Papers in Evolutionary Economic Geography

17.11

Critical links in knowledge networks – What about proximities and gatekeeper organizations?

Tom Broekel and Wladimir Mueller

Utrecht University
Urban & Regional research centre Utrecht

Critical links in knowledge networks – What about proximities and gatekeeper organizations?

Tom Broekel[*,$] and Wladimir Mueller[#]

[$]Department of Economic Geography, Utrecht University, The Netherlands
[#]Institute of Economic and Cultural Geography, Leibniz University of Hannover, Germany

29. May 2017

Abstract

The paper analyzes links in knowledge networks that are essential for the integration and knowledge diffusion properties of the entire network. By focusing on *critical links*, as defined in network science, we evaluate these links' properties from the perspective of the proximity and regional gatekeeper literature. We thereby gain insights into likely conditions of their emergence and functions. Moreover, we extend the dyadic perspective on regional gatekeeper organizations and link it more strongly to the network science and proximity framework literature.

An empirical study applies these arguments and investigates the proximity characteristics of critical links in 132 technology-specific subsidized knowledge networks in Germany. The results show that critical links tend be formed between regional gatekeepers that offer related knowledge resources. The links bridge institutional distances by utilizing the benefits of geographic and social proximity.

Keywords: proximities, knowledge networks, gatekeeper, R&D subsidies, critical links

JEL-Classifications: D85, L14, O33, R10

---

[*] Corresponding author: Tom Broekel, Department of Economic Geography, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands, e-mail: t.broekel@uu.nl

1 Introduction

Inter-organizational innovation networks have received considerable attention in the literature (Ozman, 2009; Powell et al., 1999). In the field of Economic Geography, the spatial configuration of knowledge networks and their development in space in particular has been in focus (Boschma and Ter Wal, 2007; Giuliani and Bell, 2005; Morrison, 2008). Amongst the most prominent theoretical frameworks for studying network emergence, evolution, and their effects in this literature is the proximity concept, developed by the French school of proximity dynamics (Torre and Gilly, 2000). Its popularity further increased through the work of Boschma (2005). This framework highlights the relevance of various proximity dimensions influencing the likelihood of knowledge link formation. In complementarity with network structural factors, such as triadic closure and preferential attachment, it has been shown to provide a powerful basis for studying network evolution and for analyzing the emergence or dissolving of links in general (Balland, 2012; Broekel and Hartog, 2013a).

However, in network science it is argued that not all links in networks are of similar importance. While some links are essential for network wide knowledge diffusion, others are merely of local relevance or even a matter for individual nodes at best (Cassi and Plunket, 2013). This paper focuses on links in networks that are most important from a network science perspective, namely those frequently referred to as "bridging" or "critical" links (Burt, 2004; Granovetter, 1973). These links are crucial for integrating the complete network by linking otherwise sparsely or even completely unconnected parts of the network. With the exception of Cassi and Plunket (2013), the network related literature in the field of Economic Geography has, however, exclusively focused on explaining the emergence and effects of "average links", ignoring the heterogeneity in links' network structural importance. In other words, despite their importance, little is still known about critical links in this literature, which is the primary motivation for this paper.

We contribute to the discussion in the work of Cassi and Plunket (2013) and revise some of their arguments. Moreover, we add insights from the literature on regional gatekeepers, which discusses similar issues. However, in contrast to the dyadic concept of "critical links", the regional gatekeeper literature focuses on the node/organizational level. For instance, Giuliani and Bell (2005) and Graf (2011) emphasize that the performance of regional innovation systems crucially depends on the presence and effectiveness of a small number of "gatekeeper" organizations connecting the regional systems' internal knowledge exchange processes to global knowledge networks. Similar to critical links, gatekeeper organizations function as bridges between otherwise sparsely connected parts of knowledge networks. We argue that they frequently achieve this by being part of a critical link. By gaining a better understanding of critical links and their relations to gatekeeper organizations, the paper therefore seeks to further integrate the network science, proximity framework, and regional gatekeeper literature.

The theoretical discussions are complemented by an empirical study, which extends the empirical investigation of Cassi and Plunket (2013), which is limited to a single network in genomics. In contrast, we identify critical links in 132 inter-organizational technology-specific knowledge networks allowing for more generalizable findings.

In accordance with the hypotheses, we find critical links to be more likely than "average links" to connect socially proximate organizations operating in different institutional frameworks and that are located in geographic proximity. Moreover, critical links provide access to related knowledge resources, underlining their relevance for enriching sub-networks' and in many cases regional knowledge bases.

The paper is organized as follows. Section 2 provides the theoretical discussion on critical links. Section 3 describes the database, which covers information on collaborative R&D used to construct knowledge networks. The empirical approach and the results are presented in Section 4. Section 5 concludes the paper.

2 Theoretical background

2.1 Critical links in network science

The concepts of "weak ties" (Granovetter, 1973) and "structural holes" (Burt, 1992) are central frameworks in social network theory. The idea behind strong and weak ties is that *strong ties* are formed within densely connected sub-networks and *weak ties* span boundaries between these sub-networks. In the context of knowledge networks, densely connected groups of nodes are frequently characterized by high degrees of redundancy and rather homogenous knowledge. In order to get access to non-redundant knowledge and new information, these groups need to establish links to other groups. In the event of two groups or sub-networks being characterized by non-redundant knowledge, they are said to represent a so-called "structural hole" (Burt, 1992). By bridging such structural holes, *weak ties* increase the diffusion of non-redundant knowledge in the network, which raises the potential for novelty creation.

Another class of links that has received attention in network science are so-called "critical links".[1] Such *critical links* connect poorly or otherwise disconnected sub-networks (Burt, 2004; Granovetter, 1973). Crucially, when critical links dissolve, the network falls apart and knowledge diffusion among its members is severely reduced. Due to the fact that these links connect sparsely linked parts of the network, they represent "bottlenecks" (Sytch et al., 2012) or "bridges" (Glu¨ckler, 2007). While every *critical link* can be classified as a *weak tie*, the same is not necessarily true of the reverse. *Critical links* are crucial for the structure and integration of the complete network, while *weak ties* may only have local relevance. Figure 1 visualizes the idea of critical links.

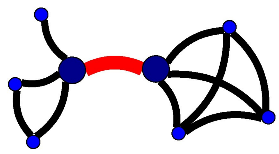


Figure 1: Critical link in a network

Contrasting their network-theoretical importance, little research exists on such *critical links* in knowledge networks. An exception is the study of Cassi and Plunket (2013). These authors compare the formation of "closure links" (intra-component links) with critical links on the basis of the French co-inventorship network in the field of genomics. Their study reveals that critical links have additional properties not directly related to their positions in networks.

2.2 The proximity framework

When assessing differences between closure and critical links, Cassi and Plunket (2013) rely on the so-called "proximity framework". A range of factors influences the creation of links and thereby the emergence of knowledge networks. This concerns structural network properties, factors at the dyad and at the node level (Boschma and Frenken, 2010; Ter Wal and Boschma, 2009). At the node level, it has been empirically shown that specific characteristics of nodes (i.e., organizations' experience) determine organizations' network embeddedness (Marín and

[1] They are also sometimes referred to as "bridging" or "gatekeeper" links.

