introduction of the BioRegio contest in the early 1990s marked the beginning of a new era of R&D funding programs. The German innovation policy experienced a paradigmatic shift away from traditional R&D funding measures towards contests between regions with a special focus on collaborative R&D projects. Central to these new competitive approaches were the stimulation of interregional competition, promoting the establishment of regional clusters and the improvement of the functionality of the regional innovation system (Eickelpasch and Fritsch 2005; Staehler et al. 2007). In this context, the presumed economic and technological benefits of clustering serve as a main rationale for modern cluster policies. The main current national cluster funding program – the Leading-Edge Cluster Competition (Spitzencluster-Wettbewerb) – was launched in 2007 by the German ministry for education and research (BMBF). 15 clusters were selected in three waves (2008, 2010, 2012) and have been funded for a 5-year period with up to 40 million Euro each. One of its main goals is the stimulation of regional networking as a lever for innovation and economic growth.

With the rising number of these programs, one major question arose: Does the public promotion of clusters provide an effective and/or efficient measure to achieve the defined goals? Currently, only a few studies try to provide an answer to this question by evaluating cluster policies. To fill this gap, the present chapter examines the impact of the Leading-Edge Cluster Competition (hereinafter referred to as LECC) on networking in the selected clusters. In analysing a unique dataset gathered from a survey of the beneficiaries, we are able to directly attribute the creation of linkages to policy influence. In particular, we contribute to the literature in two ways: first, we enrich the discussion on the effectiveness of policy endeavours and add to the rare empirical evidence on the impacts of cluster policies. Second, this study is one of the few which analyses the effects of a specific cluster policy on the linkages and the related network structure by means of social network analysis (SNA).

The remainder of the chapter is organized as follows: In Sect. 18.2 we provide the basic theoretical rationales for cluster policies and discuss the results of existing studies that focused on the evaluation of cluster policy impacts. Subsequently, we briefly introduce the concept and objectives of the LECC and describe the research methodology, focusing on the network aspect, in Sect. 18.3. We present our results in Sect. 18.4 and conclude in Sect. 18.5.

18.2 The Leading-Edge Cluster Competition, Clustering and Cluster Policies

In 2007 the German ministry for education and research (BMBF) followed up previous successful devices by launching the LECC, an initiative that aims at strengthening Germany's innovation potential and economic success by means of promoting regional clusters. The support of "Leading Edge Clusters" should result in the exploitation of regional innovation potentials and finally in innovation and economic growth. The program was open for all types of technologies and focused on the funding of clusters with the most promising strategies for future markets that have the potential to count among the "Leading Edge" in their respective industry (BMBF 2012).

Overall, 15 clusters were selected in three waves (2008, 2010, 2012), to be labeled as "Leading-Edge Clusters" and to be funded for a 5-year period with up to 40 million Euro each. The selection was consigned to an independent jury of publicly renowned experts from industry and academia.

Moreover, an accompanying evaluation is conducted to monitor the achievement of the declared goals and to derive concrete recommendations for the advancement of the measurement. Therefore, timely evaluations, especially of innovative funding schemes, are a crucial learning mechanism for the adaptive policy maker (Metcalfe 1995).

One main claim of the program is the support of regional networks. The idea is that the creation of an innovative environment, including intensive R&D collaboration between research institutes and industry, should boost an eminent innovative performance that allows for reaching an international leading position.

The entering of regional networks as a focal point of the national research and innovation policy rooted in the increased perception of innovative activities exhibiting a strong regional component and that embeddedness in networks is crucial to firms' innovativeness and competitiveness. Thus, theoretical concepts that account for the regional character of innovation, such as the cluster approach (Porter 1998) or the idea of the regional innovation system (Cooke and Morgan 1994; Braczyk et al. 1998), constitute the rationale for modern innovation policy.

Since the end of the ninetieth century, scholars theorize on the economic benefits that arise for firms locating in geographic agglomerations of related industries (Marshall 1890; Porter 1998). In addition, several empirical studies provide evidence on the positive effects of co-location on innovation (Audretsch and Feldman 1996; Baptista and Swann 1998; Beaudry and Breschi 2003; Aharonson et al. 2008; Lecocq et al. 2009).

The reasons for clustering are manifold. Theorists argue that firms in clusters exploit the advantages of low transaction costs as they are located close to specialized suppliers and clients and have access to a specialized labor pool or are exposed to competitive pressure which drives profitability (e.g. Porter 1998). Furthermore, the proximity to scientific institutions and firms within the same or related industries results in the existence of a common knowledge spillover pool. Nevertheless, spatial proximity per se is neither a necessary nor a sufficient condition for knowledge spillovers (Giuliani 2007; Breschi and Lissoni 2009). The exploitation of existing innovation potentials in certain regions and the efficiency of the regional innovation system depends heavily on the degree of networking among regional actors (Koschatzky 2000; Sternberg 2000; Fritsch and Eickelpasch 2005).

Innovations develop during a collective learning process of several actors in which common knowledge generation, accumulation and diffusion are crucial ingredients (Asheim and Gertler 2006). Especially in the early stages of technology development, when knowledge is specific and complex, continuous communication and face-to-face contacts are indispensable for the efficient transmission of knowledge (Feldman 1994; Breschi and Lissoni 2001). The ease and costs of linkages and knowledge exchange are in turn related to the geographical distance of the correspondent actors. Moreover, spatial proximity allows for the development of trustful relationships and decreases the social distance among related actors (Boschma 2005). Hence, a firm's integration into the regional innovation network providing access to external knowledge sources is a crucial determinant of the firm's learning process and resulting innovative capabilities (Koschatzky 2000).

Although these insights constitute the core rationale for regional cluster policies fostering joint R&D projects, potential gains from clustering do not suffice as a legitimization for political intervention. According to economic welfare theory, political interference is justified when the market coordination mechanisms are not able to result in efficient/optimal outcomes. Evolutionary economists complement these classical arguments by pinpointing to the existence of system failures. Related to this view, the malfunctioning or ineffectiveness of innovation systems provides a reason for political action. Particularly, the presence of network failures in the sense of a deficiency of an optimal degree of linkages among actors in the innovation system formulates a rationale for cluster policies (Carlsson and Jacobsson 1997; Andersson et al. 2004). Hence, the declared aim of the current German cluster policy, the LECC and related programs is the generation of value added for the region and for the national economy by stimulating the creation of regional networks.

With the expiration of the early pioneer programs and the subsequent introduction of new expanded instruments, such as the LECC in Germany, questions regarding the effectiveness and/or efficiency of the public promotion of clusters came up. Evaluation studies of cluster policies were introduced with the purpose to analyse the surplus for the region and the economy that is attributable to the funding measure. Due to the long term character of these effects and the infancy of evaluation concepts, quantitative impact studies on cluster policies are relatively rare and there have been only few attempts to apply SNA in the context of cluster policy evaluation (see Giuliani and Pietrobelli 2011 for a review). Moreover, the few existing analyses provide ambiguous results.

Martin et al. (2011) evaluate the impact of cluster policy on certain firm variables (for instance production and employment) and find no robust effects compared to non-funded firms. In fact, the policy measure which was included in their examination, the French "Local Productive Systems" program, focused rather on the idea of the industrial districts and merely interfirm collaboration than on the concept of the regional innovation system. Nishimura and Okamuro (2011) find that mere participation in the Japanese Industrial Cluster Project has no significant effect on the R&D productivity of firms. Only if cluster participants collaborate with national universities in the same cluster region positive effects were observed.

In a more general framework, Fornahl et al. (2011) evaluate how R&D subsidies, network embeddedness, and locational factors are related to the innovative performance of biotech firms in Germany. Their findings suggest that location in a cluster, even after controlling for embeddedness into knowledge networks, has a positive effect on patent performance. In contrast, R&D subsidies have no effect when given to single firms, and only a slight effect when R&D collaborations are supported. Counterfactual analyses of specific cluster funding programs in Germany show that the success of BioRegio and related programs is grounded above all on the mobilization of long-term cooperations that would not have existed without the program. In this process, primarily collaborations between firms and research institutions were initiated (Staehler et al. 2007). Similar results are obtained by Falck et al. (2010), who find that firms in targeted industries of a regional cluster initiative are more likely to become innovators despite a reduction of their R&D expenditures. Engel et al. (2012) compare the performances of winning regions to non-winning regions in the BioRegio and BioProfile contest in terms of patents and public R&D projects. They find strong short-term effects, but these effects seem to diminish in the long run.

Overall, it appears that only cluster policies that lead to increased and/or intensified collaboration have an impact on innovative and economic performance of funded actors. It remains unclear how policies change the structure of interaction in form of collaboration networks and how these changes influence knowledge flows and subsequent performance. Since we evaluate an on-going program, we focus on the former, i.e. on the policy effect on the structure and intensity of interaction as an intermediate outcome rather than on economic impacts. With the application of SNA, we are able to observe the underlying network structures in the selected clusters and the ramifications originated by political influence. This allows us to provide a hint whether first politically desired effects occurred.

18.3 Data and Research Methodology

Our empirical analysis is based on a survey of actors (benefiting firms and public research organizations) of four clusters (labelled A to D) that were chosen as "Leading-Edge Clusters" in the first wave of the competition at the end of 2008.[1] The survey was conducted in late summer of 2011, almost 3 years after the

[1] The response rate, especially of firms, in one cluster was too low for a meaningful analysis. For reasons of confidentiality, we have to refrain from characterizing the clusters in more detail. Even

announcement of the winning cluster regions of the first wave, to capture first effects on the network structure. Additionally, in autumn 2011 face-to-face interviews were conducted with a small sample (6) of actors per cluster (24 in sum) in order to add to our understanding and complement the interpretation of the results from the survey.

We construct R&D networks on the basis of survey data by means of a free recall method with a fixed choice design (Guliani and Pietrobelli 2011). Thereby, beneficiaries (firms and research institutes) were asked to list the names and address of their up to ten (strategically) most important R&D cooperation partners. The address information was used to assign actors to be located in the cluster region, in the rest of Germany, in the rest of Europe, or outside Europe. The cluster regions are defined as those regions which host the majority of the respective beneficiaries. All clusters span several NUTS 3 regions (Kreise) and some cross boundaries of NUTS 2 regions (Länder). Therefore, the cluster regions are individually defined as combinations of NUTS 3 regions.

Even though it is argued that the roster recall method is to be preferred (ter Wal and Boschma 2009; Giuliani and Pietrobelli 2011), we chose the free recall design for mainly two reasons. First, the generation of a fixed list of actors (roster) would have led to large differences in the size of the clusters (imposed by the empirical design), since the cluster managements define their boundaries in quite different ways (e.g. only funded actors, only formal members of the cluster association, all actors that somehow participate in cluster activities). Secondly, with a roster recall linkages to R&D partners who are not cluster actors could not be observed. However, such extra local (and extra cluster) linkages are of high relevance for cluster success (Bathelt et al. 2004). Our decision for the fixed choice approach in limiting the number of partners to the ten most important ones followed primarily two considerations. On the one hand the acquisition effort of sufficient data for the network analysis is still within the bounds of feasibility for the respondents. On the other hand, the focus on the most important R&D partners allows us to assume an equal weight of the mentioned linkages and prevents the overestimation of linkages with lower intensity.

The formation of R&D cooperations is based on the expected benefits of both partners arising from collaborative activities. These benefits can arise in different ways depending on the type of strategies partners pursue.

To grasp in more detail the nature of the observed network and to understand the underlying motivations that lead to the choice of the partner or the maintaining of a link, we collected information on attributes of these linkages, namely the reason for the strategic importance of the link. Motives to cooperate are manifold: collaboration partners might be chosen as a valuable source "of applied knowledge" or "of basic knowledge". In both cases, learning from the partners' competencies is a central rationale for collaboration. Cooperations might also be formed because

---

though the clusters differ with respect to technological specialization, age, and location, we cannot make use of this information in our analysis.

partners supply their specific capabilities to a common task, i.e. "complementary competences" are the source of strategic importance of a partnership. Partners might also be valuable because of their specific "research infrastructure" not present in firm's own facilities. To account for these different motives for partner choice, we asked the firms[2] to indicate, for each partnership, the motives that qualify it as strategically important.

Furthermore, to attribute the observed network dynamics to the influence of the policy, the actors were explicitly asked, whether the mentioned relations have existed before 2007 (date of the announcement of the LECC and if they were initiated or intensified by the cluster initiative). Hence, our analysis relies on the comparison of the network structure before and after the policy started. We have to acknowledge that this is only an artificial dynamism since we do not have the information about the most important R&D partners in 2007, but can only observe a subset of those that were active at that time, namely those that were still present at the time of the survey.

18.4 How Policy Influences Cluster Structures

18.4.1 Actor Structures

Describing the actor structures in the four clusters, we distinguish four groups. First, *beneficiaries* are those organizations that receive subsidies from the LECC. Second, those beneficiaries who replied to our survey are the *respondents*. Third, *actors* are all the nodes in the network, i.e. all respondents and all organizations that were named by the respondents. Fourth, *cluster actors* refer to those actors that are members of the respective cluster association. This group encompasses all beneficiaries but also organizations that receive no direct funding.

