Proximity, network formation and inventive performance: in search of the proximity paradox

Cassi, Lorenzo and Plunket, Anne

Université Paris 1, CES - Centre d'Economie de la Sorbonne, OST - Observatoire des Sciences et des Techniques, Paris, ADIS Paris sud

15 January 2013

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

January 2013

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 (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; Gilsing et al., 2008). Organizational proximity also facilitates innovation since it reduces uncertainty, limits the risk of opportunism, and supports communication between actors and increases performance.

While a high degree of proximity, whatever defined, 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 empirically investigate the proximity paradox. In order to do so, we contrast how various forms of 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. Doing so, we test if these two sets of variables play similar or opposite roles. Moreover, we intend to investigate if there is an optimal level of proximity, in the sense that too much proximity could harm actors' innovation performance and, if it is the case, to check if complementarity between different proximity dimensions could allow economic actors to cope with this.

We address these issues through the analysis of co-inventor networks in the field of genomics in Europe between 1990 and 2006. 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). Our results provide partial support for the proximity paradox. Although proximity always increases the likelihood of technological collaborations, only technological and organizational proximity explain innovative performance. Network position alone and geographical proximity do not contribute to performance.

In contrast, the actors' network position moderates how technological proximity affects innovative performance.

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 network is 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, 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). Second, 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.

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 importance 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). The proponents of cohesion argue that closure ties provide economic actors 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 unconnected actors that are more difficult to mobilize and coordinate (Fleming et al. 2007). Nevertheless, this disadvantage may be compensated through some form of organizational proximity, that is, a governance structure that reduces uncertainty and favors 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 in next section, on the optimal level of technological proximity (Gilsing, et al. 2008, Rowley et al. 2000, Sorenson, et al. 2006).

2.3. Technological proximity and innovative performance

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; Rowley et al. 2000). 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 (Gilsing, et al. 2008, Sorenson et al. 2006). 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.

Finally, the relationship between technological proximity and innovation also depends on the actors' position within networks. Social proximity should enable to explore distant sources of knowledge since cohesion facilitates coordination and innovation. However, Gilsing et al. (2008) find that performance rapidly drops when central actors operate at larger technological distance, simply because they are not able to overcome dissimilarity and absorb knowledge effectively. On the contrary, they find that highly peripheral actors have a higher level of performance when technological distance is large. As a consequence, our hypothesis is that inventors will choose to collaborate with actors they know (indirectly), that is, they will choose closure ties when they search for similar competences that may be found in their close neighborhood. They will choose to collaborate with partners with whom they have no social link when they search for distinct skills that may not be found within their organization or own environment.

3. Data and network formation method

3.1. Data

The dataset is composed of all the genomic patents published at the European Patent Office between 1990 and 2010 and extracted form the Worldwide Patent Statistical Database (PATSTAT); it was built using a specific search strategy involving genetics and genomics keywords in order to define the genomic filed (Laurens, Zitt and Bassecoulard, 2010)[1]. 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

[1] 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).

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

patents with at least one inventor reporting a European postal address (EU15, Switzerland and Norway) and their co-inventors, independently of their localization. However, to avoid biasing the network relationships and distances between European actors, we have included the direct link between two non-European co-inventors, if it exists[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. The patent offers also information on applicants and technological fields based on IPC – International Patent Classification – codes. 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 the patent is filed and not necessarily the employer.

The database also provides information on the number of forward citations received from subsequent patents, which is commonly used as an indicator of the patent technological and economic value since it is related to its technological importance. This is a quite standard measure whose advantage and limitations are known (Albert et al. 1991, Trajtenberg, 1990, Harhoff et al. 2003, Gambardella et al. 2008). However, using EPO patents reduces the possible bias related to the fact of citing patents coming from friends or acquaintances since as opposed to the USPTO, EPO citations are manly reported by examiners and not by the inventors themselves. Thus, a citation received can be considered as independent of social relations between inventors. Finally, since citing patents can come from offices other than EPO (European Patent Office), we consolidate them considering only citations coming from other EPO patents and other offices with an equivalent European patent as suggested by the 2009 OECD patent manual (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. We consider the number of citations received in a five-year window.

3.2. 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 dyads within 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.

