Papers in Evolutionary Economic Geography

# 12.17

**The co-evolution of proximities - a network level study**

Tom Broekel

Utrecht University
Urban & Regional research centre Utrecht

# The co-evolution of proximities - a network level study[☆]

Tom Broekel

*Institute of Economic and Cultural Geography,*
*Leibniz University of Hanover,*
*Schneiderberg 50, Hanover, Germany,*
*broekel@wigeo.uni-hannover.de*

*August 29, 2012*

**Abstract**

Despite the growing number of studies, still little is known about how network structures and proximity relations between linked actors evolve over time. Arguments are put forward for the existence of co-evolution dynamics between different types of proximity configurations within networks. An empirical investigation tests these arguments using information on the development of 280 networks.
Amongst others, it is shown that institutional and cognitive proximity configurations co-evolve in the short as well as in the long-run. While institutional and social proximity configurations are only related in the long run. Moreover, temporal auto-correlation dynamics characterizes the development of cognitive proximity configurations.

*Keywords:* proximities, co-evolution, R&D subsidies, knowledge networks, network evolution
JEL-classification: R10, R11, D85

## 1. Introduction

The evolution of knowledge networks has received considerable attention in the field of Economic Geography (Glückler, 2007; ter Wal and Boschma, 2009). Besides trying to understand the fundamental dynamics of network development, economic geographers are particularly interested in whether geographic proximity remains a significant factor for knowledge link formation when taking other types of proximity into account. These other proximity types include social, cognitive, institutional, and organizational proximity that are argued to be substitutes for geographic proximity in this respect (Boschma, 2005). Empirical evidence confirms the relevance of all proximity types being significant drivers of network evolution (cf. Balland, 2012).
While temporal changes in the relative importance of proximities for knowledge network formation have been studied in detail (Ter Wal, 2011; Balland *et al.*, 2012b) little is still

[☆]I would like to thank Pierre-Alexandre Balland for his valuable comments and suggestions on earlier drafts of this paper. Of course, all remaining errors are mine.

known about how exactly these networks develop over time. This particularly regards the evolution of proximity configurations within these networks, which are shown to be essential for firms' innovative performance (cf. Fornahl *et al.*, 2011; Broekel and Boschma, 2012). The present paper aims at contributing to this research field. It builds upon the proximity conception of Boschma (2005) and focuses on five different types of proximity: cognitive, social, organizational, institutional, and geographic proximity. These proximities are known to be interrelated and correlated (Boschma and Frenken, 2010). In the paper it is argued that proximity configurations within networks do not only systematically change in the course of network evolution but that the change in one proximity configuration can be related to the change in another. This may give rise to co-evolution dynamics between (some) proximity configurations. Three different types of such co-evolution dynamics are put forward: simultaneous (short-term) and long-term co-evolution as well as temporal autocorrelation.
In the second part of the paper these arguments are empirically tested by identifying factors explaining the temporal change of proximity configurations in networks. Particular attention is hereby paid to the identification of co-evolution dynamics. The investigation utilizes relational data on R&D cooperation that have been subsidized by the German federal government. On this basis 280 networks are constructed, which are observed for 2 to 13 years. A reduced-form vector autoregression (VAR) model is employed to identify the three types of co-evolution dynamics.
The empirical results confirm the existence of all three types of co-evolution dynamics. For instance, a close link exists between the short-term change in the geographic and cognitive configuration of networks. Networks expanding in the cognitive dimension tend to shrink geographically. Not surprisingly, the opposite holds for geographic and social proximity: networks with increasing social proximity are likely to consolidate geographically as well. In the long run in particular institutional and cognitive proximity configurations tend to co-evolve. The latter configuration is moreover characterized by temporal autocorrelation dynamics.
The paper is structured as follows. The literature on proximities is briefly reviewed and the five types of proximity are introduced. In addition, theoretical arguments for the existence of co-evolution dynamics between proximity configurations within networks are put forward. Section 3 gives an overview on the employed data of cooperative R&D subsidies used to approximate knowledge networks. The empirical approach is presented in Section 4. Section 5 summaries the findings and Section 6 provides a concluding discussion on some of the empirical study's shortcomings.

## 2. Theory

### *2.1. The proximity approach*

Researchers increasingly try to understand the mechanisms behind networks' emergence, development, and their structural change (cf. Cantner and Graf, 2006; Glückler, 2007; Balland, 2012). In the field of Economic Geography the proximity concept has become a popular and powerful theoretical basis for approaching these issues. Based on the French school of proximity dynamics (Torre and Rallet, 2005) and popularized by the article of Boschma (2005), it has stimulated an increasing number of studies that evaluate the role of geographic proximity in comparison to other proximity types. While geographic proximity has been seen (often implicitly) as the key force for the establishment of knowledge exchange

relations in this literature for a long time (cf. Feldman and Florida, 1994), it has become just one among others in recent years.
According to Boschma (2005) there are at least four other types of proximity that are highly relevant for knowledge link formation and for links' impact on organizations' innovative success.
The first is cognitive proximity that refers to the degree of overlap in two actors' knowledge bases. Actors need to have a complementary absorptive capacity to identify, interpret, and exploit knowledge of other actors (Cohen and Levinthal, 1990). In other words, overlap in actors' knowledge bases is essential for efficient communication. However, if the overlap is too strong, the likelihood that the interaction will result in an innovative knowledge recombination is lower than when they have dissimilar knowledge bases (Nooteboom, 2000).
Actors may also be close in organizational terms (organizational proximity). Boschma (2005) defines organizational proximity "*... as the extent to which relations are shared in an organizational arrangement, either within or between organizations*" (p. 65). It can be seen as a continuous scale going from autonomy to control. It is very low for independent actors and very high for actors that are part of the same hierarchical system. Accordingly, organizational proximity helps to manage knowledge exchange and reduces transactions costs.
In contrast, institutional proximity can be seen as the degree to which organizations are subject to the same institutional framework at the macro level. This refers to reward schemes, norms, and values of conduct. Frequently, researchers make a distinction between profit and non-profit organizations in this respect (Broekel and Boschma, 2012).
Social proximity describes the social embeddedness of actors in terms of friendship, kinship, and experience at the micro-level (Boschma, 2005). Of particular interest is the role of trust, which is likely to be positively influenced by social proximity and is frequently argued to foster knowledge exchange (Nooteboom, 2002).
The role of geographic proximity, defined as the physical distance between organizations, is mainly seen as facilitator of other types of proximity in this framework. However, there are also arguments for an independent role of this proximity type. For instance, Broekel and Binder (2007) argue for a direct impact of geographic proximity on knowledge exchange. They put forward that geography influences individuals' motivation and search heuristics and can thereby bias them towards spatially close knowledge sources.

### *2.2. Proximities, proximity configurations, and their relations*

Substantial empirical evidence exists for these proximity types impacting knowledge link formation. Prominently, Jaffe *et al.* (1993) analyze patent citations and find that they tend to come overproportionnally more frequent from the same geographical area as the inventors of the cited patent.[1] Breschi and Lissoni (2003) repeat and extent the analysis and show that geographical proximity looses its predictive power for patent citations when controlling for social proximity. The importance of cognitive proximity is confirmed by Mowery *et al.* (1998), Cantner and Graf (2006), and Cantner and Meder (2007). In contrast to the above, Balland (2012) and Broekel and Boschma (2012) take into account wider sets of proximity types. Still, both studies confirm the simultaneous relevance of all types of proximities for network formation. Based on the cooperation network of the world wide video

[1] See also Thompson and Fox-Kean (2005) on this.

game industry between 1987-2007, Balland *et al.* (2012b) further establish that the impact of proximities on network development is not constant over time. They find that geographic and cognitive proximity become more relevant as the industry matures.

