Proximity, Network Formation and Inventive Performance:

In Search of the Proximity Paradox

Lorenzo Cassi
lorenzo.cassi@univ-paris1.fr
Université Paris 1, CES - Centre d'Economie de la Sorbonne &
OST - Observatoire des Sciences et des Techniques, Paris

Anne Plunket
anne.plunket@u-psud.fr
Université Paris Sud 11, ADIS

Abstract

This paper investigates how network relations, proximity and their interplay affect collaboration and their inventive performance. Using patent citations as a proxy for patent quality, we investigate how the network and proximity characteristics of co-inventors enable them to access different sources of knowledge, in different geographical and organizational contexts, and finally affect the quality of inventive collaboration. Our findings enable to address the proximity paradox, which states that proximity facilitates collaboration and knowledge sharing, but it does not necessarily increase innovative performance, too much proximity may even harm innovation (Boschma and Frenken, 2009; Broekel and Boschma, 2011).

Keywords: Social networks, geographical proximity, technological proximity, co-patenting, network formation.

JEL codes: O31, O33, R11, D85, L65

1. Introduction

The predominance of geographical proximity for knowledge creation and interactive learning is largely explained by the characteristics of knowledge underlying the processes of innovation (Audrestch and Feldman, 1996; Bathelt et al., 2004). First, spatial propinquity facilitates information and knowledge sharing through frequent interactions, especially when knowledge is tacit, complex and sticky. Second it contributes to solving coordination problems through trust building and inter-organizational learning.

However, geographical proximity *per se* only explains part of the story. Since knowledge is not in the air, actors need to be embedded in local networks and occupy central positions in order to gain access to information and resources that influence innovation (Whittington et al., 2009). And indeed, social proximity within networks is increasingly acknowledged as a key mechanism to understand knowledge flows underlying interactive learning and innovation (Agrawal et al. 2008; Breschi and Lissoni, 2009).

Besides geography and networks, two other factors mediate knowledge flows and interactive learning, one technological and one organizational. Because of absorptive capacity, some level of technological proximity is needed between actors in the network for interactive learning and knowledge exchange to take place (Jaffe, 1989; Cohen and Levinthal, 1990; Nooteboom, et al. 2007). Organizational proximity also facilitates innovation since it reduces uncertainty, limits the risk of opportunism, supports communication between actors and increases performance.

While a high degree of proximity and network embeddedness are main drivers of network formation and knowledge diffusion, the impact on innovative performance is rather ambiguous, "since proximity between actors does not necessarily translate into higher innovative performance" (Boschma and Frenken, 2009). The so-called "proximity paradox" argues that the drivers of network formation should be distinguished from the determinants of innovative performance. If proximity and network embeddedness clearly explain the formation of network relations (Autant-Bernard, et al. 2007; Cassi and Plunket, 2012) interactive learning and knowledge flows (Agrawal et al. 2006; Breschi and Lissoni, 2009), they may not necessarily benefit innovative performance, they may even be harmful for interactive learning (Boschma, and Frenken, 2009).

This paper aims to provide further evidence addressing the so-called proximity paradox. In order to do so, we contrast how proximity and network relations, separately or in combination, affect the formation of collaborations on the one hand, and their inventive performance, on the other hand. The formation of collaborations is studied through co-inventor dyads and the inventive performance is measured by the patent forward citations, as a proxy for the value of inventions or patent quality (Trajtenberg, 1990; Harhoff, et al. 2003). We analyze the networks and proximity characteristics of co-inventor ties raising two questions. First, is there an optimal level of proximity, in the sense that too much or too little proximity may harm the invention process? Second, how the network position combined with proximity enable actors to increase the quality of their inventions.

We address these issues through the analysis of co-inventor networks in the field of genomics in Europe between 1990 and 2006. Our results provide partial support for the proximity paradox as geographical and organizational proximity play little role for explaining innovation performance. In contrast, the actors' network position and technological proximity are prominent.

The paper is organized as follows: section 2 reviews the recent literature on the topic positing our contribution. Section 3 provides a description of data and an explanation of how networks have been built up. Section 4 describes the estimation design; results are discussed in section 5. Section 6 concludes.

2. Networks, proximity and the performance of innovation: the proximity paradox

In this section, we consider how network and proximity determinants, as well as their interplay may differ in explaining innovative activities and performance.

2.1. Geographical proximity and networks

Knowledge diffusion and innovation are known to be highly localized and embedded in industrial clusters, and a large literature has investigated localized knowledge externalities and their impact on knowledge creation. Since a decade or so, an intense debate has been going on to understand under what conditions individuals and firms benefit from these knowledge externalities. However and besides the death of distance, a number of arguments have been advanced to show that, 1) knowledge is far from being in the air and freely accessible to all, and 2) permanent proximity may not be necessary (Torre and Rallet, 2005).

The first argument emphasizes the various channels through which knowledge diffusion occurs. A strong result is that individuals and firms need to be embedded in local networks in order to benefit from knowledge diffusion, especially when knowledge is sticky, complex or tacit, because then, individuals need frequent interactions that facilitate communications, interactive learning and trust building. Although networks are non-geographical mechanisms, they strongly interact with it (Ter Wal, 2011). Two reasons for that: first, knowledge diffusion follows inter-personal channels such as labor mobility, professional acquaintances, co-ethnicity, friendship or kinship (Agrawal et al. 2006). Second, these knowledge flows are localized to the extent that individuals and inter-personal networks are also localized, essentially because individuals are not very mobile in space (Breschi and Lissoni, 2009). This contributes to explain why geography end-up playing little or even no role, once social networks are accounted for (Autant-Bernard et al. 2007; Maggioni et al., 2007).

Concerning the second argument of “non-permanent proximity”, the role of networks is once again emphasized. First, other forms of proximity, such as belonging to the same organization, may play similar roles to geographical proximity in sharing tacit knowledge and solving coordination problems. In a recent paper, Cassi and Plunket (2012) have indeed shown that organizational, social and geographical proximity endorse similar roles and act as substitutes in explaining the formation of co-inventor collaborations. In other words, organizational and social proximity compensate geographical distance when explaining technological collaborations. Second, individuals and firms need access to external and non-local sources of knowledge. In other words, the innovative performance also depends on the extent to which they are also embedded in a global network comprising more distant partners (Bathelt et al., 2004; Witthington et al. 2009).

Regarding the proximity paradox, geographical and organizational proximity play a major role for the formation of collaboration and networks. Once these networks exist, they may play little or even no role for subsequent collaborations. Regarding the performance of innovation, the need to access external knowledge may limit the role of proximity and rather favor networks.

2.2. Network positions, knowledge flows and innovation

Once the role of networks has been accepted as being prominent, it may be that some positions within these networks are more favorable for accessing knowledge and for innovating. The relationship between network positions and performance is usually considered through closure and bridging positions (Coleman, 1988; Burt, 1992; Fleming et al., 2007; Baum, et al., 2012).