Siotis, 2008; Powell et al., 1996). The structural network level matters as well, as, for example, the preferential attachment argument (Baraba´si and Albert, 1999) suggests that central nodes are more likely to get linked to new nodes than others.

The "proximity framework" focuses on the third level, the dyad level, which deals with the relationship between two nodes in the network. At this level, scholars have paid much attention to the proximity framework popularized by the work of Boschma (2005). He argues that five different proximity types crucially influence the creation of links in knowledge networks. Based on the French school of proximity dynamics (e.g., Torre and Rallet, 2005), the framework highlights that organizations are more likely to get connected when they are cognitively, geographically, institutionally, organizationally, or socially proximate. Hence, the proximity framework adopts the homophily concept from sociology (Powell et al., 2005) by explicitly specifying a number of relevant dimensions of similarity, i.e., the different proximity types. In addition to stimulating link formation, all types of proximity also play a role for the effectiveness of knowledge transfer and novelty generation simultaneously.

The analytical power and empirical relevance of the proximity framework has been repeatedly shown. This concerns the simultaneous relevance of the different proximity dimensions for knowledge link creation and network evolution (Balland, 2012; Broekel and Hartog, 2013a) and the importance of proximity structures in knowledge links for innovation (Fornahl et al., 2011; Broekel and Boschma, 2012). However, the proximity framework and the empirical studies employing it, explain knowledge link formation and their impact in general. That is, their focus is on the emergence and impact of the "average" link in inter-organizational knowledge networks. While this is a valid strategy in most instances, it ignores the heterogeneity among links, which amongst others, is highlighted by Cassi and Plunket (2013). By comparing closure and critical links, Cassi and Plunket (2013) show that the emergence of critical links is largely driven by organizational and technological diversity as well as some geographic proximity. Without explicitly pointing it out, these authors thereby link the concept of critical links to the ideas of the "proximity framework" and underline that proximity structures differ significantly among links. Following this line of thinking, we will argue that the proximity framework can provide further insights into critical links: their functions as well as likely conditions of emergence. The same is true for the literature on regional gatekeepers, introduced next.

2.3 The regional gatekeeper literature

Critical links connect poorly or otherwise disconnected sub-networks and bridge structural holes. This mirrors the role so-called "broker" organizations are argued to play in the innovation literature (Hargadon and Sutton, 1997). Gould and Fernandez (1989) define a "gatekeeper" as an actor who holds a brokering position between an actor group's internal and external partners. This is analogous to the role of specific boundary spanning individuals, who link their organization to the external environment and are therefore called gatekeepers (Tushman and Katz, 1980). For instance, Allen (1977) defines "technological" gatekeeper as R&D professionals with the intellectual ability to absorb external information and make it accessible for other employees of their firm. Their intermediary position does not only impact on gatekeepers' own performance (Hargadon and Sutton, 1997), but also matters for all organizations who rely on them for access to external knowledge (Hargadon, 1998). For the present paper, it is crucial that the functions attributed to gatekeepers at the node level mirror those of critical links on the link level. More precisely, the two concepts relate to each other as at least one node of the two nodes connected by a critical link is a gatekeeper node. One may even say that a gatekeeper node becomes a gatekeeper because of the existence of a critical link. The concept of gatekeeper was transferred to the spatial context by Giuliani and Bell (2005). These authors argue that regional gatekeepers are important for embedding regional innovation systems into inter-regional and global knowledge networks. They thereby serve as important sources for new knowledge and ideas from outside the region that are diffused into the regional innovation system (Gertler, 1997; Bathelt et al., 2004). Regional gatekeepers help to avoid lock-in situations and allow organizations to exploit the advantages of strong local embeddedness without the negative aspects of long-term loss of diversity (Bathelt et al., 2004, Glückler, 2007). Due to their ability to access specific knowledge, they are also referred to as "knowledge gatekeepers" (Malipiero et al., 2012; Morrison et al., 2013) and play crucial roles for network wide knowledge diffusion. In this context, it is crucial to differentiate between gatekeepers in a network sense and regional gatekeepers. While the former are defined on the basis of a complete network, the latter are related to an organization connecting a regional network to an outside network. While regional gatekeepers are always gatekeepers from a network perspective, the same does not necessarily apply the other way around. This raises the question of how critical links are related to regional gatekeepers.

We suggest that in addition to the proximity literature, insights from the literature on regional gatekeepers can be helpful for understanding the emergence and function of critical links in networks. Crucially, in addition to enriching the analysis of critical links, we also see benefits from extending the strongly node-level oriented perspective on regional gatekeepers to a dyadic perspective. While this literature has generated a considerable amount of knowledge about the nodal characteristics of gatekeepers, little is known about the properties of their links. We will show in the following that relating critical links to the proximity framework and the regional gatekeeper yields benefits for all three strands of literature.

2.4 An integrating perspective

The network science literature defines critical links as links not located within densely linked parts of networks, as they are weak ties and bridge structural holes. Concerning the emergence of weak and strong ties, the social network literature focuses on the main determinants of community formation (Sytch et al., 2012). This particularly refers to the homophily effect: Organizations with similar characteristics are more likely to link (Powell et al., 2005). The homophily effect tends to reinforce the emergence of densely connected network communities; that is, the emergence of strong ties (McPherson et al., 2001). Accordingly, we hypothesize that critical links (as types of weak ties) are less likely to be characterized by homophily than average links. The proximity framework extends this idea and explicitly proposes a number of characteristics (proximities) organizations need to share in order for their link probability to increase. Following Boschma (2005), it is argued that geographical, cognitive, social, institutional, and organizational proximity in particular matter in this respect.

Nooteboom et al. (2007) highlight the prominent role of cognitive proximity in the establishment of successful R&D alliances. Cognitive proximity defines the technological overlap between two organizations. At small cognitive distances, organizations have the necessary absorptive capacities for efficient communication and learning from each other. Moreover, and in contrast to the other proximity types, cognitive proximity creates the potential for novel ideas resulting from two organizations' knowledge exchange and learning (Nooteboom, 2000). Cognitive proximity holds an outstanding role among the proximity types for these reasons (Boschma and Frenken, 2010). As argued above, critical links span 'structural holes' and enable the transfer of novel information (e.g., Burt, 1992, 2004). In this sense, critical links connect non-redundant knowledge pieces and establish links between groups of organizations with different experiences, skills, and knowledge bases (McEvily and Zaheer, 1999). From the proximity framework perspective, it can therefore be expected that critical links tend to bridge significant degrees of cognitive distances.

Nooteboom et al. (2007) and Boschma (2005) point out that large cognitive distances are difficult to bridge. We therefore argue that in order to be effective and relevant for knowledge diffusion in networks as well as providing access to novel information (Burt, 1992, 2004; Cassi and Plunket, 2013), critical links are unlikely to be characterized by the most unfavorable conditions for knowledge transfer and communication. More precisely, we hypothesize that either **a)** the cognitive distance bridged by critical links does not exceed the necessary proximity for communication and learning, or **b)** that some other conditions allow organizations to overcome this distance.

With respect to **a)** Teece et al. (1994) emphasize the importance of related knowledge bases for innovation. On the one hand, it offers a basis for novel knowledge re-combinations and on the other, it allows for sufficient communication (Nooteboom, 2000). If the cognitive proximity bridged by critical links corresponds to that of related knowledge, critical links are not the most likely to be formed but they can still bridge structural holes and bring together non-redundant related knowledge. Hence, we refine the argument of Cassi and Plunket (2013) that critical links are likely to bridge large cognitive distances (i.e., connecting variety) inasmuch as critical links are more likely to connect related knowledge.