In order to avoid simultaneity biases, we consider all determinants with a lag of one period. For this reason, we may only investigate links among already *active actors,* which implies we may only consider two types of ties: intra-component and bridging ties, i.e., inter-component ties. For intra-component ties, the network distance is equal or greater than 2. The network distance is measured as the geodesic distance, that is, the shortest path between any two inventors. This distance is infinite for bridging ties, since they occur between previously unconnected inventors. Another

[3] The definition of network population boundaries is a tricky issue. Although we use only European inventors for the regressions, the network is built using all inventors, European and non-European. Doing so, we are neither sub-estimating the degree (i.e. we consider all their partners) and the clustering (i.e. their link) of European inventors, nor do we over-estimate their social distance (i.e. we include all paths of two and three length existing between European inventors via non-European inventors).

[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.

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[5] and technological field (IPC codes). For this reason, this information cannot only be used to highlight organizational or technological determinants with a lag.

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 82% of all intra-component ties[6].

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 co-inventor dyads in terms of proximity regarding geography and organizational arrangements (Table 1). The descriptive statistics for all variables are given in the appendix (Table 1a).

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 mainly through bridging ties in 65% of the cases (180/275).

Table 1. Geographical proximity and organizational arrangements

	Total	(1)	**Closure**	(2)	**Bridge**	(2)
Total	1988	100 %	820	41 % (1)	980	49 % (1)
Geographical proximity						
Same region	699	35 %	345	42 %	284	29 %
Cross-country	275	14 %	73	9 %	180	18 %
Organizational arrangements						
Same organization	831	42 %	490	60 %	251	26 %
Between firms	648	32 %	192	23 %	387	39 %
Between public institutions	76	4 %	43	4 %	32	4 %

Note: 1988 observations for the realized dyads - 820 closure ties - 980 bridging ties – 188 intra-component ties other than closure
(1) % in the total sample – Closure and Bridge do not sum to 100 % - the difference come from the non closure intra-component ties
(2) % for each sub-sample (Closure and Bridge)

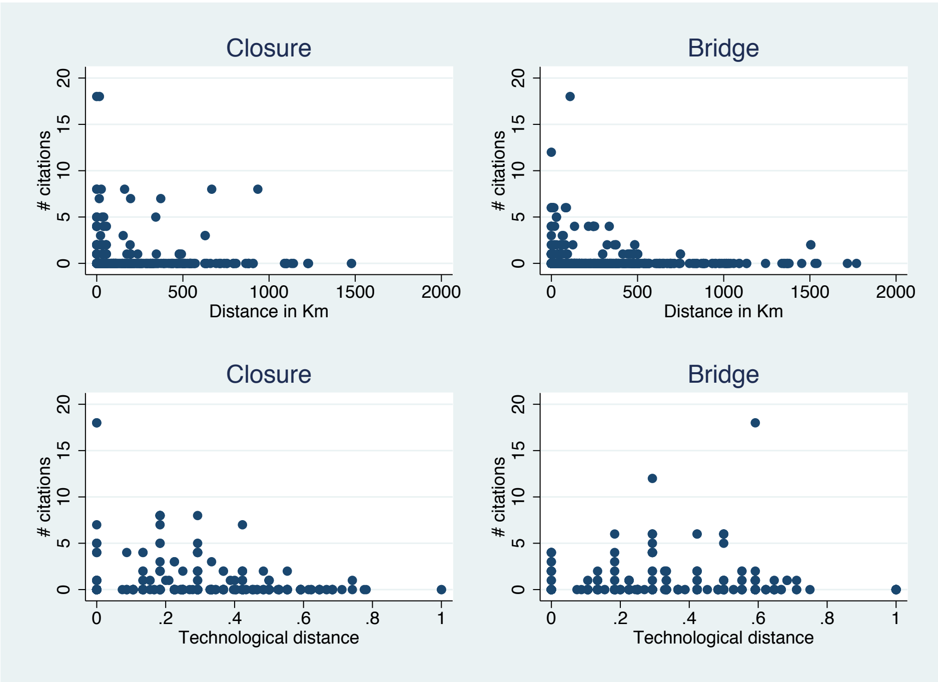
[5] Even for industry-university collaborations, most of the time there is only one affiliation for a given patent, for this reason inventors of a given patent have the same affiliation even if the applicant designated in the patent does not employ them.
[6] Among all intra-component ties, 66% occur within a geodesic distance of 2 (i.e., with the partner of one's partner), and 16% within a geodesic distance of 3.