A first view at the composition of the networks of strategically important R&D partners in the four clusters (Table 18.1) reveals that the network size as measured by the number of nodes (actors) varies between 44 (cluster B) and 97 (cluster C). Some of this variation can be attributed to the different number of respondents, which ranges from 12 (clusters B and D) to 17 (clusters A and C).

Regarding the regional distribution of actors, it can be seen that the majority is located within the cluster or national boundaries. Only a small fraction of actors is located outside Germany, with some differences between the clusters. The consideration of the distribution of linkages exposes an almost similar picture. Most of the linkages are directed into the cluster region, followed by national linkages. Nevertheless, the clusters display remarkable differences concerning the focus on intraregional linkages and the embeddedness in international networks. It is

[2] We did not ask the research institutes since the motives to cooperate differ between the private and the public sphere.

Table 18.1 Composition of the clusters and their networks of strategically important R&D partners

Cluster	A	B	C	D
Beneficiaries: no. of organizations that received a questionnaire	24	19	33	35
Respondents: no. of organizations that provided information about their R&D partners	17	12	17	12
Response rate (2)/(1)	71 %	63 %	52 %	34 %
Actors: no. of nodes in the network	61	44	97	48
Cluster actors: no. of nodes that are members of the cluster association	24	20	41	25
Share of actors located in cluster region	36.1 %	50.0 %	45.4 %	47.9 %
In Germany	50.8 %	20.5 %	37.1 %	47.9 %
In Europe	8.2 %	11.4 %	7.2 %	4.2 %
Outside Europe	4.9 %	18.2 %	10.3 %	0.0 %
Number of linkages	101	43	126	58
Into cluster region	53.5 %	48.8 %	55.6 %	55.2 %
To Germany	38.6 %	20.9 %	31.0 %	41.4 %
To Europe	5.0 %	11.6 %	5.6 %	3.4 %
To outside of Europe	3.0 %	18.6 %	7.9 %	0.0 %

noticeable that while cluster B seems to find a number of R&D partners internationally, cluster D is almost exclusively cooperating on a regional and national scale.

18.4.2 Network Structure and Effects of the Leading-Edge Cluster Competition

In Table 18.2, structural indicators and their changes in the course of the LECC are presented; in Fig. 18.3 (Appendix) network visualizations are displayed. To infer on the effect of the cluster policy, we compare the measures for the network based on all reported linkages with those for the network consisting only of those linkages that were present before 2007 (when the LECC was announced).

One of the first important findings from the network analysis is that the policy has a significant positive impact on the intensity of networking.[3] On average, more than half (52.5 %) of the existing linkages were affected by the LECC in the sense of initiation or intensification, with a minimum of 42.9 % in cluster C and a maximum of 65.3 % in cluster A. The majority of these links (35.6 %) was initiated by the program, indicating a strong impact of the policy measure on networking.

[3] Since we cannot observe the whole network in 2007, one could expect that some past linkages dissolved and the policy effect on the intensity is overestimated. However, being asked about the change in total number of cooperation partners, 80 % of the beneficiaries reported an increase.

Table 18.2 Structural indicators for each network with and without policy impact

Cluster	A	B	C	D	Ø
Linkages initiated by cluster program	45.5 %	41.9 %	20.6 %	34.5 %	35.6 %
Linkages intensified by cluster program	19.8 %	11.6 %	22.2 %	13.8 %	16.9 %
Linkages initiated or intensified by cluster program	65.3 %	53.5 %	42.9 %	48.3 %	52.5 %
Density (among respondents)	0.154	0.068	0.132	0.106	0.115
Density (among respondents before 2007)	0.063	0.023	0.081	0.030	0.049
Components (weak)	1	3	1	3	
Centralization (indegree)	0.141	0.024	0.081	0.104	0.088
Centralization (before 2007)	0.053	0.034	0.042	0.048	0.044
Mean outdegree (only respondents)	5.941	3.583	7.412	4.833	5.645
Mean indegree (whole network)	1.656	0.977	1.278	1.208	1.304

Accounting only for the linkages among respondents, network density (all active linkages divided by the number of possible linkages) increased in all four clusters (on average from 4.9 % to 11.5 %). In cluster C, the increase from 8.1 % to 13.2 % is the lowest in relative terms, indicating that the cluster was already well connected before participation. According to face-to-face interviews with some of the actors, this increase of linkages is mainly a consequence of the increased visibility of potential partners and synergy potential triggered by the LECC; i.e. the policy measure mitigates the problem of intermediation within the clusters (Cantner et al. 2011). Furthermore, new partners entered projects via reputational advice from already known partners. The newly established contacts were initiated with the expectation to cooperate in the long run and beyond the own core competences.

Besides this policy effect on the intensity of collaboration between actors, we also observe a structural change with respect to the concentration of partnerships on few central actors. Attributable to the public funding, the extent of the centralization (based on the indegree) (Freeman 1979) increases in three of the four clusters and on average from 4.4 % to 8.8 %. This suggests that the newly established ties are preferentially formed with actors who were already central before the clusters decided to participate in the LECC.

The clusters exhibit certain differences concerning their interior network structure. Cluster A and C form in each case a connected network since their network consists of only one component. That is to say that each actor is directly or indirectly connected to the network. The remaining clusters display a more fragile network topology. Moreover, clusters A and D seem to be more concentrated on few central actors, while cluster B displays a less hierarchical structure. The average number of connections also shows some differences between the clusters. In cluster B, the average respondent named 3.6 important cooperation partners (outdegree) while in cluster C more than twice this number (7.4) was reported. The mean indegree tells us how often the average actor is being named as a R&D partner. In cluster B this measure is below one (0.98), indicating that some actors are not named at all (of course, these can only be respondents). The maximum is observed in cluster A, in which actors are named 1.66 times on average.

Table 18.3 Policy affected linkages to cluster actors (percentages)

	A	B	C	D
Share of policy initiated linkages to cluster actors	71.7	66.7	84.6	90.0
Share of policy intensified linkages to cluster actors	65.0	80.0	82.1	75.0

In Table 18.3, we report the share of policy initiated (intensified) linkages to cluster actors in all policy initiated (intensified) linkages. For the induced (intensified) linkages, these shares range between 67 % and 90 % (65 % and 82 %), indicating that new cooperations are mainly established among cluster members. However, these figures also show that the cluster policy also mobilizes partnerships beyond the cluster boundaries.

In summary we find that the LECC has proven successful in meeting the objective to foster the networking activities in the regions. The basis for an intensified and broader knowledge transfer is founded, which may lead to a higher innovative performance of the system in the future.

18.4.3 Geographic Reach

A clear-cut direction of the policy influence becomes evident when analysing the geographical reach of the cooperation links. Although certain cluster specificities in the regional focus of the ties exist (see Table 18.1 and the discussion in Sect. 18.4.1), the overall picture reveals a strong effect on regional and national linkages. Table 18.4 compares policy induced linkages with non-induced linkages for each cluster and in total. In all clusters we observe a significantly higher share of local linkages for the induced linkages compared to the non-induced links. In most cases this goes hand in hand with lower shares of linkages at higher geographical distance. Exceptions are worldwide linkages in cluster A and national linkages in cluster B. A comparison of the regional distribution of all linkages reveals that roughly 75 % of induced linkages are local, while only 44 % of non-induced linkages are local. The majority of the remaining induced linkages are national with few international linkages being triggered by policy. The shares for the non-induced linkages to the rest of Germany and to outside Europe are significantly higher, while the difference for linkages to European partners is large but not significant.

Consequently, and corresponding to the declared aim of the policy, the LECC primarily stimulates local connections among actors and affects to a lower extent the creation of ties on a national and international level. Hence, in a first instance the LECC is effective in fostering intraregional networks.

Table 18.4 Regional distribution of policy induced vs. non-induced linkages (percentages)

Cluster	A		B		C		D		Total		
Share of linkages induced by LECC	45.5		41.9		20.6		34.5		33.5		
	Induced		Induced		Induced		Induced		Induced		
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	t-statistic
Geographic reach thereof											
Into cluster region	67.4	41.8	72.2	32.0	80.8	49.0	85.0	39.5	74.5	43.6	(−5.78)
To Germany	23.9	50.9	22.2	20.0	15.4	35.0	15.0	55.3	20.0	40.8	(4.10)
To Europe	4.3	5.5	5.6	16.0	3.8	6.0	0.0	5.3	3.6	6.9	(1.31)
To outside Europe	4.3	1.8	0.0	32.0	0.0	10.0	0.0	0.0	1.8	8.7	(2.99)
Pearson's Chi-squared	8.4 (df = 3)		10.2 (df = 3)		9.1 (df = 3)		11.1 (df = 2)		29.1 (df = 3)		

18.4.4 Science-Industry Interaction

Another important goal of the LECC is to connect industry and science to increase the speed of transfer of scientific discoveries into marketable products (BMBF 2012). Figure 18.1 shows the shares of all linkages within and between industry and science in the first bar for each cluster while the respective shares in the second bar are restricted to the linkages induced by the LECC. In three of the four clusters, research cooperations between firms and public research dominate. The connections that were induced by the LECC show a relatively stronger focus on interactions between firms, which is actually quite surprising given the stated goal of the policy. Across all clusters, 25 % of the non-induced linkages are between firms compared to 35 % firm-firm linkages among the induced linkages. Accordingly, linkages among public research as well as linkages between firms and public research are less frequent among the induced linkages then among the non-induced partnerships.[4] Overall, the differences between clusters imply that the motives to cooperate with specific partners are to be found in the regional and technological environment rather than in some (presumed) requirements stated by the policy maker. At the same time, the policy seems to favour market oriented research collaborations between firms rather than science-industry interactions.

18.4.5 Relevance of Linkages

To grasp the nature of the existing and newly established links, we asked the beneficiaries to substantiate the strategic importance of their links according to the four motives discussed in Sect. 18.3. With respect to cluster specificities in the motives to cooperate, we observe some generalities but also some notable differences. The responses are summarized in Fig. 18.2 for each cluster distinguishing between all partnerships (dark grey) and only those that were initiated by the cluster policy (light grey). This allows us to identify differences between clusters in their motivation to cooperate and also gives us the opportunity to observe any systematic deviations of policy induced linkages from the overall picture.

First of all, access to sources of applied knowledge is, with one exception, the most important reason for the strategic importance of R&D collaborations. This is followed by the technical infrastructure that is available with the R&D partners. The acquisition of basic knowledge is especially important in cluster A, while complementary capabilities are of high importance in cluster D.

In general, the policy induced linkages are not biased towards any of these motives. A statistically significant difference only arises for the use of research infrastructure, which shows to be of lower strategic importance for policy induced

[4] A Chi-squared test comparing the two distributions shows a significant difference at the 10 %-level.

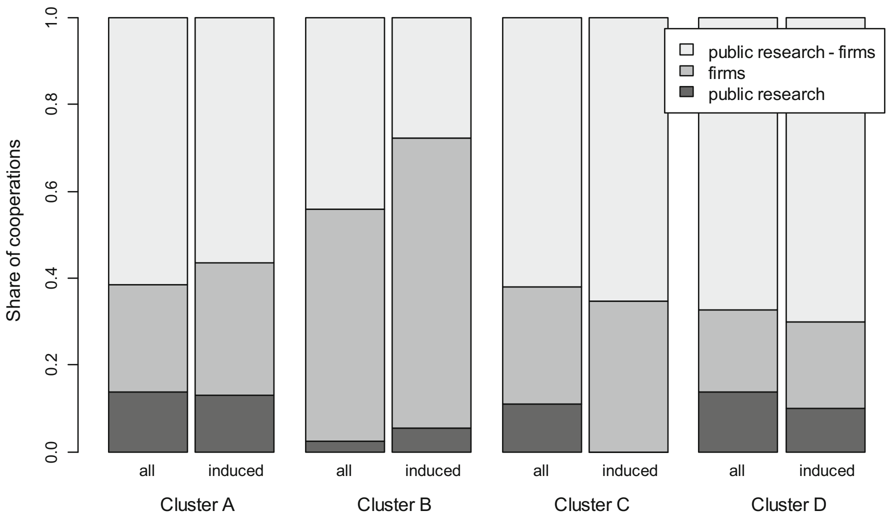


Fig. 18.1 Interaction between science and industry

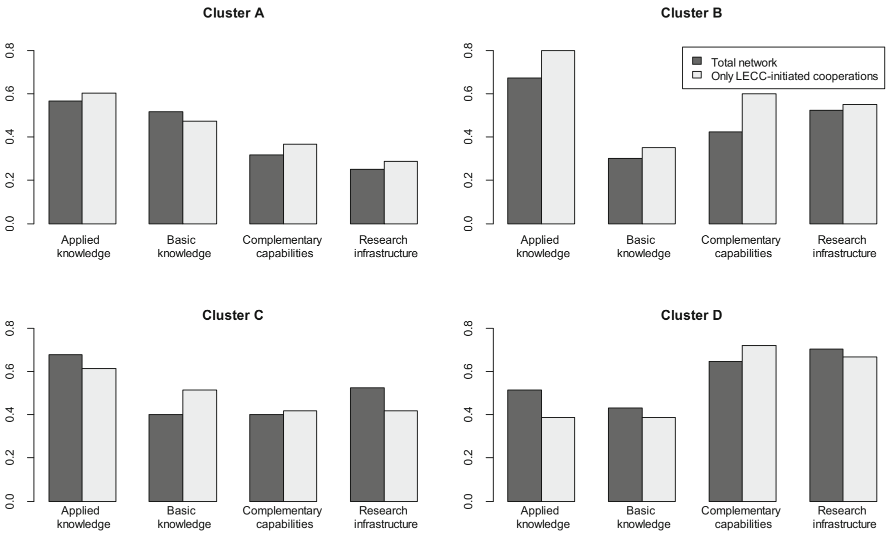


Fig. 18.2 Reasons for strategic importance of R&D partners

cooperations.[5] In cluster B, it seems that the LECC managed to bring together actors with complementary capabilities and strengthened the exchange of applied

[5] For 53.2 % of the pre-existing partnerships and 38.2 % of the policy-induced partnerships, the use of research infrastructure was mentioned as a strategic asset. A t-test shows that this difference is significant at the 5 % level.

knowledge. In cluster C the acquisition of basic knowledge was reinforced. From an evaluation perspective, this result reflects the high flexibility of the policy measure since it is open for various types of partnerships.