Concerning **b)**, among the conditions that may allow organizations to bridge large cognitive distances are the cases in which other proximities substitute for missing cognitive proximity (Boschma, 2005). Cassi and Plunket (2015) suggest that the relatively larger cognitive[2] distances of critical links may be compensated for by other types of proximity (social, geographic, institutional, and organizational).

The regional gatekeeper literature offers an idea about a particular proximity type that is likely a

[2] Cassi and Plunket (2015) also argue for critical links being characterized by large social distances. However, in the context of inter-organizational knowledge networks, we think that the relationship between social proximity and critical links is much less clear.

substitute for a lack of cognitive proximity in this context. Giuliani and Bell (2005) and Graf (2011) underline the importance of regional gatekeepers' absorptive capacity. It allows them to establish long-distance links, which offer knowledge that regional organizations do not provide. Accordingly, they bridge a cognitive gap existing between regional organizations and region-external knowledge networks. Moreover, Morrison (2008) shows that gatekeepers primarily engage with region external organizations that are active in similar or complementary technologies. Cognitive proximity to these organizations allows them to overcome (substitute) the geographic distance. We therefore expect that the cognitive distance between gatekeepers and their regional partners is larger than their distance to their region-external partners. For our hypotheses on the characteristics of critical links, this observation implies the following: if critical links involve at least one regional gatekeeper organization and are characterized by significant cognitive distance, critical links are more likely to be found among a regional gatekeeper's links to other regional organizations than among its links to region-external organizations. Put differently, critical links are likely to link geographically proximate but cognitive distant organizations because geographic proximity helps in dealing with the cognitive gap. This argument finds some empirical support in the study by Broekel (2015) who shows that growing cognitive distances between linked organizations tend to be correlated with decreasing geographic distances. Moreover, it is frequently shown that large firms and universities primarily act as gatekeepers (Graf, 2011). These organizations' distinct technological profiles are usually unmatched by other regional organizations.

We summarize the above discussion in a number of hypotheses that describe our expectations of critical links' properties in relation to average links.

Hypotheses

(1) Critical links frequently involve regional gatekeeper organizations.

(2) Critical links connect related knowledge bases.

> (2a) When critical links are characterized by large cognitive distances at least one other type of proximity (geographic, institutional, organizational, and social) will be present.
>
> (2b) Critical links tend to be characterized by large cognitive but small geographic distances.

Besides deriving expectations on characteristics of critical links in knowledge networks, the discussion presents an example of how the three different literature strands (network science, proximity framework, gatekeeper literature) relate to and can be used to enrich one another. In the remainder of the paper, we will test these hypotheses.

3 Data

3.1 Subsidized R&D database

The database utilized in our empirical studies contains information on networks created through subsidized R&D collaborations. Such similar data are frequently used to model and analyze inter-organizational knowledge networks in the field of Economic Geography (Scherngell and Barber, 2009). Balland (2012), and Broekel and Hartog (2013b) show that knowledge networks based on subsidized R&D collaboration have similar properties and determinants as networks not based on subsidized collaborations. Consequently, subsidies data on joint R&D projects is suitable for constructing innovation networks.

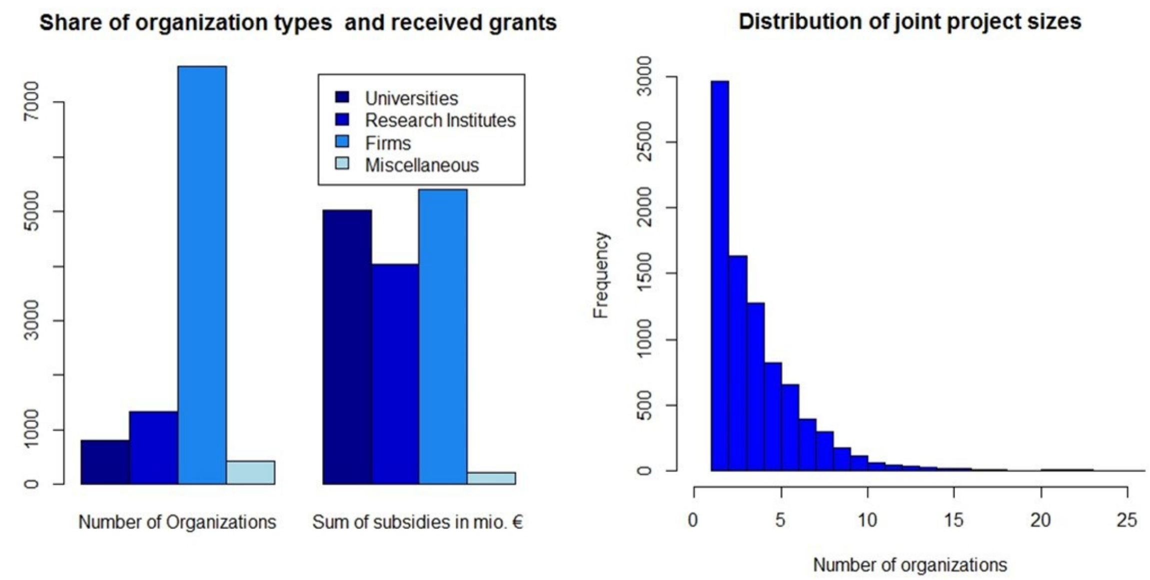
We employ data on subsidized joint R&D projects funded by the German federal government. The Federal Ministry of Education and Research (BMBF) manages the largest share of these projects. To a smaller extent, other federal ministries contribute as well. The database includes detailed information on more than 160,000 individual funds granted since 1960, which includes information on the grant period, the name and postal code (location) of the receiving as well as executing organization, the granting sum, and the classification number. An important differentiation is made between the executing and receiving organization. For instance, a university can be the grant receiving organization, while a faculty or an institute is specified as the executing unit. Hence, large organizations like multinational companies, non-university research institutes (e.g., The Max-Planck-Society), and universities are frequently split into different executing units. Another important piece of information concerns the technological content of projects. This is reflected in a technological hierarchical classification scheme, called "Leistungsplansystematik", which includes over 20 main classes.

The classes represent different technological areas, which cover energy research, biotechnology etc. In addition, the main classes are split hierarchically into different sub-classes. For example, we can distinguish between bionic (L07534) or adaptronic (L07533). The research areas also cover non-technological areas, as, for example, perspectives for rural areas (DB0300) are listed as well. A more detailed description of this data is carried out by Broekel and Graf (2012). It should be pointed out that the data has the drawback of potential policy biases (Broekel, 2015).

3.2 Descriptive statistics

We base the empirical analysis on all subsidized joint R&D projects running between 2003 and 2012. This period covers 8,604 joint projects with 35,264 individual grants. In total, 8,903 distinct receiving organizations are listed. In order to construct comparable organizations, we follow Broekel (2015) and combine the name of the receiving unit with the municipality code of the executing unit. As a result, 10,215 organizations are distinguished. These are classified into four types: Universities, firms, research institutes, and miscellaneous organizations. The left plot in Figure 2 gives an overview of the distribution of received grants as well as the number of funded organizations for each organizational type. More than 70 per cent of all subsidized organizations are firms. However, the total value of subsidies for firms (over EUR 5 billion) is only slightly above the received grants for universities. This is mainly due to the higher co-financing rate for non-profit organizations. The right plot in Figure 2 highlights that most projects have two participants.