Closure positions refer to a situation in which actors are at least indirectly connected within the same sub-network component. When actors are embedded in networks in which they are densely connected, collaborations and knowledge flows are facilitated for the following reasons. First, individuals belong to the same community; they know each other, at least indirectly. This social proximity contributes to generate common language codes, behavioral routines and group norms, which promotes trust and collaboration (Coleman, 1988). Second, cohesive networks also facilitate communication between individuals, and exchanges of tacit and complex knowledge (Reagans and McEvily, 2003). For both of these reasons, increased cohesion solves in part coordination problems by reducing uncertainty as well as the risks of opportunism. Thus, it reduces the cost of coordination and increases the likelihood of collaboration.

As closure ties occur within a close and dense network of actors already indirectly linked, they use redundant information paths within a close community of inventors with similar knowledge bases and technological skills. The proponents of cohesion argue that closure ties provide organizations with two main advantages (Fleming et al. 2007). First, the redundancy facilitates exchanges and mobilization of knowledge. Second, the cost of coordination is largely reduced since actors share a high level of trust. These two reasons explain why closure ties contribute to strengthen the innovation process and promote performance. However, being embedded in very dense and strongly cohesive networks may also harm individuals in their search of new knowledge and their learning processes. In fact, Burt (1992) argues that knowledge accessing is more efficient when individuals occupy structural holes that enable the link of unconnected actors. Individuals positioned in structural holes are able to broker knowledge flows across unconnected groups (e.g. Gargiulo and Benassi, 2000).

Bridging ties provide a brokerage position by connecting actors that have no network link in common since they belong to separate network components. These ties allow establishing a channel across clusters, which facilitates access to non-redundant and even novel sources of information and knowledge (Burt, 1992). This is even reinforced when bridging ties enable actors to access distant clusters endowed with diverse sources of knowledge and resources (Bathelt, et al. 2005).

However, the impact of bridging ties on innovation performance is also ambiguous. On the one hand, bridging positions are inherently “weak” and fragile ties, since they link together unconnected actors that are more difficult to mobilize and coordinate (Fleming et al. 2007). However, this disadvantage may be compensated through some form of organizational proximity, that is, a governance structure that reduces uncertainty and favours coordination in the innovation process (Boschma, 2005). On the other hand, bridging ties promote creativity and provide opportunities for novel combination and recombination of ideas (Obstfeld, 2005; Fleming et al. 2007), which might increase the performance of innovation.

In sum, although closure and bridging ties play different roles, they may both promote innovative performance depending on the *type of knowledge* created and, as will be discussed now, on the optimal level of technological proximity (Nooteboom, et al. 2007, Rowley et al. 2000, Sorenson, et al. 2006).

2.3. Technological proximity and innovative performance

When it comes to knowledge diffusion and especially creation, technological or cognitive proximity plays a key role compared to other forms of proximity, such as social, organizational or geographical boundaries, which aim is primarily that of coordination and learning “facilitator” (Broekel and Boschma, 2011).

Technological proximity means that actors share the same knowledge base or technology. Actors are more likely to collaborate when they have very similar knowledge bases, since it makes communication, learning processes and knowledge sharing easier. However, too much cognitive proximity may yield diminishing and even negative returns since the learning process may not be very rich if actors have a very high degree of technological overlap. As a result, there is an inverted-U relationship between technological proximity and the formation of collaborations (Mowery et al. 1998). A similar relationship is also expected when considering the innovative performance. Some degree of dissimilarity in the knowledge base is needed since resource heterogeneity provides an opportunity for learning and innovation. This is true up to a certain "optimal" level of cognitive distance since after this point, uncertainty and complexity become too important for collaboration to be coordinated and managed (Nooteboom et al. 2007). In the specific case of the Dutch aviation industry, Broekel and Boschma (2011) do not find evidence of an inverted u-shape relationship between technological proximity and innovation performance. Instead, they find a true negative impact supporting the proximity paradox arguments.

However the relationship between technological proximity and innovation depends on the type of knowledge, and the underlying learning process, as well as the level of proximity between actors. When knowledge is complex, its diffusion is more complicated when actors are distant from a geographical, organizational or social point of view. Diffusion and recombination of knowledge is easier when knowledge is of moderate complexity and actors are close to some extent, organizationally or socially (Sorenson et al. 2006). Thus, when knowledge is complex, closure ties may play a more prominent role for promoting innovative performance as compared to bridging ties.

The industrial and technological environments and the underlying learning regimes do also moderate the relationship between technological proximity and innovative performance. If competition is driven by innovations based on exploitative learning, it is expected that innovative performance will be higher when actors are technologically close, and form closure ties (Rowley et al. 2000). In both of these situations, actors share similar knowledge bases, they may benefit from specialization effects and cumulative knowledge. However, technological proximity should not be too close either as already explained. In contrast, if innovation is rather based on exploration, actors may have higher innovative performances if they use different and non-redundant knowledge bases, which means technological proximity should not be too close and actors could benefit from forming bridging ties. However as claimed by Rowley et al. (2000), closure and bridging ties are not necessarily contradictory. It may indeed be the case that in some environments, actors benefit from relying on both situations, closure and bridging ties at the same time, in order to be able to benefit and combine the advantages of exploitation and exploration.

We explore the role of various forms of proximity and network positions and their interaction in determining network tie formation (bridging or closure) and innovation performance, measured in terms of technological value of the patents (i.e. forward citations). We intend to test if these two sets of variables play similar or opposite roles in this context. Moreover, we try to provide some empirical support to the proximity paradox: proximity makes knowledge sharing easier but too much proximity could harm innovation performance.

3. Data and network formation method

The dataset under investigation is composed of all the genomic patents published at the European Patent Office between 1990 and 2006, with at least one inventor reporting a European postal address (EU15, Switzerland and Norway, see Appendix for number of patents and inventors distinguished by country of residence) and the citations received by those patents over the period 1994-2010 by other patents independently of the technological content.

The database was built during a recent research project carried out by ADIS-Paris Sud, LERECO-INRA and the OST – Observatoire des Sciences et des Techniques - supported by the French national research agency (ANR – Agence National pour la Recherche). The EPO Worldwide Patent

Statistical Database (PATSTAT)[1] was searched using a specific strategy involving genetics and genomics keywords in order to define the genomic field (Laurens, Zitt and Bassecoulard, 2010). "Genetics *stricto sensu* is the science of gene heredity and variation of organisms by looking at single genes […] in contrast, genomics typically looks at all the genes or at least at large fractions of a genome as a dynamic system, over time, to determine how they interact and influence biological pathways, networks and physiology, in a much more global sense" (ibid, p.649). A number of experts were asked to validate the lexical query for filtering genomics out of genetic and finally the field delineation and the border areas.

Our final database is a sub-sample of 12,968 patents filed by 4,406 distinct applicants and 24,708 inventors[2]. The data include all patents with at least one inventor reporting a European postal address (EU15, Switzerland and Norway). Thus the population investigated includes all the European inventors and their co-inventors, independently of their localization. Moreover we have included the *direct* link between two non-European co-inventors, if it does exist.[3] Every patent provides information on the inventors, their name and postal address, which enables to define their geographical location at the NUTS 3 level for European inventors and the geographical distance between them. The patent offers also information on applicants, for which we have determined whether they are private companies, research institutes and universities, non for-profit organizations or individuals. For each patent, we also know their IPC – International Patent Classification – codes, which identify their technological fields. We use all these information in order to define the inventor's individual characteristics such as geographical location, technological specialization and affiliation. The affiliation is in this case the organization for which

[1] The original project used PATSTAT October 2007 version. The citation data refer to a more recent version of PATSTAT, April 2011.

