# JENA ECONOMIC RESEARCH PAPERS

2017 – 016

Regional Innovator Networks – A Review and an Application with R

by

Holger Graf

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Regional Innovator Networks – A Review and an Application with R*

Holger Graf

Friedrich Schiller University Jena, Economics Department
email: holger.graf@uni-jena.de

November 7, 2017

Abstract

The article serves as an introduction to the empirical analysis of innovation or knowledge networks based on patent data with a particular focus on regional networks. I provide a review of the literature of innovation networks and how it connects to systemic approaches within the field of innovation studies. The SNA methodology is introduced by performing a comparative regional network study based on the publicly available OECD patent databases.

Keywords: Regional Innovation; Network Analysis; Patent Data

JEL Classification: L14, O31, R11

1 Introduction

A few years ago, the term 'network' could easily be considered a buzz word in the field of economics. By now, it has entered the mainstream literature and concepts, such as small worlds, structural holes, or preferential attachment can almost be considered as general knowledge. Networks have entered models ranging from strategic interaction (Goyal 2007; Jackson 2008, 2011) to macroeconomic analyses (Fagiolo 2016), and are used to study phenomena, such as financial networks (Markose, Giansante, and Shaghaghi 2012; Acemoglu, Ozdaglar, and Tahbaz-Salehi 2015), labour markets (Calvó-Armengol and Jackson 2004), knowledge diffusion (Valente 1995; Cowan and Jonard 2004), clusters (Giuliani 2013), venture capital (Sorenson and Stuart 2001), economic development (Hidalgo et al. 2007), and technological trajectories (Verspagen 2007; Martinelli 2012), just to name a few.

In this chapter, I provide an introduction to the empirical analysis of innovation or knowledge networks based on patent data with a particular focus on regional networks. Since there are already several review articles on innovation networks in general (Ozman 2009; Cantner and Graf 2011;

*Prepared for the Handbook of Research Methods and Applications in Industrial Dynamics and Evolutionary Economics.

Phelps, Heidl, and Wadhwa 2012) and some with a particular geographical focus (ter Wal and Boschma 2009; Broekel et al. 2014), the present article is complementary with an emphasis on applications of the method. As such, the chapter serves two main purposes. First, to provide a review of the literature of innovation networks that shows its connections to systemic approaches within the field of innovation studies. Second, to introduce the SNA methodology by performing a comparative regional network study and providing a replicable source code for an empirical analysis in R on the basis of the publicly available OECD patent databases.

I will proceed to pinpoint the relevance of interaction in innovation processes in Section 2, establish the connection between the innovation system approach and networks in Section 3 and provide a short review of the literature on innovation networks in Section 4. The empirical application follows in Sections 5 and 6 where we first set up the data and second perform the descriptive analysis. Conclusions are presented in section 7.

2 Innovation, interaction, and learning

Processes of innovation and learning are decisive factors for economic development and the maintenance of wealth in economies (Schumpeter 1912, 1942). An in-depth understanding of the functioning of processes of innovation and learning is therefore at the core of economics. A fundamental insight of past research is that innovation is generally not pursued in isolation. Already in 1890, Alfred Marshall identified informational spillovers as one of the factors leading to the agglomeration of economic activity. He argued that clustered firms have a better production function than isolated firms (see Krugman 1991). In models of endogenous growth, R&D spillovers are a major source of endogenous growth (Romer 1990; Aghion and Howitt 1992). Zvi Griliches devoted much of his research effort to identify and measure these positive externalities to provide us with a full account of the social returns to innovation (see Griliches 1992, for a review of this line of research). However, it was not until the early 1980's that the black box of the innovation process attracted more research. Studies on collective invention showed that knowledge exchange is not just an externality, but might occur due to an active decision of economic agents (Allen 1983). Kline and Rosenberg (1986) emphasize interactive features of the innovation process within the firm and highlight the importance of heterogeneous external knowledge sources and feedback mechanisms.

The innovation system approach places interaction between heterogeneous actors at the core of a learning economy (Lundvall 1992). Studies on national innovation systems strengthened the view that the functioning of innovation systems is the basis for innovation-based economic development (Nelson 1993). Success stories of regional innovation (Cooke and Morgan 1994; Braczyk, Cooke, and Heidenreich 1998; Keeble et al. 1999) and in particular the prominent example of Silicon Valley (e.g., Saxenian 1994) highlighted the beneficial effects of regional interaction and this argument was strengthened by findings on the regional dimension of knowledge flows (Jaffe, Trajtenberg, and Henderson 1993).

However, recent attempts to identify channels of knowledge spillovers challenge the argument by Jaffe, Trajtenberg, and Henderson (1993) that knowledge flows are localized. Thompson and Fox-Kean (2005), for example, find no intra-national localization effects when using a finer level of technological aggregation in their sample. Breschi and Lissoni (2009) argue that there are more and possibly other relevant dimensions than just geographical proximity and introduce a measure of social proximity between inventors to the 'experiment' by Jaffe, Trajtenberg, and Henderson (1993). They find that social proximity explains most of the identified spillovers and argue that geographical proximity merely facilitates these face-to-face contacts, but geographical proximity is certainly not a sufficient condition for knowledge transmission. This finding is supported by other studies stressing the importance of labor mobility for knowledge flows (Zucker, Darby, and Armstrong 1998; Almeida and Kogut 1999; Møen 2005). The consequence of these findings is that to benefit from knowledge spillovers, actors have to establish relations to others and thereby position themselves within a social network.

Besides – often involuntary – labor mobility, firms can establish such linkages through cooperative R&D to benefit from external knowledge. There is a long tradition in the industrial organization literature that highlights the benefits of cooperation and analyzes the circumstances under which cooperations are most fruitful (for an overview, see Hagedoorn 2002). Some time ago, the focus of economists on cooperative R&D and the importance of external knowledge sources shifted from a bilateral perspective to a multilateral or network view (e.g., Powell, Koput, and Smith-Doerr 1996). A highly recognized special issue in *Research Policy* provides a good demonstration of this shift and in their introduction to that volume, DeBresson and Amesse (1991) bring together the different strands of the literature in various disciplines to make a point for why the concept of networks is so important for understanding the process of innovation. In the same issue, Freeman (1991, p. 499) points out that networks are a "characteristic of national and international innovation systems". His review also shows that while in that special issue the main theme was networks of innovators, hardly any study actually analyzed the structural properties of the networks but still focussed on the bilateral relations in counting shares of firms cooperating or providing information about which types of arrangements (joint ventures, customer–supplier relations, etc.) dominate. However, it was clear that there was a need for analyzing networks not just in a conceptual but also in an empirical way.

3 Innovation systems and innovation networks

The systemic view of innovation processes (e.g., Lundvall 1992; Nelson 1993; Edquist 1997) emphasizes the importance of a division of innovative labor and knowledge transfer between innovative actors. These interactions are embedded within an environment of specific institutions guiding the innovation process. More generally, a system might be defined according to the following three characteristics (Edquist 2004, p. 187):

i) A system consists of two kinds of constituents: there are, first, some kinds of components and, second, relations among them. The components and relations should form a coherent whole (which has properties different from the properties of the constituents).

ii) The system has a function, i.e., it is performing or achieving something.

iii) It must be possible to discriminate between the system and the rest of the world; i.e., it must be possible to identify the boundaries of the system. If we, for example, want to make empirical studies of specific systems, we must, of course, know their extent.[1]

[1]Only in exceptional cases is the system closed in the sense that it has nothing to do with the rest of the world (or because it encompasses the whole world).

Table 1: Five forms of proximity

	Key dimension	Too little proximity	Too much proximity
Cognitive	knowledge gap	misunderstanding	lack of sources of novelty
Organizational	control	opportunism	bureaucracy
Social	trust (based on social relations)	opportunism	no economic rationale
Institutional	trust (based on common institutions)	opportunism	lock-in and inertia
Geographical	distance	no spatial externalities	spatial lock-in

Source: Boschma (2005).

An innovation system is then defined as a network of actors who interact in the processes of the generation, diffusion, and utilization of new, economically useful knowledge under a distinct institutional framework (Cantner and Graf 2003). Institutions are understood as 'sets of common habits, norms, routines, established practices, rules or laws that regulate the relations and interactions between individuals, groups and organizations' (Edquist and Johnson 1997, p. 46). Thus, an innovation system is a group of actors who are related by know-how flows, subject to an institutional environment. These linkages can be built on purpose or emerge unintendedly; they can be incorporated in materials, products or persons or they can develop disembodied through informal knowledge exchange between the actors. In contrast to market exchange, the reciprocity in such a system of informal contacts is less strict, meaning that there is no direct payment for every bit of information, but if someone is free-riding, there are punishments such as exclusion from information channels.

The functioning of innovation systems is based on the information and know-how exchange between actors. The success of the exchange depends on the frequency of interaction, the compatibility of know-how components, and the degree of mutual understanding. These criteria highlight the specific types of *proximity* between actors and constitute the different concepts of innovation systems. Proximity covers a number of dimensions and not just geographical closeness (Torre and Gilly 2000). Boschma (2005) provides an overview of the different concepts of proximity (see table 1) and its impact on innovation, which will be helpful for further discussion. Besides the positive effects of proximity, special attention is paid to the dangers of too much proximity, as the risk of lock-in rises with neglect of knowledge inputs from outside the system.

Cognitive proximity refers to the distance in knowledge space, meaning that people sharing the same knowledge base and expertise may learn from each other (Nooteboom 2000). Firms usually search for new knowledge in close proximity to their existing knowledge-base. In following different trajectories, they develop specific capabilities, which makes them different from other firms. Knowledge from external sources can thus only be absorbed if the cognitive gap is not too large. Boschma (2005) brings forward three points why actors wanting to learn from each other should not be too close in knowledge space: first, the generation of knowledge requires dissimilar but complementary bodies of knowledge in order to trigger creativity and new ideas. Second, too much cognitive proximity might lead to a cognitive lock-in as routines obscure the view on new technologies and opportunities. The third reason is related to involuntary spillovers, as direct competitors with no specific capabilities are very reluctant to share knowledge or learn from each other and would rather not co-locate (Cantwell and Santangelo 2002).

Organizational proximity refers to the organization of production. It is based on similarity and adherence, which means that actors are close if they belong to the same relational framework (i.e., the way interaction and coordination is organized) or share common knowledge and capacities (Torre and Gilly 2000). Often, this type of proximity is associated with networks of reciprocal and trust-based relationships (Boschma 2005). Organizational arrangements, such as networks, reduce uncertainty in the transfer and exchange of knowledge but differ in the extent to which the related actors are autonomous or the degree of control that can be exerted on each other. Low organizational proximity then refers to independent actors on spot markets while high organizational proximity refers to hierarchical relationships within a firm. With respect to the effects on learning and knowledge creation, Boschma (2005) summarizes that organizational proximity is required to control uncertainty and opportunistic behavior, but that too much proximity is often not sufficiently flexible to allow for novelty.