18.5 Discussion and Conclusion

Policies aiming at the promotion of clusters are frequently conducted but only seldom evaluated (Martin and Sunley 2003; Brenner and Schlump 2011). The aim of this study was to add to our understanding of the effects and mechanisms of cluster policies by analysing the impact of the German Leading-Edge Cluster Competition on the underlying network structure. Since the LECC is an on-going initiative, we could only report intermediate effects on networking within the funded clusters. By means of Social Network Analysis on the basis of a carefully constructed questionnaire it was possible to identify effects on the network of strategically important R&D partners within the clusters that are attributable to the policy instrument.

Our results show a significant effect on the network structure in terms of density, centralization and geographical reach. Measures on structural effects in terms of number (breadth), weight (intensity) and distribution of linkages (centralization) indicate policy influences already 3 years after starting the funding.

First, on average more than half of the existing linkages were either initiated or intensified by the LECC with the consequence of an increased density of the networks. Second, since the majority of these policy-affected linkages are within the cluster regions, the LECC shifted the focus of collaboration towards local networking. While such an effect is quite natural for a cluster oriented policy, it is not to be judged without some scepticism. Experiences of a Japanese cluster initiative show that local firms have a higher R&D productivity if they collaborate with partners outside the cluster (Nishimura and Okamuro 2011). Moreover, path-dependencies for firms and regions which can lead to spatial lock-in in the long run inhere in the mere search for internal collaborations (Sternberg 2000). These concerns have also been brought up in the discussion on local buzz and global pipelines (Bathelt et al. 2004) and have been related to the stage of the cluster within its life-cycle by Brenner and Schlump (2011). They suggest that a network renewal by means of increased cluster external linkages is especially important in more mature phases of cluster development. Since the four clusters analysed in this paper differ considerably with respect to age or maturity of technology, the dimension "stage in a cluster life cycle" requires further scrutiny.

A third result is concerned with the distribution of linkages within the networks. In three out of four cases the network becomes more centralized, i.e. it exhibits a stronger orientation towards a few, central actors. Interviews with selected beneficiaries in the clusters suggest that this development is rated particularly important for the integration of SMEs within the cluster. For small firms, which in general struggle with difficulties to get in contact with large firms, the LECC offers opportunities to connect to these; the firm representatives value these contacts of crucial importance for their long term integration into the network and finally their innovative performance. However, more centralized networks are also more vulnerable, since their dependence on the functioning of single actors is higher as compared to other network structures. With respect to the rate of knowledge diffusion, Cowan and Jonard (2004) could show that small world structures are the superior form of organization. The results of Schilling and Phelps (2007) on the structure of industry networks add to the difficulties in evaluating this development towards increased centralization. They find negative effects of network centralization on future patenting in the short run but positive effects in the long run.

Fourth, with respect to the interaction between science and industry, we find that the majority of connections that were affected by policy link firms with universities or research institutes. However, the LECC does not increase the relative frequency of science-industry linkages but slightly favours linkages within industry. We interpret the differential policy impact among the clusters as a sign of flexibility of the policy measure as it leaves the choices of partnership to the beneficiaries.

With respect to our research design, we have to acknowledge some limitations. While we can observe cooperations that were established as a consequence of the LECC, we are unable to make statements about linkages that were present before the policy started and have become obsolete. We cannot exclude that newly formed partnerships substituted previous relationships, which would imply that we overestimate the impact of the LECC on the interaction intensity. However, this problem is somehow mitigated since additional sources of information indicate an overall increase in collaboration intensity.

Overall, while we can state that the LECC has met its objective to intensify collaboration among innovative actors, our intermediate evaluation does not allow us to infer, that this will lead to a better performance of the selected clusters in the future. At this stage, we are unable to provide evidence on correlations between the observed structural changes and the innovative performance of the cluster regions. Statements in this direction will require a subsequent long term analysis including comparisons to non-funded clusters.

Appendix

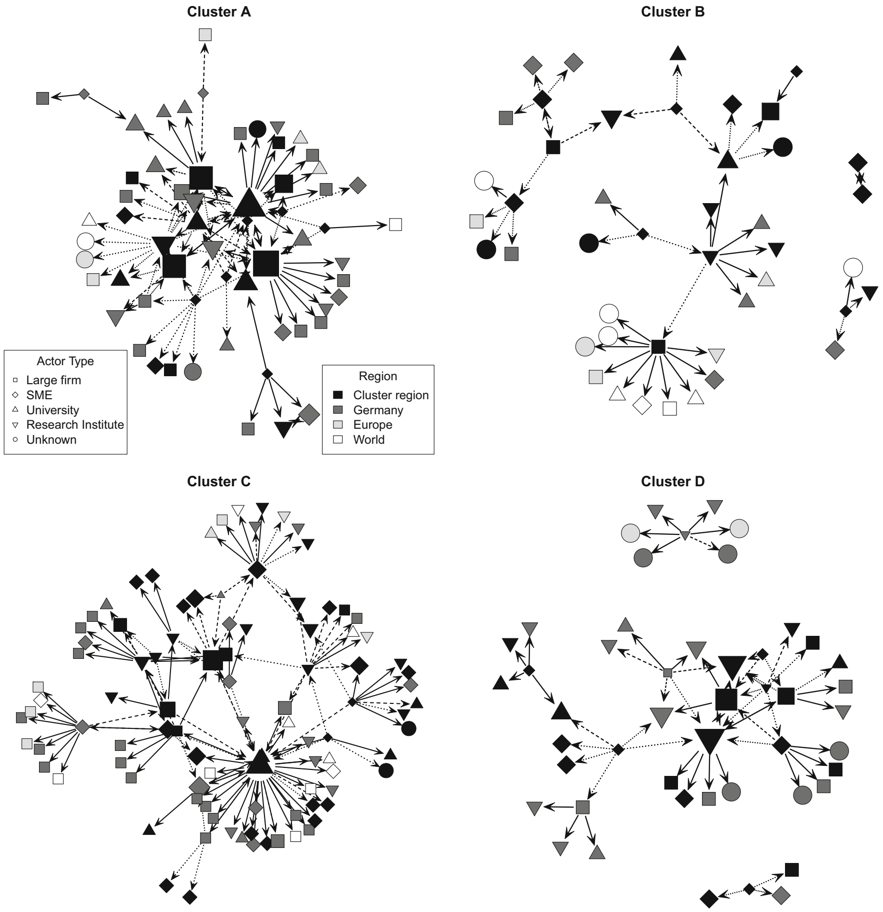


Fig. 18.3 Networks of strategically important R&D partners in clusters A to D. Arrows indicate a partnership from the respondent to one of the most important R&D partners. Dotted arrows indicate that the partnership was initiated through participation in the LECC, dashed blue arrows indicate that the partnership was intensified through the policy, and solid arrows indicate partnerships that were not influenced by the policy. Node size is proportional to indegree, i.e. to the frequency of being named as a partner. The colours and the shapes of the nodes indicate the actor's geographic location and type according to the legend.

References

Aharonson BS, Baum J, Plunket A (2008) Inventive and uninventive clusters – the case of Canadian biotechnology. Res Policy 37(6–7):1108–1131

Andersson T, Hansson E, Schwaag Serger S, Sörvik J (2004) The cluster policies Whitebook. IKED-International Organisation for Knowledge Economy and Enterprise Development. Malmö

Asheim B, Gertler M (2006) Regional innovation systems and the geographical foundations of innovation. In: Fagerberg J, Mowery DC, Nelson RR (eds) The Oxford handbook of innovation. Oxford University Press, Oxford, pp 291–317

Audretsch DB, Feldman M (1996) R&D spillovers and the geography of innovation and production. Am Econ Rev 86(3):630–639

Baptista R, Swann P (1998) Do firms in clusters innovate more? Res Policy 27:525–540

Bathelt H, Malmberg A, Maskell P (2004) Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. Prog Hum Geogr 28(1):31–56

Beaudry C, Breschi S (2003) Are firms in clusters really more innovative? Econ Innov New Technol 12(4):325–342

BMBF (2012) http://www.hightech-strategie.de/en/468.php. Accessed 15 May 2012

Boschma RA (2005) Proximity and innovation: a critical assessment. Reg Stud 39(1):61–74

Braczyk H-J, Cooke P, Heidenreich M (1998) Regional innovation systems: the role of governances in a globalized world. UCL Press, London

Brenner T, Schlump C (2011) Policy measures and their effects in the different phases of the cluster life cycle. Reg Stud 45(10):1363–1386

Breschi S, Lissoni F (2001) Knowledge spillovers and local innovation systems: a critical survey. Ind Corp Change 10(4):975–1005

Breschi S, Lissoni F (2009) Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. J Econ Geogr 9(4):439–468

Cantner U, Meder A, Wolf T (2011) Intermediation, reciprocity and innovation cooperations of firms. Pap Reg Sci 90(2):313–330

Carlsson B, Jacobsson S (1997) Search of useful public policies: key lessons and issues for policy makers. In: Carlsson B (ed) Technological systems and industrial dynamics. Kluwer Academic Publishers, Dordrecht

Cooke P, Morgan K (1994) The regional innovation system in Baden-Württemberg. Int J Technol Manage 9(3/4):394–429

Cowan R, Jonard N (2004) Network structure and the diffusion of knowledge. J Econ Dyn Control 28(8):1557–1575

Eickelpasch A, Fritsch M (2005) Contests for cooperation – a new approach in German innovation policy. Res Policy 34(8):1269–1282

Engel D, Mitze T, Patuelli R, Reinkowski J (2012) Does cluster policy trigger R&D activity? Evidence from German biotech contests. Eur Plann Stud (forthcoming) http://ssrn.com/abstract=2007931

Falck O, Heblich S, Kipar S (2010) Industrial innovation: direct evidence from a cluster-oriented policy. Reg Sci Urban Econ 40(6):574–582

Feldman MP (1994) The geography of innovation. Kluwer Academic Publishers, Dordrecht

Fornahl D, Broekel T, Boschma R (2011) What drives patent performance of German biotech firms? The impact of R&D subsidies, knowledge networks and their location. Pap Reg Sci 90 (2):395–418

Freeman LC (1979) Centrality in social networks: conceptual clarification. Soc Netw 1:215–239

Giuliani E (2007) The selective nature of knowledge networks in clusters: evidence from the wine industry. J Econ Geogr 7:139–168

Giuliani E, Pietrobelli C (2011) Social network analysis – methodologies for the evaluation of cluster development programs. Technical notes no IDB-TN-317. Inter-American Development Bank

Koschatzky K (2000) The regionalisation of innovation policy in Germany – theoretical foundations and recent experience. Working papers firms and region no R1/2000. Fraunhofer Institute for Systems and Innovation Research

Lecocq C, Leten B, Kusters J, Van Looy B (2009) Do firms benefit from being present in technology clusters? Evidence from a panel of biopharmaceutical firms. Open Access Publications from Katholieke Universiteit Leuven. Katholieke Universiteit Leuven

Marshall A (1890) Principles of economics: an introductory volume. Macmillan, London

Martin R, Sunley P (2003) Deconstructing clusters: chaotic concept or policy panacea? J Econ Geogr 3(1):5–35

Martin P, Thierry M, Mayneris F (2011) Public support to clusters – a firms level study of French “Local Productive Systems”. Reg Sci Urban Econ 41:108–123

Metcalfe JS (1995) Technology systems and technology policy in an evolutionary framework. Camb J Econ 19(1):25–46

Nishimura J, Okamuro H (2011) R&D productivity and the organization of cluster policy: an empirical evaluation of the industrial cluster project in Japan. J Technol Transf 36(2):117–144

Porter M (1998) Clusters and the new economics of competition. Harv Bus Rev November–December:77–90

Schilling M, Phelps CC (2007) Interfirm collaboration networks: the impact of large-scale network structure on firm innovation. Manag Sci 53(7):1113–1126

Staehler T, Dohse D, Cooke P (2007) Evaluation der Fördermaßnahmen BioRegio und BioProfile. http://www.e-fi.de/fileadmin/Evaluationsstudien/Evaluation_der_Foerdermass-nahmen_BioRegio_und_BioProfile.pdf. Accessed 18 May 2012

Sternberg R (2000) Innovation networks and regional development – evidence from the European Regional Innovation Survey (ERIS): theoretical concepts, methodological approach, empirical basis and introduction to the theme issue. Eur Plan Stud 8(4):389–407

ter Wal ALJ, Boschma R (2009) Applying social network analysis in economic geography: framing some key analytic issues. Ann Reg Sci 43(3):739–756

Chapter 19
Effects of Competence Centres on Regional Knowledge Production: An Agent-Based Simulation of the Vienna Life Sciences Innovation System

Manuela Korber and Manfred Paier

Abstract Competence centres have gained high recognition as a policy instrument for improving science-industry collaboration. With the requirement for longer-term, institutionalized and geographically concentrated R&D, competence centres provide an environment for joint learning processes and transfer of "sticky" knowledge. They can thus be interpreted as spatially focused R&D networks linking academia and industry. The objective of this chapter is to investigate in a dynamic perspective how a public competence centres programme affects knowledge production in its environment – the regional innovation system. In order to address this issue, we draw on a simulation approach and develop an agent-based model of the Vienna Life Sciences innovation system. Heterogeneous agents representing companies, research organisations and universities are endowed with knowledge and create output, thus generating system performance in terms of scientific publications, patents as well as high-tech jobs. Simulations refer to different long-term scenarios regarding public funds for competence centres. Thus, we explore agent-based simulation as a potential way to address the complexities of knowledge interaction in the context of the "local buzz" versus "global pipelines" discussion in the geography of innovation literature. First results with the empirically calibrated model, e.g. on long-term effects, indicate the potential of the approach for ex-ante impact assessment of network-related measures in R&D policy.