[2] The disambiguation of inventors' homonymies has been dealt following the methodology proposed in Carayol and Cassi (2009).

[3] Definition of network population boundaries is a tricky issue. We have decided to include non-European co-inventors and their link between each other. Doing so, we are not sub-estimating the degree (i.e. we consider all their partners) and the clustering (i.e. their link) of European inventors, and we are to not over-estimating the social distance (i.e. we include paths of two and three length existing between European inventors via non-European inventors).

the patent is filed and not necessarily the employer. For instance, it may happen in a number of cases that academic inventors file patents for a private company instead of their own university, but we are not able to identify them.

For the patent produced by the dyads we study, we consider the forward citations they receive. Since citing patent can come from a patent office different from EPO, we consolidate them considering only citing patents having an equivalent EPO patent as suggested by OECD patent manual (OECD, 2009). We consider only citations coming from other EPO patents and from patents of other offices with an equivalent European patent (Martinez, 2010). Moreover, we identify and eliminate self-citations defined at the level of inventors, i.e. a citation is considered as a self-citation if the two patents have at least one inventor in common. For each patent belonging to the dataset, we consider the number of citations received in a five-year window.

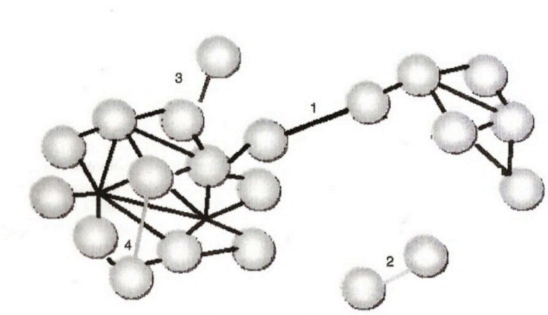
The number of forward citations received from subsequent patents is commonly used as an indicator of patent quality since it is related to its technological importance (Albert et al. 1991, Trajtenberg, 1990, Harhoff et al. 1999 & 2003, Gambardella et al. 2008). We use number of citation received as a proxy of quality of each patent. This is a quite standard measure whose advantage and limitations are known (REFERENCE). What is it worth to emphasize in this context is that using EPO patents make less likely the possible bias related to citations strategy (e.g. to cite patent invented by people I already know). Differently from USPTO, in EPO citation are manly reported by examiners and not by the inventors themselves. Thus a citation received can be considered as independent on the social relation between the inventors involved.

3.1. Network building and tie definitions

In order to build the network,[4] we assign a link (edge) between any two inventors (nodes) who file a patent together. The actors that co-patent form small components that increase over time and eventually connect to other components through new co-patenting activities. Networks may thus be described as bundles of actors that are connected but all the actors within a network are not necessarily linked. Networks evolve over time through the formation of dyads between co-inventors. Following Amburgey et al. (2008), we classify each new link[5] according to the connectivity to the overall network. Four categories of links may be distinguished, as represented in Figure 1: (1) a link bridging two components; (2) a link determining the creation of a new component; (3) a pendant to an existing component; or (4) an intra-component link.

[Figure1]

Figure 1. Type of network ties



Source: Fig. 10, Amburgey et al. (2008)

These new links are explained by the network structure and the inventors' individual characteristics. In order to avoid simultaneity biases, we consider all determinants with a lag of

[4] Social Network Analysis computation has been programmed by the authors themselves with SAS. The SPAM modules developed by James Moody (2000) have been extremely helpful.

[5] We consider only new link between European co-inventors, because, by assumption, the information about social relations of non-European inventors is partial.

one period. For this reason, we may only investigate links among already active actors, that is, bridging and intra-component ties. Another reason for investigating these links comes from the specificity of patents as compared to publications (Fafchamps et al., 2010, Ponds et al., 2007); co-inventors of a given patent have, by definition, the same affiliation[6] and technological field (IPC codes). For this reason, this information cannot be used to highlight organizational or technological determinants.

Bridging and intra-component ties have very different consequences on network structure. Bridging allows for the linking of separate groups of inventors and establishing channels that facilitate the access to resources or other assets. Intra-component ties allow for the establishment of a direct link between actors already (indirectly) connected and the increase of cohesion. In order to consider the impact of cohesion on network formation and innovation, we focus on closure ties that occur within a close social proximity when geodesic distance is equal to 2 or 3. These ties represent 83% of all intra-component ties[7].

Finally, since ties may die out after a certain period of time, we use a five-year moving window to get a more realistic picture of the network for any given year. So, for instance, the network in 1994 is built up considering all the patents published between 1990 and 1994. Accordingly, an inventor is considered as active (e.g. in 1994), if she has at least one patent over the 1990-1994 period.

3.2. Proximity and networks: descriptive statistics

In this section, we briefly describe the characteristics of the co-inventor dyads in terms of proximity regarding geography, organizational arrangements (Table 1 a and b) and technology by the types of ties (Figure 2).

[6] Even for industry-university collaborations, most of the time there is only one affiliation for a given patent, for this reason inventors of a given patent have the same affiliation even if the applicant designated in the patent does not employ them.

[7] Among all intra-component ties, 66% occur within a geodesic distance of 2 (i.e., with the partner of one's partner), and 17% within a geodesic distance of 3.

More than half of all inventors are located in the three largest European countries, that is, Germany, United Kingdom and France. Regarding geography, 86% of all ties occur within the same country and 35% within the same NUTS3 region. Closure ties occur more frequently within regions. When dyads occur across countries, they happen through bridging ties in 65% of the cases (180/274). Regarding organizational arrangements, 42% of all ties occur within the same organization. But, these are mainly closure ties in 59% of those cases. When collaboration happens outside the organization, they occur mainly between organizations with similar types, specifically between firms (in 91% of the cases) and through bridging ties (60%). This illustrates the differences between closure and bridging ties. Closure ties are occur mainly within the same region and country and within the same organization as opposed to bridging ties that occur rather outside regions but within countries and between separate companies.

Figure 2 illustrates the link between forward citations and geographical as well as technological distance by types of ties. We see that closure ties receive more citations than bridging ties, and that bridging ties are more geographically spread. Citations occur for a smaller technological distance for closure ties and larger technological distance for bridging ties.