Regarding organizational arrangements, 42% of all ties occur within the same organization, moreover, these are mainly closure ties in 60% of those cases. When collaborations are inter-organizational, they occur mainly through bridging ties. This illustrates the differences between closure and bridging ties; closure ties 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 organizations.

Table 1a in appendix shows that closure ties receive more citations than bridging ties that are more geographically spread. Figure 2 also illustrates the link between forward citations and geographical as well as technological distance by types of ties. Citations occur for a smaller technological distance for closure ties and larger technological distance for bridging ties.

In summary, and as expected, bridging ties occur at higher geographical, technological and organizational distance than closure ties.

Figure 2. Citations given geographical and technological distance



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 comparison.

4.1. Explaining tie formation[7]

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 1 and it is 0 otherwise. All realized and possible dyads between any two pairs of inventors represent more than 305 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 provides 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 0.0021%[8]), 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)[9]. 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). 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[10]

The inventive performance is assessed through the number of forward citations. Since the dependent variable is a count measure, a Poisson process is used to estimate the model (Hausman et al., 1984). However, the underlying assumption is that the mean as well as the variance of the number of events during a time period should be equal. In our case, the variance exceeds the means as shown by the highly significant dispersion parameter for this reason, we implement a Negative Binomial procedure (see results in table 2 – alpha – over dispersion test, p<0.01 in all cases) (Cameron and Trivedi, 2005).

[7] Dyadic data are typically not independent since residual involving the same individual are likely to be correlated, that is $Cov(\varepsilon_{ij},\varepsilon_{ik})\neq 0$. In consequence, standard errors must correct for cross-observation in the error terms involving the same inventors. The quadratic assignment procedure enables to handle this non-independence using a permutation procedure. Given the number of possible dyads, the procedure is difficult to apply on the whole sample. As a robustness check, we have applied the Netlogit procedure of the SNA R package on the final sample. The results provided in as a supplementary material, lead to similar results (Supplement 1).

[8] If n is the number of active inventors, n*(n-1) is the number of potential ties between these inventors. We estimate this number to be approximately 14190 European inventors. The probability for an event to occur in the sample is approximately equal to 2133/((14190*14189)/2) given that the sample is composed of 2133 realized links.

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

[10] We only observe the performance of dyads for which patents have been applied for and this induces a selection bias. A solution would be to employ a two-stage Heckman selection model, but this requires specifying independent variables that affect the probability of a dyad to patent but that do not affect the level of citations. Since we lack a theory that would suggest such variables, we are not totally confident in this two-stage procedure. However, estimates have been implemented after bootstrapping and the Mills Ratio is not significant.

Errors may be correlated across dyads when there are more than two inventors on a patent. In order to cope with this issue, robust standard errors are adjusted for intra-group correlations (clustered by patents). The model controls for year fixed effects to capture the possible correlation of the dependent variable with omitted time-invariant variables inducing differences in citations across years.

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 and cohesion; it is equal to 1 when the geodesic distance prior to collaboration is equal to 2 or 3, and 0 otherwise. The variable "Bridge" estimates the impact of connecting two separate components. It takes the value 1 when geodesic distance is infinite between inventors and 0 otherwise, and it tests the impact of brokerage. We expect these variables to have a positive impact on performance.

The other proximity variables are tested using geographical, technological and organizational proximity. "*Geographical proximity"* is the inverse of the geographical distance in kilometers which is computed using the longitude and latitude coordinates of each inventor's NUTS 3 centroid[11]; it is thought to 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. "*Technological proximity"* is computed as the Jaffe's (1989) index t_{ij}, an uncentered correlation of two vectors f_{ik} and f_{jk} representing each inventor *i* and *j* technological position defined in terms of *k* IPC 4 digit – International Patent Classification – classes. The index is ranging from zero and one, depending on the degree of overlap between the co-inventors' prior patent IPC codes.

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

As discussed before, some technological distance is needed to go beyond 'local search' but too much distance may limit communication, knowledge exchange and recombination (Gilsing et al. 2008). As a consequence, we expect innovative performance to display an inverted u-share in response to technological proximity, which is introduced in a quadratic form, namely "*technological proximity sq*".