19.1 Introduction

There is broad agreement on the fact that innovation is necessary to achieve and sustain competitiveness of regions (e.g. Cooke 2002b). In this regard, public funding for research and development is considered an important contribution to

M. Korber (✉) • M. Paier
Innovation Systems Department, AIT Austrian Institute of Technology, Donau-City-Straße 1, 1220 Vienna, Austria
e-mail: manuela.korber@ait.ac.at

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_19,

secure an effective and efficient regional innovation system, in the recent past increasingly shifting attention to the support of networking activities and collaborative knowledge production (see also Chap. 18 of this volume by Cantner et al.). However, evidence on effects of policy instruments supporting regional R&D networks on a region´s innovative performance is scarce, since the particular effects are attributable to specific policy interventions only to a very limited extent. The reason is that interventions in self-organized systems – like regional innovation systems – trigger effects which are not easily predictable (Fischer and Fröhlich 2001). Small changes in subsystems might affect the overall system in different ways, ranging from no effect at all to far-reaching system transitions. The research into self-organized systems investigates phenomena such as non-linearity, path dependency, bifurcations and emergence. Such phenomena particularly challenge systematic empirical research on how policy induced R&D networks affect the performance of regional innovations systems, which is – as described in various chapters of this volume – on top of the research agenda in economic geography and regional science.

This chapter focuses on the effects of a specific type of a policy programme supporting geographically localised networking activities, the Austrian competence centres programme. Competence centres are institutionalized research ventures that gather academic and industry partners at a certain location. These are dedicated to high-level, application-oriented research on a particular topic of broad societal interest, thus aiming at critical mass and international visibility. This public funding scheme has a high relevance for the Austrian innovation system and draws on a considerable budget. All in all, EUR 675 million of public funds are devoted to the promotion of competence centres in the framework of the COMET[1] programme (The Austrian Research Promotion Agency – FFG 2010, pp. 3–7).

The objective of this chapter is to investigate the effects of competence centres in a specific sectorial, regional context – the Vienna life sciences innovation system – on the evolution of knowledge output in the long run. Hereby, we apply agent-based modelling techniques to explore knowledge production in different long-term scenarios (30 years) regarding competence centre funding in an empirically calibrated model.

The model comprises a set of heterogeneous agents representing companies, research organisations and universities in the life sciences innovation system. They are able to create knowledge through own R&D, exchange knowledge with others through R&D networks and other exchange mechanisms, and produce output using public and private funding sources. At this point, it is important to mention that agent-based simulation does not provide forecasts, but gives insights into basic aspects of different scenarios. Thus, the simulation model can be seen as an experimental laboratory for testing policy measures ex ante. This has the potential to improve the current practice of relying mainly on past experience and expert

[1] COMET (Competence Centers for Excellent Technologies) programme: funding period from 2008 to 2019.

knowledge, which is hardly able to avoid bias on individual perceptions and strategic behaviour.

The chapter is organized as follows: The theoretical foundation of this chapter is given in Sect. 19.2, which leads to the formulation of hypotheses regarding effects of competence centres. Section 19.3 is dedicated to the empirical test environment, i.e. the Vienna life sciences innovation system. The main components of the agent-based model are described in Sect. 19.4. Then, Sect. 19.5 presents details on the different simulation scenarios and their respective results. The paper closes with a discussion of the method, its limitations and an outlook on further research.

19.2 Public Funding of Localised R&D Networks and Regional Knowledge Production

The importance of R&D cooperation for innovation in regions is well established in scientific literature (e.g. Castells 2000; Fischer 2006; Chap. 18 of this volume by Cantner et al.). Regions have been attracting the interest of both scientists and policymakers as the designated sites of innovation and competitiveness. Numerous empirical studies have focused on regional clusters, starting with the famous success story of Silicon Valley (Saxenian 1991). Some clusters prosper, while others do not – the reasons for that difference in progress are still unclear (Carter 2007, p. 24). Studies on clusters usually draw on the common rationale that territorial agglomeration provides the best context for an innovation-based globalizing economy due to localized learning processes and “sticky” knowledge grounded in social interaction (Asheim and Coenen 2005). It is a general lesson from research on R&D networks that open and interconnected networks with links to different knowledge categories foster knowledge diffusion, exploit synergies, and bridge social closure (Uzzi 1997).

R&D networks provide channels for communication, interaction, and mutual knowledge exchange inside and beyond the regional boundaries, transcending sectors as well as industries. They facilitate collective learning and the exploitation of complementarities on the basis of non-market exchange relations, which are so important for the generation of novelty (Pyka 2002). Thus, R&D networks are a prime vehicle to balance regional economic focus, technological specialization and diversity as sources of the innovative performance of the region and thus of its economic prosperity (Karlsson et al. 2009).

Competence centres are a special type of policy-induced R&D networks as they gather academic and industrial partners in a certain location. They support the role of universities in applied research networks, not only as an important pool for highly skilled labour (Lambooy 2004; Betts and Lee 2004, p. 2; Graf 2006, p. 40; Jaffe 1989, p. 957) but also as important partners for small and medium sized companies (McMillan et al. 2000; Liebeskind et al. 1996). Tacit knowledge can generally be referred to as a prime reason for cooperation (Fischer 2003, p. 346).

In this regard, competence centres support innovation by enabling access to implicit (or tacit) knowledge[2] both geographically localized, but also to distant knowledge via internationally networked partners. Thus, competence centres can be regarded as public instrument to stimulate the combination of the local buzz versus global pipelines structure for regional knowledge production (see Chap. 16 of this volume by Breschi and Lenzi). They may serve as additional levers to accelerate regional knowledge diffusion by supporting local networking augmented with channels to tap international knowledge sources, and, by this, foster overall regional innovation output.

Especially in the life sciences, industrial and academic agents operate in a dynamic environment characterised by fast-expanding scientific knowledge and scattered expertise. High development costs are often associated with long time lags in the commercialization of scientific results (Cooke 2002a). Therefore, agents operate under high uncertainty, and, in order to keep pace with innovation, they engage in research cooperation networks (Powell et al. 2005), typically involving both small dedicated biotechnology firms and large diversified, often transnational, firms with access to global markets and world class academic institutions. The demand for tacit knowledge leads the organizations to cluster geographically and to engage in R&D networks, especially relevant for small dedicated biotech firms (Korber 2012). We are thus led to expect that (regarding the number of SMEs and high-tech jobs) *the SME sector benefits more from the competence centres programme than large companies do (Hypothesis 1).*

The extraordinary importance of scientific knowledge in the life sciences is to a large extent associated with high collaboration intensity between academia and industry. Knowledge spillovers between scientific organizations and companies in cooperative projects as well as during co-publications are most intense if they are based on dense and frequent face-to-face contact (Schartinger et al. 2002). Firms which cooperate with universities are more often involved in basic research, they have access to higher quality ideas and their invention process is more effective in general (Fabrizio 2009). We therefore hypothesise that due to close scientific cooperation of companies and universities, *the competence centres programme strongly favours regional patenting activity (Hypothesis 2).*

According to recent research policy, universities are generally required to increase cooperation with industry partners. In this regard, worries arise regarding an undesired drift to applied research and, consequently, a decrease in publication output. On the other hand, empirical evidence has been produced indicating that researchers that accomplish both patenting and publication activities, are likely to produce more publications. In fact, even reinforcement effects of both activities have been observed (Van Looy et al. 2006). Consequently, we expect that *the number of publications in the region rises due to the existence of the competence centres programme (Hypothesis 3).*

[2] Tacit knowledge is uncodified, maybe even uncodifiable and varies individually (Fischer 2003, p. 345).

19.3 Competence Centres and the Vienna Life Sciences Innovation System

In the following, we give a brief overview of the life sciences innovation system in the Vienna region before we turn to the Austrian competence centres programme, which plays an important role therein. The mid-1980s saw the foundation of a joint venture of two large international firms,[3] which sparked off dynamic activities in the life sciences sector in Vienna (IMP 2011). By 2011, more than 400 life sciences companies were located in Vienna, almost a quarter is involved in biotechnology and medical technology with 9,100 employees, generating about EUR 1.7 billion in sales (2010). 14,100 persons were employed in academic research for life sciences (LISAvienna 2011b, pp. 3–7). The Vienna life sciences innovation system is especially known for medical biotechnology such as oncology, immunology and neurobiology with expertise in analytical methods and services, diagnostics and diagnostic technologies, microbiology or pharmaceuticals. Another focus of life sciences in Vienna that gained momentum recently is the research on medical technology and devices (Austrian Life Sciences Directory 2009).

From a policy perspective, funding activities specifically dedicated to the life sciences were rather underdeveloped until the late 1990s. Only in 1999, an Austrian biotech programme was introduced which has led to the setup of the life sciences cluster initiative in 2002 (LISAvienna 2011a). Since 2003, the focus of regional research policy lies on life sciences (WWTF 2011) and in 2004, more than 5 % of the Austrian public research budget was invested in biotechnology, covering all parts of the innovation system with a combination of generic and biotech-specific instruments and a focus on education, research and fiscal policy[4] (Reiss et al. 2005, pp. 74–75).

By 2004, science-industry linkages have been identified as the major weakness of the Austrian innovation landscape in international comparison (OECD 2004). First attempts to cope with this issue remained unsuccessful and only the new Austrian university law provided incentives for universities to engage in cooperative projects with industry partners (Schibany et al. 2013). The development of the competence centres programmes can be seen as a milestone in this regard, but also the increased number of thematic programmes made an important contribution. In this context, it is worth noting, that Austrian research policy in general emphasises indirect funding, i.e. tax incentives for research. Institutional funding is to a large extent used by universities, while the number of public non-profit research institutes is rather low compared with OECD countries. Almost fifty percent of the direct public funds are directed to cooperative research projects, particularly to competence centres (Schibany et al. 2009, pp. 73–74).

---

[3] These companies were Boehringer Ingelheim and Genentech.

[4] Reiss et al. (2005, pp. 74–75) used historical data (1994–2002) on policy activities and national performance in biotechnology and benchmarked data regarding biotech policies in the year 2004.

Meanwhile, Austria's culture of cooperation is of high renown worldwide. A broad range of different public funding instruments aims at the promotion of science-industry relations, thus supporting the fast commercialization of scientific findings. Especially, the enlargement of technology and structural programmes since the 1990s has been important in this regard, and competence centres and networks are among the most prominent and successful examples of these funding measures. Companies play an essential role in knowledge and technology transfer, this is reflected even in a strong and sustainable increase in national R&D expenditures and provoked by an increasing number of researching firms and their respective research intensity (Schibany et al. 2009, p. 81).

The Austrian competence centres programme aims at improving science-industry collaboration, the efficiency of industrial research and at building up human capital through a temporary boost of research endeavours. In its current make, the Austrian competence centres programme, COMET, comprises three different types of activities with increasing size, resource endowments and duration: The K-Projects, K1- and K2-Centres (The Austrian Research Promotion Agency – FFG 2010). In K-Projects, running for 3–5 years, participation of at least one academic and three industrial partners is required. These projects are one million EUR per year in maximum size and offer up to 45 % public subsidy. K1-Centres support science-industry collaboration of at least one academic and five industrial partners. For a 7-year period, a maximum annual budget of 1.5 million EUR is foreseen with up to 50 % public support. Compared with this, the K2-Centres are the largest and most challenging scientific endeavours: With at least one academic and five industrial partners, a K2-Centre obtains up to 55 % support over a 10-year period, with a maximum of five million EUR per year. Hereby, international significance and visibility of the research is demanded. As of 2012, about 25K-Projects, 16K1-Centres, and five K2-Centres have been funded through the Austrian competence centres programme (The Austrian Research Promotion Agency – FFG 2010).