Accordingly, all five proximity types matter for the formation of knowledge links. Simply put, this means that proximate actors have higher chances to link for knowledge exchange. However, it is also frequently argued that proximities are interrelated. Boschma and Frenken (2010) put forward that despite being "*analytically orthogonal*", in practice and empirics proximity types often "*turn out to be correlated*" (p. 131). For example, when an actor connects to a cognitively proximate actor the latter might also be geographically close. As a result, the realized link will be characterized by cognitive and geographic proximity. This has significant implications. If for instance, cognitive proximity is a main determinant for network formation and this correlation is relatively strong, the network will become more and more proximate in the cognitive as well as in the geographic dimension. One could also say that as a result of proximities being correlated at the relational actor level, the network's *configurations* with respect to geographic and cognitive *proximity* co-evolve.

The *network formation process* can additionally cause proximity configurations to co-evolve. Boschma (2005) argues that proximities are related in a substitutive way meaning that being proximate in one dimension can help to overcome missing proximity in another dimension. For instance, Singh (2005) shows that links become more relevant among cognitively distant researchers if they are located in geographic vicinity. Similarly, Ponds *et al.* (2007) find that geographical proximity helps to overcome institutional distance. Given a substitutive relation between proximities, links will be primarily realized that connect actors proximate in one dimension but distant in another. On the network level one can therefore expect a negative correlation between the changes of the according proximity configurations.

However, Broekel and Boschma (2011) also present evidence that (at least in case of small firms) the relation between geographic and cognitive proximity is complementary in character. This is, links characterized by geographic and cognitive proximity are more likely being realized than links that are solely characterized by geographic proximity. In accordance to the above, a complementary relation between proximities will yield a positive correlation between the developments of different proximity configurations at the network level.

It may also be the case that it is sufficient to be proximate in just one dimension to link, being proximate in another dimension may not yield further effects (Boschma and Frenken, 2010). In this case, the network formation process will only convey the distribution of proximities at the relational actor level into the network.

Understanding these relations and how they influence the development of proximity configurations within networks is crucial for a number of reasons. For instance, it is not just embeddedness into knowledge networks that matters for actors' innovation and economic performance. Boschma and Frenken (2010) point out that it is essential with whom actors interact and especially in what type and degree of proximity actors' cooperation partners are. This is confirmed in recent empirical studies highlighting that actors need to offer complementary knowledge in order to be valuable cooperation partners (cf. Fornahl *et al.*, 2011; Broekel and Boschma, 2012). In other words, to be beneficial, it is essential that knowledge networks have a particular type of proximity configuration.

In addition, the development of networks cannot be fully understood if changes in proximity configurations are ignored. Balland *et al.* (2012a) argue that current configurations can have an impact on actors' future cooperation behavior. A simple example in this respect is that actors, which frequently interact, are likely to become more similar in the cognitive dimension. This in turn increases the likelihood that they will continue to interact in the future as their bilateral communication capability improves.

Few studies address the changing of proximity configurations in networks. A notable exception in this respect is the study by Ter Wal (2011), who investigates the time-changing relevance of proximity types for network formation based on the German biotechnology patent co-invention network between 1970 and 1995. In addition, he plots the average geographic distance of links as well as the network's tendency of triadic closure[2], which can be seen as a measure of social proximity, over time (Figures 5 and 6 on page 37). While he just wants to describe the *configuration of the network* with respect to geographic and social proximity, it is interesting to note that, with the latter measure's growth being quite erratic, both measures tend to increase as the network matures and growths. Accordingly, it appears to be the case that the two proximity configurations follow a similar development path, i.e. that they co-evolve.

The present paper takes this up and asks if this observed "co-evolution" of the two proximity configurations is mere accident or an indication of a systematic *co-evolutionary* relation induced by the network formation process? And, in case of the latter, do such co-evolution dynamics exist for other proximity types as well? Balland *et al.* (2012a) highlight that changes of proximity configurations and actor characteristics take place over varying time periods. So, it is interesting to analyze if, and if so, which proximity configurations co-evolve in the short run and which rather show long-term co-evolution patterns. Giving answers to these questions is the primary objective of the present paper. To accomplish this I will first lay out some arguments speaking for the existence of such systematic relations in the (co-)evolution of proximity configurations in networks. In a second step, these are tested in an empirical investigation.

### *2.3. Dynamic relations between proximity configurations*

On the basis of the above, I put forward three different ways in which proximity configurations can co-evolve as networks change over time. The first one is *simultaneous co-evolution*. It describes the correlation between the changes of two proximity configurations in a network in one period. Given the short-term dimension, actor and population characteristics remain stable. In contrast, over longer time periods, actor and population characteristics may change which, amongst others, may give rise to *long-term co-evolution* dynamics. Lastly, *temporal autocorrelation* dynamics imply that when the configuration of one proximity type changes in one period, the same configuration will also change in the subsequent period. It may exist as short-term as well as long-term process. In case of the first, actor and population characteristics remain unchanged, while they may vary in the latter. I will provide arguments for why each of these dynamics might exist. It is however beyond the scope of the paper to discuss the potential relevance of the three dynamics for each of the five proximity types.

*Simultaneous co-evolution dynamics* between proximity configurations are likely to exist in

[2] The measure "*is expressed as the ratio of the observed number of closed triads over the number of random expected closed triads*" Ter Wal (2011).

several instances. The most obvious cases that come to mind are the relations between social and geographic proximity, as well as between institutional and organizational proximity. Social contacts are naturally more frequent among individuals living at closer geographic distances. "*Geographical proximity is most likely to stimulate social proximity, because short geographical distances favour social interaction and trust building*" (Boschma, 2005, p. 67). Accordingly, if links span smaller geographic distances the probability is high that they will also connect socially close actors and the other way around.
Sorenson (2003) puts forward another relation: Actors are more likely to establish social links with other actors if they share the same cognitive background, or as he in the style of Lazarsfeld and Merton (1954) colloquially puts it: "*birds of a feather, flock together*". Put differently, social and cognitive proximity of links are quite likely to be correlated.
If such patters occur frequently and the network formation process is not dominated by substitutive relations among proximities, a change in the configuration of links with respect to one proximity type will be most likely mean a simultaneous change in the respective other proximity configuration(s), i.e. both proximity configurations co-evolve simultaneously as the network develops.

*Long-term co-evolution* will automatically come into existence when (short-term) simultaneous co-evolution dynamics remain significant over longer time periods. In contrast to the letter they involve changes in the actor characteristics or the actor population. Balland *et al.* (2012a) point out that: "*in the short run, proximity creates knowledge networks, in the long run, knowledge networks create proximity*" (p. 9). Hence, through participating in knowledge networks actors' characteristics change over time. For instance, technological relatedness is a main criterion in merger and acquisition decisions (Hussinger, 2010). Accordingly, firms that interacted frequently and thereby became more similar in terms of their knowledge bases are likely to become subsidiaries of the same mother organization. Growing cognitive proximity is relating to increasing organizational proximity in this case.
In the "regional lock-in" scenario it is argued that economic relations in a region, which include knowledge networks, become so rigid and focused on a particular economic activity that new developments and ideas from outside the region are ignored. This goes hand in hand with actors becoming unaware of, or incapable to respond to, technological and structural change in the long run (Grabher, 1993). In other words, a regional lock-in is likely to involve a long-term co-evolution of geographic and cognitive proximity configurations.