Social proximity might be defined, "[...]in terms of socially embedded relations between agents at the micro-level. Relations between actors are socially embedded when they involve trust, that is based on friendship, kinship and experience" (Boschma 2005, p. 66). This type of proximity draws on the embeddedness literature which states that economic relations are embedded in social context and these social ties influence economic action (Granovetter 1985). While social relations positively influence the degree of knowledge exchange and interactive learning (Breschi and Lissoni 2003), they might lead to a lock-in and an underestimated risk of opportunistic behavior (Uzzi 1997; Boschma 2005).

While social proximity describes relations on the micro-level, *institutional proximity* is associated with the institutional framework at the macro level (norms and values of conduct). Formal institutions, such as laws, rules and language are distinguished from informal institutions, such as values, cultural norms or habits. Depending on the formulation of these institutions, they can support or hamper innovation, e.g., the pharmaceutical industry would clearly suffer if no effective patent system existed. While institutional proximity provides the basis for co-ordination and interactive learning, too much proximity can lead to inertia, and therefore provides insufficient opportunities for change.

Geographical proximity describes the spatial distance between economic actors. While this usually accounts for physical distance, it is influenced by transport and communication infrastructure which might decrease the costs for face-to-face interaction. Geographical proximity enhances innovation as it facilitates knowledge sharing while the institutions co-evolve during the frequency of interaction. However, geographic proximity cannot be considered a sufficient condition for the exchange of tacit knowledge. Blanc and Sierra (1999) describe how large multinationals face difficulties in getting access to the local networks when setting up subsidiaries in these locations. Locations which get too narrowly focussed on certain types of activities might run into a spatial lock-in, so that they are unable to follow new developments. To reduce the risk of such a type of lock-in, it is necessary to either diversify and profit from Jacobs externalities or establish and foster non-local linkages to get access to external knowledge.

Depending on the definition of an innovation system, different types of proximity are pronounced. Within concepts of innovation systems, where the system boundaries are spatial, in terms of national, regional or local innovation systems, aspects of geographical proximity are highlighted. Geographical proximity allows frequent face-to-face contact to exchange information and build a stock of common knowledge. Regional agglomerations of economic activity, such as clusters (Porter 1990), industrial districts (Lazerson and Lorenzoni 1999) or innovative milieu (Camagni 1991, 1995a, 1995b) are said to possess these conditions favorable to knowledge exchange. Focussing on a common culture, language and legislation (institutional proximity), one ends up at national systems of innovation (Freeman 1987; Lundvall 1988; Nelson 1993).

The focus of technological systems (TS) (Carlsson and Stankiewicz 1991) or sectoral innovation systems (SIS) (Breschi and Malerba 1997; Malerba 2002) is on the compatibility of the knowledge-stock. Here, the actors' cognitive proximity is highlighted, which fosters the successful exchange of knowledge. Depending on the type of the knowledge-base, spatial agglomeration of economic activities is not always necessary (Breschi and Malerba 1997).

4 Social network analysis in innovation research

As pointed out in the preceding section, networks are a vital part of an innovation system. However, the innovation system literature is not the only area of application of social network analysis within the field of economics. Actually, the bulk of empirical studies on networks is not set in an innovation system context but is concerned with the measurement of knowledge spillovers (Singh 2005; Breschi and Lissoni 2009), the influence of network position on firm success (Powell, Koput, and Smith-Doerr 1996; Uzzi 1997; Ahuja 2000; Owen-Smith and Powell 2004; Funk 2014), or, more recently, with network evolution (Riccaboni and Pammolli 2002; ter Wal 2014; Tomasello et al. 2017).

But before I provide a short overview of the types of network studies[2], let me introduce some of the most important terms and definitions.[3] A network is a set of nodes between any two of which there exists or does not exist a link (also tie or edge). In undirected networks all links are reciprocal (two firms that cooperate), in directed networks this does not have to be the case (the one I ask for advice does not necessarily have to ask me for advice). Links are either weighted to measure the intensity of interaction or unweighted if we are only interested in the existence of a link. The degree of a node is the number of links that involve that node, and is a natural measure of the node's popularity. Another common measure of node centrality is betweenness which accounts for the position in the whole network (whereas degree is only based on the ego network[4]). It is calculated as the frequency with which an actor is positioned between pairs of other actors on the shortest path connecting them. The distance between two nodes is the number of links in the shortest path between them. Networks might be partitioned into components such that two nodes are members of the same component if and only if there exists a path between them. If there is a path between every pair of nodes, the network consists of only one component.

4.1 Types of networks

Within the innovation literature, we find various types of networks and the following examples are surely not exhaustive. Depending on the research question, nodes might be firms, individuals, regions, patents, etc. and linkages can be any type of relation between these nodes. Table 2

[2] A more comprehensive review is found in Cantner and Graf (2011).

[3] Comprehensive overviews of network concepts are found in Wassermann and Faust (1994), Jackson (2008), or Borgatti, Everett, and Johnson (2013). More details are provided in Section 6.

[4] Ego networks consist of a focal node ("ego") and the nodes to whom ego is linked (the "alters") plus the ties among the alters (Borgatti, Everett, and Freeman 2002).

Table 2: Examples of networks in innovation studies

Network type	Nodes	Edges	Application
Cooperation networks	Firms	Cooperations	Powell, Koput, and Smith-Doerr (1996)
Co-funding networks	Organizations	Cooperations	Broekel and Graf (2012)
Regional networks	Regions	Cooperations	Wanzenböck, Scherngell, and Lata (2015)
Co-authorship networks	Individuals	Publications	Barabasi et al. (2002)
Inventor networks	Individuals	Patents	Fleming, King, and Juda (2007)
Innovator networks	Patent applicants	Inventors	Cantner and Graf (2006)
Citation networks	Patents, publications	Citation	Sorenson, Rivkin, and Fleming (2006)
Product space	Product classes	Co-occurence	Hidalgo et al. (2007)
Knowledge base	Industries	Labour flow	Neffke, Henning, and Boschma (2011)

provides a short list of typical studies concerned with innovation and knowledge exchange. Firms are most commonly linked through joint research projects (cooperation networks). The idea is that collaborating firms exchange knowledge and learn from each other. While in reality the knowledge flows need not be symmetric, it is usually impossible to observe the actual knowledge flows. Information on public funding of joint research is frequently used to study innovation policy or regional embeddedness. Individuals are most often linked through documented joint work. In the case of co-authorship networks the information is based on publication records and two authors are linked if they have jointly written a paper. Inventor networks are a special type of co-authorship networks where linkages are established through joint patents instead of publications. The underlying assumption is that co-authors or co-inventors know each other and have exchanged some information or even learned from each other (Lissoni 2010). Innovator networks are an aggregation of inventor networks. Inventors are the individuals responsible for the development of the patent, but the rights attached to the patent belong to the applicant. Since only economically exploitable knowledge is patentable, we use the term innovator for the applicant even though the patent (the invention) only serves as an input for a possible innovation. Typically these innovators are firms, but they might also be individuals, research organizations or universities. Inventors who appear on patents belonging to distinct innovators are assumed to be mobile inventors who might establish linkages across organizations and serve as a channel for information and/or knowledge transmission. Citation networks constitute directed networks in which the nodes are citing and cited entities and the link is established through the citation. Here, the citation is assumed to represent some indication of knowledge flowing from the cited to the citing node. Patents that cite other patents are used to identify spillovers (Jaffe, Trajtenberg, and Henderson 1993) or technological trajectories (Verspagen 2007). Publications that cite other works can provide information about the importance of certain concepts and their interconnection later authors draw on (Cantner and Graf 2011). Finally, some approaches make use of technological or industry classifications to infer on the relatedness, coherence, or diversity of economic and innovative activities in regions, sectors, or countries.

4.2 Network boundaries

Social network analysis is subject to strong data requirements. Data needs to be more or less complete in terms of actors and relations constituting the system and dynamic studies require the same kind of data over longer time periods. Information derived from patents or publications has these properties and is commonly used to study networks. For example, networks of co-authorship are used to analyze the development of scientific communities (Barabasi et al. 2002; Moody 2004), networks of co-invention can help us understand the evolution of local clusters (Fleming, King, and Juda 2007; Fleming and Frenken 2007), investigate university-industry relations (Balconi, Breschi, and Lissoni 2004), or identify channels for knowledge spillovers (Breschi and Lissoni 2009). The structure and characteristics of clusters and regional networks is explored in a number of studies (Cantner and Graf 2006; Graf and Henning 2009; Graf 2011; ter Wal 2013).

However, one has to be aware of the network boundaries induced by the data choice. Regional boundaries are usually determined by statistical offices so that a region might be defined on the NUTS-3 level or as a Metropolitan Statistical Area (MSA). Technological boundaries might be based on patent (IPC classes) or industry classifications (SIC, NACE) but there are also more sophisticated approaches where search strategies combine classifications and keywords to determine the boundaries of a technological field. The time window of the analysis is a critical assumption regarding link duration and decay. Regarding the appropriate length of time window, there is no consensus among network researchers regarding the correct length. Some assume only the publication year (Wagner and Leydesdorff 2005), others three (Li et al. 2014), five (Li, Liao, and Yen 2013), or seven years (Fleming and Marx 2006), and some do not account for a link decay at all (Breschi and Catalini 2010). While this decision certainly influences the level of network metrics, it should not affect the direction of change in dynamic analyses. Therefore, it is up to the researcher to balance the trade-off between networks of higher density and connectedness on the one side and more observations over time on the other.

Boundaries also include the level of activity where links might only be observed above some threshold level of output. E.g. with co-inventor or co-authorship data, only collaborations that lead to patented or published outcomes can be observed. By focussing on e.g. triadic patents or publications in top journals, this threshold can be varied. Firm strategies towards patenting might differ between sectors. Services are clearly under-represented and important actors, such as venture capitalists or government agencies are not observed. Irrespective of the valuable insights generated by patent networks, these problems of patent data especially apply to comparative sectoral studies but also with regions if they are specialized in industries with different patenting propensities.

A different approach in terms of data collection is taken in a variety of studies on clusters and industrial districts (Giuliani and Bell 2005; Giuliani 2007; Boschma and ter Wal 2007). Here, smaller groups of relevant actors are asked about different types of relations to co-located actors, but due to the high efforts related to this research approach, these studies are usually constrained to specific cases rather than large technological or even national systems. ter Wal and Boschma (2009) discuss the pros and cons of the various forms of data collection for the application of social network analysis in the context of economic geography. Data on R&D alliances is frequently employed to measure the influence of firms' positions within industry networks on some measure of innovative or economic success (Ahuja 2000; Rowley, Behrens, and Krackhardt 2000; Schilling and Phelps 2007). A less frequently used source of relational data can be retrieved from information on public R&D subsidies. Since many research projects are funded as collaborations relations between funded firms and research organizations can be identified. In contrast to data on R&D alliances, such information reveals interaction not only between firms but also involves public research, besides that it usually covers collaborations at an earlier stage of the innovation process (Broekel and Graf 2012).