Table 1a. Descriptive statistics

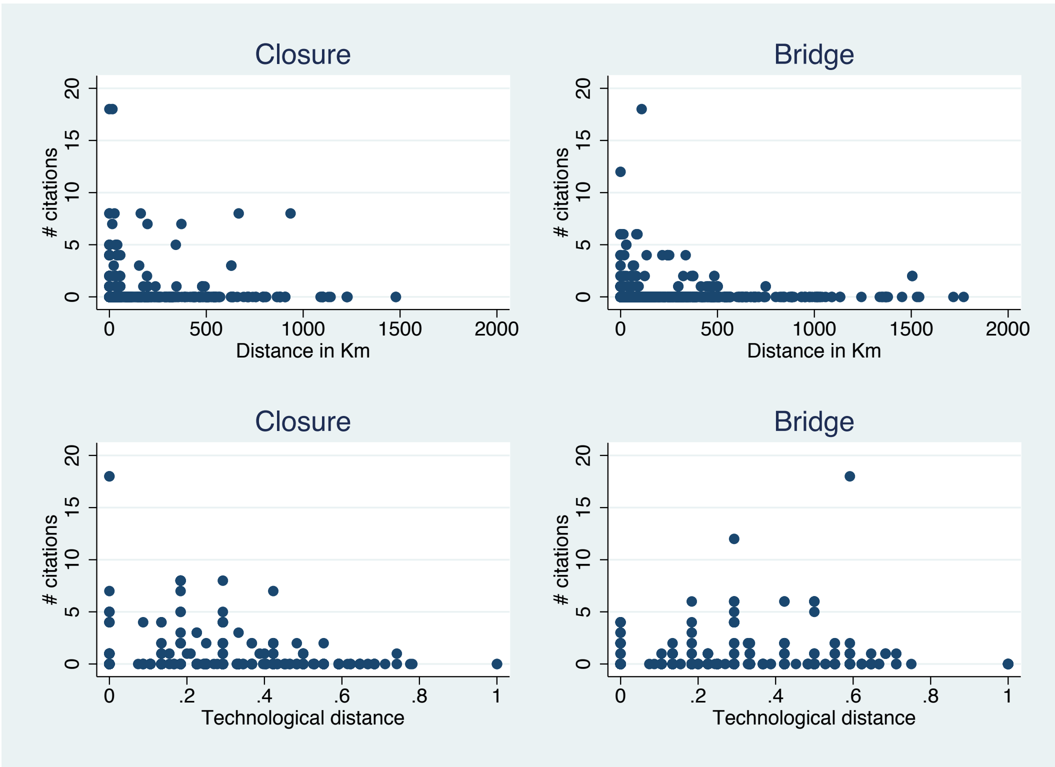
	All ties				Bridging ties				Closure ties			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
# citations	0.25	1.18	0.00	18.00	0.31	1.38	0.00	18.00	0.23	1.07	0.00	0.25
Geographical proximity (log)	-2.98	2.49	-7.99	0.00	-2.55	2.44	-7.30	0.00	-3.43	2.48	-7.99	-2.98
Distance in Km	154.6	273.6	0.00	2944.2	120.33	238.2	0.00	1478.3	191.05	303.8	0.00	154.6
Technological distance	0.72	0.21	0.00	1.00	0.74	0.20	0.00	1.00	0.71	0.21	0.00	0.72
Same applicant	0.42	0.49	0.00	1.00	0.60	0.49	0.00	1.00	0.26	0.44	0.00	0.42
Same type	0.35	0.48	0.00	1.00	0.28	0.45	0.00	1.00	0.42	0.49	0.00	0.35
Degrees - Avrg (log)	1.96	0.53	0.69	3.60	2.15	0.53	0.69	3.51	1.80	0.48	0.69	1.96
Degrees - Abs.diff. (log)	1.62	0.86	0.00	3.95	1.71	0.88	0.00	3.95	1.52	0.82	0.00	1.62
Border	0.05	0.23	0.00	1.00	0.02	0.15	0.00	1.00	0.08	0.28	0.00	0.05
# inventors per patent (log)	1.92	0.41	1.10	3.85	1.94	0.38	1.10	2.94	1.90	0.46	1.10	1.92

Note: 1988 observations for the realized dyads - 980 bridging ties and 820 closure ties – 188 intracomponent other than closure ties

Table 1b. Geographical proximity and organizational arrangements

	Total		Closure		Bridge	
Total	1988	1.00	820	0.41	980	0.49
Geographical proximity						
Same country	1713	0.86	747	0.91	800	0.81
Same region	699	0.35	345	0.42	284	0.29
Different country	274	0.14	73	0.09	180	0.18
Organizational arrangements						
Same organization	831	0.42	490	0.60	251	0.25
Same type	705	0.35	226	0.27	410	0.41
Between firms	645	0.32	192	0.23	387	0.39

Figure 2. Citations given geographical and technological distance



In summary, and as expected bridging ties occur at higher geographical, technological and organizational distance than closure ties. Does it make a difference for tie formation and inventive performance? This is estimated in the following sections.

4. Estimation and variables

In order to explore the proximity paradox, we contrast the impact of proximity and network determinants, first, on the decision to collaborate and second, on the performance of this collaboration. We use similar variables in both analyses to ease the comparison.

4.1. Explaining tie formation

In this first analysis, we estimate how prior network relations and various forms of proximity drive network tie formation. For two inventors i and j, we estimate the probability of forming a tie p_{ij}. If the tie is observed, the dependent variable takes the value of 1 and it is 0 otherwise. Three cases are considered whether we estimate the probability of forming a closure, a bridging or any type of tie as compared to not forming any tie. All realized and possible dyads between any two pairs of inventors represent more than ten million observations and that raises important difficulties of estimation. In order to handle this problem, we adopt a case-control approach: for any realized tie and its related co-inventors, we randomly select ten possible but not realized co-inventors that have filed a patent in the same year as the observed tie, which provide five controls for each co-inventor.. Because the proportion of ties in the sample (11%) is much higher than the proportion of ties in the population (around 0.0022%), logistic regressions may be biased. For this reason, rare event logistic models may be more appropriate to estimate models based on a case-control design (King and Zeng, 2001; Sorenson et al., 2006)[8]. The strategy is to select all the "cases" for which the event is realized (p_{ij}=1, we observe a realized tie in the population as well as in the sample) and consider a random selection of controls (p_{ij} =0, the tie is potential but not realized). Using this sampling method, we know the fractions of ones in the population; in our case, we have 1048

[8] Rare event logit has been implemented through the ReLogit Stata routine proposed by Tomz (1999).

bridging ties and 886 closure ties. To estimate the rare event logit, we implement the prior correction procedure, which involves computing the usual logistic regression and correcting the estimates using prior information about the fraction of ones in the population. In doing so it is possible to correct the estimation, taking in account the difference between the probability of a positive case observed in the sample and the *rarity* of the event actually observed in the population. In our case, we compute the fraction of ones in the population by dividing the number of realized ties by the number of potential ties.

For each regression, we include a full set of year dummies and we use a cluster robust procedure to adjust standard errors for intra-group correlation between realized ties and their controls.