*Temporal autocorrelation* may show as a short-term and long-term process. Actors are permanently confronted with new developments and technological progress implying that they need to constantly update their knowledge basis. They therefore adopt their position within knowledge networks such that they gain access to knowledge pieces helping to successfully deal with new technological problems. The problems actors face, i.e. the development trajectory of the technologies they are active in, therefore influence their ego-networks' configuration in terms of cognitive proximity. If a significant number of network actors faces such challenges, negative temporal autocorrelation dynamics can occur for the network's cognitive proximity configuration in a relatively short-time period.
In contrast, over longer time-periods actor characteristics change, which may also cause temporal autocorrelation. A simple example has already been presented above: actors that

frequently interact become more similar in the cognitive dimension, which in turn increases their chances to interact again in the next period. Such processes will show as positive temporal autocorrelation for the cognitive proximity configuration. However, they may be observable only over longer time periods as actors' characteristics are unlikely to change in the short-run.
Long-term autocorrelation dynamics may also exist for social proximity. Frequent interaction in cooperation stimulates social proximity by facilitating the development of trust and shared values (Ben-Porath, 1980). As social proximity enhances link creation a self-energizing effect is possible for the configuration of social proximity, which will show as positive autocorrelation.

## 3. Data

### *3.1. Data on subsidized R&D subsidies*

I employ data on research and development (R&D) projects that have been subsidized by the German federal government. The majority of these subsidies programs is initiated by the Federal Ministry of Education and Research (BMBF). A number of other ministries contribute as well but to smaller extents. Amongst others, policy aims at stimulating inter-organizational cooperation activities through subsidization. It thereby hopes to initiate technology transfer from the public to the private sector or to foster collective learning processes. The effectiveness of this approach is confirmed in many empirical studies (Scherngell and Barber, 2009, 2011; Fornahl *et al.*, 2011). For this reason, information on subsidized R&D cooperation, i.e. the subsidization of joint projects, can be used to model knowledge networks (Broekel and Graf, 2012).
Data on subsidized R&D projects has a number of features that make it very suitable for the present investigation. First of all, it allows for constructing a large number of distinct knowledge networks over multiple time periods. In this respect, the data is comparable to co-patent and co-publication data and shares their most important advantages. The data has one additional advantage, though: The actual length of a cooperation is known as the starting and ending date of joint projects are defined by the length of the subsidization.
These advantages come at some costs. Most importantly, networks based on R&D subsidies and their structures are subject to specific policy objectives. For instance, policy might aim at subsidizing particular types of cooperation in certain technological fields. This has to be taken into account in the empirical assessment.

### *3.2. Data of the "Förderkatalog"*

Comprehensive information on subsidized projects are published in the so-called "Förderkatalog" (subsidies catalog). It lists detailed information on more than 130,000 individual funds granted between 1960 and 2009. I refrain from presenting the data in detail (see on this Broekel and Graf, 2012).
For the empirical assessment I focus on the years 1999-2011 in which 37,702 projects, split into 68,746 individual funds, were granted to 23,399 German organizations. However, for the construction of one variable, I additionally consider the years 1990 to 1998. The structure of the data is identical for both data sets.
All funds are classified by an internal hierarchical classification scheme developed by the

German Federal Ministry of Education and Research called "Leistungsplansystematik". Its 16 main areas, which include biotechnology, energy research, etc. These are spitted into varying numbers of sub-classes. These are considerably fine-grained. For instance, it can be differentiated between plant genomics (areas: K04210) and micro-organic genomics (area: K024220). The data also covers non-technological activities. For instance, ecological conceptions for urban regions (area: FC1013) and the improvement of education and training conditions for women (area: RB2550) also receive "R&D" subsidies. At the highest level of disaggregation (6-digits) almost 1,300 unique *research areas* can be differentiated. However, many of these research areas include only few projects and therefore do not contain sufficient network structures. I therefore aggregate the six-digit level to the four-digit level yielding 410 research areas for which meaningful networks can be constructed.[3]

The cooperative nature of projects is indicated by funds being granted to joint projects realized by consortia of organizations ("Verbundprojekte").[4] Participants in joint projects agree to a number of regulations that guarantee significant knowledge exchange between the partners (cf. Broekel and Graf, 2012). Accordingly, two organizations are defined to cooperate if they participate in the same joint project. This implies that the original two-mode network data is transformed into a one-mode projection. On this basis, networks are constructed for each year and 4-digit research area based on the projects that are running for at least one day in the respective year. Isolates are added if organizations receive subsidies for non-cooperative projects.[5] To study these networks' evolution they need to be observed for at least two consecutive years. This is the case for 280 networks that are observed for two to thirteen years, see bottom right plot in Figure 1. They serve as unit of analysis in the following. Note that the population of organizations in one research area (nodes in the network) varies between years as I only consider those organizations that received subsidies in the respective year.

The size of the networks varies considerably.[6] The smallest network consists of three nodes in one period. The largest has 887 nodes. In average the networks have about 75 nodes. The top left plot in Figure 1 depicts the distribution of network sizes pooled over all years. In the middle right plot in Figure 1 it is further visible that most networks have density values less than 0.1.[7] Another important characteristic of networks is their degree centralization.[8] For the 280 networks it ranges from almost zero to 0.78. The distribution of this measure is depicted in the plot on the middle left of Figure 1.

Broekel and Graf (2012) highlight that R&D subsidies networks also differ with respect to the participation of the public sector in general and public research organizations in particular. Networks in research areas that are closer to the education system or that concern basic research are naturally more dominated by the presence the latter. The top left plot in

[3]The four-digit level appears to be the (subjectively) best choice given the trade-off between network size and "technological" disaggregation.

[4]Large joint projects with sub-projects in which multiple organizations participate are disaggregated at the sub-project level.

[5]Isolates: Nodes in a network that are not linked to any other node.

[6]The size of a network is commonly represented by the number of nodes, i.e. number of active organizations.

[7]Network density: The number of observed links divided by the number of potential links.

[8]Degree centralization: The extent to which a network has a star-like structure (Freeman, 1979).