4.3 Research focus

The last topic in this section concerns the focus of research within the innovation network literature[5]. Looking at performance aspects, we can distinguish two main research approaches. The first approach seeks to explain the performance of actors according to their position within the network. This line of research could be viewed as an advancement of research on R&D alliances and cooperations as it not restricted to observations of the frequency and intensity of bilateral relations in the sense of relational embeddedness (Granovetter 1973), but also makes it possible to analyse the structure of such a network and of individual positions therein captured by the notion of structural embeddedness (Burt 1992; Gulati 1998). If we allow for the possibility that knowledge can flow not just between two actors, but through several nodes of a network, the position in the network becomes of major importance for the acquisition of external knowledge. Ahuja (2000) shows the relevance of indirect linkages in addition to direct forms of interaction. While Ahuja (2000) finds a significant impact of direct ties (degree) on subsequent innovation output, Powell, Koput, and Smith-Doerr (1996) obtain rather ambiguous results concerning the influence of a central position in the network on firm growth. In their study, degree centrality is predicts employment growth, though no significant effects were found for closeness centrality or firms in the main component. In a subsequent study on an expanded dataset ranging from 1988-1997, Powell et al. (1999) conclude that centrality stimulates growth in size and internally-funded R&D and reinforces the use of R&D alliances. However, Baum, Calabrese, and Silverman (2000) show that the influence of alliances on patenting intensity depends on the partner type, and others argue that the environmental context also shapes the structure of relations and their influence on performance. Rowley, Behrens, and Krackhardt (2000) study two networks constituted by horizontal alliances in the US semiconductor and steel industries and test the differential influence of strong ties, weak ties, and the density of the neighbourhood on return on assets. They find an overall negative influence of strong ties on performance, but in the exploitation environment of the steel industry, they find a positive influence. Weak ties are overall positive, but especially so in the exploration context of the semiconductor industry. In his valuable overview of studies relating interfirm networks and innovation, Ozman (2009) concludes that the difficulty of obtaining robust results in some areas is due to the fact that empirical network studies are investigating different sectors in different environmental contexts and define networks in different ways.

The second approach is concerned with the performance of the network as a whole. As noted above, the innovation system approach is mainly interested in the functioning and performance of the system but only few studies explicitly account for the constituting network in terms of actors and relations. An early exception is the study by Leoncini, Maggioni, and Montresor (1996), who analyze sectoral input-output data in a network framework to show differences between the Italian and German technological system. Another example for such an approach is the study by

[5]A more general overview of different research foci within network studies with an emphasis of differences between social sciences and physics can be found in Borgatti et al. (2009) or Hidalgo (2016). I also abstract from the more theoretical approach taken in game theoretic network approaches (see Jackson 2014; Jackson, Rogers, and Zenou 2017)

Fleming, King, and Juda (2007), who show that differences in regional innovative performance can be traced back to the connectedness of the respective inventor networks. Breschi and Lenzi (2016) show that social proximity coupled with dense local cliques among inventors as well as links with inventors outside the city are beneficial for inventive productivity in US cities. Contrary to these findings, Lobo and Strumsky (2008) as well as Bettencourt, Lobo, and Strumsky (2007) find agglomeration effects to be more important than the structure of regional inventor networks for metropolitan innovation performance. While not empirical, Cowan and Jonard (2003; 2004; 2007; 2007) study how different network structures, e.g. small world properties, affect the flow of knowledge by means of formal modelling and subsequent simulation. Graf and Kalthaus (2016) analyse the relationship between structural properties of national research networks and performance in terms of research output and international embeddedness. They find that cohesion of the national network is beneficial, whereas centralization has detrimental effects on performance.

The overarching research question within the range of regional innovation networks is about the relationship between the structure of regional networks and indicators of their economic performance. The regional innovation system (RIS) approach, introduced in section 3, serves as the theoretical basis for such analyses. While the institutional and political framework has a prominent role within existing RIS studies (Cooke 1998), the work on regional networks highlights the interactive character of innovation and explicitly accounts for the network of relations within the RIS (Fritsch and Graf 2011). From the literature, we can derive the hypothesis that local knowledge flows are beneficial to the technological competencies of local actors, which should then translate into innovative and finally economic success. Empirical tests of these hypotheses are, however, still scarce.

5 Data and Methods

5.1 Data collection

The analysis of networks requires relational data. In this Chapter, I will use the example of innovator networks where patent applicants are linked via common inventors (Breschi and Lissoni 2004; Cantner and Graf 2006). Such links might be established because of mobile inventors or in cases of co-application where every inventor on a co-application links all applicants. In the following example, we will work with two databases published by the OECD twice a year. The REGPAT database provides amongst others information on the location of inventors and applicants (OECD 2016b). Since the cleaning of names in working with raw patent data is especially relevant for network studies (Trajtenberg, Shiff, and Melamed 2006; Raffo and Lhuillery 2009), we also rely on the HAN (Harmonized Applicant Names) database (OECD 2016a).

For the analysis of regional networks, we need to define the boundaries of the region and decide which patents are to be included. Here, we use the NUTS3 level[6] to define regions and select all patents with at least one inventor located in that region. The reason for relying on the location of the inventor instead of the applicant lies in the practice of firms with several locations and some research organizations (especially, the Fraunhofer and Max Planck societies in Germany) to assign

[6]NUTS stands for Nomenclature des unités territoriales statistiques and is a system of hierarchical regional definitions for regional statistics. The NUTS3 level corresponds to “small regions”, e.g. cities or small regions. In larger cities which are also 'Bundesländer' (e.g. Berlin or Hamburg), there is no difference between NUTS1, 2 or 3.

Table 3: Number of patents per region

	Stuttgart	Munich	Berlin	Hamburg
EPO patent applications	17,163	34,763	23,773	13,076

patents with the address of the headquarters rather than with the actual location of invention. Of course, the inventor location does not necessarily coincide with the place of invention if inventors commute. However, given that most research is performed in teams (Wuchty, Jones, and Uzzi 2007), we can expect that not all team members commute. Therefore, we should catch most patents that have some connection to the region in focus (Lychagin et al. 2016).

For the present exemplary comparative study, I chose four large cities in Germany. Stuttgart and Munich are both located in the south of Germany with a strong automotive industry, hosting the headquarters of multinational firms, such as Daimler, Porsche, and BMW. During the last years, a vibrant Biotechnology industry developed in Munich. Hamburg is located in the north of Germany with the second largest harbour in Europe. It has a long history in trade and important industries are logistics, aerospace, and food. The capital city of Berlin lost much of its industrial base during the cold war and developed strengths in the research and education sector. After reunification, it developed strengths in the service sector and creative industries but also in life sciences and the health sector.

5.2 Data setup

The various steps of the following analysis are programmed in R (R Core Team 2016). The code and sample data files to reproduce all results, figures, and tables presented here, are published as Graf (2017). I recommend to follow the steps in the file `regpat-network-example.r` while reading the remainder of this article. The following packages are used for the analysis: `igraph` for network analysis (Csardi and Nepusz 2006), `sna` and `network` for network analysis and visualization[7] (Butts 2008, 2015, 2016), `intergraph` for converting between `igraph` and `network` objects (Bojanowski 2015), `stringr` for handling strings during the cleaning of inventor names (Wickham 2012), `xtable` to produce tables for use in LaTeX (Dahl 2014), and `Matrix` for efficient matrix manipulation in sparse formats (Bates and Maechler 2015). References to elements of the source code are set in `typewriter` font.

From the REGPAT database (OECD 2016b), we use applications at the European Patent Office (EPO). After reading the data (code sections 1a or 1b), we extract all patents with at least one inventor located in the focal regions (code section 2). Table 3 gives the total number of patent applications per city (the code to reproduce tables and figures is in code section 5).

In the REGPAT and HAN databases, information on application dates, inventors, and applicants are stored in different tables which can be linked via a unique application identifier. Based on the lists of application IDs that we used for the patent count in Table 3, we create tables of priority year-patent ID (`EPO.IPC.reg`, code section 3.2) inventor-patent ID (`INV.ID.reg`, code section 3.3) and applicant-patent ID combinations (`HAN.ID.reg`, code section 3.4). Inventor names in the raw database are subject to a substantial amount of noise (Raffo and Lhuillery 2009). We

[7]For many routines, `igraph` has shown to be more efficient in terms of computing time and memory than `sna` and `network` packages.

would like to be sure that i) two patents with the same inventor name are actually invented by the same person and ii) two patents by the same person are assigned to the same inventor name. The first problem mostly arises with very common names and is difficult to solve. Typically, inventor name disambiguation is performed by checking the addresses, IPC classes, applicants, and co-inventors (Pezzoni, Lissoni, and Tarasconi 2014), however inventors might be mobile in any of these dimensions. To minimize type II errors (false negatives) due to the second problem, we run a series of data cleaning procedures before matching (code section 3.5). In particular, we change all entries to upper case and remove accents, foreign letters, corrupted strings, and common name modifiers (titles). Since the HAN database on applicant names is already cleaned in a similar way, we can use it as it is for the purpose of this exercise (code section 3.6 provides some examples for shortening applicant names). In the next step, we merge these two tables so that we have one data frame for each region with Patent IDs, clean inventor names, clean applicant names, and the year of application (`HAN.INV.reg`, code section 3.7).

To observe network dynamics, we need to make an assumption regarding tie duration. Patents are applied for at one point in time, however, the process of invention that led to the patent took most likely several years. Even when the patent has been applied for, it would be absurd to assume that co-inventors do not interact any more. Therefore, it is common to take a moving windows approach where ties typically last for three up to seven years (see section 4.2). For the present analysis, we assume seven year moving windows so that a tie established by two patents of the same year will be present for seven periods (code section 3.8). A tie established by two patents applied for in the years 2000 and 2006 respectively, will only be present in one period (the 2000–2006 network). However, we will not consider the link formed by a mobile inventor who is only mentioned on patent A of firm 1 in 2000 and again on patent B of firm 2 in 2007. We will analyse 18 seven-year periods starting with the period 1990–1996 and ending with the period 2007–2013. The number of patents that constitute each network along with the number of nodes is displayed in Figure 1. The number of patents grows in all cities throughout most of the time, Munich and Stuttgart show a sharp increase between the periods ending in 2000 and 2005 and flatten afterwards. For the number of actors, we observe peaks around the 2005 period in all regions. The order among regions is different than for the number of patents. Munich has more patents but less applicants than Berlin throughout the most recent periods and Stuttgart has more patents but the smaller network compared with Hamburg. The main reason for these differences is that both, Munich and Stuttgart, are home to several of the largest patenting companies in Germany so that the average number of patents per applicant is higher.