Figure 2: Subsidies data description of the observation period



To generate networks on this basis, we project this two-mode network data into a one-mode network by means of the method proposed by Newman (2001). That is, links emerging from large projects are weighted less than links established in small projects.[3]

To construct a significant number of sufficiently large technology-specific networks, we follow Broekel (2015) and disaggregate the data at the level of 4-digit research areas. This level provides acceptable network sizes and it is simultaneously a suitable level of technological disaggregation. We exclude small networks with less than 10 links, as network measures are not meaningful in their instance. Networks with a density of one are excluded from the empirical analysis. Moreover, we exclusively focus on the giant component of each 4-digit research area network since this part represents the core of the network with the most important links and organizations. On this basis, we are able to establish 132 networks that are observed for the period 2003-2012.

Table 1 provides an insight into the descriptives of observed network structures. It shows that the networks are characterized by high degrees of heterogeneity in their size, centralization, density, and shares of organization types, as well as in the number of different industries (two-digit NACE codes).

Table 1: Network statistics

n=132	Min	Max	Median	Mean	SD
Nodes	6	1511	51	114.7	185.6
Links	10	7709	169	484.4	894.1
Density	0.006	0.93	0.14	0.21	0.20
Degree centralization	0.06	0.86	0.28	0.30	0.15
Betweenness centralization	0.03	0.86	0.39	0.38	0.16
Projects	5	706	36	77	106.3
Industries	2	66	13	18.24	14.4
Share universities	0%	100%	24.6%	30.7%	22.5%
Share firms	0%	90%	51.2%	45%	27.3%
Share institutes	0%	75%	19.6%	22.2%	13.9%
Share miscellaneous	0%	53%	0%	2.1%	6%

[3] Newman (2001) suggests to account for the number of nodes related to an event when projecting from two-mode to one-mode networks with the following weighting: $w_{i,j} = \sum_a \frac{1}{N_a} - 1$ with N_a being the number $Na - 1$ of nodes related to an event in the two-mode network.

4 Empirical approach

4.1 Bridging centrality

Various methods have been developed in the field of social network analysis concerning the measurement of nodes' importance for the overall cohesion of networks (Wasserman and Faust, 1994). Granovetter (1973) highlights the importance of so-called bridges (i.e., critical links) for reducing the overall geodesic distance between organizations. However, the empirical identification of such links is a complex task. We employ one of the most recent approaches, which was developed by Hwang et al. (2008) for the identification of critical links.

According to the definition of Hwang et al. (2008), critical links and nodes in a graph are characterized by being critical connectors of modular regions in the network. Consequently, in contrast to other links in the network, critical links will score highly on what these authors call "bridging centrality". In order to estimate bridging centrality, we first computed the weighted betweenness centrality Φ of each link *e* in a network:

$$\Phi(e) = \sum_{s \neq t \epsilon V} \frac{\sigma_{st}(e)}{\sigma_{st}} \qquad (1)$$

where σ_{st} is the quantity of shortest weighted geodesic distances between nodes *t* and *s*. $\sigma_{st}(e)$ as a subset of σ_{st} refers to the score of shortest weighted paths passing through link *e*. Node betweenness centrality refers to nodes' frequency of being located on the shortest paths (geodesic distance) between (in)directly linked nodes (Freeman, 1977)[4] and may represent organizations' capabilities to absorb knowledge diffusing in the network (Owen-Smith and Powell, 2004).

Since betweenness centrality captures nodes' global embeddedness, we also calculate the bridging coefficient Ψ of each node *v,* which captures nodes' embeddedness into their local network surroundings:

$$\Psi(v) = \frac{1}{d(v)} \sum_{i \epsilon N, d(i) > 1} \frac{d(i)}{d(i) - 1} (2)$$

with the degree of node *v* being defined as *d*(*v*). $\delta(i)$ refers to the quantity of links leaving the adjacent sub network of node *v* among the links incident to each adjacent node *i* of node *v*. Put simply, $\delta(i)$ is the sum of links belonging to the neighbors (nodes) of node *v*. The bridging

[4] The geodesic distances in the weighted network are estimated according to the method proposed by Dijkstra (1959).

coefficient Ψ of link e is calculated on this basis by:

$$\Psi(v) = \frac{d(i)\Psi(i)+d(j)\Psi(j)}{(d(i)+d(j))(|C(i,j)|+1)}, e(i,j)\epsilon E \qquad (3)$$

where i and j are defined as the coincident nodes to link e. C (i, j) is the number of common adjacent nodes of node j and i. Finally, the bridging centrality C_{Br} of link e is defined by:

$$C_{Br}(e) = R_{\Phi(e)} + R_{\Psi(e)} \qquad (4)$$

whereby R_{Φ} refers to the rank betweenness of link e and R_{Ψ} defines the rank of link e with respect to the above bridging coefficient. Rank normalization is used to correct for potential scale differences between the bridging coefficient and the weighted betweenness centrality.

The resulting measure of bridging centrality is a suitable measure for identifying critical links in a network and empirically outperforms purely distance based measures like betweenness centrality in terms of identification, because it simultaneously considers local (bridging coefficient) and global network properties (betweenness centrality) (Hwang et al., 2008).[5] Links scoring highly on this type of centrality tend to connect different sub networks and when removed are likely to disintegrate the network. Accordingly, they act as critical links in knowledge networks.

Crucially, the measure does not allow for a binary distinction between critical and non-critical links. Rather, it represents a degree of "criticalness", i.e., links obtaining large values are more critical than those with lower values. We therefore examine the ranks of links with respect to their bridging centrality in each technology-specific network. The relative nature of our measure has the advantage of considering each network's specific characteristics, which allows for comparing critical links across various network structures.[6] Alternatively, one could define critical links in an absolute manner by, for instance, identifying links whose removal increases networks' number of components. However, it implies that one component networks will always lack critical links. Such definition also ignores other types of critical links, whose existence, for instance, doesn't reduce network fragmentation but strongly shortens geodesic distances and thereby *critically* influences network internal knowledge diffusion. In light of our previous discussion, such links equally qualify as *critical links*. Accordingly, our relative

[5] The measure is moreover significantly different from link betweenness. In the later empirical analyses, the correlation between bridging and link betweenness turns out to be just 0.281***.

[6] Note that it is our aim to identify general characteristics of critical links that apply on average in most technology-specific networks.

measure of critical links is more general than alternative (absolute) definitions that ultimately are dependent on networks' structures. In order to identify characteristics of critical links using our relative approach, we employ a number of alternative scenarios. More precisely, in the first scenario (I) the top one percent of links with the highest bridging centrality is defined as being the critical link. In the second scenario (II), we look at the 'second-tier' critical links that are defined as 2 percent percentile of links with the highest bridging centrality values. However, in this scenario we exclude the top 1 percent of links. In the third (III) and fourth (IV) scenarios we define all links being critical within the top 3 and 4 percent of bridging centrality values with the previous top 2 and 3 percent being excluded again. This approach allows us to get more detailed insights into changes in links' characteristics along the distribution of their criticalness values. The variable created on this basis, which will serve as the dependent variable in the later regressions, has a value of one if a link qualifies as a critical link and zero otherwise.