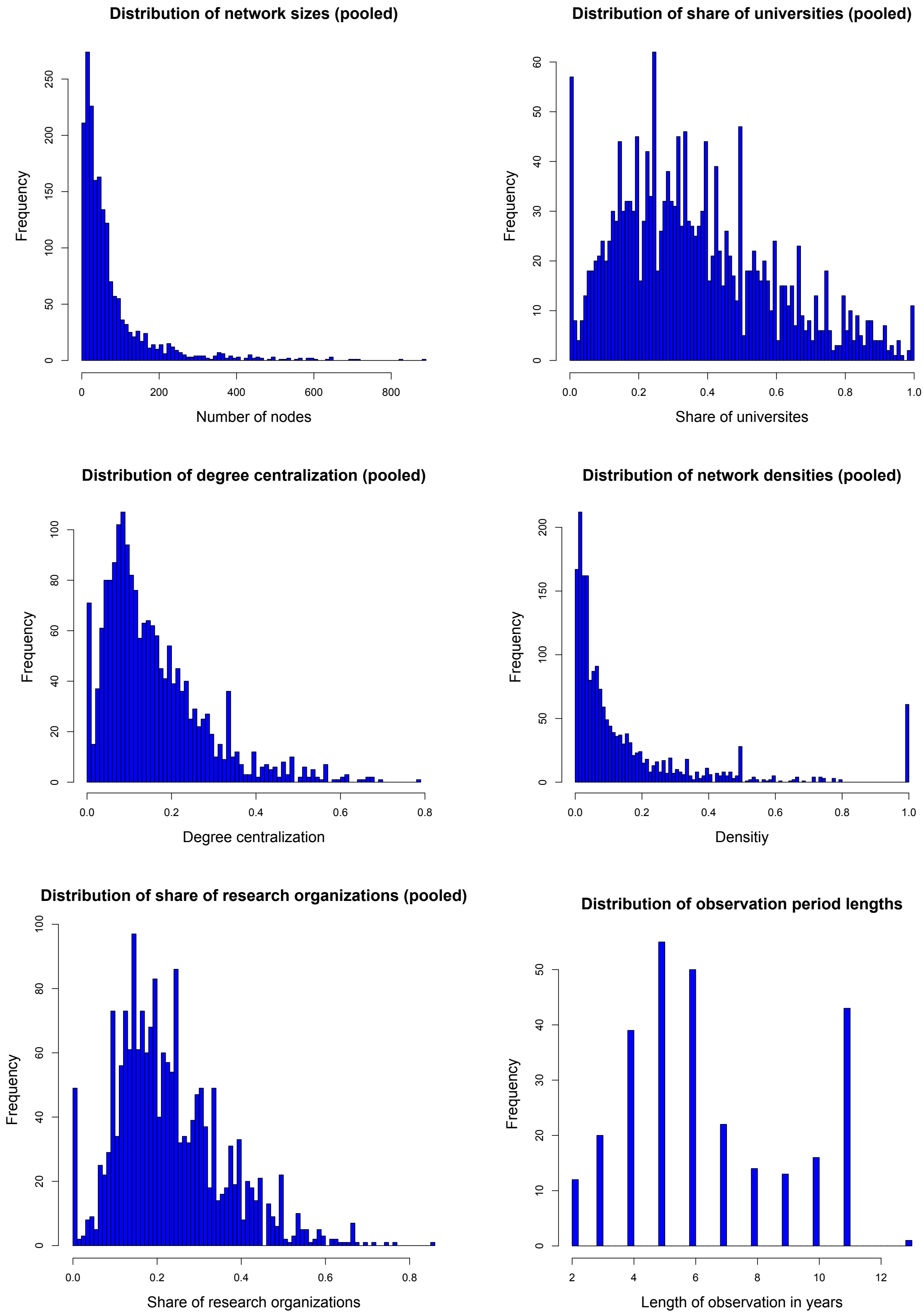


Figure 1: Some network descriptives

Figure 1 highlights that the share of universities lies about 5 to 30 percent for most networks. However, in a large number of networks universities account for more than 60 percent of the nodes. The same is true for research organizations (bottom left plot). The varying importance of universities and research organizations in networks has considerable implications for the structure of networks (Broekel and Graf, 2012).
The plots for the four mentioned network characteristics (size, density, degree centralization, structure of participating organizations) indicate significant heterogeneity among the networks, which demands consideration in the empirical assessment.

## 4. Empirical approach

### *4.1. Empirical variables*

Before the empirical variables are presented, some conceptual clarification is necessary concerning proximities and the notion of 'proximity configurations'. When investigating the role proximities play for network formation, proximity usually refers to the relation between two actors irrespectively of these being actually linked. For instance, researchers will look at the likelihood that two actors that are geographically proximate are having a higher chance of being (or getting) linked than those that are geographically distant. However, the focus of the paper is on co-evolution processes between proximities during the development of networks. For this reason, I follow Ter Wal (2011) and concentrate on the relation of *linked* actors, i.e. the focus is on the proximity structure of actually realized links. The aggregated proximity over all realized relations in a network can be seen as the structure of the network (or its configuration) with respect to one proximity type.
In his study Ter Wal (2011) focuses on the average geographic distance of links in a network to describe the configuration of the network with respect to geographic proximity. Straightforwardly, I extent this idea to the other four proximity types implying that in addition to the average geographic distance, I calculate the average cognitive, organizational, social, and institutional distance of links in the 280 networks.[9]

However, the average distance (proximity) of links in a network is dependent on the distances among actors in the underlying population. In addition, data on subsidized R&D cooperation is inherently subject to changes in policy programs and directives. Hence, when comparing such measures between networks this has to be taken into account. For this reason I compare the observed proximity configuration of networks with the according configuration in the underlying actor population. The proximity configuration is 'normalized' by setting it into a relation with the same measure estimated over all *potential* (i.e. all realized and not realized) links among actors at the same moment in time. The resulting ratio shows if linked actors (i.e. the network) are more proximate in one dimension than what can be expected given the characteristics of the underlying population of actors in that year. For the present paper it is important that it represents the result of the network formation process in previous periods controlled for changes in the underlying actor population. By studying the temporal dynamics of the *proximity configuration* measure inference can be made about

[9] The notion of proximity is unfortunate in this respect as it can be used as a synonym for 'short distance' as well as a reference the proximity concept, e.g. 'cognitive proximity'. To avoid confusion, I will use the notion of 'distance' in the empirical analysis.

the unobserved network formation process and its driving factors. For the estimation of the different distances I rely on established ways to approximate the five proximity types (cf. Balland, 2012; Broekel and Boschma, 2012).

*Organizational distance.* Information are provided for all funds which organization received the funds and, if relevant, which sub-unit is actually executing the project. This differentiation applies primarily to large firms with several sub-units and universities as well as research organizations with multiple institutes. In these cases the mother organization is listed as funded organization and its sub-units / institutes as executing units (see on this Broekel and Graf, 2012). The measure ORGA is defined being zero when two cooperating organizations (executing (sub-)organizations) share the same funded (mother) organization and one otherwise.[10] Using this measure implies that all networks are constructed on the basis of cooperation between executing 'organizations'.

*Institutional distance.* All executing organizations are classified into one of the four categories: university, extramural research organization, private firm, and miscellaneous. I define two organizations to be subject to the same institutional framework if they belong to the same category. As for organizational distance, the variable INST is defined being zero if the cooperating organizations share the same institutional background and one otherwise.

*Cognitive distance.* The subsidies database also includes information on the industry organizations belong to by means of two-digit NACE codes. For the construction of the variable COG, I follow Breschi *et al.* (2003) and Broekel and Boschma (2012) and estimate the co-occurrence of these two-digit NACE codes in the more than 1,300 six-digit research areas. In other words, the cognitive distance between two NACE codes is defined on the basis of the frequency with which their organizations receive subsidies in the same research area. In order to take into account indirect relations between NACE codes and to account for size of the NACE industries the Cosine index is employed:

$$COG\_CO = \frac{\sum_{k=1}^{n} w_{zk}, w_{gk}}{\sqrt{\sum_{k=1}^{n} w_{zk}^2 \sum_{k=1}^{n} w_{gk}^2}} \tag{1}$$

with $n$ being the number of two-digit NACE codes (84) and $g$,$k$,$z$ indicating the respective focal NACE codes. In the equation, $w_{zk}$ is the number with which NACE code $z$ and $k$ coincide in 6-digit research areas. The estimated values vary between zero and one with values close to one indicating high similarity. The variable COG is defined by the negative[11] Cosine value of the cooperating organizations' NACE codes.

*Social distance.* The social distance measure is based on past relations between organizations. More precise, the variable SOC is calculated as the number of links between two cooperating organizations in the previous ten years (moving window). All projects that are still ongoing in the considered year are excluded. To ensure comparability of the variable's value over the entire time period (1999-2011), cooperation networks based on subsidized projects are

[10] This denotation guarantees that large values indicate large organizational distances.

[11] This ensures that large values indicate large cognitive distances.

constructed for the years 1990 to 1998 as well. The variable is again multiplied by minus one for easier interpretation.[12]

*Geographic distance.* Geographic distance GEO is calculated as the physical distance between two cooperating organizations. It is estimated on the basis of the geographic coordinates of the center point of organizations' municipalities.

*Control variables.* To account for networks' heterogeneity, I consider the previously introduced network size (SIZE), network density (DENSITY), degree centralization (CENTRAL), share of universities (UNIVERSITY), and the share of research organizations (RESEARCH). In addition, I take into account the number of industries as approximated by the number of unique two-digit NACE codes (INDUSTRIES) and that are involved in the according research areas as this number might related to the measures of cognitive and institutional distance.