We interact network positions (closure and bridging) with geographical proximity and technological distance in order to test if they have substitutable or complementary impacts on network tie formation and performance. Closure and geography should be substitutes since they endorse similar roles in knowledge communication (Cassi and Plunket, 2012). Closure should complement technological proximity, reinforcing the impact of local search.

[11] 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:

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

We finally 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 centers or among private companies (Ponds et al. 2007). In order to account for different organizational settings, we have considered different categorical variables: "Same applicant" takes the value 1 when inventors have patented for the same organization prior to tie formation and 0 otherwise, and "Same type" takes the value 1 when inventors have filed a patent for organizations of similar types, either between companies or between public institutions.

Control variables

A number of control variables are introduced in the regressions. First, we account for *preferential attachment*, which refers to the tendency for the most connected actors to connect amongst themselves; popular actors tend to attach to popular actors; likewise, low degree actors do so with their peers (Newman and Park, 2003). For this reason, the rate at which actors acquire new ties is a function of the number of ties they already have, thus indicating some form of productivity. For each actor, we measure the *degree centrality*, that is, the number of direct ties (with geodesic distance equal to 1). Since the study considers the likelihood of two inventors in forming a dyad, the degree must be introduced in a symmetric way in the regression. That is, we must examine this measure for both inventors and consider the "*average degree*" $\overline{n}_{ij}$ and the "*difference in degrees*" Δn_{ij} for both inventors' degrees (Fafchamps et al. 2010).

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

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

High degree inventors may be more likely to form a tie because they are more attractive to each other. Thus, the likelihood of forming a dyad will increase with $\overline{n}_{ij}$. The sign on Δn_{ij} will differ depending on whether inventors have similar or rather dissimilar degrees. Second, we also consider the number of years since each inventor's first patent, in order to control for experience with the patent process. For the same reasons, we do also consider their stock of patents. As for the degree, variables are introduced as the difference and the average value between both inventors.

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. *Originality* is a Herfindahl type index based on backward citations made by the patent, indicating if the patent cites a large set of technologies. It is assessed as an indicator of patent value (Trajtenberg et al. 1997).

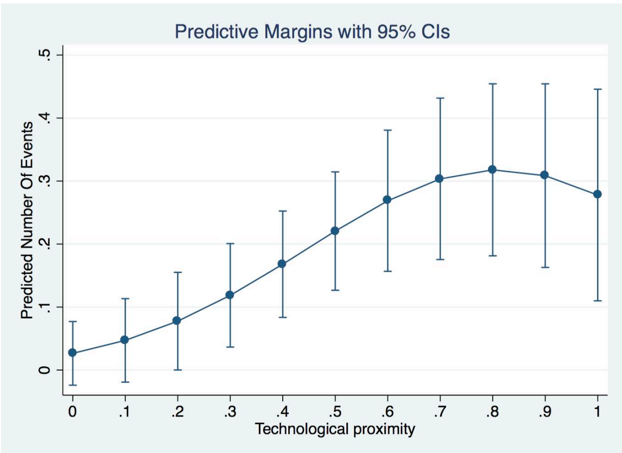
5. Estimation results and discussion

In order to test the existence of a proximity paradox, Table 2 contrasts the impact of relational and proximity variables on tie formation and quality of the patents that result from these collaborations. Regarding proximity mechanisms (model 1b and 1c[12]), technological, geographical and organizational proximity are highly significant, that is, the likelihood of tie formation is larger when co-inventors share similar technological fields, work within close spatial distance and for the same applicant. This confirms the fact that inventors patent first of all with individuals that belong to their own organization (Singh, 2005).

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 enters the model with a quadratic form (model 1c.), 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 is equal to .8[13], 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.