4.2. Explaining inventive performance

The inventive performance is assessed through patent forward citations. The citations, received by other patents in subsequent years following application, provide a measure of the technological quality and economic value of a patent (Trajtenberg, 1990; Harhoff et al. 2003, Gambardella et al. 2008). We compute citations based on equivalent patent (Martinez, 2010) without self-citations. Since the dependent variable is a count measure, a Poisson process is used to estimate the model (Hausman et al., 1984). The basic Poisson process is written as:

$$\Pr\left[p_{it}\right] = f(p_{it}) = \exp(-\lambda_{it}(x_{it}))\left[\frac{\lambda_{it}^{p_{it}}(x_{it})}{p_{it}!}\right]$$

The underlying assumption of the Poisson process is that the probability as well as the variance of the number of events during a time period is equal to the rate λ_{it}, thus making the assumption that there is no heterogeneity in the sample. Since the variance often exceeds the mean and since such over-dispersion causes the standard errors of the parameters to be underestimated, the statistical significance levels are overstated. For this reason, the negative binomial model is often preferred (Cameron and Trivedi, 2005) as it is the case here since the dispersion parameter is highly significant. The model is fitted using a conservative Huber/White/sandwich robust variance estimator and controls for year fixed effects to capture the possible correlation of the dependent variable with omitted time-invariant variables. For the negative binomial we use only the realized patents, which leaves the sample with 886 observations (dyads) for closure ties and 1048 bridging ties.

4.3. Independent Variables

The network variables are tested through the impact of prior network ties. They are assessed through closure and bridging ties. The variable "*Closure"* estimates the impact of social proximity, when the geodesic distance prior to collaboration is equal to 2 or 3. It should increase the likelihood of collaboration since inventors collaborate more easily with their partners' partners because "knowing" them facilitates trust and collaboration, and favor inventiveness. It should also ease innovation output since closure facilitates the use of complex and tacit knowledge. The 1,988 dyads are part of inventor teams that have generated 993 patents. Among those, 43 patents (and 226 dyads) are composed of teams that combine both closure and bridging ties. We introduce a dummy "*Closure and Bridging" patent* to estimate the fact that combining cohesion and brokerage may yield higher returns (Rowley et al. 2000).

The proximity variables are tested using geographical, technological and organizational proximity. In order to calculate the "*geographical proximity"* in kilometers, we locate each inventor at the NUTS 3 level based on its postal address. All European inventors are identified this way; the non-European inventors have been dropped from the regressions. The distance is calculated using the latitude and longitude coordinates of each NUTS 3 centroid[9]. Geographical proximity is thought to

[9] We adjust the latitude and longitude coordinates for the earth curvature; thus the distance in km between two points A and B is computed as:

have a positive impact on the likelihood of forming a tie and the innovative performance since frequent interaction decreases transaction costs, and enables to invest in more complex and productive learning processes.

Collaboration is easier among inventors that share similar technological interests and specializations. For this reason, we suppose that "*technological distance*" decreases the likelihood of collaboration. It is computed as the complement of the Jaffe's (1989) index t_{ij}, which is a proximity measure ranging between zero and one, depending on the degree of overlap between the co-inventors' prior patent IPC codes.

$$t_{ij} = \frac{\sum_{k=1}^{K} f_{ik} f_{jk}}{\sqrt{\sum_{k=1}^{K} f_{ik}^2 \sum_{k=1}^{K} f_{jk}^2}}$$

f_{ik} and f_{jk} represent each inventor *i* and *j* technological position. Innovative performance is an inverted u-shape of technological distance since some technological distance is needed to go beyond 'local search' but too much proximity may limit communication, knowledge exchange and recombination (Gilsing et al. 2008).

We then consider the impact of organizational proximity. Organizational proximity occurs when two inventors file a patent for the same applicant. When inventors file a patent for different organizations, two inventors may work for similar types of organizations, either among academia and public research centres or among private companies (Ponds et al. 2007). We suppose that inventors are more likely to form ties within their own organizational boundaries or with inventors belonging to similar organizational types. In order to account for different organizational settings, we consider three occurrences: "Same applicant" takes the value of 1 when inventors have patented for the same organization prior to tie formation and 0 otherwise; "Same type" takes the

---

d(A,B) = 6371 × arccos[sin(latitude(A)) × sin(latitude(B)) + cos(latitude(A)) × cos(latitude(B)) × cos(|longitude(A) – longitude(B) |)]

value of 1 when inventors have patented for the same organizational type (firms or companies) and 0 otherwise; and "Different type, different applicant" as the last occurrence, in our case university – industry relationships.

We interact closure with geographical proximity and technological distance in order to test if they may have substitutable or complementary impacts on network tie formation and performance. Our hypothesis is that inventors will choose closure ties when they require similar competences that may be found in a close neighborhood. They will choose bridging links when they need distinct skills that may not be found in their own environments. Closure and geography should be substitutes since they endorse similar roles in knowledge communication (Cassi and Plunket, 2012). The sign should therefore be negative. Closure should also moderate the negative impact of technological distance.

Control variables

A number of control variables are introduced in the regressions. First, we account for preferential attachment based on the *degree centrality measure*. Since the study considers the likelihood of two inventors in forming a tie, we must examine this measure for both inventors and consider the average $\bar{n}_{ij}$ and the difference Δn_{ij} of both inventors' degrees (Fafchamps et al. 2010).

$$\bar{n}_{ij} = \frac{(n_i + n_j)}{2}$$

$$\Delta n_{ij} = \left| n_i - n_j \right|$$

For each type of tie, we expect a different sign. In particular, we expect the average measure to be positive and the difference to be negative for closure ties and *vice versa* for bridging ties. When actors belong to the same sub-network, individuals tend to link to partners similar to themselves in terms of degree: thus the difference in the number of partners should tend to zero. This is even more important for individuals with a greater number of collaborations since they are more visible within the network. When individuals are searching for an effective collaboration that enables them to access new and different resources, it is likely that similarity is less important or even plays a negative role. In this case, a greater difference would have a positive effect on tie formation and, consequently, we should expect a negative effect of the average degree as well.

We introduce other types of controls. First, we control for the distinction between domestic and foreign inventors' collaboration. Since being a foreigner is strongly correlated with geographical distance, we prefer to consider the specific case of foreigners located in *border* countries by introducing a dummy for inventors located in two countries with a common frontier.

Second, we also consider the team size through the *number of inventors* involved in research determining the new link. We expect a positive impact on the performance, since the number of inventors could be interpreted a proxy of resources invested.

Finally, we also consider the number of years since the first tie in order to control for experience with the patent process. Again, in order to account for the symmetric relation, we introduce the difference and average value of both inventors' experiences, namely *Experience – absolute difference* and *Experience – average difference.*

5. Estimation results and discussion

In order to test the existence of a proximity paradox, Table 2 contrasts the impact of relational and proximity variables on (1) tie formation and (2) quality of the patents that result from these collaborations. The negative binomial specification includes year dummies[10] and robust standard errors adjusted for intra-group correlation of errors (clustered by patents).[11] Across models,

[10] Year dummies results to be significant (test chi2 = 65.69, df = 10, p-value = 0.000) showing different citations intensity across years.

[11] The null hypothesis of the dispersion parameter (alpha) equal to zero is rejected at 1% in all the models tested. This suggests over-dispersion of the forward citation variable and justifies the adoption of the negative binomial model instead of a Poisson specification.

variables and controls remain overall consistent in sign and magnitude, suggesting that they are rather robust to the introduction of additional variables.