### *4.2. Method*

The variables presented above are estimated for each network and year. As their dynamic relations are in the focus they are transformed into their relative annual growth rates:

$$GROWTH_{it} = \frac{\overline{DISTANCE}_t - \overline{DISTANCE}_{t-1}}{\overline{DISTANCE}_{t-1}} \tag{2}$$

By using growth rates all time invariant effects, i.e. network fixed effects, and potential unobserved heterogeneity among networks are removed. In addition, the growth rates are normalized with the annual median growth, which controls for any economy wide shocks, or annual effects that impact all networks.[13]

To study the dynamic relations between these variables I apply a popular approach from the firm growth literature. Amongst others, researchers in that field are interested in whether profit growth leads to employment growth or vice versa (cf. Coad, 2009). They thereby face a similar problem as the investigation in the present paper: the variables of interest are dynamically related in a complex and endogenous way making uni-directional analyses (standard regression) inappropriate.

One solution for this problem is the 'reduced-form' vector autoregression (VAR) model (see for a discussion Stock and Watson, 2001; Coad, 2009). It does not resolve the causality issue in the processes of network evolution. "*Instead, the results* [of a reduced-form vector autoregression] *should be interpreted merely in terms of describing the regularities that may be observed during the* [growth] *processes*" (Buerger *et al.*, 2012, p. 571).

To study the dynamics of a series of variables the following regression model is specified:

$$x_{i,t} = a + \beta x_{i,t-1} + \epsilon_{i,t} \tag{3}$$

[12] I also estimated a second measure that takes into account previous direct and indirect links between organizations as expressed by the geodesic distance between two organizations in the previous years' cooperation networks. However, this variable proofed to be highly correlated with SOC and therefore dropped.

[13] While the mean is usually used for normalization the presence of some extreme values makes the median yielding more robust results.

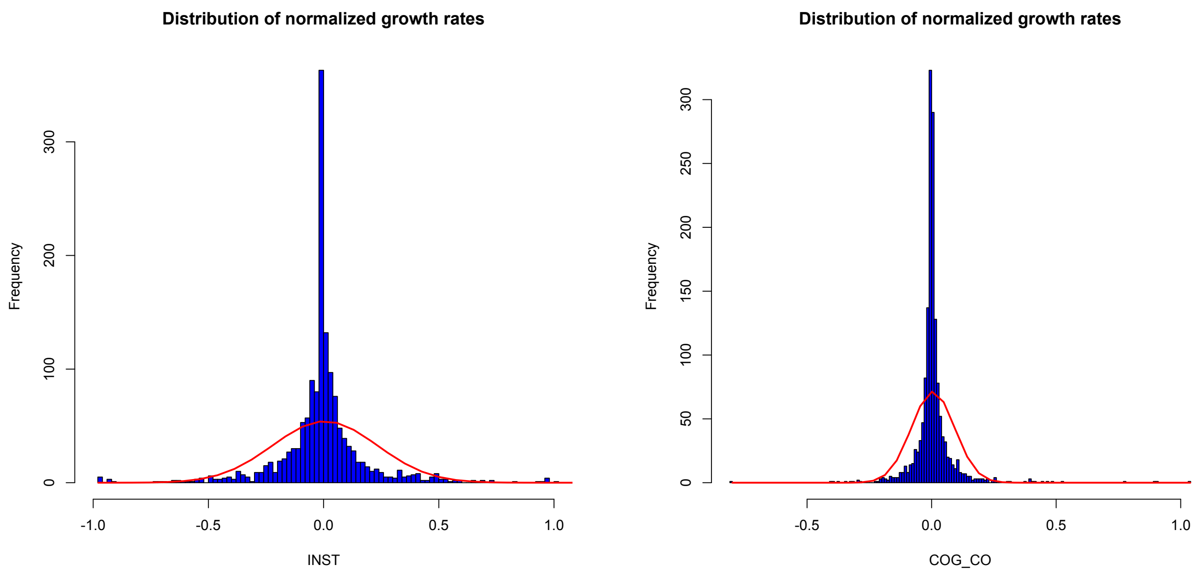


Figure 2: Examples of growth rate distributions

whereby $x_{i,t}$ is a vector of variables describing a network $i$ at time $t$. In the present study these are the five variables approximating the change in proximity configurations. $\epsilon$ is the vector of errors. To control for networks' heterogeneity in terms of levels and growth, I add control variables in normalized (centered to the annual median) growth rates as well as in levels. Accordingly, the above regression equation is extended by the matrix $C$ representing the control variables in levels and $gC$ containing the corresponding growth rates.

$$x_{i,t} = a + \beta x_{i,t-1} + C_{i,t} + gC_{i,t} + \epsilon_{i,t} \tag{4}$$

Following Stock and Watson (2001) Equation 4 can be estimated with a series of $m$ individual ordinary least-squares (OLS) regressions. However, the normalized growth rates turn out to be non-normally distributed, see Figure 2. Quantile regressions (also known as least-absolute deviation regression) are therefore used as they are less impacted by outliers and more appropriate when the dependent variables is not Gaussian (see for a discussion Koenker and Hallock, 2001; Coad and Rao, 2006). In addition, I rely on bootstrapped standard errors for robust and reliable statistical inference (cf. Elfron, 1979). To test the presence of *temporal autocorrelation* the model is estimated as given in Equation 4. However, to study the other types of dynamics the model is adapted. For the case of *simultaneous co-evolution* Equation 4 is rewritten as:

$$x_{t,z} = a + \beta x_{i,i \neq z,t} + C_{i,t} + gC_{i,t} + \epsilon_{i,t} \tag{5}$$

with $z \epsilon i$ representing the focal proximity configuration, which in contrast to Equation 4 is not considered in the set of explanatory variables.
Equation 4 and 5 are estimated using the 280 research area specific networks that are observed for at least two consecutive time periods (1,286 observations). Some descriptives and the

correlation structure between the employed variables are presented in Table 3 and Table 4 in the Appendix.
The third type of dynamics, *long-term co-evolution*, is investigated by estimating the average of variables' growth rates for all networks that are observed for at least 8 years.[14] Equation 5 is therefore slightly adapted.

$$\overline{x_z} = a + \beta \overline{x_{i,i \neq z}} + \overline{C_i} + \overline{gC_i} + \epsilon_i \tag{6}$$

Model 6 is estimated on the basis of the 95 networks for which at least 7 consecutive growth rates can be estimated.

## 5. Empirical results

The results for the first two dynamics, *simultaneous co-evolution* and *auto-correlation*, are shown in Table 1. The coefficients of the control variables already reveal some interesting processes. For instance, the negative coefficient of gIND in Model 2 (dep.: COG)[15], signals cognitive distance to decrease in networks in which the number of involved industries increases. This means that the variety of industries in the population grows faster than the knowledge variety in the network. Model 3 (dep.: SOC) is characterized by the largest number of significant coefficients. Increasing shares of research organizations (gRESEARCH) tend to go along with growing social distances . These institutes are therefore less likely to be involved in repeated interaction. Increasing density also goes hand in hand with low stability of relations (gDENSITY).
In contrast to the control variables being estimated as growth rates, fewer significant coefficients are observed for the variables expressed in levels. A notable exception being the size of networks that positively relates to the growth of organizational distance (SIZE in Model 4 (dep.: ORG)). A potential explanation might be the larger heterogeneity of organizations in big networks, which implies that new links are more likely being established between organizations that are rather distant in terms of their organizational background. While the findings for the control variables are interesting they are not in the foreground of the paper.