4.2 Characteristics of links

In order to analyze the proximity structures of critical links, we construct variables capturing the characteristics of all realized collaborations/links. Descriptive statistics are given in Table 2.

Geographic distance is conceptualized as pure physical distance between organizations. Geographically proximate organizations generally have more opportunities for interaction and meetings by chance, easier arrangements of face-to-face contacts, and generally benefit from lower transaction costs (Feldman and Florida, 1994). As a result, knowledge relations are more likely to be established between geographically proximate organizations (Jaffe et al., 1993; Breschi and Lissoni, 2001). To estimate geographic distances between organizations, we use the coordinates of their respective municipalities' centroids. Due to the higher probability of large networks stretching over larger geographic distances than small networks, we normalize the geographic distance for each link by dividing the value with the maximum value in each network. Consequently, the variable geographic distance GEODIST varies between zero and one with values close to zero indicating short geographic distances.

An alternative measure of geographic proximity is REGLINK, which captures the differentiation between regional and inter-regional links, with the variable being one if the linked organizations are located within the same region and zero otherwise. The regional delineation is based on German labor market regions (141 regions) as defined by Kosfeld and Werner (2012).

Organizational distance refers to the degree of autonomy and control, which can be performed in organizational arrangements, either between organizations or within organizations (Boschma, 2005). Scholars frequently approximate organizational distance by means of two organizations belonging to the same mother organization (Balland, 2012; Broekel, 2015). The measure ORGDIST takes a value of one if linked organizations do not share the same (mother) organization and zero otherwise.

Institutional distance defines the degree of similarity concerning formal rules and informal constraints (macro-level) shared by organizations. Thus, a certain level of institutional proximity provides a stable framework for interactive learning (Boschma, 2005). According to Broekel (2015), two organizations share the same institutional framework when they are of the same organizational type. Our data allows for distinguishing between four organizational types: private firms, research institutes, universities, and miscellaneous.[7] Similar to Ponds et al. (2007), INSTDIST is defined as one if linked organizations are not of the same type and zero otherwise.

Social distance refers to trust, friendship, and experience based on repeated interactions. Socially embedded relationships increase the likelihood of R&D collaborations among partners (Boschma and Frenken, 2010). In a common manner, social distance between organizations is approximated by their number of (joint) past collaborations. To obtain this measure, we construct the technology-specific networks on the basis of subsidized R&D joint projects for the period 1993 to 2002. To make the values comparable across networks, the number is normalized with the maximum number of collaborations in the corresponding network. Moreover, the values are multiplied by minus one in order to transform the values into distances. The resulting measure SOCDIST indicates large social distances when its value is close to minus one.

Cognitive distance captures the degree of overlap between organizations' knowledge bases. We estimate the cognitive distance between organizations on the basis of their sector membership given in the subsidies data. The data provides two-digit NACE codes for all subsidized

[7] E.g., associations, ministries, and societies.

organizations and, in the case of universities, universities of applied science, and for some large service related sectors, three-digit NACE codes. The first measure is an index of technological (i.e., cognitive) similarity between sectors. It exploits the fact that most R&D subsidization programs belong to a relative narrowly defined technological field, which is represented by a specific research area in the technological classification scheme of the subsidies data ("Leistungsplansystematik"). The more frequently that two sectors are subsidized through the same technology oriented subsidization scheme, the more likely they draw on similar technological knowledge.

For the estimation of the index, we rely on the index proposed by Teece et al. (1994) and Bryce and Winter (2009). We count the number of co-occurrences of individual subsidies grants[8] attributed to two sectors' organizations within each of the six-digit technological classes in the subsidies data and compare it to the frequency that can be expected when these grants are randomly assigned to sectors' organizations. We follow the estimation described in detail in Bryce and Winter (2009, p. 1575f.) with J_{ij} representing the number of individual grants acquired by organizations in sectors i and j classified into the same six-digit technological class. K is the number of technological classes and n_i represents the total number of individual grants that organizations in sector i have acquired. n_j is the corresponding number for sector j. The size of sectors needed in the estimation is approximated by the total number of grants acquired by organizations in a particular sector. The expected number of grants in a technological class for i and j is (x_{ij}) can be seen as hypergeometric random variable whose mean and variance can be derived. The final index τ_{ij} is then estimated as the standardized difference between the observed and expected numbers of co-occurrences. In order to avoid size bias and for easier interpretation, it is standardized and divided by the maximum similarity score. Negative values imply strong dissimilarity and hence their interpretation is the same as in the case of zero values. They are set to zero, implying that the final similarity index ranges between 0 and 1 with values close to one indicating maximal technological similarity. For easier interpretation (as distance), the negative value of this index is used to represent the cognitive distance between organizations (i.e., their respective sectors). The created variable is denoted as COGDIST. Above, we argued that COGDIST might be related to non-linear effects: Critical links are likely

[8] As cooperative projects might be formed on the basis of resource complementarity and not similarity, these are excluded.

to be characterized by relatedness, i.e., some but not very high cognitive distance. To capture this effect, we include the variable in a linear and quadratic term. Multicollinearity issues are avoided by subtracting the mean before squaring the values.

Regional gatekeepers are likely involved in critical links (see Section 2.4). For the identification of a regional gatekeeper, we use the definition of Gould and Fernandez (1989) and adopt the method of Graf (2011) for constructing a regional gatekeeper index. Firstly, organizations may qualify as regional gatekeepers if they have least one inter-regional and one regional link. Secondly, an index is constructed, reflecting the Euclidean distance to zero of an organization's numbers of inter-regional and regional links. Similar to Graf (2011), we chose a threshold of two-percent and define organizations as a regional gatekeeper if their index value is among the highest two percent of all observed values. The variable REGGATE1 obtains a value of one if just one of the link's nodes qualifies as a regional gatekeeper and zero otherwise. If both nodes are identified as a regional gatekeeper, REGGATE2 will be one and zero otherwise.[9]

Control variables are created to account for networks' heterogeneity. This includes network density (DENSITY) and size, with the latter being the number of nodes in a network (SIZE). The values of these variables are identical for all links belonging to the same network.

Table 2: Descriptive statistics

	Min	Max	Mean	Median	SD
GEODIST	0	1	0.440	0.417	0.252
SOCDIST	0	1	0.938	1	0.231
COGDIST	0	1	0.209	0.132	0.377
$COGDIST^2$	0	0.895	0.134	0.029	0.222
INSTDIST	0	1	0.506	1	0.500
ORGDIST	0	1	0.995	1	0.072
REGLINK	0	1	0.123	0	0.328
REGGATE1	0	1	0.205	0	0.404
REGGATE2	0	1	0.013	0	0.114
SIZE	6	1,511	444.738	273	450.701
DENSITY	0.007	0.938	0.069	0.027	0.116

[9] Note that this approach is prone to identify large organizations as regional gatekeepers. In the paper's context, this has no further implications

4.3 Empirical set-up

To identify the characteristics of critical links, we compare their proximity structures to those of "average" links. That is, for each network the most critical, in terms of bridging centrality, links are identified and their respective characteristics are compared to the properties of the other links in the networks. Hence, with some caution, the comparison allows for inferences to be made about the conditions under which critical links are formed. In any case, it allows for drawing conclusions on their unique characteristics and their function.