A recent ex-post evaluation of the Austrian competence centres programme hints at the remarkable increase in science-industry relations in Austria beginning from the late 1990s (Schibany et al. 2013). However, the study is quite frank about the potential contribution of the competence centres programme to the performance of the innovation system: The respective shares are less than 2 % of total public and 0.9 % of total private expenditures on R&D, which puts the role of the competence centres programme, at least in terms of financial contribution, into perspective. Nevertheless, the evaluation concedes a structuring effect on public funding and on intensifying collaboration between industry and academia. A core question that remains from this evaluation project is, *what would be the right level of public investment*, so that an effect of the competence centres programme could emerge on the system level. This issue can obviously not be addressed by standard evaluation since there is no counterfactual history. This is where our agent-based simulation comes in, and is able to contribute to the discussion by exploring different scenarios in the given innovation system.

19.4 The Model

In our modelling approach, we seek to estimate effects of regional networking induced by competence centres on the innovative performance of the Vienna life sciences innovation system. Building on the SKIN model – Simulating Knowledge Dynamics in Innovation Networks (Pyka and Saviotti 2002; Ahrweiler et al. 2004; Gilbert et al. 2001), we develop an empirically guided, case-specific agent-based model to address this question. In this model, companies, universities, public research and other relevant research organizations are treated as heterogeneous agents that make investment decisions about conducting research, exchange assets with other regional and external agents and produce innovative output.

Referring to the particular empirical case of the life sciences sector in Vienna (Austria), we investigate the long-term impact of a competence centres programme, as a more institutionalized form of regional science-industry cooperation, on innovative behaviour and knowledge-related output. Hereby, we view the life sciences sector as a localised, sectoral innovation system (Malerba 2002) and focus on the performance of both the organisations and the innovation system as a whole. In the model, firms, universities, public and private research organisations are heterogeneous agents with individual research strategies that are able to create innovative output.

The model specification used here (Fig. 19.1) builds on an existing agent-based model of the life sciences innovation system in Vienna (Korber 2012). Innovative output is modelled as an evolutionary process of generating and recombining knowledge assets among organizations. The model consists of three basic elements:

First, the input side is represented by different kinds of *Financial Resources*, both *Private Funds* and *Public Funds*, enabling and triggering the agents' research activities. Private funds include market revenues, private equity, initial public offerings, bank credits and venture capital. Public funds in the general model include direct funds, both bottom-up (initiated by the organisation) and top-down (initiated by government) as well as indirect funding (tax incentives), and institutional funding (for universities and public research organisations).

Second, the *Vienna Life Sciences Innovation System* itself – representing the core element of the model – comprises companies, private and public research organizations, universities, universities of applied sciences and government agencies. Organizations are characterized by individual knowledge endowments and are able to invest in isolated as well as collaborative research activities. In effect, the use of public funds reduces the costs of research for organizations. The *Knowledge Interactions* therein include collaborative research, labour mobility, teaching and lectures, creation of start-ups and spin-offs, licensing, consulting and contract research, as well as extra-regional relations.

The third part of the model describes the *Output* side of the system, focusing on performance in terms the number of *scientific publications*, *patents* and *high-tech jobs*. Agents produce individual or joint research results, which undergo an

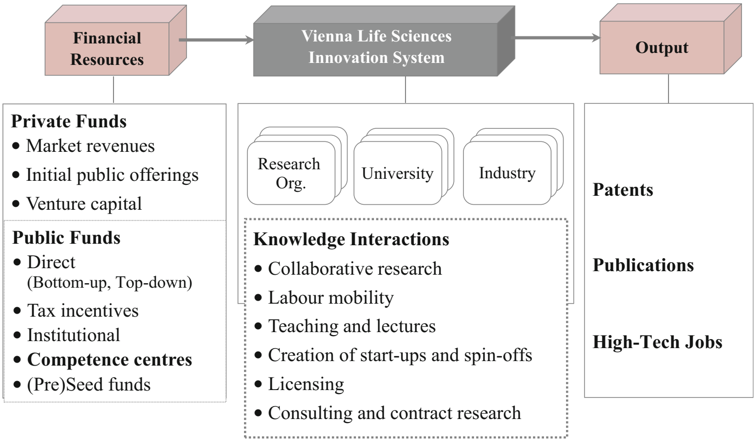


Fig. 19.1 The agent-based model

evaluation process using fitness functions (see Korber 2012 for details). Accordingly, a research result of sufficient fitness is attributed the status of a *scientific publication* or a *patent,* depending on the dominant orientation of the consortium towards science or industry. Following a positive evaluation, this output feeds back on the involved organizations in terms of an enhanced knowledge endowment and modified attributes, e.g. increased financial stock. Repeatedly unsuccessful organizations deplete their financial resources or forget their knowledge and finally exit the system.

For the purpose of our research interest, we focus on the Austrian competence centres programme COMET as additional element of the input side. Hereby, we include the three classes of competence centres differing in terms of duration and size (K, K1 and K2) as described in the previous section. Consortia with industry, university and research organizations can apply for a competence centre. The award of a competence centre grant depends mainly on the consortium composition in terms of the share of science and industry partners and on the quality of proposed research.

19.5 Simulation and Results

In our simulation approach, we focus on the long-term effect of the competence centres programme on selected output indicators in the Vienna life sciences innovation system. Thus, we aim at discerning the impact of this specific policy measure against the background of a larger set of public and private funds in the innovation system. The model was implemented in NetLogo (Wilensky 1999) and calibrated with empirical data on organizational characteristics and public funds from 1999 to

Table 19.1 The simulation scenarios at a glance

Scenarios	Implemented policy measures	Legend
Reference scenario	*Competence centres programme,* Direct funding, Indirect funding, Institutional funding, (Pre) Seed funding	Competence centres programme (perpetuated)
Termination scenario	*Competence centres programme (Year 1–15)*, Direct funding, Indirect funding, Institutional funding, (Pre)Seed funding	Competence centres programme (terminated after 15 years)
Alternative scenario	Direct funding, Indirect funding, Institutional funding, (Pre)Seed funding	No competence centres

2010.[5] Simulations are based on a standard portfolio of research funds from private and public sources as displayed in Fig. 19.1. The additional effect of funding for competence centres is explored using three different scenarios (see Table 19.1).

The *Reference scenario* represents the perpetuation of the empirical calibration period, comprising the standard portfolio of financial resources from Fig. 19.1 including the competence centres programme. The *Termination scenario* is characterized by the standard portfolio with the competence centres programme terminated after 15 years, while the *Alternative scenario* denotes the standard portfolio without competence centre funding. At this point, we assume that the availability of competence centre funding is not restricted exogenously, i.e., all consortia that pass the quality evaluation are funded. As a consequence, total demanded public funds emerge endogenously on the input side and are on an increasing path. This idealistic assumption is adopted in order to explore the critical size of a competence centres programme – an issue that was also discussed in the recent evaluation of the competence centres programme in Austria (Schibany et al. 2013, p. 7).

On the input side, the three scenarios differ with respect to the total annual amount of public funds granted (Fig. 19.2). Obviously, the Reference scenario is the most expensive one in terms of required public funds. On the other hand, we observe that in the Termination scenario funding does not completely descend to the level of the Alternative scenario after the competence centre programme is discontinued. This result might indicate that the competence centres programme has improved the ability of the agents to define high-quality projects, e.g. through newly combined knowledge assets or higher expertise levels. As compared with the Alternative scenario, the system exhibits some memory effect, and alternative public funds can be exploited to a greater extent.

On the output side, we refer to the three hypotheses formulated in Sect. 19.2. Generally speaking, the Alternative scenario reflects an increase of the overall population that results from the basic model specification and the given empirical

[5] Agent population at the setup comprises 75 organizations as given by the database on life sciences in the Vienna region. Detailed parameter settings including reference to the data sources are provided in Tables 19.2, 19.3, 19.4 and 19.5 in the Appendix. Simulations were run for 120 time steps (four time steps per year), representing an observation period of 30 years. All diagrams show mean values over ten simulation runs.

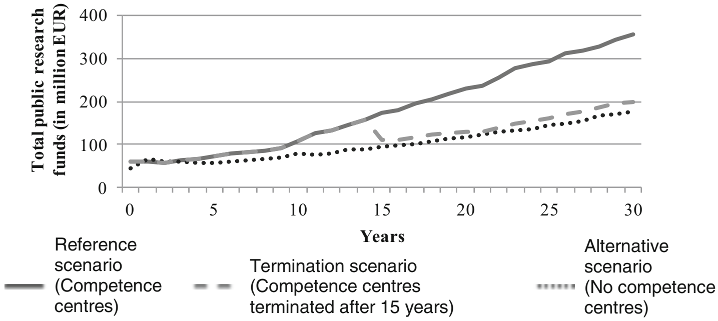


Fig. 19.2 Total public research funding in the three scenarios

calibration. While this model feature can be discussed per se, the Alternative scenario nevertheless offers a baseline for comparison with the other two scenarios.

Hypothesis 1 states that the SME sector benefits more from the competence centres programme than large firms do. In the simulation, agent demographics reveal quite different evolution for different agent types (Fig. 19.3). All three scenarios are characterized by a positive evolution of small and medium sized enterprises, but only in the long run. Especially in the Reference scenario with continuous funding for competence centres, the SME sector benefits regarding the number of firms. This may result from reduced bankruptcy exits, but also from competitive advantages due to greater heterogeneity of the knowledge pool. To a lesser extent, research organizations exhibit a positive effect from competence centres in terms of agent numbers. In comparison, the large enterprise and the university sectors do not exhibit increased growth rates. Moreover, the number of high-tech jobs (Fig. 19.4) shows a considerable impact from the competence centres programme, which is also partly due to the growing number of SMEs in the system. Thus, Hypothesis 1 is supported by the model, stating that competence centres have a positive long-term impact on the size of the SME sector, while the large enterprise sector is weakly affected.

Hypothesis 2 refers to the ability of competence centres to boost patenting activity in the region. Regarding the number of patents, the model produces differing evolutionary paths in each of the scenarios (Fig. 19.4a). While the Alternative scenario (standard portfolio without competence centres) exhibits a moderate growth, the Reference scenario with competence centres leads to a considerable increase in patenting. In the Termination scenario, patenting returns to the slower pace of the Alternative scenario quickly, when competence centre funding stops (after 15 years). Again, a memory effect can be identified. The model thus predicts that competence centres have an impact on patenting activities especially in the long run.

Hypothesis 3 is related to the ability of competence centres to increase scientific publication activity in the region. Simulation results reveal little difference between the three scenarios in this respect (Fig. 19.4b). This means that competence centres in the model do not evoke large amounts of additional scientific output. This may be

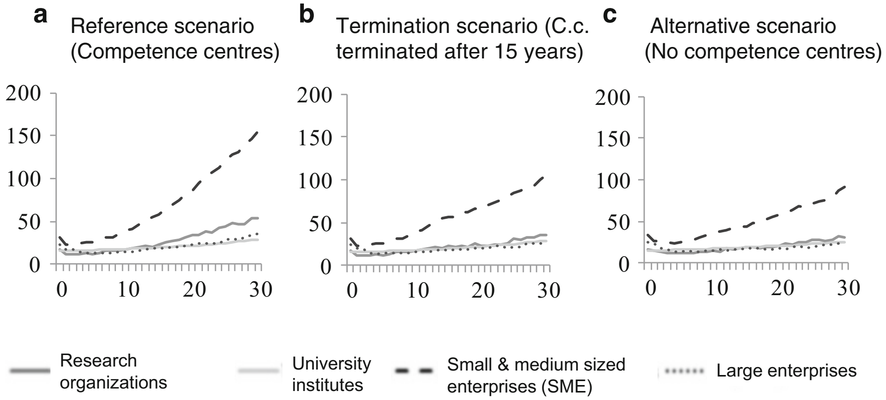


Fig. 19.3 Number of agents in the population (by agent type) (**a**) Reference scenario (Competence centres) (**b**) Termination scenario (C.c. terminated after 15 years) (**c**) Alternative scenario (No competence centres)

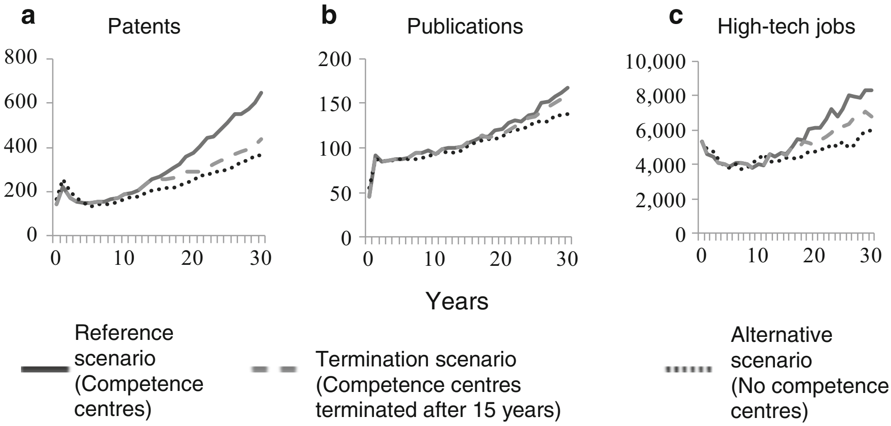


Fig. 19.4 Selected performance indicators by scenarios (**a**) Patents (**b**) Publications (**c**) High-tech jobs

attributable to the design of the competence centres that foresees a smaller share of academic partners in the consortia, and sees academic partners primarily as sources of knowledge. What is interesting, though, is the fact that in the Termination scenario, scientific publications show almost no downturn after discontinuation of the competence centres programme and remain sustainably at the pace of the Reference scenario.