*Simultaneous co-evolution.* The first simultaneous co-evolution is observed between networks' configuration with respect to organizational and geographic proximity. In Model 1 (dep.: GEO) cognitive distance (COG) obtains a significant negative coefficient suggesting that decreasing cognitive distance tends to correlate with increasing geographic distance. In other words, if organizations connect to cognitively distant organizations these are likely to be geographically proximate as well. On the network level it means that networks, which expand in space, are probable to consolidate in cognitive variety. This fits to the fact that geographic proximity substitutes for missing cognitive proximity in the establishment of knowledge links (Boschma, 2005).
In Section 2.3 I argued that simultaneous co-evolution dynamics can be expected to exist between social proximity and other proximity configurations (e.g. cognitive, geographic). At least with respect to geographic proximity this is confirmed in Model 1 (dep.: GEO) as the

[14] The threshold of seven years was chosen as a balance of 'long-termness' and number of observations.
[15] "dep.: COG" indicates the dependent variable being COG.

coefficient of SOC becomes positive significant. Accordingly, growing social distances in the network tend to involve increasing geographic distances. This is in line with the observation that social contacts are frequently located in geographical proximity (Breschi and Lissoni, 2001).

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| Dep. Var. | GEO | COG | SOC | ORG | INST |
| (Intercept) | 0.002 | 0.001 | -0.053 | 0.001 | 0.009 |
| GEO | - | -0.003 | 0.079 | 0.001 | -0.035 |
| COG | -0.441** | - | -0.151 | 0.000 | 0.766*** |
| SOC | 0.021** | -0.001 | - | 0.001 | 0.003 |
| ORG | -0.001 | 0.001 | 0.018 | - | 0.001 |
| INST | -0.049 | 0.065*** | -0.026 | 0.000 | - |
| Lag(GEO) | 0.005 | - | - | - | - |
| Lag(COG) | - | -0.026** | - | - | - |
| Lag(SOC) | - | - | 0.016 | - | - |
| Lag(ORG | - | - | - | -0.002 | - |
| Lag(INST) | - | - | - | - | -0.01 |
| SIZE | 0.000 | 0.000 | 0.000 | 0.000** | 0.000 |
| DENSITY | 0.000 | -0.004 | 0.068 | 0.005 | 0.005 |
| DEG_CENT | -0.006 | -0.003 | -0.012 | 0.005 | -0.006 |
| RESERACH | -0.001 | 0.008 | -0.021 | -0.001 | -0.021 |
| UNIVERSITY | -0.001 | -0.002 | 0.044 | -0.006 | -0.006 |
| INDUSTRIES | 0.000 | 0.000 | 0.000 | -0.001* | 0.000 |
| gSIZE | 0.02 | 0.003 | -0.379*** | -0.002 | 0.018 |
| gDENSITY | 0.015 | 0.004 | 0.357*** | 0.001 | -0.022 |
| gDEG_CENT | 0.024** | 0.000 | -0.005 | 0.001 | 0.005 |
| gRESEARCH | -0.01 | 0.003 | 0.068*** | 0.000 | -0.046 |
| gUNI | 0.016 | -0.001 | -0.011 | -0.001 | -0.030 |
| gINDUSTRIES | -0.021 | -0.009* | 0.046 | 0.001 | -0.033 |

* refers to a significance level of 0.1, ** to a significance level of 0.05, and *** to 0.01.

Table 1: Results for co-evolution dynamics

Another simultaneous co-evolution exists between institutional and cognitive proximity configurations. INST is characterized by a positive significant coefficient in Model 2 (dep.: COG). The same is true for COG in Model 5 (dep.: INST). Accordingly, the dynamic relation is bi-directionally confirmed, which can be seen as a sign of robustness. Organizations that share the same institutional framework are probable to be cognitively proximate as well implying that changes in organizational distance correlate with changes in cognitive distance. This has two quite obvious implications. Firstly, variables approximating the two proximity types tend to be correlated in empirical studies, which makes isolating their effects difficult. Secondly, the correlation between the two variables matters for the design of R&D subsidies programs. If policy preferably stimulates cooperation between organizations with divers institutional backgrounds (e.g. cooperation between public research organizations and private firms) it directly influences the cognitive configuration within knowledge networks and thereby impacts the innovation success of embedded organizations (cf. Fornahl *et al.*, 2011).

*Temporal auto-correlation dynamics.* Temporal auto-correlation dynamics are identified for just one proximity configurations, namely cognitive proximity. The lagged variant of cognitive distance (Lag(COG)) gains negative significance in Model 2 (dep.: COG). An increase in cognitive distance in one year is therefore frequently followed by a decreasing distance in the subsequent year, or the other way around. A potential explanation was put forward in Section 2.3: Firms face changing technological problems, which require the permanent adoption of the cognitive/technological structure in their ego-networks (direct cooperation partners). However, permanently changing partners would show as a negative autocorrelation for the social proximity configuration, which remains insignificant, though. It suggests that the relation is more complex and deserves more research in the future.

*Long-term co-evolution.* When comparing Table 1 and 2 some overlap in the results obtained for the control variables in the long-term and short-term investigation can be observed. In contrast to the latter, in which the growth in density correlates with increasing social distance, such is the case for the level of density (DENSITY) in the long-term analysis, Model 8 (dep.: SOC). In addition, social distance tends to increase as the network grows in size (gSIZE) in both specifications.
Some results are specific for the long-term investigation, though. For instance, when networks grow in geographic reach, the share of research institutes (gRESEARCH) increases as well, see Model 6 (dep.: GEO). The positive coefficient of UNIVERSITY in Model 8 (dep.: SOC) signals that social distance tends to increase with more universities participating in the network. Again, I refrain from discussing these relations in more depth as the focus is on proximities.
There is only one short-term relation that translates to the long-term perspective, which may be due to the smaller number of observations (95 vs. 1,286) making significant relations less likely in the long-term investigation.
The positive correlation between cognitive and institutional distance in the long-run investigation is the (short-term) simultaneous co-evolution dynamic persisting in the long run. It becomes visible in Model 7 (dep.: COG) in which INST gains a positive significant coefficient. Hence, the finding confirms the impression of this being a very robust dynamic. It can be concluded that organizations sharing the same institutional framework are also likely to be cognitively proximate, which causes a strong co-evolution between the two proximity configurations.
A dynamic that is only visible in the long-run is the co-evolution of the social and institutional proximity configuration. In contrast to all other relations in this analysis, it is characterized by a bilateral significance of the according variables: growth in institutional distance (INST) is found to be positively related to growth in social distance, see Model 8 (dep.: SOC). The same holds the other way around, see Model 10 (dep.: INST). The meaning of the finding is straightforward: If social distance is once more interpreted as the opposite of the persistence of links (low social distance implies high persistence), the results suggests that higher persistence tends to go along with increased institutional distance. In

| | Model 6 | Model 7 | Model 8 | Model 9 | Model 10 |
|---|---|---|---|---|---|
| Dep. Var. | GEO | COG | SOC | ORG | INST |
| (Intercept) | -0.019 | 0.006 | -0.126 | 0.045 | 0.012 |
| GEO | - | -0.049 | -0.568 | 0.429 | 0.052 |
| COG | -0.251 | - | 0.164 | -0.046 | 0.662 |
| SOC | 0.016 | 0.01 | - | 0.074 | 0.133* |
| ORG | 0.017 | -0.005 | 0.042 | - | 0.031 |
| INST | -0.067 | 0.083** | 0.993*** | -0.028 | - |
| SIZE | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| DENSITY | 0.05 | -0.007 | 1.021** | 0.164 | -0.165 |
| DEG_CENT | -0.054 | -0.003 | -0.418 | 0.015 | 0.088 |
| RESERACH | 0.136* | 0.005 | -0.591 | 0.087 | 0.098 |
| UNIVERSITY | 0.005 | -0.003 | 0.391** | -0.053 | -0.071 |
| INDUSTRIES | 0.001 | 0.000 | 0.004 | -0.001 | 0.000 |
| gSIZE | 0.039 | 0.013 | -0.651** | 0.056 | 0.089 |
| gDENSITY | 0.005 | 0.009 | 0.07 | 0.018 | -0.039 |
| gDEG_CENT | 0.024 | -0.007 | 0.358 | -0.218 | 0.001 |
| gRESEARCH | 0.039 | 0.003 | -0.01 | 0.076 | -0.01 |
| gUNI | -0.007 | 0.000 | 0.084 | 0.006 | -0.018 |
| gINDUSTRIES | -0.023 | -0.018 | -0.222 | -0.172 | 0.077 |

* refers to a significance level of 0.1, ** to a significance level of 0.05, and *** to 0.01.