We now proceed by extracting 2-mode edgelists (combinations of n applicants and m inventors) for each period and region from the data frames in `HAN.INV.reg`. These edgelists are stored in a new object (a list named `HAN.INV`) which contains r (number of regions) elements (again lists) on the first level and i (number of periods) dataframes for each region (code section 3.10). These 2-mode edgelists are converted to adjacency matrices in two steps. First, they are expanded to 2-mode sociomatrices via cross-tabulation (code section 3.11). Second, these $n \times m$ matrices are converted into $n \times n$ adjacency matrices via matrix multiplication ($A = II'$, where A is the adjacency matrix and I is the 2-mode sociomatrix, code section 3.12). We use the `igraph` package to read the sparse adjacency matrices and convert them into a format for network analysis, i.e. an `igraph` object (code section 4.1). The sparse matrix format is used because of its shorter computing time and

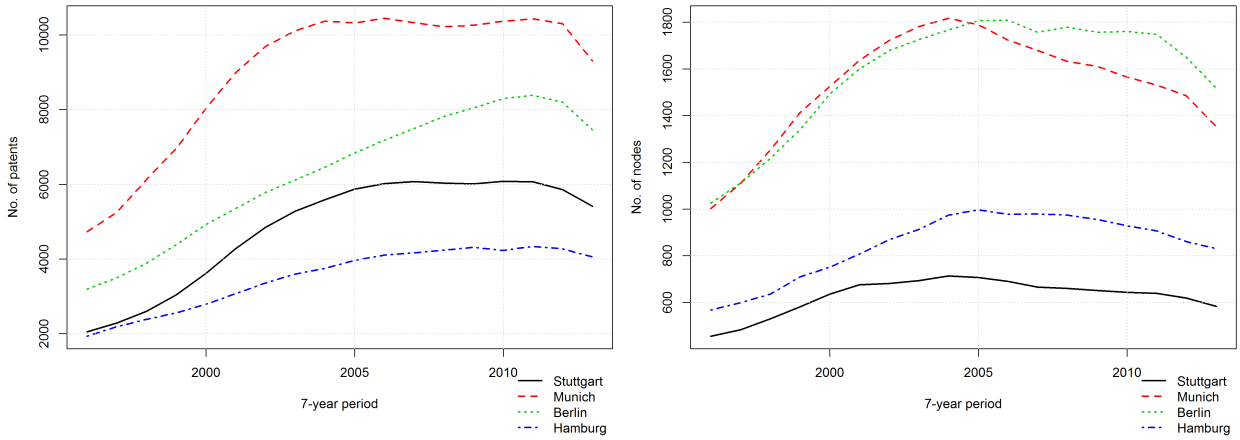


Figure 1: Number of patents and nodes per seven-year period

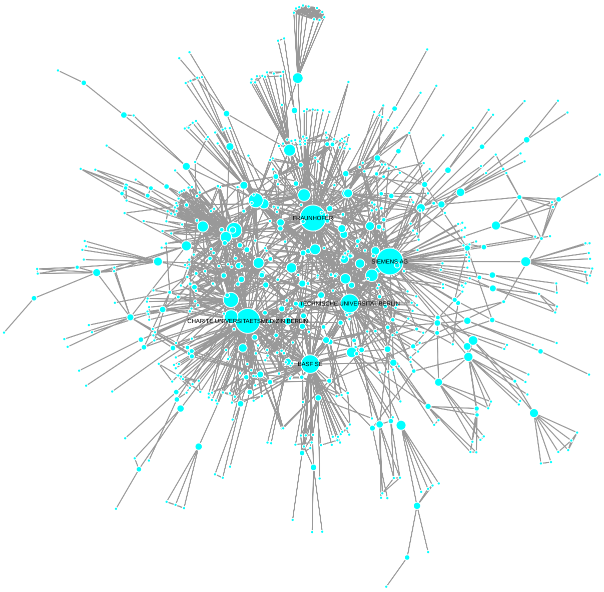
less memory occupation in comparison to the standard format. The resulting object, `network.ig`, is an object of class 'list' with the same dimensions as `HAN.INV` which contains igraph objects for every region and period. in the following section, we proceed to describe some of the characteristic properties and developments of these networks (code section 4.3).

6 Results

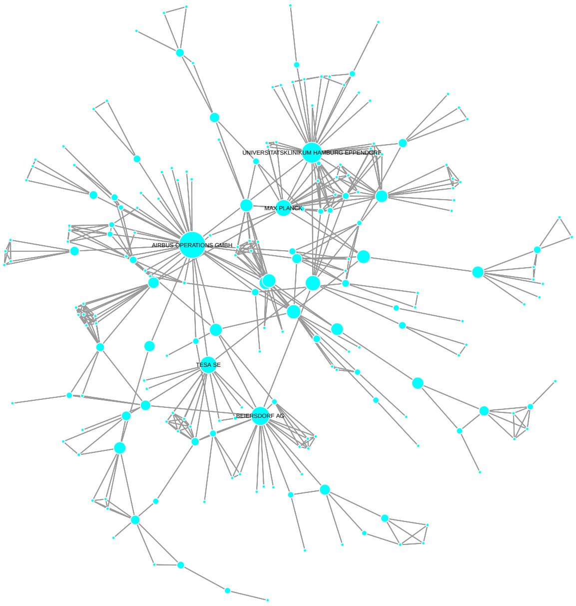
One of the first steps in network analysis is to look at the visualizations. On the one hand it serves as a check if the various steps of data manipulation have been correct. On the other hand, it provides a rough overview of the structure of the respective networks and the importance of individual nodes. Figure 2 shows the main components of the innovator networks for the period 2006–2012[8]. For reasons of clarity, only the names of the five most central nodes according to betweenness centrality are displayed. As a first impression, we observe more tightly knit networks in Berlin and Munich. Regarding the central actors, we find at least two universities or public research institutes (FRAUNHOFER, MAX PLANCK) among the top five in all regions. Siemens is among the top five in all regions except for Stuttgart where Bosch is the most central actor. Lists of the ten most important actors according to betweenness centrality are provided in Table A.1 in the appendix. The names of these actors show the strength of the automotive industry in Stuttgart (Bosch, Daimler, Behr, Porsche) and Munich (BMW, Audi), a dominance of public research in Berlin (seven out of ten), and a rather diverse industry in Hamburg.

As researchers, we do not rely on visual inspection but rather on statistics. Social network analysis provides us with a number of measurement tools for different structural characteristics of networks (see e.g. Wassermann and Faust 1994; Borgatti, Everett, and Johnson 2013). Figure 3 presents comparisons between some measures of cohesion of the regional networks over time. The concept of cohesion is related to how tightly knit a network is. There are several network measures related to cohesion. *Density* is defined as the number of ties expressed as percentage of the number

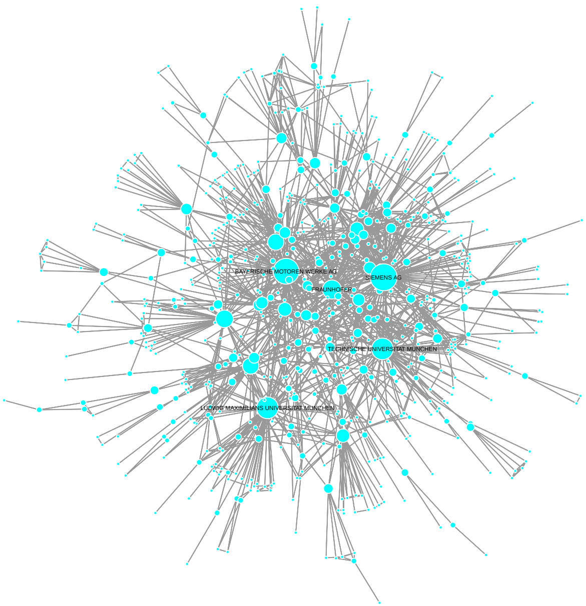
[8]In the example code, the full networks and the main components are plotted for all regions and periods (code section 5, 'Plot networks').



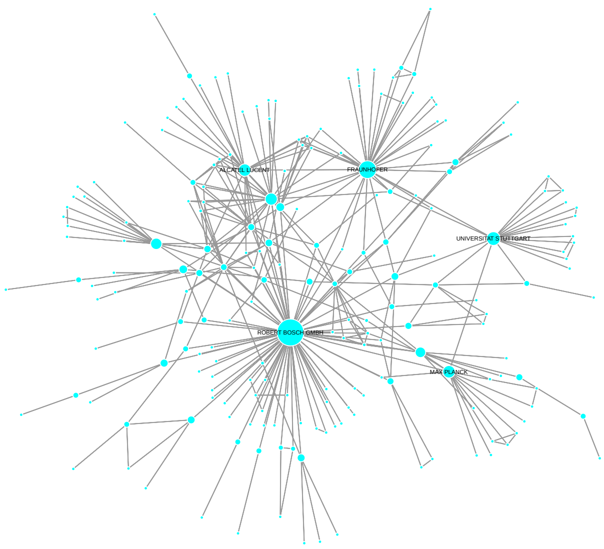
(a) Berlin



(b) Hamburg



(c) Munich



(d) Stuttgart

Figure 2: Network visualizations: main components of innovator networks for the period 2006–2012. Nodes are patent applicants, edges are common inventors. Node size is proportional to betweenness centrality.

of pairs, i.e. potential ties.

$$D = \frac{\sum_{i=1}^{n} d_i}{(n^2 - n)},$$

with n as the size of network and $d_i = \sum_j a_{ij}$ as the degree, i.e. the number of edges incident to a node. For comparisons between networks of the same size, density is highly informative. When comparing networks of different sizes, it might be more meaningful to compare the average number of connections, which is given by the *mean degree* (weighted or unweighted) ($\bar{d}$):

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

Neither density nor mean degree tell us anything about the distribution of links, i.e. high values might be because there are few highly active nodes. Connectedness measures use the distribution of links to measure cohesion. If we assume that knowledge and information can only flow between actors that are connected by any path which means they are members of the same component, the *share of actors in the main component* tells us something about the functionality of a network (or innovation system) to integrate actors in the innovation process. With similar reasoning, we might be interested in actors who do not interact, measured by the *share of isolates*. *Connectedness* is the proportion of pairs that can reach each other, i.e. which are members of the same component. As such, it also accounts for connections within the rest of the component distribution.

$$Connectedness = \frac{\sum_{i \neq j} r_{ij}}{n(n-1)}$$

where r_{ij} is 1 if i and j are in the same component and 0 otherwise. *Fragmentation* is the opposite, i.e. the number of pairs that cannot reach each other

$$Fragmentation = 1 - Connectedness$$

Average distance (L) uses the concept of path length to measure cohesion:

$$L = \frac{\sum_{i \neq j} d_{ij}}{n(n-1)}$$

Assuming that knowledge is more easily transferred face-to-face a lot will be lost when travelling across many actors so that cohesive networks are characterized by low path length. However, distance can only be calculated within components, so that a fragmented network with many small components will have a shorter average distance than a network with a large main component. The *clustering coefficient* measures the presence of high-density substructures within the network.

$$CC_i = \frac{\sum_{j \in N_i} d_j^i}{(k_i^2 - k_i)},$$

with N_i as the neighbourhood of i, $k_i = |N_i|$, and d_j^i the degree of j within N_i. For the network as a whole, we get

$$CC = \frac{1}{n} \sum_{i=1}^{n} CC_i,$$

which is the same as *transitivity*, the proportion of triples with three ties as a proportion of triples with two or more ties. The latter two measures are relevant for the identification of small world networks which are characterized by high clustering and short average path length (Watts and Strogatz 1998). Small world networks can be considered an ubiquitous phenomenon as they appear in man-made, biological, ecological, and technological systems (Uzzi, Amaral, and Reed-Tsochas 2007) and have attracted a great deal of attention across disciplines since the 1950s (Schnettler 2009).