Regarding *proximity mechanisms*, there are some striking differences between model 1a and 1b. Technological and geographical proximity are highly significant, that is, the likelihood of tie formation is larger when co-inventors share similar technological fields and work within close spatial distance. Organizational proximity is also strongly significant and positive; the likelihood of forming a tie increases when inventors patent for the same applicant. This confirms the fact that inventors patent first of all with individuals that belong to their own organization (Singh, 2005). In summary, collaborations mainly occur when inventors are located in close geographical distance to each other, work in similar technological areas and presumably patent for the same organization. While all the sources of similarity impact the formation of collaborations, two variables seem to behave differently when it comes to patent quality, the outcome of the collaboration. First, geographical proximity does not seem to play any role, as in Fornahl et al. (2011), suggesting that the result of collaboration does not suffer from the geographical distance. Technological proximity appears to have a quadratic form, which means that it displays diminishing returns. This supports in part the proximity paradox since it is harmful for the quality of patents when actors have too similar knowledge bases. The optimal level of technological proximity seems to be equal to .8[12], that is 44% of all ties (47% for closure ties and 41% for bridging ties) have a technological proximity equal or greater than .8 (see figure 3). This result could be explained by the redundancy of knowledge that may occur and reduce the performance of the patent in terms of technological quality.

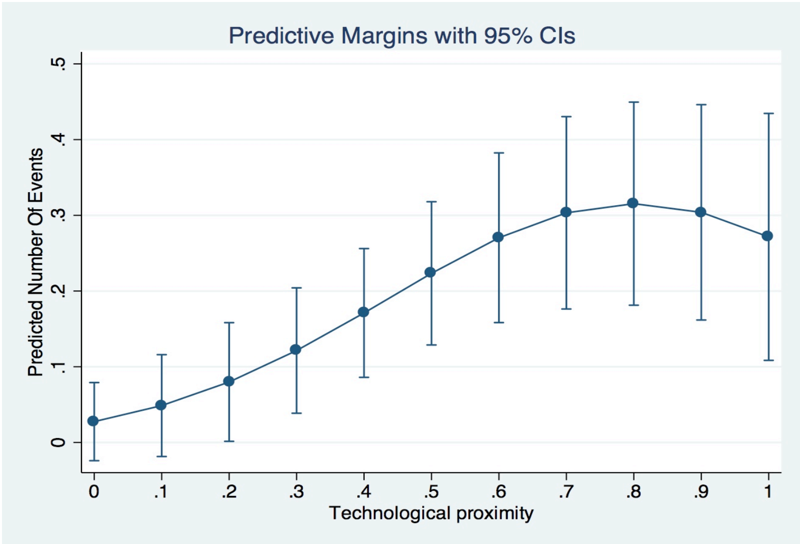
[12] The turning point is equal to coefficient of technological proximity/2*coefficient of technological proximity sq.

Table 2 – Estimation of Tie formation and Citations

VARIABLES	(1a) Tie Formation	(1b) Citations	(2a) Tie Formation	(2b) Citations	(3a) Tie Formation	(3b) Citations	(4a) Tie Formation	(4b) Citations
Closure			2.575***	0.057	2.302***	-2.029*	2.577***	0.070
			[20.15]	[0.20]	[4.47]	[-2.15]	[20.09]	[0.25]
Geographical proximity	0.591***	-0.004	0.540***	-0.005	0.540***	-0.006	0.540***	-0.000
	[37.52]	[-0.09]	[32.30]	[-0.12]	[32.44]	[-0.14]	[32.27]	[-0.01]
Technological proximity	1.197+	6.128*	1.095	6.088*	1.110	6.082*		4.152
	[1.67]	[2.13]	[1.47]	[2.12]	[1.50]	[2.07]		[1.29]
Technological proximity sq	0.702	-3.808+	0.540	-3.798+	0.496	-4.558*	1.104***	-4.335*
	[1.32]	[-1.82]	[0.97]	[-1.82]	[0.89]	[-2.09]	[6.19]	[-2.08]
Same type	-0.072	1.105***	-0.127*	1.104***	-0.127*	1.133***	-0.510*	-0.282
	[-1.17]	[3.77]	[-2.03]	[3.77]	[-2.02]	[4.00]	[-2.29]	[-0.26]
Same applicant	2.473***	0.700*	1.656***	0.695*	1.657***	0.717*	1.442**	-2.559*
	[21.47]	[2.30]	[13.09]	[2.26]	[13.10]	[2.35]	[3.19]	[-2.15]
Closure x Technological proximity					0.379	2.832*		
					[0.56]	[2.27]		
Same type x Technological proximity							0.548+	1.966
							[1.78]	[1.35]
Same applicant x Technological proximty							0.295	4.427**
							[0.48]	[2.87]
Degrees - Avrg	-0.043	0.862**	-0.422***	0.830**	-0.422***	0.767*	-0.412***	0.875**
	[-0.59]	[3.05]	[-5.29]	[2.66]	[-5.27]	[2.55]	[-5.16]	[2.89]
Degrees - Abs.diff.	0.010	-0.383*	0.102+	-0.373*	0.103+	-0.327*	0.101+	-0.357*
	[0.21]	[-2.52]	[1.88]	[-2.35]	[1.88]	[-2.09]	[1.86]	[-2.28]
Border	-1.264***	-0.815+	-1.175***	-0.818+	-1.177***	-0.772+	-1.172***	-0.792+
	[-13.49]	[-1.80]	[-12.21]	[-1.80]	[-12.18]	[-1.72]	[-12.13]	[-1.72]
# inventors per patent	-0.995**	-0.643	-1.150***	-0.647	-1.152***	-0.671	-1.161***	-0.703+
	[-3.14]	[-1.53]	[-3.51]	[-1.54]	[-3.51]	[-1.60]	[-3.55]	[-1.68]
Originality		1.630*		1.624*		1.276		1.458+
		[1.99]		[1.99]		[1.52]		[1.80]
Constant	-7.966***	-6.922***	-7.413***	-6.869***	-7.391***	-6.079**	-6.939***	-5.157*
	[-12.13]	[-3.76]	[-10.99]	[-3.80]	[-10.92]	[-3.27]	[-10.75]	[-2.44]
Alpha (overdispersion)		2.577***		2.576***		2.552***		2.558***
		[12.39]		[12.38]		[12.30]		[12.44]
Observations	23,206	1,988	23,206	1,988	23,206	1,988	23,206	1,988
Log Likelihood	.	-855.5	.	-855.5	.	-852.9	.	-851.6
D.F.	19	20	20	21	21	22	21	23
Chi2	.	63.63	.	64.08	.	72.75	.	89.61

Robust z-statistics in brackets *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