It is our goal to identify the general characteristics of critical links applying (on average) in most instances. We therefore pool the links of all networks, implying that the final dataset has as many observations as the sum of the links of all networks. Small networks will contribute smaller numbers of observations (links) to the final dataset. To reduce this bias, we consider a dummy variable for each network in the estimations.

Our dependent variable takes a value of 1 if a link has been identified as a critical link and zero otherwise. The number of identified critical links depends on the scenarios introduced in Section 4.1. Given the generally low thresholds (2-4 percent), their number is very small compared to that of all remaining links in each scenario, resulting in the dependent variable being characterized by an excessive number of zeros. We therefore rely on rare event logistic regression as specified by King and Zeng (2001).

4.4 Results

All models meet the necessary statistical requirements and can therefore be interpreted.[10] Table 3 shows the results. The coefficient of the regional gatekeeper variable REGGATE1 is negative and highly significant in the first three models. Consequently, links are less likely to be critical links when just one of their nodes is a regional gatekeeper. In contrast, the coefficient of REGGATE2, with the variable being one when both link's nodes are regional gatekeepers and zero otherwise, obtains a positive and highly significant value in model I and II. In line with **hypothesis (1)**, it indicates that critical links are especially formed between two regional gatekeepers. Notably, model I is superior to the other models in terms of fit, as indicated by Akaike information criterion (AIC). Hence, the most critical links (2 percent threshold) differ

[10] The variance inflation factor (VIF) test indicates the absence of multicollinearity in each model. Maximum VIF scores are depicted in table 3. The use of robust standard errors also does not yield different results.

more significantly in their characteristics from average links than when somewhat less critical links are also considered (4 percent threshold).

Our results give some insights on the proximity characteristics of critical links. The coefficient of organizational (ORGDIST) distance is insignificant in all models. No particularity of critical links is found with regard to this proximity type.

This is different for cognitive proximity. In the first, second and third model, the coefficients of COGDIST become significant when the squared version of COGDIST is included. In this case, the linear term obtains a significant positive value and $COGDIST^2$ is negatively significant. The AIC values suggest preferring three models including $COGDIST^2$ over those excluding this variable. Accordingly, we find an inverted U-shaped relationship between cognitive distance and the probability of observing a critical link. The finding suggests that critical links are characterized by positive but not too large values of technological similarity. In other words, they link technologically related organizations and potentially yield significant benefits for innovation (Frenken et al., 2007). It confirms our **hypothesis (2)** regarding critical links connecting related knowledge and supports the findings of Cassi and Plunket (2015) for a larger set of knowledge networks.

A positive significant coefficient is also observed for INSTDIST. Consequently, critical links bridge institutional borders by connecting organizations embedded into distinct institutional frameworks. The finding corresponds with the results for cognitive proximity as institutional borders frequently go along with cognitive differences. Critical links bring together related knowledge separated by institutional boundaries. Given the way INSTDIST is empirically constructed, critical links particularly bridge institutional boundaries between public (basic) research and (application-oriented) R&D in the private sector. This finding is consistent with Petruzzelli et al. (2010) who, in three case studies, especially observe knowledge gatekeepers (universities) as establishing weak ties to private organizations, when these organizations aim at diversifying their technological competencies. Moreover, the finding suggests that institutional proximity is an unlikely substitute for a lack of cognitive proximity, which is in line with Broekel (2015), who finds cognitive and institutional proximity to be positively correlated.

Geographic distance (GEODIST) obtains significant negative coefficients in the first and second model. Hence, critical links tend to cross shorter geographic distances than average links. Interestingly, the alternative measure of geographic proximity REGLINK (two organizations being co-located in the same region) remains insignificant in most models. Taken together, critical links do not tend to connect organizations in the same, but rather in neighboring regions. Critical links appear to be characterized by a certain degree of geographical proximity, which, however, exceeds the average radius of German labor market regions.

The variable measuring social distance, SOCDIST, is characterized by significantly negative coefficients in the first two models. That is, critical links are more likely established between organizations which have a shared past of collaborating. This finding adds support to **hypothesis 2a** and **2b** regarding proximities other than cognitive easing the establishment of critical links. However, we detect such an effect exclusively for geographic and social, but not for institutional and organizational, proximity.

In summary, our results clearly support our expectations on the characteristics of critical links: Critical links in technology-specific networks connect socially proximate regional gatekeeper organizations that are located in relative geographic vicinity, offering each other access to related knowledge.

5 Discussion and Conclusion

Research on network evolution is still at an early stage (Powell et al., 2005). This also concerns research from a geographical perspective. For instance, Ter Wal and Boschma (2009, p.741) stress that "[...] particularly in the application to regional issues, network research is still in its infancy." In the field of Economic Geography, prominent research seeks to understand the formation of networks from the perspective of the proximity framework. Studies particularly compare the relative relevance of different types of proximity for the creation of network links. Another flourishing research strand analyzes the roles, characteristics, and importance of regional gatekeeper organizations in, and for, regional innovation networks. While both literature strands deal with knowledge networks, they are still relatively unconnected, as the first focuses on the link (dyadic) and the second on the node level.

Table 3: Rare event logistic regression results for predicting critical links

DEPENDENT VAR.	Model I 1 % Top-links		Model II 2 % Top-links		Model III 3 % Top-links		Model IV 4 % Top-links		VIF (max)
GEODIST	-0.392**	-0.392**	-0.170**	-0.167**	-0.213*	-0.210*	-0.081	-0.081	
	(0.194)	(0.1954)	(0.140)	(0.140)	(0.115)	(0.115)	(0.100)	(0.100)	1.376
SOCDIST	-0.642**	-0.641**	-0.286**	-0.290**	-0.133	-0.137	-0.096	-0.096	
	(0.147)	(0.147)	(0.117)	(0.118)	(0.101)	(0.101)	(0.088)	(0.088)	1.053
COGDIST	-0.106	0.375**	-0.084	0.300***	0.078	0.51**	0.064	0.100	
	(0.125)	(0.215)	(0.092)	(0.240)	(0.079)	(0.206)	(0.069)	(0.176)	2.718
$COGDIST^2$		-0.322**		-0.686**		-0.780**		-0.066	
		(0.122)		(0.392)		(0.338)		(0.294)	2.621
INSTDIST	0.487***	0.488***	0.494***	0.492***	0.538***	0.536***	0.507***	0.507***	
	(0.094)	(0.094)	(0.068)	(0.068)	(0.056)	(0.056)	(0.049)	(0.049)	1.351
ORGDIST	-0.408	-0.409	-0.424	-0.416	-0.634*	-0.624*	-0.586*	-0.585*	
	(0.511)	(0.510)	(0.363)	(0.363)	(0.278)	(0.278)	(0.247)	(0.247)	1.016
REGLINK	0.112	0.112	0.172*	0.170*	0.195**	0.194**	0.092	0.092	
	(0.132)	(0.132)	(0.097)	(0.097)	(0.079)	(0.079)	(0.071)	(0.071)	1.364
REGGATE1	-1.031***	-1.032***	-0.604***	-0.600***	-0.324***	-0.319***	-0.036	-0.036	
	(0.138)	(0.138)	(0.086)	(0.086)	(0.064)	(0.064)	(0.052)	(0.052)	1.261
REGGATE2	0.685**	0.686**	0.391**	0.381**	0.419*	0.408*	0.44*	0.439*	
	(0.386)	(0.386)	(0.296)	(0.296)	(0.227)	(0.227)	(0.182)	(0.182)	1.252
SIZE	-0.017	-0.018	-0.007	-0.007	-0.006	-0.006	-0.003	-0.003	
	(0.012)	(0.012)	(0.010)	(0.010)	(0.008)	(0.008)	(0.007)	(0.007)	1.233
DENSITY	0.461	0.460	0.449	0.451	-0.256	-0.236	-0.693	-0.692	
	(1.478)	(1.479)	(1.371)	(1.368)	(1.194)	(1.193)	(1.091)	(1.091)	1.224
NETW. DUMMIES	....	...	...	...	...	...	...	...	
INTERCEPT	-2.355**	-2.352**	-2.933***	-2.933***	-2.402***	-2.422***	-2.325***	-2.326***	
	(0.985)	(0.985)	(0.832)	(0.831)	(0.692)	(0.692)	(0.619)	(0.619)	
Log-Likelihood	-3,624	-3,424	-6,215	-6,014	-8,511	-8,108	-10,575	-10,874	
AIC	7,491	7,093	12,673	12,272	17,264	16,861	21,393	21,995	
N	63,941	63,941	63,276	63,276	62,653	62,653	62,018	62,018	