Table 2: Results for long term co-evolution

other words, organizations with the same institutional background are more likely to cooperate if they have already been cooperating before. Given institutional differences it seems reasonable that private-public links, e.g. university - firm links, are especially characterized by low persistence. However, it could also be the case that links within the same institutional sphere are more persistence as institutional barriers are lower and individuals with the same background might find it easier to re-establish links.

## 6. Discussion and conclusion

The paper contributes to the literature on proximities as well as to the literature on knowledge network evolution. It builds on the proximity approach of Boschma (2005) and argues that proximity structures among connected actors (proximity configurations) are subject to systematic change as knowledge networks evolve. Specifically, arguments are put forward for the existence of distinct co-evolution dynamics between proximity configurations that differ in their temporal dimensions as well as in the involved proximity types.

The arguments are empirically tested using a unique data set of subsidized R&D cooperation disaggregated into 280 networks that are observed for 2 to 13 years. The econometric model is estimated using a reduced-form vector autoregression (VAR).

The results indicate the existence of a significant simultaneous co-evolution of cognitive and geographic proximity configurations. Similar relations are found to exist between cognitive and institutional as well as between social and geographical proximity configurations. The simultaneous co-evolution dynamic between institutional and cognitive proximity configurations is found to persist in the long-run.

While the empirical findings by and large confirm the theoretical predictions, the empirical analysis suffers from a number of shortcomings that need to be discussed. First of all, in order to empirically model proximity configurations, I estimated the average geographic, organizational, social, institutional and cognitive distance of links in networks of subsidized R&D cooperation. In other words, I focused on just one specific measure to approximate the structure of proximities in networks: the average proximity/distance of links. One may argue that this is just one, and maybe not even the most important, measure to describe these structures. For example, the chosen approach does not allow differentiating between networks with many spatially proximate and many spatially far-reaching links on the one side, and networks that are primarily characterized by links spanning medium geographic distances on the other. Future research therefore needs to broaden the approach and take into account different measures, e.g. the minimum, maximum or variance.

Moreover, many potentially existing long-term co-evolution dynamics cannot be identified in the present study because they involve changing actor characteristics. These are inadequately covered in the data or even eliminated in a necessary normalization of the proximity configuration measure. For example, this means that actors are modeled to have the same cognitive profile in all periods. In fact, social embeddedness is the only attribute which change is explicitly captured by the empirical approach. This might explain why the largest number of empirically identified long-term co-evolution dynamics involves this type of proximity configuration.

The employed data is responsible for another shortcoming. The networks in the study are based on subsidized R&D cooperation. While this yields a number of advantages it also implies that all observed networks did not "freely" develop over time and space. To the contrary, their development might have been significantly impacted by policy. Since the empirical investigation includes a large number of networks formed by many diverse policy programs and accounts for a wide range of network characteristics, this is unlikely to cause systematic biases in the empirical results. Nevertheless, a comparison with findings obtained on the basis of network constructed from other data (co-publication, co-patenting, etc.) might provide additional confirmation or reveal interesting differences.

For modeling the long-term dynamics of networks, I rely on 95 networks observed for at least 8 and a maximum of 13 years. It is surely debatable if this qualifies as long-term development. Future research might be able to follow networks' development for much longer periods (cf. Balland *et al.*, 2012b). Such will improve the adequacy of the empirical approach and the underlying theories, which become more relevant in time periods exceeding 8 years.

Despite these shortcomings the present study provides new insights into the mechanisms of network evolution. The empirical analysis shows that (a number of) different proximity configurations in networks are not independent of each other but co-evolve as networks develop over time. This clearly supports a dynamic approach to proximities (Balland *et al.*, 2012a). The study also raises many new questions. For instance, what are the precise reasons and mechanisms for proximity configurations to co-evolve? How and why do knowledge networks and proximity structures between linked actors change over time and what processes are at work in the short and long run? Hence, despite the flourishing research in this field, we are still far from being proximate to fully understand the mechanisms, causes, and consequences of knowledge network evolution.


Balland P. A. (2012) Proximity and the evolution of collaborative networks: Evidence from r&d projects within the gnss industry, *Regional Studies* **46**(6), 741–756.

Balland P. A., Boschma R. A. and Frenken K. (2012a) Proximity and innovation: From statics to dynamics, unpublished working paper.

Balland P.-A., de Vaan M. and Boschma R. (2012b) The dynamics of interfirm networks along the industry life cycle: The case of the global video games industry 1987-2007, *Journal of Economic Geography* **forthcoming**.

Ben-Porath Y. (1980) The f-connection: Families, friends, and firms and the organization of exchange, *Population and Development Review* **6**(1), 1–30.

Boschma R. and Frenken K. (2010) The spatial evolution of innovation networks. a proximity perspective, in Boschma R. and Martin R. (Eds.) *Handbook of Evolutionary Economic Geography*, Edward Elgar, Cheltenham, UK.

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

Breschi S. and Lissoni F. (2001) Knowledge Spillovers and Local Innovation Systems: A Critical Survey, *Industrial and Corporate Change* **10**(4), 975–1005.

Breschi S. and Lissoni F. (2003) Mobility and Social Networks: Localised Knowledge Spillovers Revisited, *CESPRI Working Paper, No. 142* .

Breschi S., Lissoni F. and Malerba F. (2003) Knowledge-relatedness in Firm Technological Diversification, *Research Policy* **32**, 69–87.

Broekel T. and Binder M. (2007) The Regional Dimension of Knowledge Transfers - A Behavioral Approach, *Industry and Innovation* **14**(2), 151–175.

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 Boschma R. A. (2011) The cognitive and geographical composition of ego-networks of firms – and how they impact on their innovation performance, *Papers in Evolutionary Economic Geography* **11.18**.

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.

Buerger M., Broekel T. and Coad A. (2012) Regional dynamics of innovation - investigating the co-evolution of patents, r&d, and employment, *Regional Studies* **46**(5), 565–582.

Cantner U. and Graf H. (2006) The Network of Innovators in Jena: An Application of Social Network Analysis, *Research Policy* **35**(4), 463–480.

Cantner U. and Meder A. (2007) Technological Proximity and the Choice of Cooperation Partners, *Journal of Economic Interaction and Coordination* **2**(1), 45–65.

Coad A. (2009) *The Growth of Firms: A Survey of Theories and Empirical Evidence*, Edward Elgar, Cheltenham, UK.

Coad A. and Rao R. (2006) Innovation and Market Value: A Quantile Regression Analysis, *Economic Bulletin* **15**(13), 1–10.

Cohen W. and Levinthal D. (1990) Absorptive Capacity: A New Perspective on Learning and Innovation, *Administrative Science Quarterly* **35**(1), 128–152.

Elfron B. (1979) Bootstrap Methods: Another Look at the Jackknife, *Annals of Statstics* **7**, 1.26.

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. A. (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. C. (1979) Centrality in Social Networks - Conceptual Clarification, *Social Networks* **1**, 215–239.

Glückler J. (2007) Economic geography and the evolution of networks, *Journal of Economic Geography* pp. 1–16.