19.6 Summary and Concluding Remarks

In this chapter, we have presented an agent-based model of the Vienna life sciences innovation system, and have simulated long-term effects of a competence centres programme – regarded as a policy induced realisation of geographically localised R&D networks – on regional knowledge production. Our explorative study intends to suggest potential ways how agent-based models could shed light on the debate in theoretical and empirical literature how policy induced local and global networking may affect the innovative performance of regional innovation systems.

In a policy context, such simulation approaches show the potential to support the ex-ante impact assessment of intended policy measures, such as the competence centre programme. Several of the existing competence centres in Austria are in the life sciences field; current programme evaluations have triggered lively debates about their impact. Facing the vagueness of empirical evidence and the open issue of learning from ex-post evaluations, we adopt an ex-ante perspective and use agent-based simulation for experimenting with alternative scenarios. Our study explores potential effects of competence centres on selected output indicators, namely scientific publications, patents and high-tech jobs by comparing three different scenarios: (i) A Reference scenario, where competence centres are continuously funded for 30 years, (ii) a Termination scenario, where funds for competence centres are terminated after 15 years and (iii) an Alternative scenario, where no competence centres exist at all.

The model shows that small and medium sized enterprises benefit more than all other agent types from the competence centres programme. This is represented by a continuous increase referring to the number of small and medium sized enterprises in the agent population during the Reference scenario, particularly in the long run. We assume this is due to an augmented heterogeneity in the joint knowledge pool and to a reduced number of SMEs that have to exit the system due to bankruptcy. The simulation results show only minimal effects of competence centres on the large enterprise sector.

Regarding patenting activities in a region, we find a strong growth of the number of patents during the last 15 years of the Reference scenario, where competence centres are funded throughout the whole 30-years-period. A memory effect is observed, i.e. if funding is stopped after 15 years agents are still able to benefit from raised expertise levels and thus patent and publish more than if competence centres have never existed. Referring to the publication activity of the agents in the region, the simulation results of the three scenarios reveal only very little differences. In contrast, simulated scientific output is not deeply impacted by the competence centres programme. We presume that this lies in the requirement of competence centres to include a smaller share of academic partners that serve mainly as a source of knowledge.

In terms of the literature investigating the geography of R&D networks, the results seem to confirm findings in related empirical works (see, e.g., Chap. 16 of this volume by Breschi and Lenzi) that support the complementary character of local and global networking as central factor for regional innovative performance.

Competence centres can be seen as important policy induced platform to stimulate such a combination of local and global networking activities. The results point to long-term effects in terms of learning and improved innovative performance of industry, especially the SME sector. Competence centres may indeed facilitate science-industry collaboration and commercialization of scientific results, especially in the long run, through institutionalized and long-term R&D networks concentrated in space.

It has to be admitted that the general limitations of the present approach are at least twofold. First, the trade-off between an account of the complexity of the innovation system on the one hand and the quest for analytic clarity and simplicity on the other, will limit the prediction capability of policy-related agent-based models in quantitative terms. Although agent-based modelling gains ground in social sciences and economics, the question regarding when we can consider a model as reliable – and empirically validated – is still not fully solved. Second, accounting for novelty in the presented knowledge generation model is restricted to re-combinations of existing knowledge endowments, which may limit the coverage of radical innovation. Evaluating the quality of the produced knowledge without reference to semantics, remains another fundamental challenge, which we have addressed by using fitness functions.

Future research will introduce statistical analysis for interpretation of the simulation results, and explore alternative fitness functions that are used for evaluating the simulated knowledge outputs. With regard to the spatial dimension of the model, a more explicit account of extra-regional R&D networks needs to be considered. Moreover, other sectors with different knowledge production regimes like machinery, information and communication technology, energy and others will help to endorse our methodological approach. Hereby, the relevance of the simulation results will be improved by involving policymakers, thus relying on the companion-modelling approach (Barreteau et al. 2003), also referred to as expert validation.

Acknowledgements This chapter reports results of research carried out in the framework of the Innovation Economics Vienna – Knowledge and Talent Development Program. The first author gratefully acknowledges the scholarship provided by this program. Both authors gratefully acknowledge support from the Austrian Science Fund (FWF): [I 886-G11] within the project "Innovation networks for regional development – An agent-based simulation approach". Furthermore, the authors would like to thank Manfred M. Fischer, Andreas Pyka and Michael Barber for valuable comments and discussions.

Appendix

The presented model is programmed with NetLogo, version 5.0.3 (Wilensky 1999). The program code for the NetLogo model on which this chapter is based is available from the authors on request. The simulation runs described in Sect. 19.5 are based on the parameter settings given in Tables 19.2, 19.3, 19.4 and 19.5.

Table 19.2 Agent parameter settings (all three scenarios)

Parameter	Value
Initial agent population	75 agents: agent number and structure according to life sciences organisations in the Vienna region as per 1999 (Austrian Life Sciences Directory 2011)
Organization type Differentiated org. type	Industry (SME, LE), University (university, university of applied sciences), Research organization (public, private): according to life sciences organisations in the Vienna region as per 1999 (Austrian Life Sciences Directory 2011), Aurelia (Bureau van Dijk 2010), organisation (business reports, web pages, etc.)
Research fields	Calibrated according to life sciences organisations in the Vienna region as per 1999 (Austrian Life Sciences Directory 2011), complemented by own inquiry
Core competencies	Calibrated according to life sciences organisations in the Vienna region as per 1999 (Austrian Life Sciences Directory 2011), complemented by own inquiry
Expertise level	Uniform random distribution from 0 to 9
Financial stock	According to yearly turnover or budget of particular organisations (Austrian Federal Ministry for Science and Research 2008, p. 62; Bureau van Dijk 2010), organisation (business reports, web pages, etc.)
Employees	Austrian Life Sciences Directory (2011), Aurelia (Bureau van Dijk 2010), organisation (business reports, web pages, etc.), complemented by own inquiry
Researchers	Austrian Life Sciences Directory (2011), Aurelia (Bureau van Dijk 2010), organisation (business reports, web pages, etc.), complemented by own inquiry
Foundation year	Austrian Life Sciences Directory (2011), Aurelia (Bureau van Dijk 2010), organisation (business reports, web pages, etc.), complemented by own inquiry
Research orientation	No research, Basic research, Applied research (Austrian Life Sciences Directory 2011), Aurelia (Bureau van Dijk 2010), organisation (business reports, web pages, etc.), complemented by own inquiry
Share of agents	67 % incremental research attitude
	34 % go-it-alone research strategy (perform own research)
	49 % conservative partner search strategy
	58 % imitative collaboration strategy (during cooperation)

Table 19.3 System parameter settings and empirical calibration (all three scenarios, per quarter)

Parameter	Value	
Random seeds	20, 25, 30, 40, 45, 65, 70, 75, 80, 90	
Receipts and expenditures		
Costs of own research	Research organization agents	EUR 100,207 multiplied by total no. of researchers of the agent
	University agents	EUR 32,702 multiplied by total no. of researchers of the agent
	Research expenses scientific sector and number of employees 2007 (Schibany et al. 2010, p. 143 and p. 149)	
	Industry agents	
	Total no. of employees <50, per quarter	EUR 25,543 multiplied by total no. of researchers of the agent
	Total no. of employees 50–249, per quarter	EUR 44,220 multiplied by total no. of researchers of the agent
	Total no. of employees ≥250, per quarter	EUR 43,250 multiplied by total no. of researchers of the agent
	Research expenses business sector and number of employees 2007 (Schibany et al. 2010, p. 140)	
Costs of cooperative research per partner	Uniform random distribution from	EUR 0–15,000
Labour mobility	Probability of 3 %	
Royalties	Uniform random distribution from	EUR 0–600,000
Remuneration for consulting projects	Uniform random distribution from	EUR 0–60,000
Remuneration for contract research	Uniform random distribution from	EUR 0–60,000
Remuneration for extra-regional relations	Uniform random distribution from	EUR 0–10,000
Private equity	Uniform random distribution from	EUR 0–25,000
	Probability per quarter of (Schibany et al. 2010, p. 141)	5 %
Bank credit	Uniform random distribution from	EUR 0–100,000
Initial financial stock of start-ups	SME: Uniform random distribution from	EUR 0–70,000
	minimum nominal capital for Austrian GmbH: EUR 35,000 (GmbHG 2012)	
	LE: Uniform random distribution from	EUR 0–100,000

(continued)

Table 19.3 (continued)

Parameter	Value	
	minimum share capital for Austrian AG: EUR 50,000 (AktG Aktiengesetz 1965)	
	University: Uniform random distribution from	EUR 0–20,000
	Research org.: Uniform random distribution from	EUR 0–20,000
Revenue per sold innovation	Uniform random distribution from	EUR 0–100,000
Venture capital	Venture capital is limited to investment in seed, start-up, early development and expansion stages. Later stage replacement and buy-out investments are excluded. Venture capital investment data for the life sciences for 2007, Average size per investment per year: USD 791,600: as of 20 January 2012: EUR 611,524 (OECD 2009, p. 96 and p. 101)	
Cash flow from issued shares	EUR 100,000	
SMEs go public	Probability of 0.1875 % (Schibany et al. 2010, p. 141)	

Table 19.4 Switches to determine the agents' option for action (all three scenarios)

Parameter	Parameter settings
Private funds	
Market revenues	ON
Private equity	ON
Initial public offering	ON
Bank credits	ON
Venture capital	ON
Knowledge interactions	
Collaborative research	ON
Labour mobility	ON
Teaching and lectures	ON
Creation of start-ups and spin-offs	ON
Licensing	ON
Consulting and contract research	ON
Extra-regional relations	ON
Agent exits	
Exit due to bankruptcy	ON
Exit because expertise is forgotten	ON

Table 19.5 Empirical calibration of government funds in the three scenarios

Parameter	Parameter settings	*Reference scenario*	*Termination scenario*	*Alternative scenario*
Direct funding	Bottom-up direct funding: Share of funded project expenses: Uniform random distribution from 0 % to 50 % Top-down direct funding: ZIT Calls: calibrated volume and probability referring to particular research fields and core competencies (The Technology Agency of the City of Vienna – ZIT 2010)	ON	ON	ON
Indirect funding	Depends on research expenses, 10 % research premium to industry agents, ceiling amount: 100,000 per agent per year (Legal Information System of the Republic of Austria (RIS) 2012)	ON	ON	ON
Institutional funding	Institutional funding volume per quarter: Uniform random distribution from 0 to 86,953 (Univ.report 08: Austrian Federal Ministry for Science and Research 2008, p. 62)	ON	ON	ON
Creation of competence centres	Calibrated volume, probability and duration, unlimited total funds (i.e. unlimited total number of competence centres) (FFG criteria: K (p. 38), K1 (p. 22), K2 (p. 12): The Austrian Research Promotion Agency – FFG 2010)	ON	Years 1–15: ON, Years 16–30: **OFF**	**OFF**
(Pre)Seed funding	Calibrated volume of (pre)seed funds; Average volume per funded project = 484,800; Uniform random distribution from 0 to 947,000 (AWS funding portfolio: Austria Wirtschaftsservice 2013)	ON	ON	ON

References

Ahrweiler P, Pyka A, Gilbert N (2004) Simulating knowledge dynamics in innovation networks (SKIN). Volkswirtschaftliche Diskussionsreihe, Universität Augsburg, Institut für Volkswirtschaftslehre, No. 267, Dec 2004

AktG Aktiengesetz (1965) current version, § 7, Gesetz 9

Asheim BT, Coenen L (2005) Knowledge bases and regional innovation systems: comparing Nordic clusters. Res Policy 34(8):1173–1190. doi:10.1016/j.respol.2005.03.013

Austria Wirtschaftsservice (2013) aws Life science portfolio – Austrian Life Science Companies 2013. https://www.awsg.at/Content.Node/files/sonstige/LISA_Porfolio_2013.pdf