Without going into all details, Figure 3 gives an impression of the four networks in which Berlin and Munich show high levels of cohesion in comparison with Stuttgart and especially Hamburg. All regions experienced a peak in terms of mean degree during the 2000s with a period of less interaction in recent years. Except for Stuttgart, connectivity remains high but the general trend towards more interaction seems to have stopped, as can also be seen by looking at the recently increasing share of isolated actors. With the exception of Hamburg, there also seems to be a trend towards more open networks with declining transitivity.

Regarding power or prominence of individual actors, there are several concepts to measure individual centrality (Freeman 1978). Degree centrality defines the most central actor as the one with most relations, i.e. degree. Degree centrality of i as share of all possible relations

$$C_i^D = d_i/(n-1)$$

As the name says, closeness centrality assumes that central actors are in close distance to many other actors. It is defined as the inverse of the sum of distances to all other nodes

$$C_i^C = \left(\sum_{i \neq j} d_{ij} \right)^{-1}$$

or normalized

$$C_i^{Cn} = \frac{n-1}{\left(\sum_{i \neq j} d_{ij} \right)}$$

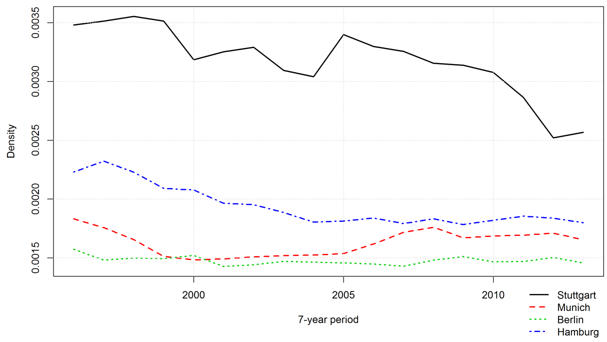
Since distance between nodes in different components is infinite, closeness is only defined for connected networks or within the main component. However, it is possible to assume a finite distance between unconnected nodes and calculate closeness for all networks.

Betweenness centrality is about the control of information flows. High-betweenness vertices lie on a large number of non-redundant shortest paths between other vertices, i.e. they can be thought of as ‘boundary spanners’.

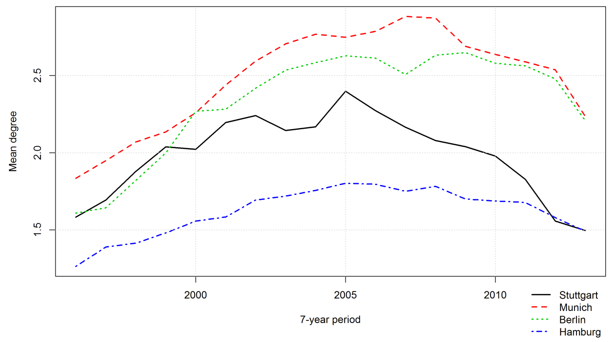
$$C_i^B = \sum_{j<k} \frac{g_{jik}}{g_{jk}}, \forall i \neq j, k$$

with g_{jk} as the number of geodesics between j and k and g_{jik} as the number of geodesics between j and k through i.

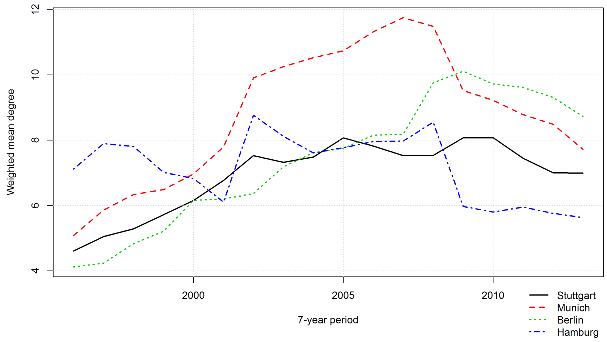
Eigenvector centrality can be viewed as a weighted version of degree centrality, where links to central nodes count more than relations to peripheral actors. Centrality of each actor is proportional to the sum of the centralities of those actors to whom he or she is connected. Using the adjacency



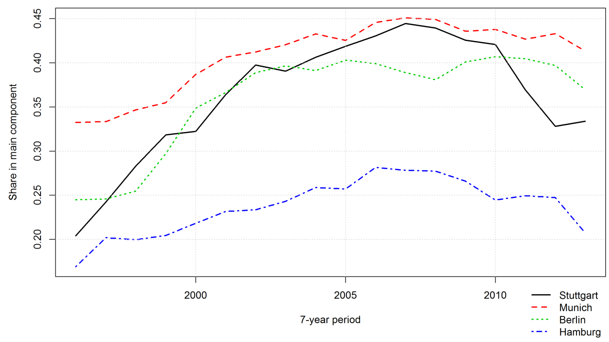
(a) Density



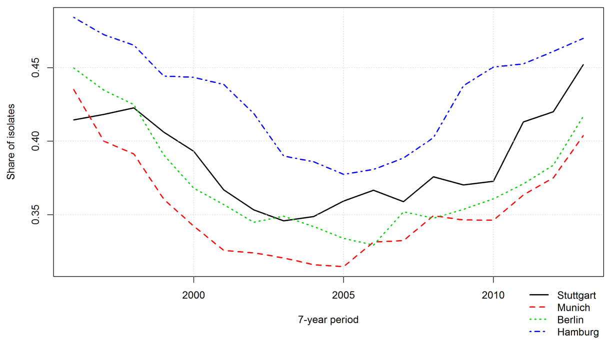
(b) Mean degree



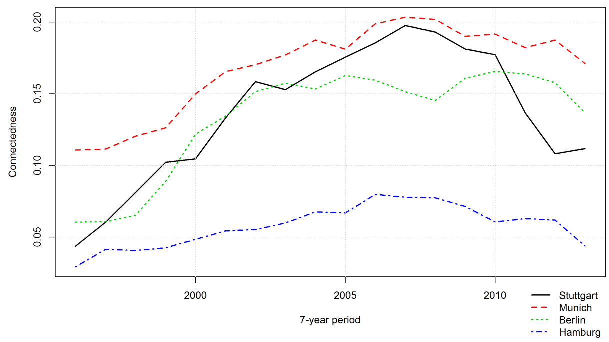
(c) Weighted mean degree



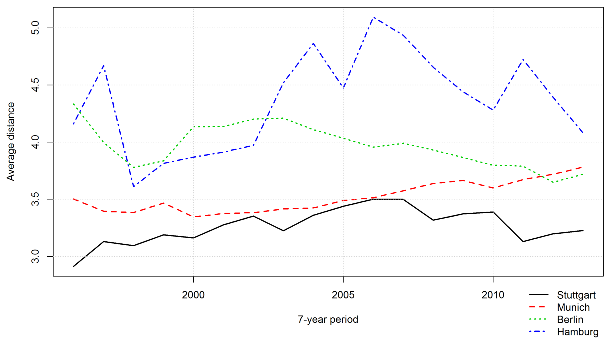
(d) Share in main component



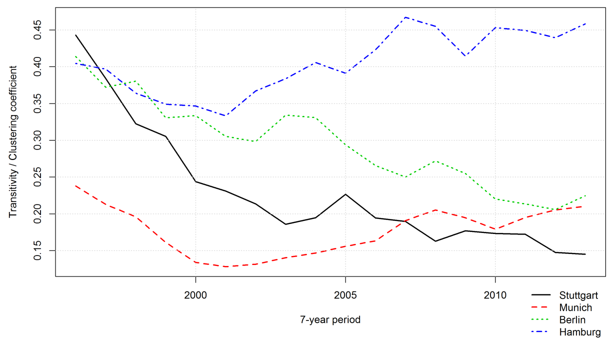
(e) Share of isolates



(f) Connectedness

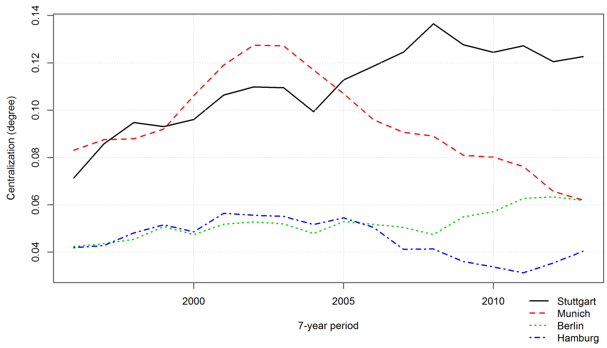


(g) Average distance

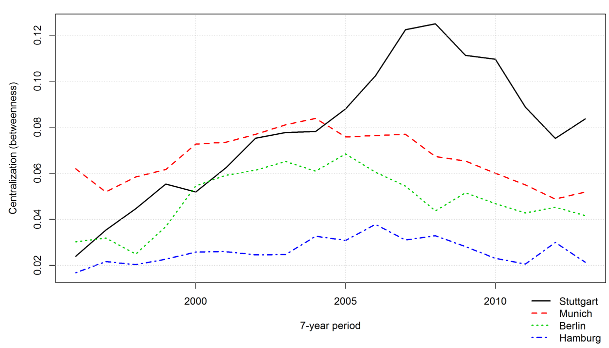


(h) Transitivity / Clustering coefficient

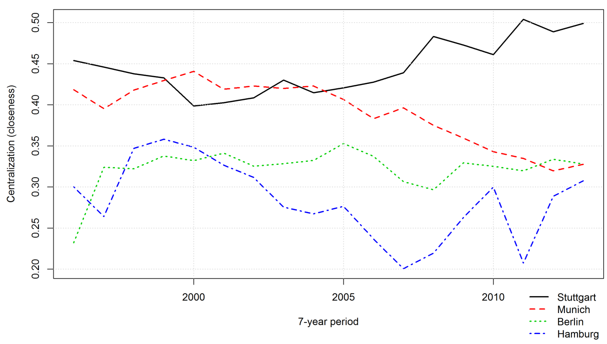
Figure 3: Measures of cohesion.