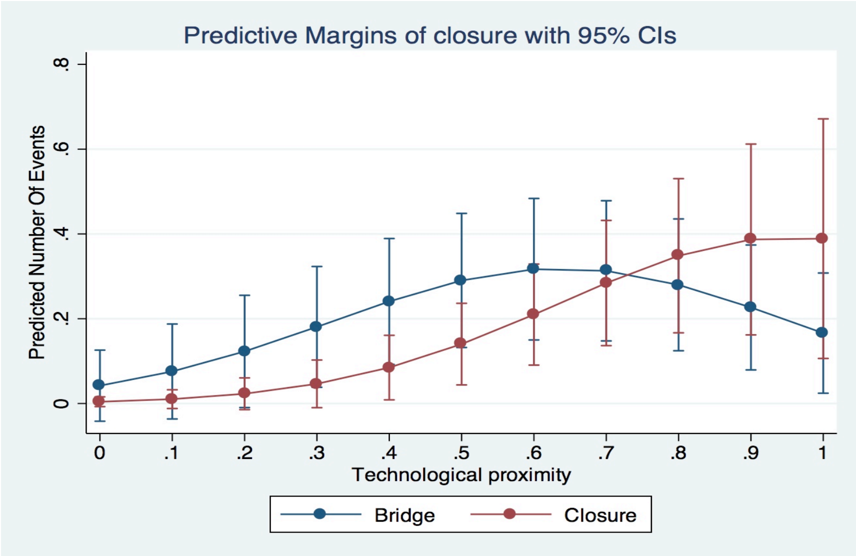
Figure 3. Predicted number of forward citations



Model 2 considers *relational mechanisms* by introducing the variable *closure,* that is, *social proximity,* which indicates that the link is formed between two inventors that are already indirectly connected at a geodesic distance of 2 or 3. This model tests whether the position of actors in the network has an impact on the quality of the patent. The question raised here is whether redundancy in local networks harms the patent as opposed to more distant ties that could bring more novelty and creativity? The first estimation (model 2b) does not support our expectations, since the variable is not significant. Which suggests that the actors network position *per se* does not impact the performance of patents. In order to investigate further the role of networks, we interact in the model 3, social proximity with technological proximity.

Because of the interaction term, the coefficient of closure, which is negative and significant, is now the effect of closure when technological proximity is equal to zero. This confirms the idea that actors are better off seeking partners out of their close network when they need access to technologically different resources, which they may presumably find outside they close social network.

Figure 4. Predicted number of forward citations - Closure versus Bridging



The interaction term is positive and significant which means that the impact of technological proximity depends on social proximity. Figure 4 displays how the network position (i.e. closure vs. bridging) has a different impact of forward citations whether inventors are technologically close or rather distant. Looking at the curve for social connected individuals, it clearly emerges that the two types of proximities are complements in the sense that technological proximity yields higher patent quality when inventors are socially close. Probably these inventors, very similar in terms of technological profiles, can obtain better results by exploiting further their common specialization. This is possible because their social connections facilitate coordination in a profitable way thanks to higher trust and control. This multiplicative effect contradicts the proximity paradox, since it suggests that more proximity is better[13].

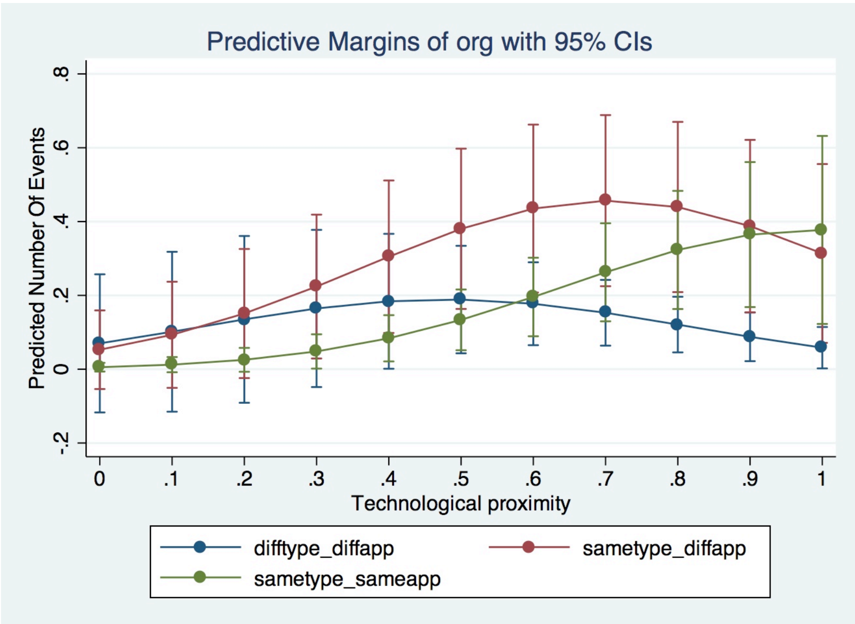
Concerning bridging ties, the figure illustrates that socially distant links yield higher patent quality as long as there is some technological distance, but as technological proximity increases, the performance reaches a peak (technological proximity = .6), after that closure ties yield higher

[13] The marginal effect of closure is tested with a Wald tests the null hypothesis that closure=0 & closure*technological proximity=0, we reject the test with a p-value of 0.0581 (chi2(2)=5.69).

quality. This confirms our understanding of bridging ties that enable collaborations to combine different knowledge bases as opposed to closure ties that occur with a high technological overlap, and provide partial support for the proximity paradox highlighted in prior findings by Broekel and Boschma (2011) as well as Fornahl et al. (2011).

These results find a confirmation if, instead of social proximity, we consider organizational proximity (model 4b and Figure 5).

Figure 5. Predicted number of forward citations – Organisation Proximity



Legend : difftype_diffapp = Different applicant and different organizational type ; sametype_diffapp = Different applicant and same organizational type ; Sametype_sameapp = Same applicant

The figure 5 shows the predicted number of forward citations estimated for links involving: (1) different applicant and type (e.g. firm vs university), (2) different applicant and same type (these are collaborations between different companies in 91%) and finally (3) same applicant. If the first two curves display a similar U-inverted shaped curve, the latter shows a very different pattern. *Same applicant* curve shows that the number of citations increases, as individuals become closer and closer in terms of technological profiles. Collaborations established within the same organization allow individuals to get better results when they can exploit their common specialization. That is not the case for collaborations taking place across organizational boundaries (i.e. different applicant). In this case, the proximity paradox holds and an optimal level of technological distance balancing complementarity and specialization can be easily identified. This is particularly true for collaborations between individuals belonging to different companies. In summary, the same mechanisms of complementarity seem to work between organizational and technological proximity exactly as well as social proximity.

6. Conclusions

The aim of the paper is to investigate how the network and proximity characteristics of co-inventors enable them to access different sources of knowledge in different geographical and organizational contexts, and finally affect the quality of inventive collaboration. Doing so, we are able to compare the effect of the proximity and network variables in establishing collaborations between individuals and the quality of this collaboration. The main conclusion is that both kinds of variables play a different role in the two contexts: what facilitates collaboration does not necessarily yield higher performance.

Our findings partly support the proximity paradox emphasized by Frenken and Boschma (2009). First, we find that proximity variables, in every declination tested - geographical, technological and organizational – have a positive impact on establishing collaborations. Second, when the quality of the collaboration output is considered, geographical distance is not significant. Third, there are main differences if we consider the inventors network position, that is closure versus bridging ties. In the former case, only technological proximity is playing a positive role. That is not the case for bridging ties that are able to manage effective collaborations at an optimal technological distance and if the institutional set is clearly defined, as in intra-organizational collaborations or between individuals sharing the same set of incentives and rules. The same is true when organizational proximity is considered instead of social proximity.