* refers to a significance level of 0.1, ** to a significance of 0.05, and *** to 0.01., standard errors in parentheses

The present paper concentrated on so-called "critical links", which are crucial for network wide knowledge diffusion by integrating the whole network and connecting parts of networks that are otherwise relatively distant. We applied the proximity framework in the analysis of these links' characteristics, which (with the exception of Cassi and Plunket (2015)) have been exclusively studied from a network structural perspective. In doing so, the paper contributes to the existing literature in three ways.

Firstly, the paper has shown that the proximity approach, which has primarily been employed to investigate the creation and dissolving of "average links", provides a powerful basis for studying the heterogeneity of links' importance in knowledge networks as well. By focusing on inter-organizational knowledge networks, we were able to revise some of the arguments by Cassi and Plunket (2015).

Secondly, we extended the discussion on links' structural importance in general and that of critical links in particular, by considering insights from the regional gatekeeper literature. The paper therefore not only contributes to the ongoing integration of network science and proximity framework, it also presents links to the literature on (regional) gatekeeper. In particular, it connects the proximity framework to research on gatekeeper organizations by extending the analysis of the latter with a dyadic perspective.

Thirdly, we presented an empirical study on the proximity structures of critical links in 132 technology-specific inter-organizational knowledge networks emerging from the subsidization of joint R&D projects in Germany. The empirical results confirm that regional gatekeeper organizations are the most likely creators of critical links. Critical links connect regional gatekeepers that tend to be located in relatively short geographic distances (however in different regions), which offer related knowledge resources. The establishment of critical links that bridge cognitive and institutional distances is further stimulated by geographic and social proximity. While geographic proximity's impact on network formation is well-established, the findings suggest that geographic proximity is particularly relevant in the formation of critical links and hence shapes the knowledge diffusion efficiency of networks.

Our empirical findings substantiate our discussion on relating gatekeeper and proximity literature to each other. Regional gatekeeper organizations were shown to be especially capable of utilizing their proximate surroundings. They are more prone to establish and maintain relations with other gatekeeper organizations offering related knowledge. That is, they take advantage of geographic and social proximities' benefits for tapping into knowledge bases that are cognitively non-redundant with sufficient overlap allowing for efficient knowledge transfer and learning. Accordingly, regional gatekeeper organizations structure their knowledge relations in a more beneficial way than other organizations in their geographical surroundings.

These findings have to be seen in the light of a number of empirical limitations. First, the empirical analysis exclusively focuses on intra-national networks and ignores long-distance and international relations. Second, networks are based on subsidized R&D collaboration. While we control for potential biases related to this, we cannot completely rule out the potential effects of subsidization policy. Third, we identify the proximity structures of critical links in a cross-sectional set-up. Therefore, any inference of the conditions of their emergence has to be done with severe reservations. Only future research using dynamic approaches might show whether our interpretations are fully justified.

The paper suggests a number of policy implications. Our empirical findings support the promotion of R&D collaboration between public and private organizations located in adjacent regions. Collaborating organizations should further be related in their knowledge bases. Crucially, strengthening collaboration among organizations with related knowledge in neighboring regions will not only facilitate their own innovation activities, but will also yield benefits for all other organizations embedded in these organizations' knowledge networks by increasing these networks' cohesion and knowledge diffusion efficiency.

References

Allen, T. J. (1977). *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information within the R&D Organization*. MIT Press, Harvard, MA.

Balland, P. (2012). Proximity and the evolution of collaboration networks: evidence from research and development projects within the global navigation satellite system (GNSS) industry. *Regional Studies*, 46(6):741–756.

Barabasi, A.-L. and Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512.

Bathelt, H., Malmberg, A., and Maskell, P. (2004). Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, 28(1):31–56.

Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies*, 39(1):61–74.

Boschma, R. and Frenken, K. (2010). The spatial evolution of innovation networks. A proximity perspective. In Boschma, R. and Martin, R., editors, *The Handbook of evolutionary economic geography*, pages 120–136. Edward Elgar, Cheltenham, UK.

Boschma, R. and Ter Wal, A. L. J. (2007). Knowledge Networks and Innovative Performance in an Industrial District: The Case of a Footwear District in the South of Italy. *Industry & Innovation*, 14(2):177–199.

Breschi, S. and Lissoni, F. (2001). Localised knowledge spillovers vs. innovative milieux: Knowledge "tacitness" reconsidered. *Papers in Regional Science*, 80(3):255–273.

Broekel, T. (2015). The co-evolution of proximities - a network level study. *Regional Studies*, 49(6):921-931.

Broekel, T. and Boschma, R. (2012). Knowledge networks in the Dutch aviation industry: the proximity paradox. *Journal of Economic Geography*, 12(2):409–433.

Broekel, T. and Graf, H. (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.

Broekel, T. and Hartog, M. (2013a). Explaining the Structure of Inter-Organizational Networks using Exponential Random Graph Models. *Industry & Innovation*, 20(3):277-295.

Broekel, T. and Hartog, M. (2013b). The Geography of Networks and R&D Collaborations. In Scherngell, T., editor, *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, pages 49–70. Springer International Publishing, Cham.

Bryce, D. J. and Winter, S. G. (2009). A General Interindustry Relatedness Index. *Management Science*, 55(9):1570–1585.

Burt, R. (2004). Structural holes and good ideas. *American journal of sociology*, 110(2):349–399.

Burt, R. S. (1992). *Structural Holes: The Social Structure of Competition*. Harvard University

Press, Cambridge, MA.

Cassi, L. and Plunket, A. (2015). Research Collaboration in Co-inventor Net- works: Combining Closure, Bridging and Proximities. *Regional Studies*, 49(6): 936-954.

Dijkstra, E. W. (1959). A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik 1*, pages 269–271.