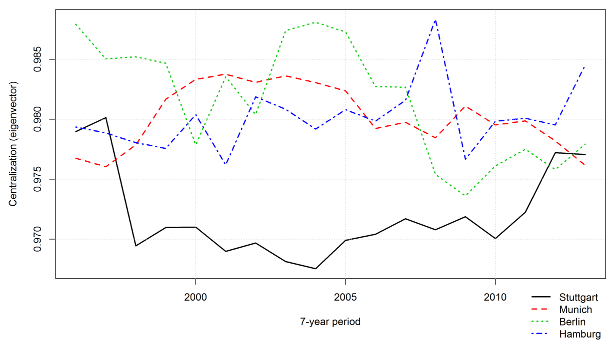
(a) Degree



(b) Betweenness



(c) Closeness (within main component)



(d) Eigenvector

Figure 4: Measures of centralization.

matrix to find eigenvector centrality, the centrality score of vertex i can be defined as:

$$C_i^{Ev} = \lambda \sum_{j \in N(i)} C_j^{Ev} = \lambda \sum_{j=1}^{n} a_{i,j} C_j^{Ev}$$

where C_i^{Ev} is the eigenvector centrality score and λ is the eigenvalue.

We use these concepts to derive structural properties of the whole graph. *Centralization* measures should "[...]index the tendency of a single point to be more central than all other points in the network [...and are ...] based on differences between the centrality of the most central point and that of all others" (Freeman 1978, p.227). These measures of network centralization are calculated by summing deviations of all individual centrality scores from the score of the most central actor and dividing it by the maximum possible value (usually a star network). For eigenvector centrality the most centralized structure is the graph with a single edge (and potentially many isolates).

Applying these four centralization indices to the regional networks, we see that these conceptual differences only partly lead to different interpretations (see Figure 4). Centralization based on Degree, Betweenness, and Closeness, all show that Stuttgart – especially towards the end – is most and Hamburg is least centralized. Munich became less centralized during the later periods and now has a similar structure as Berlin in that respect. Centralization based on eigenvector centrality leads to a different picture, with Stuttgart emerging as the least centralized network throughout most periods.

While the analysis in the present chapter stops at this point, research certainly does not. In any analysis of regional performance it would be important to include the interplay between internal and external social proximity (Graf 2011; Fritsch and Graf 2011; Breschi and Lenzi 2016) and control for the cohesion and diversity of regional knowledge bases, e.g. by including measures of related and unrelated variety (Frenken, Oort, and Verburg 2007; Neffke, Henning, and Boschma 2011). In order to explain processes of tie formation and network evolution, as in ter Wal (2014), the `R-Siena` package (Ripley et al. 2017) provides well developed routines.

7 Conclusions

The purpose of this Chapter was twofold. First, it showed the value of applying methods of social network analysis for the analysis of innovation systems by reviewing and linking the relevant literatures. Second, as part of a Handbook on methodology, it provided an application of social network analysis tools in a comparative study of four large German cities. We used information on patent application at the EPO as published in the REGPAT database published by the OECD. The OECD HAN database with harmonized applicant names proved valuable in disambiguating applicants. Innovator networks were reconstructed by linking applicants via common inventors and a series of cleaning steps were performed to increase the reliability of inventor name matching. The availability of patent data and the use of automated cleaning show their benefits in an easy to use and comparably fast analysis of regional or other types of innovation systems over time. Data setup, preparation, and analysis were all performed with R (R Core Team 2016) and are documented in form of a tutorial which can be accessed by interested readers. The tutorial is supposed to provide young researchers or researchers new to the topic an easy entry to the analysis of innovator networks. For more fine grained analyses that account for different actors types, actor location, or other node and edge characteristics, additional effort needs to be supplied.

References

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi. 2015. "Systemic risk and stability in financial networks". *American Economic Review* 105 (2): 564–608. doi:10.1257/aer.20130456.

Aghion, P., and P. Howitt. 1992. "A model of growth through creative destruction". *Econometrica* 60 (2): 323–351. doi:10.2307/2951599.

Ahuja, G. 2000. "Collaboration networks, structural holes, and innovation: A longitudinal study". *Administrative Science Quarterly* 45 (3): 425–455. doi:10.2307/2667105.

Allen, R. C. 1983. "Collective invention". *Journal of Economic Behavior and Organization* 4:1–24. doi:10.1016/0167-2681(83)90023-9.

Almeida, P., and B. Kogut. 1999. "Localization of knowledge and the mobility of engineers in regional networks". *Management Science* 45 (7): 905–917. doi:10.1287/mnsc.45.7.905.

Balconi, M., S. Breschi, and F. Lissoni. 2004. "Networks of inventors and the role of academia: An exploration of Italian patent data". *Research Policy* 33:127–145. doi:10.1016/s0048-7333(03)00108-2.

Barabasi, A., H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. 2002. "Evolution of the social network of scientific collaborations". *Physica* A 311 (3–4): 590–614. doi:10.1016/s0378-4371(02)00736-7.

Bates, D., and M. Maechler. 2015. *Matrix: Sparse and Dense Matrix Classes and Methods.* R package version 1.2-0. http://CRAN.R-project.org/package=Matrix.

Baum, J. A. C., T. Calabrese, and B. S. Silverman. 2000. "Don't Go It Alone: Alliance Network Composition and Startups' Performance in Canadian Biotechnology". *Strategic Management Journal* 21 (3): 267–294. http://www.jstor.org/stable/3094188.

Bettencourt, L. M. A., J. Lobo, and D. Strumsky. 2007. "Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size". *Research Policy* 36 (1): 107–120. doi:10.1016/j.respol.2006.09.026.

Blanc, H., and C. Sierra. 1999. "The internationalisation of R&D by multinationals: a trade-off between external and internal proximity". *Cambridge Journal of Economics* 23 (2): 187–206. doi:10.1093/cje/23.2.187.

Bojanowski, M. 2015. *intergraph: Coercion Routines for Network Data Objects.* R package version 2.0-2. http://mbojan.github.io/intergraph.

Borgatti, S. P., M. G. Everett, and L. C. Freeman. 2002. *Ucinet for Windows: Software for Social Network Analysis.* Harvard: Analytic Technologies.

Borgatti, S. P., M. G. Everett, and J. C. Johnson. 2013. *Analyzing Social Networks.* London: Sage.

Borgatti, S. P., A. Mehra, D. J. Brass, and G. Labianca. 2009. "Network Analysis in the Social Sciences". *Science* 323 (5916): 892–895. doi:10.1126/science.1165821.

Boschma, R. A. 2005. "Proximity and innovation: A critical assessment". *Regional Studies* 39 (1): 61–74. doi:10.1080/0034340052000320887.

Boschma, R. A., and A. L. J. ter Wal. 2007. "Knowledge networks and innovative performance in an industrial district: The case of a footwear district in the south of Italy". *Industry and Innovation* 14 (7): 177–199. doi:10.1080/13662710701253441.

Braczyk, H.-J., P. Cooke, and M. Heidenreich. 1998. *Regional Innovation Systems: The Role of Governances in a Globalized World.* London: UCL Press.

Breschi, S., and F. Lissoni. 2004. "Knowledge Networks from Patent Data: Methodological Issues and Research Targets". Chap. 28 in *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*, ed. by H. Moed, W. Glänzel, and U. Schmoch, 613–643. Dordrecht: Springer Netherlands.

Breschi, S., and C. Catalini. 2010. "Tracing the links between science and technology: An exploratory analysis of scientists' and inventors' networks". *Research Policy* 39 (1): 14–26. doi:10.1016/j.respol.2009.11.004.

Breschi, S., and C. Lenzi. 2016. "Co-invention networks and inventive productivity in US cities". *Journal of Urban Economics* 92:66–75. doi:10.1016/j.jue.2015.12.003.

Breschi, S., and F. Lissoni. 2003. *Mobility and Social Networks: Localised Knowledge Spillovers Revisited.* Working Paper 142. Cespri.

— . 2009. "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows". *Journal of Economic Geography* 9 (4): 439–468. doi:10.1093/jeg/lbp008.

Breschi, S., and F. Malerba. 1997. "Sectoral innovation systems: Technological regimes, Schumpeterian dynamics, and spatial boundaries". In *Systems of Innovation: Technologies, Institutions and Organizations*, ed. by C. Edquist, 130–156. London: Pinter.

Broekel, T., P.-A. Balland, M. Burger, and F. van Oort. 2014. "Modeling knowledge networks in economic geography: a discussion of four methods". *The Annals of Regional Science* 53 (2): 423–452. doi:10.1007/s00168-014-0616-2.

Broekel, T., and H. Graf. 2012. "Public research intensity and the structure of German R&D networks: a comparison of 10 technologies". *Economics of Innovation and New Technology* 21 (4): 345–372. doi:10.1080/10438599.2011.582704.

Burt, R. S. 1992. *Structural Holes: The Social Structure of Competition.* Cambridge, MA: Harvard University Press.

Butts, C. T. 2015. *network: Classes for Relational Data.* R package version 1.13.0. The Statnet Project (http://statnet.org). http://CRAN.R-project.org/package=network.

— . 2016. *sna: Tools for Social Network Analysis.* R package version 2.4. https://CRAN.R-project.org/package=sna.

— . 2008. "Social network analysis: A methodological introduction". *Asian Journal of Social Psychology* 11 (1): 13–41. doi:10.1111/j.1467-839X.2007.00241.x.

Calvó-Armengol, A., and M. O. Jackson. 2004. "The effects of social networks on employment and inequality". *American Economic Review* 94 (3): 426–454. doi:10.1257/0002828041464542.

Camagni, R. 1995a. "Global network and local milieu: Towards a theory of economic space". In *The Industrial Enterprise and its Environment: Spatial Perspectives*, ed. by G. Conti, E. Malecki, and P. Oinas. Avebury: Aldershot.

— . 1991. "Local milieu, uncertainty and innovation networks: Towards a new dynamic theory of economic space". In *Innovation Networks, Spatial Perspectives*, ed. by R. Camagni. London, New York: Belhaven Press.

— . 1995b. "The concept of innovative milieu and its relevance for public policies in European lagging regions". *Papers in Regional Science* 74:317–340. doi:10.1111/j.1435-5597.1995.tb00644.x.

Cantner, U., and H. Graf. 2011. "Innovation Networks: formation, performance and dynamics". Chap. 15 in *Handbook on the Economic Complexity of Technological Change*, ed. by C. Antonelli, 366–394. Cheltenham, UK: Edward Elgar. http://www.e-elgar.com/bookentry_main.lasso?id=13391.

— . 2003. "Innovationssysteme und kollektive Innovationsprozesse". In *Strukturen und Strategien in einem Innovationssystem: Das Beispiel Jena*, ed. by U. Cantner, R. Helm, and R. Meckl, 21–44. Sternenfels: Verlag Wissenschaft & Praxis.

— . 2006. "The network of innovators in Jena: An application of social network analysis". *Research Policy* 35 (4): 463–480. doi:10.1016/j.respol.2006.01.002.

Cantwell, J. A., and G. D. Santangelo. 2002. "The new geography of corporate research in information and communications technology (ICT)". *Journal of Evolutionary Economics* 12 (1–2): 163–197. doi:10.1007/978-3-7908-2720-0_17.

Carlsson, B., and R. Stankiewicz. 1991. "On the nature, function and composition of technological systems". *Journal of Evolutionary Economics* 1 (2): 93–118. doi:10.1007/bf01224915.