Austrian Federal Ministry for Science and Research (2008) University report 2008. 2nd corrected edn, III-22 of the annex and III-22 of annexes XXIV. GP – Report – Main document. http://www.parlament.gv.at/PAKT/VHG/XXIV/III/III_00022/imfname_146833.pdf
Austrian Life Sciences Directory (2009) http://www.lifesciencesdirectory.at/. Accessed 13 Jun 2009
Austrian Life Sciences Directory (2011) http://www.lifesciencesdirectory.at/. Accessed 3 Dec 2011
Barreteau O, Antona M, d'Aquino P, Aubert S, Boissau S, Bousquet F et al (2003) Our companion modelling approach. J Artif Soc Soc Simul 6(2):1, http://jasss.soc.surrey.ac.uk/6/2/1.html
Betts JR, Lee CWB (2004) Universities as drivers of regional and national innovation: an assessment of the linkages from universities to innovation and economic growth. Paper presented at the John Deutsch Institute conference on "Higher Education in Canada", Queen's University, Kingston, 13–14 Feb 2004
Carter AP (2007) Measurement of the clustering and dispersion of innovation. In: Polenske KR (ed) The economic geography of innovation. Cambridge University Press, Cambridge/New York/Melbourne/Madrid/Cape Town/Singapore/São Paulo
Castells M (2000) The rise of the network society, 2nd edn. Blackwell, Malden
Cooke P (2002a) Biotechnology clusters as regional, sectoral innovation systems. Int Reg Sci Rev 25(1):8–37
Cooke P (2002b) Regional innovation systems: general findings and some new evidence from biotechnology clusters. J Technol Transf 27(1):133–145
Fabrizio KR (2009) Absorptive capacity and the search for innovation. Res Policy 38:255–267
Fischer MM (2003) The new economy and networking. In: Jones D (ed) New economy handbook. Academic Press/Elsevier, San Diego, pp 343–367
Fischer M (ed) (2006) Innovation, networks, and knowledge spillovers: selected essays. Springer, Berlin/Heidelberg
Fischer M, Fröhlich J (2001) Knowledge, complexity and innovation systems: prologue. In: Fischer M, Fröhlich J (eds) Knowledge, complexity and innovation systems. Springer, Berlin/Heidelberg/New York, pp 1–17
Gilbert N, Pyka A, Ahrweiler P (2001) Innovation networks – a simulation approach. J Artif Soc Soc Simul 4(3):8
GmbHG (2012) Gesetz über Gesellschaften mit beschränkter Haftung. *§ 6 Abs. 1 Satz 2 GmbHG*
Graf H (2006) Networks in the innovation process – local and regional interactions. Edward Elgar, Cheltenham
IMP (2011) Research Institute of Molecular Pathology: IMP. http://www.imp.ac.at/the-imp/history/. Accessed 25 Jan 2011
Jaffe AB (1989) Real effects of academic research. Am Econ Rev 79:957–970
Karlsson C, Andersson A, Cheshire P, Stough R (2009) Innovation, dynamic regions and regional dynamics. In: Karlsson C, Andersson AE, Cheshire PC, Stough RRD (eds) New directions in regional economic development. Edward Elgar, Northampton, pp 1–33
Korber M (2012) Agent-based modelling of complexity in life sciences – with a special emphasis on the impact of public funding on research activities. Book series of the innovation economics Vienna – knowledge and talent development programme PhD- & Master-Theses, Doctoral work, Vol 16, AIT Austrian Institute of Technology GmbH, Vienna University of Economics and Business, PhD- & Master-Theses Series 16
Lambooy JG (2004) The transmission of knowledge, emerging networks, and the role of universities: an evolutionary approach. Eur Plann Stud 12(5):643–657
Legal Information System of the Republic of Austria (RIS) (2012) Einkommensteuergesetz EStG 1988, Prämien für Forschung und Bildung (Forschungsprämie, Bildungsprämie). *§ 108c*
Liebeskind J, Oliver A, Zucker L, Brewer M (1996) Social networks, learning, and flexibility: sourcing scientific knowledge in new biotechnology firms. Organ Sci 3:783–831
LISAvienna (2011a) 1. Oktober 2011 – LISA VR wird zu LISAvienna. http://www.lisavr.at/siteLayout.php. Accessed 25 Nov 2011

LISAvienna (2011b) Vienna life science report: sector survey: facts and directory 2011/2012. LISAvienna, Life Science Austria, Biocom AG, Vienna/Berlin

Malerba F (2002) Sectoral systems of innovation and production. Res Policy 31(2):247–264

McMillan G, Narin F, Deeds D (2000) An analysis of the critical role of public science in innovation: the case of biotechnology. Res Policy 29(1):1–8

OECD (2004) Public-private partnerships for research and innovation: an evaluation of the Austrian experience. OECD, Paris

OECD (2009) OECD biotechnology statistics – 2009. OECD, Paris

Powell W, White D, Koput K, Owen-Smith J (2005) Network dynamics and field evolution: the growth of inter-organizational collaboration in the life sciences. Am J Sociol 110 (4):1132–1205

Pyka A (2002) Innovation networks in economics: from the incentive-based to the knowledge-based approaches. Eur J Innov Manage 5(3):152–163

Pyka A, Saviotti P (2002) Innovation networks in the biotechnology-based sectors. In: Pyka A, Küppers G (eds) Innovation networks: theory and practice. Edward Elgar, Cheltenham/Northampton, pp 75–107

Reiss T, Lacasa ID, Mangematin V, Corolleur F, Enzing C, van der Giessen A et al (2005) Benchmarking of public biotechnology policy. Fraunhofer Institute Systems and Innovation Research, European Commission Enterprise Directorate General, Final report

Saxenian A (1991) The origins and dynamics of production networks in Silicon Valley. Res Policy 20(5):423–437. doi:10.1016/0048-7333(91)90067-Z

Schartinger D, Rammer C, Fischer M, Fröhlich J (2002) Knowledge interactions between universities and industry in Austria: sectoral patterns and determinants. Res Policy 31(3):303–328

Schibany A, Berger M, Dachs B, Falk R, Gassler H, Huber P et al (2009) The Austrian innovation system: appraisals and perspectives. Austrian research and technology report 2009, pp 60–84

Schibany A, Berger M, Buchinger E, Dachs B, Dinges M, Ecker B et al (2010) Life sciences in Österreich. Austrian research and technology report 2010, pp 138–165

Schibany A, Dinges M, Reiner C, Reidl S, Hofer R, Marbler F et al (2013) Ex-post Evaluierung der Kompetenzzentrenprogramme Kplus und Kind/net. Preliminary final report

The Austrian Research Promotion Agency – FFG (2010) FFG Fokus: Kompetenz und Exzellenz – Die Comet-Zentren und -Projekte: Österreichs Flaggschiffe in der Spitzenforschung. Österreichische Forschungsförderungsgesellschaft mbH, 1st edn

The Technology Agency of the City of Vienna – ZIT (2010) Call archive. http://www.zit.co.at/en/funding/current-calls/call-archive.html. Accessed 10 Mar 2010

Uzzi B (1997) Social structure and competition in interfirm networks: the paradox of embeddedness. Adm Sci Q 42:35–67

Bureau van Dijk (2010) Aurelia company database. http://www.bvdep.com/tw/AURELIA.html. Accessed 8 Aug 2010

Van Looy B, Callaert J, Debackere K (2006) Publication and patent behavior of academic researchers: conflicting, reinforcing or merely co-existing? Res Policy 35(4):596–608, http://dx.doi.org/10.1016/j.respol.2006.02.003

Wilensky U (1999) NetLogo. http://ccl.northwestern.edu/netlogo. Accessed 2 Mar 2012

WWTF (2011) Programme & Ausschreibungen – life sciences. http://www.wwtf.at/programmes/ls/. Accessed 15 Jun 2011

Part V
Epilogue

Chapter 20
Synopsis and Outlook

Thomas Scherngell

20.1 Placement of the Volume in the Literature and Contribution

R&D Networks and collaborative knowledge production have attracted increasing attention in the recent past, both in the scientific realm and in the policy sector. They can be characterised as organisational forms supporting knowledge production and diffusion processes by allowing participating actors to get access to new knowledge more rapidly, to learn from each other and to explore and exploit synergies. Today it is widely agreed that such networks play a crucial role in developing and integrating new knowledge in the innovation process, and, thus, have a profound impact on the innovative and economic competitiveness of firms, regions and countries.

Spatial studies of innovation have also increasingly shifted attention to the exploratory and explanatory investigation of the spatial dimension of networks. One of the fundamental questions raised by the theoretical and empirical research concerns the analysis of the complex relationships between R&D networks and geography, as well as the interdependence of network dynamics, spatial economic development and the innovative behaviour of organisations (see, for instance, Autant-Bernard et al. 2007). This recent focus on the geography of R&D networks has made clear the need for methodological advancements in the domain of spatial econometrics and spatial statistics, on the one hand, and, on the other hand, in applying and adjusting methods originally introduced in other disciplines for spatial network analyses.

From a historical perspective, networks – formally and far more generally defined as a set of nodes inter-linked by a set of edges – have long been a subject

T. Scherngell (✉)
Innovation Systems Department, AIT Austrian Institute of Technology, Vienna, Austria
e-mail: thomas.scherngell@ait.ac.at

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_20,

of research in different disciplines. Barthelemy (2011) considers two main research threads – showing distinct disciplinary roots – that are now increasingly coming closer to each other for the spatial analysis of networks in general. The first one broadly covers the realm of Complex Network Analysis (CNA) and Social Network Analysis (SNA), comprising a set of tools stemming from graph theory. In this context, mathematics, mathematical sociology, physics and computer science have initially focused on random networks, starting with the seminal work of Erdös and Rényi (1959) on random graph models. In the 1990s, structures and dynamics of real-world networks have come into focus; Watts and Strogatz (1998) observe real-world network properties to propose new models of random networks. This has been the starting point for investigations of structure and dynamics of all kinds of real-world networks, and the development of a significant number of software tools for network analysis. However, spatial aspects of such networks have usually been touched on only rudimentarily.

The second research stream originates from quantitative geography where structure and drivers of interactions between discrete spatial entities, such as commodity or transportation flows, have been studied since the 1970s. Haggett and Chorle (1969) discuss the role of geographical space for network formation and describe tools and models to characterise such networks. These are similar questions that are nowadays relevant in the context of the spatial analysis of R&D networks that – driven by the focus on the Geography of Innovation (Feldman 1994) – have drawn attention in spatial studies of innovation in the recent past due to recognition of the important role of networks for generating successful innovation (see, for instance, Bergman 2009; Scherngell and Barber 2009).

A great deal of theoretical and empirical research has investigated the spatial dimension of R&D networks over the past decade. The growing research activities in this direction are also related to the increasing availability of datasets and computer capabilities for the analysis of large-scale networks featuring a large number of nodes and links. Today, we have both the tools and the computing capacity to investigate such large-scale networks, including geo-referencing of nodes and edges. However, in methodological terms, studies that investigate spatial structures of R&D networks greatly differ (see also Chap. 2 by Autant-Bernard and Hazir). Also a variety of conceptual approaches to interpreting results from spatial analyses of R&D networks can be identified in these different works. There remains a need for integrating and comparing different methodologies and underlying conceptual models to be used for addressing research questions that are both on the scientific and on the policy agendas.

The present volume – as a joint product of scholars analysing the geography of R&D networks from different angles, from distinct disciplinary backgrounds, using a diverse set of methodologies and producing a range of policy conclusions in diverse spatial and sectoral environments – clearly addresses this research gap in a quite significant and fruitful way. It constitutes – on the one hand – a unique collection of articles presenting methodological advancements for the analysis of R&D networks from different disciplines, and – on the other hand – a distinguished anthology of novel empirical contributions on the relationship between geography and network structures as well as the impact of such networks on knowledge creation and innovative performance of firms, regions or countries.

In line with the three parts of the book, the main findings may be categorised into, *first*, methodological advances for spatial studies of R&D networks; *second*, novel empirical insights into spatial R&D network structures; and, *third*, the impact of R&D network structures and dynamics on knowledge creation and diffusion in a Science, Technology and Innovation (STI) policy context. The first category clearly contributes to integrating and combining methods originating from different disciplines—such as mathematics, computer science or complex systems science—with more traditional spatial analysis methods from geography and regional science. The second and third category is realised by a selection of new empirical works employing cutting-edge methods – some of them presented in this book in an abstract manner – and using novel, systematic datasets on different types of R&D networks at different spatial, sectoral or technological scales. These novel empirical insights significantly contribute to the theoretical debate in the Geography of Innovation literature, such as, for instance, the debate on the interplay between geographically localised and geographically dispersed knowledge flows for regional knowledge creation (see Feldman and Kogler 2010).

In this synopsis, a sketch of the main findings is presented in the following, distinguishing between the methodological and the empirical realm. After that, the section closes with some ideas – derived from the different chapters – for a future research agenda.

20.2 Main Methodological Findings

From the articles gathered in Part II of the volume, we have learnt that methods from CNA and SNA, rooted in graph theory, provide promising and powerful tools that complement traditional spatial analysis methods for spatial studies of R&D networks (for an overview see Chaps. 2 and 3 of the volume). The methodological and analytical approaches usually rely on different conceptual views on how and under which conditions networks are formed. While in CNA, real-world networks are usually considered as one potential realisation of a random graph, game theoretical approaches consider networks as equilibrium of choices or utilities of participating individuals (see Chap. 2). Common analysis tools to study the geography of R&D networks, such as spatial interaction models, take up both views. They relate observed interaction intensities within a network to different influential factors, including origin-specific, destination-specific and (spatial) separation variables, such as geographical distance, and some notion of randomness (see Chap. 2); in newer approaches spatial interaction models also take into account the issue of spatial dependence among flows, referred to as network autocorrelation (see Chaps. 6 and 8). Thus, from a network analytic perspective, they focus on factors at the node and the dyad level (see Chap. 4). However, such approaches often neglect the structural network level referring to the influence of indirect linking structures on tie formation, as for instance, preferential attachment mechanisms. Exponential random graph modelling (ERGM) is presented as a promising methodological approach to analyse the geography of R&D networks, actually taking into account the structural network dimension and its effects on network formation (see Chaps. 2, 4 and 12 for an overview and illustrative applications). ERGM are stochastic models viewing an observed network with a fixed number of nodes as a specific realisation of multiple hypothetical networks with similar properties (see Chap. 4). In analysing the geography of R&D networks, an ERGM can explain the formation of a network by means of regular, local network patterns. Among the set of possible network configurations, higher probability is given to those that are similar to the observed network in terms of these local structures (see Chap. 2).