Feldman, M. P. and Florida, R. (1994). The Geographic Sources of Innovation: Technological Infrastructure and Product Innovation in the United States. *Annals of the Association of American Geographers*, 84(2):210–229.

Fornahl, D., Broekel, T., and Boschma, R. (2011). What drives patent performance of German biotech firms? The impact of R&D subsidies, knowledge networks and their location. *Papers in Regional Science*, 90(2):395–418.

Freeman, L. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41.

Freeman, L. (1979). Centrality in social networks conceptual clarification. *Social networks*, 1:215–239.

Frenken, K., Van Oort, F., and Verburg, T. (2007). Related Variety, Unrelated Variety and Regional Economic Growth. *Regional Studies*, 41(5):685–697.

Gertler, M. S. (1997). The invention of regional culture. In Lee, R. and Wills, J., editors, *Geographies of Economies*, pages 47–58. Edward Arnold, London.

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

Glückler, J. (2007). Economic geography and the evolution of networks. *Journal of Economic Geography*, 7(5):619–634.

Gould, R. and Fernandez, R. (1989). Structures of mediation: A formal approach to brokerage in transaction networks. *Sociological methodology*, 19(1989):89–126.

Graf, H. (2011). Gatekeepers in regional networks of innovators. *Cambridge Journal of Economics*, 35(1):173–198.

Granovetter, M. (1973). The strength of weak ties. *American journal of sociology*, 78(6):1360–1380.

Hargadon, A. and Sutton, R. (1997). Technology brokering and innovation in a product development firm. *Administrative science quarterly*, 42(4):716–749.

Hargadon, A. B. (1998). Firms as Knowledge Brokers: Lessons in Pursuing Continuous Innovation. *California Management Review*, 40(3):209–227.

Hwang, W., Kim, T., Ramanathan, M., and Zhang, A. (2008). Bridging Centrality: Graph Mining from Element Level to Group Level. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*, pages 336–344, New York, New York, USA. ACM Press.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localisation of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, *10*, 577–598.

King, G. and Zeng, L. (2001). Logistic regression in rare events data. *Political analysis*, 9:137–163.

Kosfeld, R. and Werner, A. (2012). Deutsche Arbeitsmarktregionen - Neuabgrenzung nach den Kreisgebietsreformen 2007-2011. *Raumforschung und Raumordnung*, 70(1):49–64.

Malipiero, A., Munari, F., and Sobrero, M. (2012). Absorptive capacity and localized spillovers: focal firms as technological gatekeepers in industrial districts. *Industrial and Corporate Change*, 21(2):429–462.

Marın, P. L. and Siotis, G. (2008). Public policies towards Research Joint Venture: Institutional design and participants' characteristics. *Research Policy*, 37(6-7):1057– 1065.

McEvily, B. and Zaheer, A. (1999). Bridging ties: A source of firm heterogeneity in competitive capabilities. *Strategic management journal*, 20:1133–1156.

McPherson, M., Smith-Lovin, L., and Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27:415–444.

Morrison, A. (2008). Gatekeepers of Knowledge within Industrial Districts: Who They Are, How They Interact. *Regional Studies*, 42(6):817–835.

Morrison, A., Rabellotti, R., and Zirulia, L. (2013). When Do Global Pipelines Enhance the Diffusion of Knowledge in Clusters? *Economic Geography*, 89(1):77–96.

Newman, M. E. J. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64(1):1–7.

Nooteboom, B. (2000). *Learning and innovation in organizations and economies*. Oxford University Press, Oxford.

Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., and van den Oord, A. (2007). Optimal cognitive distance and absorptive capacity. *Research Policy*, 36(7):1016–1034.

Owen-Smith, J. and Powell, W. W. (2004). Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community. *Organization Science*, 15(1):5–21.

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

Petruzzelli, A. M., Albino, V., Carbonara, N., and Rotolo, D. (2010). Leveraging learning behavior and network structure to improve knowledge gatekeepers' performance. *Journal of Knowledge Management*, 14(5):635–658.

Ponds, R., van Oort, F., and Frenken, K. (2007). The geographical and institutional proximity of research collaboration. *Papers in Regional Science*, 86(3):423–443.

Powell, W., Koput, K., and Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative science quarterly*, 41(1):116–145.

Powell, W., Smith-Doerr, L., and Koput, K. (1999). Network Position and Firm Performance: Organizational Returns to Collaboration in the Biotechnology Industry. In Andrews, S. and Knoke, D., editors, *Research in the Sociology of Organizations*, pages 129–159. Greenwich, CT.

Powell, W. W., White, D. R., Koput, K. W., and Owen-Smith, J. (2005). Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences. *American Journal of Sociology*, 110(4):1132–1205.

Scherngell, T. and Barber, M. J. (2009). Distinct spatial characteristics of industrial and public research collaborations: evidence from the fifth EU Framework Programme. *The Annals of Regional Science*, 46(2):247–266.

Sytch, M., Tatarynowicz, A., and Gulati, R. (2012). Toward a theory of extended contact: the incentives and opportunities for bridging across network communities. *Organization Science*, 23:1658–1681.

Teece, D., Rumelt, R., Dosi, G., and Winter, S. (1994). Understanding corporate coherence: Theory and evidence. *Journal of Economic Behavior and Organization*, 23:1–30.

Ter Wal, A. L. J. and Boschma, R. (2009). Applying social network analysis in economic geography: framing some key analytic issues. *The Annals of Regional Science*, 43(3):739–756.

Torre, A. and Rallet, A. (2005). Proximity and Localization. *Regional Studies*, 39(1):47–59.

Torre, A. S. and Gilly, J.-P. (2000). On the Analytical Dimension of Proximity Dynamics. *Regional Studies*, 34(2):169–180.

Tushman, M. L. and Katz, R. (1980). External Communication and Project Performance: An Investigation into the Role of Gatekeepers. *Management Science*, 26(11):1071–1085.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*, volume 8. Cambridge University Press, Cambridge, UK.

Appendix

Table 4: Correlation matrix Model I

		1	2	3	4	5	6	7	8	9	10	11
(1)	INSTDIST	1.0										
(2)	ORGDIST	0.068	1.0	-	-	-	-	-	-	-	-	-
(3)	COGDIST	0.437	0.085	1.0	-	-	-	-	-	-	-	-
(4)	COGDIST2	0.408	0.084	0.623	1.0	-	-	-	-	-	-	-
(5)	GEODIST	-0.027	0.026	-0.032	-0.021	1.0	-	-	-	-	-	-
(6)	SOCDIST	0.023	0.034	0.114	0.100	-0.022	1.0	-	-	-	-	-
(7)	REGLINK	0.064	-0.050	0.035	0.025	-0.484	0.012	1.0	-	-	-	-
(8)	REGGATE1	0.165	-0.016	0.033	0.039	0.021	-0.108	0.016	1.0	-	-	-
(9)	REGGATE2	0.015	-0.025	-0.0167	-0.015	0.041	-0.114	-0.015	0.227	1.0	-	-
(10)	DENSITY	0.016	-0.015	-0.061	-0.033	0.117	-0.043	-0.054	-0.022	0.121	1.0	-
(11)	SIZE	-0.069	0.017	0.058	0.034	-0.059	0.042	0.019	0.109	-0.005	-0.403	1.0