A number of limitations must be raised. Considering the impact of collaboration on innovative performance through co-inventor dyads reduces the scope of the study, since it does not enable to easily consider the network position and characteristics of the organizations in which they invent. The performance of an invention is not only due to the inventor characteristics and abilities; it also depends on their respective organizations. Finally, the dyad approach does not enable us to consider the innovative performance at the team level rather than at the dyad level. This limits also the scope of the results since we are left with a dichotomous approach that does not consider all the complexity of the collaboration.

References

Agrawal, A. & Kapur, D. & McHale, J., (2008), How do spatial and social proximity influence knowledge flows? Evidence from patent data, Journal of Urban Economics, 64(2), 258-269.

Albert, M.B., D. Avery, F. Narin, P. McAllister. (1991). Direct Validation of Citation Counts as Indicators of Industrially Important Patents. *Research Policy*. 20(3) 251- 259.

Audretsch, D. and Feldman, M. (1996), R&D Spillovers and the Geography of Innovation and Production, American Economic Review, 86, 630-640.

Autant-Bernard C., Billand P., Frachisse D. and Massard N. (2007), Social distance versus spatial distance in R&D cooperation: Empirical evidence from European collaboration choices in micro and nanotechnologies, Papers in Regional Science, 86(3), pages 495-519

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: 31-56

Baum J.A.C, McEvily B., and Rowley T.J. (2012), Better with Age? Tie Longevity and the Performance Implications of Bridging and Closure, Organization Science, 23:529-546

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

Boschma, R. A. and Broekel, T. (2009), Knowledge networks in the Dutch aviation industry: the proximity paradox. Papers in Evolutionary Economic Geography# 09.15.

Broekel T. and Boscham R. (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

Boschma, R. A. 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 and Innovation, 14(2), 177–199.

Boschma, R.A. and Frenken, K. (2009): The Spatial Evolution of Innovation Networks. A Proximity Perspective. In: Boschma, R.A., Martin, R. (Eds.), Handbook of Evolutionary Economic Geography, Cheltenham: Edward Elgar, chapter 5.

Breschi, S. and Lissoni, F. (2009), Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows, Journal of Economic Geography 9(4), 439-468.

Broekel, T. and Boschma, R. (2011), Knowledge Networks in the Dutch Aviation Industry - The Proximity Paradox, Journal of Economic Geography, (forthcoming)

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

Carayol N. and Cassi L. (2009), Who’s Who in Patents. A Bayesian approach, Université paris 1 Panthéon-Sorbonne, hal-00631750, HAL.

Cassi L. and Plunket A. (2010), The determinants of co-inventor tie formation: proximity and network dynamics, Papers in Evolutionary Economic Geography (PEEG), 1015

Cameron, A. C. and Trivedi, P. K. (2005), Microeconometrics: Methods and Applications, Cambridge University Press, N. Y

Czarnitzki D., Hussinger K. Schneider C. (2011), Commercializing academic research: the quality of faculty patenting Industrial and Corporate Change, 20(5): 1403-1

Cohen, W.M., Levinthal, D.A., 1990. Absorptive capacity: a new perspective on learning and innovation. Administrative Science Quarterly 35, 128–152.

Coleman, J. S. (1988), Social Capital in the Creation of Human Capital, The American Journal of Sociology 94, S95--S120.

Fafchamps, M., Goyal, S. and van der Leij, M. J. (2010), Matching and Network Effects, Journal of the European Economic Association 8(1), 203—231

Fleming, L., Mingo S. and Chen D. (2007), Collaborative brokerage, generative creativity, and creative success. Administrative Science Quarterly 52(3): 443-475.

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

Gambardella, A., Harhoff, D., Verspagen, B. (2008). The Value of European Patents, European Management Review, 5(2), 69-84

Gargiulo, M. and Benassi, M. (2000), Trapped in Your Own Net? Network Cohesion, Structural Holes, and the Adaptation of Social Capital, Organizational Science 11(2), 183-196.

Harhoff, D., F. Narin, F. Scherer, K. Vopel. (1999). Citation Frequency and the Value of Patented Inventions. *The Rev* Trajtenberg, M. (1990). A Penny for Your Quotes: Patent Citations and the Value of Innovations. *RAND Journal of Economics*. 21(1) 172-187. *iew of Economics and Statistics*. 81(3) 511-515.

Harhoff, D., F. Scherer, K. Vopel. (2003). Citations, Family Size, Opposition and the Value of Patent Rights - Evidence from Germany. *Research Policy*. 32(8) 1343-1363.

Jaffe, A. B. (1989), Real Effects of Academic Research, The American Economic Review 79(5), 957-970.

King, G., Zeng, L., 2001. Logistic regression in rare events data. Political Analysis 9, 137–163

Harhoff, D., Scherer, F.M., Vopel, K. (2003). Citations, Family Size, Opposition and the Value of Patent Rights - Evidence from Germany. Research Policy 32: 1343-1363.

Laurens, P., Zitt, M. and Bassecoulard, E. (2010), Delineation of the genomics field by hybrid citation-lexical methods: interaction with experts and validation process, Scientometrics 82(3), 647-662.

Maggioni, M. A. Nosvelli, M. and Uberti, T.E. (2007), Space versus networks in the geography of innovation: A European analysis, Papers in Regional Science, 86(3), 471-493

Martinez, C. (2010), “Insight into Different Types of Patent Families”, OECD Science, Technology and Industry Working Papers, 2010/02, OECD Publishing

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

Moody, J. (2000), SPAN: SAS Programs for Analyzing Networks. University of North Carolina at Chapel Hill

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

Obstfeld D. (2005), Social Networks, the Tertius Iungens Orientation, and Involvement in Innovation, Administrative Science Quarterly, 50, 100-130

OECD (2009), OECD Patent Statistics Manual,

http://www.oecdbookshop.org/oecd/display.asp?sf1=identifiers&st1=9789264056442

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.

Reagans R. and McEvily B . (2003) Network Structure and Knowledge Transfer: The Effects of Cohesion and Range, Administrative Science Quarterly, 48, 240-267

Rowley T., Behrens D. and Krackhardt D. (2000) Redundant governance structures: an analysis of structural and relational embeddedness in the steel and semiconductor industries, Strategic Management Journal, 21 (3), 369–386

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

Sorenson O., Rivkin J.W. and Fleming L. (2006), Complexity, networks and knowledge flows, Research Policy, 35, 994-1017

Ter Wal A. (2011), The Dynamics of the Inventor Network in German Biotechnology: Geographical proximity and Triadic Closure, Papers in Evolutionary Economic Geography, 11.02

Tomz, M., 1999. Relogit (Stata ado file). Available at

http://www.stanford.edu/~tomz/software/software.shtml

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

Trajtenberg, M. (1990). A Penny for Your Quotes: Patent Citations and the Value of Innovations. RAND Journal of Economics 21(1): 172-187.

Wittington K.B., Owen-Smith J. and Powell W.W. (2009), Networks, Propinquity, and Innovation in Knowledge-intensive industries, Administrative Science Quarterly, 54 :90-122