Another crucial issue in analysing the geography of R&D networks is the identification of so-called communities in networks, roughly defined as sub-networks whose members are more tightly linked to one another than to other members of the network. The volume comprehensively discusses community detection approaches (see Chap. 3) and presents an application to project-based European R&D networks (see Chap. 9). Community detection provides a promising pathway to describing the structural organisation of R&D networks with respect to the existence of relevant substructures that may exert influence on the spatial structure of the whole network. While initially only available for unweighted networks, newer methods in community detection are able to take into account weighted networks that are often more intuitive for describing and analysing R&D collaboration networks (see Chap. 3).

An additional methodological finding of the volume lies in the discussion of SNA related techniques for the analysis of spatial aspects of R&D collaboration networks (see Chaps. 5, 7, 13, 15, 16, and 18). In this context, the volume significantly contributes to the recent debate in economic geography and regional science on the usage of SNA techniques for the spatial analysis of different kinds of complex systems, such as R&D networks (Bergman 2009; Ter Wal and Boschma 2009). To gain deeper understanding of R&D network formation mechanisms, a critical point concerns the value of knowledge that is made available to an innovating actor when gaining a specific network position. A new measure presented in this volume (see Chap. 5), labelled Ego Network Quality (ENQ) index, constitutes a promising approach for investigations in this direction. The measure is intended to capture the value of knowledge available from a node's position in a given network taking into account different structural network dimensions in form of direct and indirect contacts, and – at the same time – accounts for individual characteristics of the nodes in these neighbourhoods. The intrinsic innovation of the measure is that it represents an integrated measure of network position and node characteristics.

However, at the global network level, aspects such as assortativity and hierarchy may play an essential role for explaining the geography of R&D networks that cannot be captured by this index in its present form. Two SNA-related measures, degree distribution and degree correlation, are discussed as promising ways to highlight certain structural network properties that affect the performance of geographically localised R&D networks (see Chap. 7). The degree distribution is interpreted as a measure for the degree of hierarchy in the R&D network under consideration, while the degree correlation serves as a proxy for the level of assortativity, i.e. the propensity of knowledge to flow between more central and more peripheral actors.

Reviewing the methodological findings, the contribution of the articles in this volume is substantial. They outline the potential for integrating and combining methods from different disciplines, and for applying these methods to address prominent research questions. ERGM has been raised as a reasonable pathway for analysing the relationship between network structure and R&D collaboration intensities in geographical space, while SNA-related measures have been proposed to better understand the spatial structure of R&D networks at the global and the local level. Concerning more traditional spatial interaction modelling approaches – which are without doubt still a powerful instrument to describe the geography of R&D networks – new statistical approaches to deal with spatial dependence issues among network links have been discussed and applied.

20.3 Main Empirical Findings and Policy Implications

Besides the methodological contributions described above, the volume includes an extensive selection of empirical works on the geography of R&D networks and collaborations, using cutting-edge methodologies – some of them discussed in an abstract manner in Part II of the volume – and new, systematic datasets on different types of R&D networks at different spatial and sectoral scales. These articles significantly advance theoretical considerations on the geography of R&D networks as well as relevant conclusions in a STI policy context.

At the European level, significant efforts have been devoted to the creation of an European Research Area (ERA), an attempt to overcome fragmentation in European research systems (see Chap. 14). From this perspective, the analysis of the spatio-temporal evolution of R&D networks is of great relevance. The articles presented in Part III of the volume contribute to this research direction in many different aspects. From a longitudinal perspective, Chap. 8 is one of the first contributions providing evidence on integration processes in two distinct European R&D networks over the time period 1999–2006. Using Poisson spatial interaction models accounting for spatial dependence among network links, the results show that the geographical dynamics of progress towards greater integration is higher in policy-induced networks within the European Framework Programmes (FPs) than in co-patenting networks. But such results may differ significantly across different technological domains, as evidenced by the differing spatial structure of thematically homogenous communities identified in FP networks (Chap. 9). It is shown that the degree and evolution of integration may differ across technological areas and that specific technological characteristics should be considered when assessing progress towards ERA. While these comparative contributions are both compelling and politically informative, it seems essential to monitor and evaluate the geographical effects of ERA policy efforts in future studies (see Chap. 14), in particular putting emphasis on the interplay between the policy goal to decrease fragmentation of research systems, on the one hand, and the policy goal to increase and sustain research excellence, on the other hand.

To some extent, the volume provides some indications that these goals are not necessarily contradictory (see Chap. 17), showing a joint positive effect of agglomeration and networking – proxied by participation in the FPs – on regional knowledge production at the level of European NUTS-2 regions. Results of a panel version of the Spatial Durbin Model (SDM) confirm the prevalence of agglomeration effects for regional knowledge production, and, by this, the importance of co-location of 'excellent' actors in terms of innovation capability. However, the study also produces statistical evidence that inter-regional R&D networks in the FPs significantly contribute to regional knowledge production. In this way, lagging regions may indeed increase their innovation capability by participating in FP networks.

Concerning the geography of R&D networks constituted under the FPs, the volume also provides relevant insights at the organisational level and in different technological domains. In terms of inter-organisational R&D collaboration, it is shown that crossing national borders indeed shows a significantly positive rather than negative effect on scientific knowledge generation, measured in terms of reported co-publication activities resulting from FP projects (see Chap. 11). This corresponds to related findings at the regional level, showing that region-pairs located in different countries significantly produce more co-publications after they have jointly participated in the FPs with organisations located in these regions (see Hoekman et al. 2013). However, it must be taken into account that different kinds of knowledge prevalent at specific stages of technological development influences the geographical and structural organisation of networking in a specific technological field. This is demonstrated for FP networks in the field of Global Navigation Satellite Systems (GNSS) (see Chap. 13), also showing that the viability of a technological field might depend on the existence of a cohesive network involving geographically dispersed and distant organisations.

In addition, the involvement of different types of organisations, mainly distinguishing between universities and firms, produces different (spatial) network structures. For the case of multi-lateral collaborations in the biotechnology field (see Chap. 12), it is shown – using ERGM methods – how different types of organisations behave in terms of connectivity, and how they form collaborative arrangements. The results suggest that universities and research organisations tend to participate in more consortiums than firms, and, by this, constitute important bridges for inter-consortium learning.

The importance of these aspects of network structural characteristics for knowledge diffusion raises the question of which factors influence network position. At the regional level, the volume provides evidence – employing a spatial econometric perspective in form of a SDM relationship – that financial R&D resources, human capital and the level of socio-economic development are important general determinants of a region's network positioning in different thematic fields, including

Information and Communication Technologies, Sustainable Development and Life Sciences (see Chap. 15). This implies important conclusions in terms of priority setting in a regional innovation policy context.

The volume significantly contributes to the debate on the local buzz versus global pipelines nature of knowledge creation, providing evidence that knowledge located further away in geographical space becomes increasingly important for regional knowledge production. This idea is further investigated in relation to the inventive productivity of US cities (see Chap. 16). The indicators developed in this chapter capture the propensity of US cities to engage not only in local networking (local buzz) but, more importantly, to entertain knowledge exchanges with actors located in other places (global pipelines), proposing a classification of US cities into four categories, that is *global cities*, *networking cities*, *isolated cities* and *local cities*. Linking these features to the knowledge production performance of the cities allows for estimating the relative importance of local buzz and global pipelines. Networking cities showing both a high level of local and global networking exhibit the highest performance. The volume confirms that external connections indeed are a vital complementary mechanism to enhance local knowledge production and diffusion.

Finally, the volume presents new empirical works on the geography of R&D networks, at a global scale in the form of R&D investment flows between countries and at a local scale in the form of networks in clusters and regional innovation systems. At the level of global R&D interactions, specific home- and host-country characteristics are identified that are conducive or obstructive to cross-border R&D investment flows of foreign affiliates (see Chap. 10). Using a novel data set on firms located in different OECD countries, the results point to the pivotal role of geographical distance, cultural and technological proximity, as well as the availability of human capital in the host country. Thus, strengthening domestic R&D capabilities and raising tertiary enrolment rates are crucial to attracting international R&D flows, which may be an important impetus to tap international knowledge sources.

Such policies have also come into practice at the local level, for instance in the form of so-called competence centres in Austria (see Chap. 19). Such competence centres are policy-induced platforms for supporting geographically localised networking while at the same time stimulating international networking via joint R&D projects with international partners. In an agent-based simulation of the Vienna Life Sciences Innovation System, the potential capability of such policy programs to increase the innovative output of the innovation system is demonstrated. Another example of policy programmes aiming to induce networking activities is the German Leading-Edge Cluster Competition (see Chap. 18). It is shown by means of different SNA measures that the program was quite effective in initiating new cooperations and in intensifying existing linkages. However, the majority of the collaborations which are influenced by the program are between local actors. Furthermore, the results show that small and medium sized enterprises used the chance to connect with local hubs, but not as much among each other.

20.4 A Future Research Agenda

The contribution of the articles presented in this volume to the literature on the geography of networks and R&D collaborations is substantial and manifold, both in terms of methodological advancements, as well as novel empirical insights and policy implications. At the same time, the volume raises urgent questions and issues for future research endeavours in the short and in the long term. In many aspects, the articles of this volume represent the starting point for different future research directions that should be followed.

In concluding from the work presented here, three main future research directions may be distinguished and summarised as follows: *First*, further methodological considerations on how to use methods from CNA and SNA for the spatial analysis of R&D networks, for instance for the projection of networks to the regional level (see Chap. 4), are at the top of the agenda for future research. The criticisms that mainly local rather than global topological network structures are considered when analysing relationships between geography and network formation may be addressed more thoroughly by using the strong portfolio of CNA instruments such as ERGM (see Chaps. 2, 3 and 4).

Second, a stronger focus on network dynamics is necessary to get a deeper understanding on how and why spatial arrangements of R&D networks change over time. Efforts in this direction are partly presented in some articles of the volume that investigate network structures at different points in time (see for instance Chap. 8). However, most applied works view the R&D network under consideration as a static object, where neither new nodes are added nor network links are created or dissolved (see Chap. 2), i.e. the space-time interdependence in the observation of network structures is far from explored. From a CNA perspective, temporal extensions to ERGM for investigating network evolution as a discrete time Markov process (see, for instance, Hanneke et al. 2010) may be a promising step to account for network dynamics (see also Chaps. 2 and 12). In an econometric context, the integration of dynamic panel econometrics in spatial econometric models, in particular for Spatial Durbin Model (SDM) relationships, should be applied more extensively (see, for instance, Elhorst 2012).

Third, investigations on the impact of R&D networks need to be further developed by means of more thorough and precise statistical instruments and, in particular, by increased integration with information sources on the outcome of networks. The need for better data on structures and outcome of R&D networks at different levels of aggregation, in particular the organisational level, is striking. Systemic evidence on the impact of R&D network structures on network outcome in the form of new knowledge and knowledge diffusion is still scarce, though some articles presented here provide some valuable starting points (see Chaps. 16 and 17). A critical point clearly is the difficulty to identify whether the network structure implies the outcome or the vice versa (see Chap. 2); the temporal extensions mentioned above are a crucial element to deal with this causality problem.

References

Autant-Bernard C, Mairesse J, Massard N (2007) Spatial knowledge diffusion through collaborative networks. Pap Reg Sci 86:341–350

Barthelemy M (2011) Spatial networks. Phys Rep 499:1–101

Bergman E (2009) Embedding network analysis in spatial studies of innovation. Ann Reg Sci 43:559–565

Elhorst P (2012) Dynamic spatial panels: models, methods, and inferences. J Geogr Syst 14:5–28

Erdös P, Rényi A (1959) On random graphs. Publicationes Mathematicae 6:290–297

Feldman M (1994) The geography of innovation. Kluwer Academic Publishers, Boston

Feldman M, Kogler DF (2010) Stylized facts in the geography of innovation. In: Hall BH, Rosenberg N (eds) Handbook of the economics of innovation, vol 1. North Holland, Amsterdam, pp 381–410

Haggett P, Chorle RJ (1969) Network analysis in geography. Edward Arnold, London

Hanneke S, Fu W, Xing EP (2010) Discrete temporal models of social network. Electron J Stat 4:585–605

Hoekman J, Scherngell T, Frenken K, Tijssen R (2013) Acquisition of European research funds and its effect on international scientific collaboration. J Econ Geogr 13:23–52

Scherngell T, Barber MJ (2009) Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU framework programme. Pap Reg Sci 88:531–546

Ter Wal ALJ, Boschma R (2009) Applying social network analysis in economic geography: framing some key analytical issues. Ann Reg Sci 43:739–756

Watts DJ, Strogatz DH (1998) Collective dynamics of small-world networks. Nature 393:440–442