Cooke, P. 1998. "Introduction: Origins of the concept". In *Regional Innovation Systems: The Role of Governances in a Globalized World*, ed. by H.-J. Braczyk, P. Cooke, and M. Heidenreich, 2–25. London: UCL Press.

Cooke, P., and K. Morgan. 1994. "The regional innovation system in Baden-Wurttemberg". *International Journal of Technology Management* 9 (3/4): 394–429. https://www.inderscienceonline.com/doi/abs/10.1504/IJTM.1994.025582.

Cowan, R., and N. Jonard. 2007a. "Merit, approbation and the evolution of social structure", *Journal of Economic Behavior and Organization* 64 (3-4): 295–315. doi:10.1016/j.jebo.2006.11.005.

— . 2004. "Network structure and the diffusion of knowledge". *Journal of Economic Dynamics and Control* 28 (8): 1557–1575. doi:10.1016/j.jedc.2003.04.002.

— . 2007b. "Structural holes, innovation and the distribution of ideas". *Journal of Economic Interaction and Coordination* 2 (2): 93–110. doi:10.1007/s11403-007-0024-0.

— . 2003. "The dynamics of collective invention". *Journal of Economic Behavior and Organization* 52 (4): 513–532. doi:10.1016/S0167-2681(03)00091-X.

Csardi, G., and T. Nepusz. 2006. "The igraph software package for complex network research". *InterJournal* Complex Systems:1695. http://igraph.org.

Dahl, D. B. 2014. *xtable: Export tables to LaTeX or HTML*. R package version 1.7-4. http://CRAN.R-project.org/package=xtable.

DeBresson, C., and F. Amesse. 1991. "Networks of innovators: A review and introduction to the issue". *Research Policy* 20 (5): 363–379. doi:10.1016/0048-7333(91)90063-v.

Edquist, C. 1997. "Systems of innovation approaches – their emergence and characteristics". In *Systems of Innovation: Technologies, Institutions and Organizations*, ed. by C. Edquist, 1–35. London: Pinter.

— . 2004. "Systems of innovation – perspectives and challenges". In *Oxford Handbook of Innovation*, ed. by J. Fagerberg, D. C. Mowery, and R. R. Nelson. Oxford, UK: Oxford University Press.

Edquist, C., and B. Johnson. 1997. "Institutions and organisations in systems of innovation". In *Systems of Innovation: Technologies, Institutions and Organizations*, ed. by C. Edquist, 41–63. London: Pinter.

Fagiolo, G. 2016. "The Empirics of Macroeconomic Networks: A Critical Review". In *Complex Networks and Dynamics: Social and Economic Interactions*, ed. by P. Commendatore, M. Matilla-García, L. M. Varela, and J. S. Cánovas, 173–193. Cham: Springer International Publishing. doi:10.1007/978-3-319-40803-3_7.

Fleming, L., and K. Frenken. 2007. "The evolution of inventor networks in the Silicon Valley and Boston regions". *Advances in Complex Systems* 10 (1): 53–71. doi:10.1142/S0219525907000921.

Fleming, L., I. King Charles, and A. I. Juda. 2007. "Small Worlds and Regional Innovation". *Organization Science* 18 (6): 938–954. doi:10.1287/orsc.1070.0289.

Fleming, L., and M. Marx. 2006. "Managing creativity in small worlds". *California Management Review* 48 (4): 6–27. doi:10.2307/41166358.

Freeman, C. 1991. "Networks of innovators: A synthesis of research issues". *Research Policy* 20 (5): 499–514. doi:10.1016/0048-7333(91)90072-x.

— . 1987. *Technology and Economic Performance: Lessons from Japan.* London: Pinter.

Freeman, L. C. 1978. "Centrality in social networks conceptual clarification". *Social Networks* 1 (3): 215–239. doi:10.1016/0378-8733(78)90021-7.

Frenken, K., F. van Oort, and T. Verburg. 2007. "Related Variety, Unrelated Variety and Regional Economic Growth". *Regional Studies* 41 (5): 685–697. doi:10.1080/00343400601120296.

Fritsch, M., and H. Graf. 2011. "How sub-national conditions affect regional innovation systems: The case of the two Germanys". *Papers in Regional Science* 90 (2): 331–353. doi:10.1111/j.1435-5957.2011.00364.x.

Funk, R. J. 2014. "Making the Most of Where You Are: Geography, Networks, and Innovation in Organizations". *Academy of Management Journal* 57 (1): 193–222. doi:10.5465/amj.2012.0585.

Giuliani, E. 2013. "Network dynamics in regional clusters: Evidence from Chile". *Research Policy* 42 (8): 1406–1419. doi:10.1016/j.respol.2013.04.002.

— . 2007. "The selective nature of knowledge networks in clusters: evidence from the wine industry". *Journal of Economic Geography* 7 (2): 139–168. http://joeg.oxfordjournals.org/cgi/content/abstract/7/2/139.

Giuliani, E., and M. Bell. 2005. "The micro-determinants of meso-level learning and innovation: evidence from a Chilean wine cluster". *Research Policy* 34 (1): 47–68. doi:10.1016/j.respol.2004.10.008.

Goyal, S. 2007. *Connections: An Introduction to the Economics of Networks.* Princeton University Press.

Graf, H. 2011. "Gatekeepers in regional networks of innovators". *Cambridge Journal of Economics* 35 (1): 173–198. doi:10.1093/cje/beq001.

— . 2017. *Replication Data for: Regional Innovator Networks - A Review and an Application with R.* doi:10.7910/DVN/L2IOZO. http://dx.doi.org/10.7910/DVN/L2IOZO.

Graf, H., and T. Henning. 2009. "Public Research in Regional Networks of Innovators: A Comparative Study of Four East German Regions". *Regional Studies* 43 (10): 1349–1368. doi:10.1080/00343400802251460.

Graf, H., and M. Kalthaus. 2016. *International Research Networks: Determinants of Country Embeddedness.* Jena Economic Research Papers 2016-016. Friedrich Schiller University Jena. https://ideas.repec.org/p/jrp/jrpwrp/2016-016.html.

Granovetter, M. S. 1985. "Economic action and social structure. The problem of embeddedness". *American Journal of Sociology* 91 (3): 481–510. doi:10.17323/1726-3247-2002-3-44-58.

— . 1973. "The strength of weak ties". *American Journal of Sociology* 78:1360–1380. doi:10.1016/b978-0-12-442450-0.50025-0.

Griliches, Z. 1992. "The search for R&D spillovers". *Scandinavian Journal of Economics* 94 (Supplement): 29–47. doi:10.2307/3440244.

Gulati, R. 1998. "Alliances and networks". *Strategic Management Journal* 19 (4): 293–317. doi:10.1002/(sici)1097-0266(199804)19:4<293::aid-smj982>3.0.co;2-m.

Hagedoorn, J. 2002. "Inter-firm R&D partnerships: An overview of major trends and patterns since 1960". *Research Policy* 31 (4): 477–492. doi:10.1016/s0048-7333(01)00120-2.

Hidalgo, C. A., B. Klinger, A.-L. Barabasi, and R. Hausmann. 2007. "The Product Space Conditions the Development of Nations". *Science* 317 (5837): 482–487. doi:10.1126/science.1144581.

Hidalgo, C. A. 2016. "Disconnected, fragmented, or united? a trans-disciplinary review of network science". *Applied Network Science* 1 (1). doi:10.1007/s41109-016-0010-3.

Jackson, M. O. 2011. "Chapter 12 - An Overview of Social Networks and Economic Applications". In *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. O. Jackson, 1:511–585. Supplement C. North-Holland. http://www.sciencedirect.com/science/article/pii/B9780444531872000127.

— . 2014. "Networks in the Understanding of Economic Behaviors". *Journal of Economic Perspectives* 28 (4): 3–22. doi:10.1257/jep.28.4.3.

— . 2008. *Social and Economic Networks.* Princeton: Princeton University Press. http://press.princeton.edu/titles/8767.html.

Jackson, M. O., B. W. Rogers, and Y. Zenou. 2017. "The Economic Consequences of Social-Network Structure". *Journal of Economic Literature* 55 (1): 49–95. doi:10.1257/jel.20150694.

Jaffe, A. B., M. Trajtenberg, and R. Henderson. 1993. "Geographic localization of knowledge spillovers as evidenced by patent citations". *Quarterly Journal of Economics* 108 (3): 577–598. doi:10.2307/2118401.

Keeble, D., C. Lawson, B. Moore, and F. Wilkinson. 1999. "Collective learning processes, networking and 'institutional thickness' in the Cambridge region". *Regional Studies* 33 (4): 319–332. doi:10.1080/713693557.

Kline, S., and N. Rosenberg. 1986. "An overview of innovation". In *The Positive Sum Strategy*, ed. by R. Landau and N. Rosenberg, 275–305. Washington: National Academy Press.

Krugman, P. 1991. "Increasing returns and economic geography". *Journal of Political Economy* 99:483–499. doi:10.1086/261763.

Lazerson, M. H., and G. Lorenzoni. 1999. "The firms that feed industrial districts: A return to the Italian source". *Industrial and Corporate Change* 8 (2): 235–266. doi:10.1093/icc/8.2.235.

Leoncini, R., M. A. Maggioni, and S. Montresor. 1996. "Intersectoral innovation flows and national technological systems: Network analysis for comparing Italy and Germany". *Research Policy* 25 (3): 415–430. doi:10.1016/0048-7333(95)00843-8.

Li, E. Y., C. H. Liao, and H. R. Yen. 2013. "Co-authorship networks and research impact: A social capital perspective". *Research Policy* 42 (9): 1515–1530. doi:10.1016/j.respol.2013.06.012.

Li, G.-C., R. Lai, A. D'Amour, D. M. Doolin, Y. Sun, V. I. Torvik, A. Z. Yu, and L. Fleming. 2014. "Disambiguation and co-authorship networks of the U.S. patent inventor database (1975–2010)". *Research Policy* 43 (6): 941–955. doi:10.1016/j.respol.2014.01.012.

Lissoni, F. 2010. "Academic inventors as brokers". *Research Policy* 39 (7): 843–857. doi:10.1016/j.respol.2010.04.005.

Lobo, J., and D. Strumsky. 2008. "Metropolitan patenting, inventor agglomeration and social networks: A tale of two effects". *Journal of Urban Economics* 63 (3): 871–884. doi:10.1016/j.jue.2007.07.005.

Lundvall, B.-Å. 1988. "Innovation as an interactive process: From user-producer interaction to the national systems of innovation". In *Technical change and Economic Theory*, ed. by G. Dosi, C. Freeman, R. Nelson, G. Silverberg, and L. Soete, 349–369. London: Pinter.

— , ed. 1992. *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. London: Pinter Publishers.

Lychagin, S., J. Pinkse, M. E. Slade, and J. Van Reenen. 2016. "Spillovers in Space: Does Geography Matter?" *The Journal of Industrial Economics* 64 (2): 295–335. doi:10.1111/joie.12103.