Grabher G. (1993) The Weakness of Strong Ties: The Lock-in of Regional Development in the Ruhr Area, in Grabher G. (Ed.) *The Embedded Firm - On the Socioeconomics of Industrial Networks*, pp. 255–277, Routledge, London, New York, Reprinted in 1994.

Hussinger K. (2010) On the importance of technological relatedness: Smes versus large acquisition targets, *Technovation* **30**, 57–64.

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

Koenker R. and Hallock K. F. (2001) Quantile Regression, *Journal of Economic Perspectives* **15**(4), 143–156.

Lazarsfeld P. F. and Merton R. K. (1954) Friendship as a social process: a substantive and methodological analysis, in Berger M., Abel T. and Page C. H. (Eds.) *Freedom and control in modern society*, Van Nostrand, New York, NY.

Mowery D. C., Oxley J. E. and Silverman B. S. (1998) Technological overlap and interfirm cooperation: implications for the resource-based view of the firm, *Research Policy* **27**(5), 507–523.

Nooteboom B. (2000) *Learning and Innovation in Organizations and Economics*, Oxford University Press, Oxford.

Nooteboom B. (2002) *Trust*, Edward Elgar, Cheltenham, UK, Northampton, MA USA.

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.

Scherngell T. and Barber M. J. (2009) Spatial interaction modelling of cross-region R&D collaboration. empirical evidence from the 5th eu framework programme, *Papers in Regional Science* **88**(3), 531–546.

Scherngell T. and Barber M. J. (2011) Distinct spatial characteristics of industrial and public research collaborations: evidence from the fifth eu framework programme, *Annals of Regional Science* **46**, 247–266.

Singh J. (2005) Collaborative Networks as Determinants of Knowledge Diffusion Patterns, *Management Science* **51**(5), 756–770.

Sorenson O. (2003) Social networks and industrial geography, *Journal of Evolutionary Economics* **13**, 513–527.

Stock J. H. and Watson M. W. (2001) Vector auoregressions, *Journal of Economic Perspectives* **15**, 101–115.

Ter Wal A. (2011) The dynamics of inventor networks in German biotechnology: geographic proximity versus triadic closure, *Papers in Evolutionary Economic Geography* **11.02**.

ter Wal A. and Boschma R. (2009) Applying Social Network Analysis in Economic Geography: Framing Some Key Analytical Issues, *Annals of Regional Science* **43**, 739–756.

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

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

## Appendix

| | N | MEAN | SD | MEDIAN | MIN | MAX | SKEW |
|---|---|---|---|---|---|---|---|
| GEO * | 1550 | 0.025 | 0.309 | 0.000 | -1.000 | 6.735 | 9.725 |
| COG * | 1553 | -0.006 | 0.121 | 0.000 | -3.179 | 0.756 | -14.535 |
| SOC * | 1523 | -0.251 | 1.002 | 0.000 | -9.335 | 1.051 | -4.111 |
| ORG * | 1200 | -0.055 | 0.786 | 0.000 | -9.616 | 1.000 | -6.774 |
| INST * | 1549 | -0.054 | 0.593 | 0.000 | -9.117 | 1.001 | -10.937 |
| SIZE | 1844 | 75.458 | 100.311 | 43.000 | 3 | 887.000 | 3.231 |
| DENSITY | 1844 | 0.147 | 0.212 | 0.066 | 0.000 | 1.000 | 2.59 |
| DEG_CENT | 1844 | 0.157 | 0.120 | 0.127 | 0.000 | 0.781 | 1.457 |
| RESEARCH | 1844 | 0.231 | 0.128 | 0.208 | 0.000 | 0.857 | 0.891 |
| UNI | 1844 | 0.362 | 0.225 | 0.327 | 0.000 | 1.000 | 0.638 |
| INDUSTRIES | 3064 | 8.166 | 8.206 | 6.000 | 1.000 | 53.000 | 2.086 |
| gSIZE* | 1553 | 0.236 | 1.054 | 0.000 | -0.915 | 14.475 | 6.344 |
| gDENSITY* | 1553 | 0.150 | 1.107 | 0.000 | -0.973 | 20.124 | 9.414 |
| gDEG_CENT* | 1553 | 0.108 | 0.663 | 0.000 | -1.038 | 9.250 | 4.467 |
| gRESEARCH* | 1553 | 0.22 | 1.973 | 0.000 | -1.006 | 40.000 | 12.656 |
| gUNI * | 1553 | 0.200 | 1.901 | 0.000 | -1.002 | 40.650 | 13.353 |
| gINDUSTRIES* | 2640 | 0.166 | 0.765 | 0.000 | -0.900 | 12.000 | 6.144 |

* Variables are median-centered

Table 3: Descriptives of variables

| | GEO | | COG | SOC | ORG | INST | SIZE | DENSITY | DEG_CENT |
|---|---|---|---|---|---|---|---|---|---|
| COG | −0.15*** | | | | | | | | |
| SOC | 0.1*** | −0.08*** | | | | | | | |
| ORG | 0.04 | 0.01 | 0.17*** | | | | | | |
| INST | −0.07*** | 0.19*** | −0.04 | −0.01 | | | | | |
| SIZE | −0.02 | 0 | −0.09*** | −0.06** | 0.04* | | | | |
| DENSITY | 0.04* | 0.01 | 0.13*** | 0.06** | 0 | −0.35*** | | | |
| DEG_CENT | −0.03 | 0.02 | 0.07** | 0.04 | 0.01 | −0.25*** | 0.15*** | | |
| RESEARCH | 0.05* | 0.02 | 0.06** | −0.02 | −0.02 | −0.14*** | 0 | −0.02 | |
| UNI | 0.04 | 0.01 | −0.05** | −0.02 | −0.02 | −0.06** | −0.11*** | −0.02 | |
| INDUSTRIES | −0.04 | −0.01 | −0.08*** | −0.05* | 0.04* | 0.79*** | −0.37*** | −0.24*** | |
| gSIZE | 0 | −0.05** | −0.44*** | −0.09*** | −0.08*** | 0.06** | −0.09*** | −0.01 | |
| gDENSITY | 0.1*** | 0.02 | 0.33*** | 0.07** | −0.15*** | −0.1*** | 0.12*** | 0.05** | |
| gDEG_CENT | 0.02 | −0.02 | 0.11*** | 0.05* | −0.16*** | −0.07*** | 0.05** | 0.26*** | |
| gRESERACH | 0.07*** | 0.01 | −0.07*** | −0.04 | −0.02 | −0.05** | 0.02 | 0.04 | |
| gUNI | −0.03 | 0.03 | −0.05* | −0.06** | 0 | −0.05* | 0.05* | 0.07*** | |
| gINDUSTRIES | −0.05* | −0.06** | −0.33*** | −0.08*** | −0.11*** | −0.04 | 0.02 | 0.03 | |

| | RESEARCH | UNI | INDUSTRIES | gSIZE | gDENSITY | gDEG_CENT | gRESERACH | gUNI |
|---|---|---|---|---|---|---|---|---|
| UNI | −0.07*** | | | | | | | |
| INDUSTRIES | −0.19*** | −0.35*** | | | | | | |
| gSIZE | −0.06** | 0.02 | 0.09*** | | | | | |
| gDENSITY | 0.07*** | 0 | −0.1*** | −0.23*** | | | | |
| gDEG_CENT | 0.03 | 0 | −0.07*** | 0.06** | 0.39*** | | | |
| gRESERACH | 0 | −0.01 | −0.05** | 0.21*** | −0.04 | 0.09*** | | |
| gUNI | −0.03 | −0.05* | −0.04 | 0.26*** | −0.05* | 0.12*** | 0.09*** | |
| gINDUSTRIES | −0.01 | −0.06** | 0.06*** | 0.79*** | −0.24*** | 0.04 | 0.18*** | 0.25*** |

Table 4: Correlation matrix