Malerba, F. 2002. "Sectoral systems of innovation and production". *Research Policy* 31:247–264. doi:10.1016/s0048-7333(01)00139-1.

Markose, S., S. Giansante, and A. R. Shaghaghi. 2012. "'Too interconnected to fail' financial network of US CDS market: Topological fragility and systemic risk", *Journal of Economic Behavior and Organization* 83 (3): 627–646. doi:10.1016/j.jebo.2012.05.016.

Marshall, A. 1890. *Principles of Economics*. London: Macmillan. http://www.econlib.org/library/Marshall/marP.html.

Martinelli, A. 2012. "An emerging paradigm or just another trajectory? Understanding the nature of technological changes using engineering heuristics in the telecommunications switching industry". *Research Policy* 41 (2): 414–429. doi:10.1016/j.respol.2011.10.012.

Møen, J. 2005. "Is Mobility of Technical Personnel a Source of R&D Spillovers?" *Journal of Labor Economics* 23 (1): 81–114. doi:10.1086/425434.

Moody, J. 2004. "The Structure of a Social Science Collaboration Network". *American Sociological Review* 69:213–238.

Neffke, F., M. Henning, and R. Boschma. 2011. "How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions". *Economic Geography* 87 (3): 237–265. doi:10.1111/j.1944-8287.2011.01121.x.

Nelson, R. R., ed. 1993. *National Innovation Systems: A Comparative Analysis*. New York: Oxford University Press.

Nooteboom, B. 2000. "Learning by Interaction: Absorptive Capacity, Cognitive Distance and Governance". *Journal of Management and Governance* 4:69–92.

OECD. 2016a. *HAN database, September 2016*.

— . 2016b. *REGPAT database, February 2016*.

Owen-Smith, J., and W. W. Powell. 2004. "Knowledge networks as channels and conduits: the effects of spillovers in the Boston biotechnology community". *Organization Science* 15 (1): 5–21. doi:10.1287/orsc.1030.0054.

Ozman, M. 2009. "Inter-firm networks and innovation: a survey of literature". *Economics of Innovation and New Technology* 18 (1): 39–67. doi:10.1080/10438590701660095.

Pezzoni, M., F. Lissoni, and G. Tarasconi. 2014. "How to kill inventors: testing the Massacrator© algorithm for inventor disambiguation". *Scientometrics* 101 (1): 477–504. doi:10.1007/s11192-014-1375-7.

Phelps, C., R. Heidl, and A. Wadhwa. 2012. "Knowledge, Networks, and Knowledge Networks: A Review and Research Agenda". *Journal of Management* 38:1115–1166. doi:10.1177/0149206311432640.

Porter, M. E. 1990. *The Competitive Advantages of Nations*. London: Macmillan.

Powell, W. W., K. W. Koput, and L. Smith-Doerr. 1996. "Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology". *Administrative Science Quarterly* 41 (1): 116–145. doi:10.2307/2393988.

Powell, W. W., K. W. Koput, L. Smith-Doerr, and J. Owen-Smith. 1999. "Network position and firm performance: Organizational returns to collaboration in the biotechnology industry". In *Research in the Sociology of Organizations*, ed. by S. B. Andrews and D. Knoke, 129–159. Greenwich, CT: JAI Press.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Raffo, J., and S. Lhuillery. 2009. "How to play the "Names Game": Patent retrieval comparing different heuristics". *Research Policy* 38 (10): 1617–1627. doi:10.1016/j.respol.2009.08.001.

Riccaboni, M., and F. Pammolli. 2002. "On firm growth in networks". *Research Policy* 31 (8-9): 1405–1416. doi:10.1016/s0048-7333(02)00071-9.

Ripley, R., T. A. B. Snijders, Z. Boda, A. Vörös, and P. Preciado. 2017. *Manual for SIENA version 4.0*. Oxford: University of Oxford, Department of Statistics. http://www.stats.ox.ac.uk/siena/.

Romer, P. M. 1990. "Endogenous technological change". *Journal of Political Economy* 98:71–102. doi:10.1086/261725.

Rowley, T., D. Behrens, and D. Krackhardt. 2000. "Redundant Governance Structures: An Analysis of Structural and Relational Embeddedness in the Steel and Semiconductor Industries". *Strategic Management Journal* 21 (3): 369–386. http://www.jstor.org/stable/3094192.

Saxenian, A. 1994. *Regional Advantage*. Cambridge: Harvard University Press.

Schilling, M. A., and C. C. Phelps. 2007. "Interfirm Collaboration Networks: The Impact of Large-Scale Network Structure on Firm Innovation". *Management Science* 53 (7): 1113–1126. doi:10.1287/mnsc.1060.0624.

Schnettler, S. 2009. "A structured overview of 50 years of small-world research". *Social Networks* 31 (3): 165–178. doi:10.1016/j.socnet.2008.12.004.

Schumpeter, J. A. 1942. *Capitalism, Socialism and Democracy.* London: Unwin, reprint (1987).

— . 1912. *Theorie der wirtschaftlichen Entwicklung, 5.Auflage (1935).* Berlin: Duncker & Humblot.

Singh, J. 2005. "Collaborative Networks as Determinants of Knowledge Diffusion Patterns". *Management Science* 51 (5): 756–770. doi:10.1287/mnsc.1040.0349.

Sorenson, O., and T. Stuart. 2001. "Syndication Networks and Spatial Distribution of Venture Capital Investments". *American Journal of Sociology* 106 (6): 1546–1588. doi:10.1086/321301.

Sorenson, O., J. W. Rivkin, and L. Fleming. 2006. "Complexity, networks and knowledge flow". *Research Policy* 35 (7): 994–1017. doi:10.1016/j.respol.2006.05.002.

ter Wal, A. L. J. 2013. "Cluster Emergence and Network Evolution: A Longitudinal Analysis of the Inventor Network in Sophia-Antipolis", *Regional Studies* 47 (5): 651–668. doi:10.1080/00343401003614258.

— . 2014. "The dynamics of the inventor network in German biotechnology: geographic proximity versus triadic closure". *Journal of Economic Geography* 14 (3): 589–620. doi:10.1093/jeg/lbs063.

ter Wal, A. L. J., and R. A. Boschma. 2009. "Applying social network analysis in economic geography: framing some key analytic issues". *Annals of Regional Science* 43 (3): 739–756. doi:10.1007/s00168-008-0258-3.

Thompson, P., and M. Fox-Kean. 2005. "Patent citations and the geography of knowledge spillovers: a reassessment". *American Economic Review* 95 (1): 450–460. doi:10.1257/0002828053828509.

Tomasello, M. V., M. Napoletano, A. Garas, and F. Schweitzer. 2017. "The rise and fall of R&D networks". *Industrial and Corporate Change* 26 (4): 617–646. doi:10.1093/icc/dtw041.

Torre, A., and J. P. Gilly. 2000. "On the analytical dimension of proximity dynamics". *Regional Studies* 34 (2): 169–180. doi:10.1080/00343400050006087.

Trajtenberg, M., G. Shiff, and R. Melamed. 2006. *The "Names Game": Harnessing Inventors' Patent Data for Economic Research.* NBER Working Papers 12479. National Bureau of Economic Research, Inc. http://ideas.repec.org/p/nbr/nberwo/12479.html.

Uzzi, B. 1997. "Social structure and competition in interfirm networks: the paradox of embeddedness". *Administrative Science Quarterly* 42 (1): 35–67. doi:10.2307/2393808.

Uzzi, B., L. A. N. Amaral, and F. Reed-Tsochas. 2007. "Small-world networks and management science research: a review". *European Management Review* 4 (2): 77–91. doi:10.1057/palgrave.emr.1500078.

Valente, T. W. 1995. *Network models of the diffusion of innovations.* Cresskill New Jersey: Hampton Press.

Verspagen, B. 2007. "Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research". *Advances in Complex Systems* 10 (1): 93–115. doi:10.1142/s0219525907000945.

Wagner, C. S., and L. Leydesdorff. 2005. "Network structure, self-organization, and the growth of international collaboration in science". *Research Policy* 34 (10): 1608–1618. doi:10.1016/j.respol.2005.08.002.

Wanzenböck, I., T. Scherngell, and R. Lata. 2015. "Embeddedness of European Regions in European Union-Funded Research and Development (R&D) Networks: A Spatial Econometric Perspective". *Regional Studies* 49 (10): 1685–1705. doi:10.1080/00343404.2013.873119.

Wassermann, S., and K. Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Watts, D. J., and S. H. Strogatz. 1998. "Collective dynamics of "small-world" networks". *Nature* 393 (4): 440–442.

Wickham, H. 2012. *stringr: Make it easier to work with strings.* R package version 0.6.2. http://CRAN.R-project.org/package=stringr.

Wuchty, S., B. F. Jones, and B. Uzzi. 2007. "The Increasing Dominance of Teams in Production of Knowledge". *Science* 316 (5827): 1036–1039. doi:10.1126/science.1136099.

Zucker, L., M. Darby, and J. Armstrong. 1998. "Geographically localised knowledge: Spillovers or markets?" *Economic Inquiry* 36:65–86. doi:10.1111/j.1465-7295.1998.tb01696.x.

A Tables

Table A.1: Top 10 nodes in the 2006–2012 innovator networks according to betweenness centrality

	Stuttgart	Munich	Berlin	Hamburg
1	ROBERT BOSCH GMBH	SIEMENS AG	SIEMENS AG	AIRBUS OPERATIONS GMBH
2	FRAUNHOFER	BAYERISCHE MOTOREN WERKE AG	FRAUNHOFER	UNIVERSITATSKLINIKUM HAMBURG EPPENDORF
3	UNIVERSITAT STUTTGART	LUDWIG MAXIMILIANS UNIVERSITAT MUNCHEN	CHARITE UNIVERSITAETSMEDIZIN BERLIN	BEIERSDORF AG
4	MAX PLANCK	TECHNISCHE UNIVERSITAT MUNCHEN	TECHNISCHE UNIVERSITAT BERLIN	TESA SE
5	ALCATEL LUCENT	FRAUNHOFER	BASF SE	MAX PLANCK
6	BEHR GMBH & CO KG	MAX PLANCK	BAYER INTELLECTUAL PROPERTY GMBH	CENTRUM FUER ANGEWANDTE NANOTECHNOLOGIE CAN GMBH
7	DAIMLER AG	F HOFFMANN LA ROCHE AG	MAX PLANCK	SIEMENS AG
8	BASF SE	GIESECKE & DEVRIENT GMBH	HUMBOLDT UNIVERSITAET ZU BERLIN	NORDEX ENERGY GMBH
9	KARLSRUHER INSTITUT FUR TECHNOLOGIE	AUDI AG	FREIE UNIVERSITAET BERLIN	TECHNISCHE UNIVERSITAET HAMBURG HARBURG
10	DR ING H C F PORSCHE AG	BSH BOSCH & SIEMENS HAUSGERAETE GMBH	FORSCHUNGSVERBUND BERLIN EV	TUTECH INNOVATION GMBH