Figure 3. Predicted number of forward citations



[12] Model 1a introduces technological proximity in a quadratic way, since it is not significant, we do not replicate this specification in the following logit models.
[13] 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 ReLogit (1a)	(1b) Tie Formation ReLogit (1a)	(1c) Citations NegBin (1b)	(2a) Tie Formation ReLogit (2a)	(2b) Citations NegBin (2b)	(3a) Tie Formation ReLogit (3a)	(3b) Citations NegBin (3b)
Technological proximity	1.301+	2.030***	6.128*	1.701***	6.076*	1.682***	6.075*
	[1.75]	[13.11]	[2.10]	[9.94]	[2.09]	[9.89]	[2.05]
Technological proximity sq	0.553		-3.778+		-3.765+		-4.523*
	[1.00]		[-1.79]		[-1.79]		[-2.05]
Geographical proximity	0.592***	0.592***	-0.007	0.554***	-0.009	0.554***	-0.010
	[38.35]	[38.33]	[-0.18]	[33.01]	[-0.21]	[33.06]	[-0.23]
Closure				3.471***	0.070	3.302***	-1.999*
				[24.38]	[0.25]	[5.76]	[-2.12]
Closure x Techno. proximity						0.236	2.809*
						[0.30]	[2.25]
Same applicant	2.458***	2.462***	0.726*	1.674***	0.720*	1.675***	0.742*
	[20.32]	[20.37]	[2.37]	[12.12]	[2.33]	[12.07]	[2.43]
Same type	-0.046	-0.045	1.116***	-0.070	1.116***	-0.069	1.146***
	[-0.71]	[-0.69]	[3.80]	[-1.01]	[3.81]	[-1.00]	[4.04]
Degrees - Avrg	0.072	0.072	0.886**	0.233***	0.846**	0.233***	0.782**
	[1.47]	[1.48]	[3.13]	[4.12]	[2.71]	[4.11]	[2.59]
Degrees - Abs.diff.	-0.026	-0.027	-0.383*	-0.526***	-0.371*	-0.526***	-0.325*
	[-0.32]	[-0.33]	[-2.51]	[-5.84]	[-2.33]	[-5.83]	[-2.08]
Border	-1.323***	-1.323***	-0.811+	-1.196***	-0.816+	-1.196***	-0.769+
	[-13.35]	[-13.36]	[-1.78]	[-11.32]	[-1.79]	[-11.31]	[-1.71]
Originality			1.613*		1.605*		1.261
			[1.97]		[1.97]		[1.50]
# inventors per patent			-0.572		-0.578		-0.594
			[-1.51]		[-1.53]		[-1.57]
Experience - Avrg	-0.151**	-0.151**		-0.147*		-0.147*	
	[-2.80]	[-2.81]		[-2.53]		[-2.52]	
Experience - Abs.diff	0.221**	0.218**		0.449***		0.448***	
	[2.81]	[2.78]		[4.95]		[4.95]	
Stock - Abs.diff	-0.124*	-0.125*		-0.182**		-0.182**	
	[-2.06]	[-2.08]		[-2.76]		[-2.76]	
Stock - Avrg	-0.096	-0.097		-0.153		-0.151	
	[-0.81]	[-0.82]		[-1.11]		[-1.09]	
Constant	-4.224***	-4.431***	-7.246***	-4.261***	-7.182***	-4.247***	-6.414***
	[-14.47]	[-22.36]	[-4.14]	[-18.87]	[-4.18]	[-18.88]	[-3.64]
Observations	22,854	22,854	1,988	22,854	1,988	22,854	1,988
Log Likelihood	-3831.02	-3831.48	-855.5	-3359.7	-855.5	-3359.62	-852.9
D.F.	22	21	20	22	21	23	22
Chi2			64.09		64.58		73.38
Alpha (overdispersion test)			2.576***		2.576***		2.551***
			[12.42]		[12.41]		[12.32]

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1
Note: Tie formation are estimated through Rare Event logistic with robust standard errors in parenthesis (clustered over inventors and controls)
Citations are estimated with a negative binomial model with robust standard errors in parenthesis (clustered over patents)
Regressions include year dummies
Except for categorical variables, all variables are in logs.

Table 2 – Estimation of Tie formation and Citations						
	(4a)	(4b)	(5)	(6)	(7a)	(7b)
	Tie Formation	Citations	Citations	Citations	Tie Formation	Citations
VARIABLES	ReLogit	NegBin	NegBin	NegBin	ReLogit	NegBin
Technological proximity	1.676***	6.125*	6.149*	9.093**	1.826***	4.398
	[9.74]	[2.11]	[2.11]	[2.59]	[6.78]	[1.34]
Technological proximity sq		-3.800+	-3.773+	-4.725*		-4.355*
		[-1.80]	[-1.79]	[-2.05]		[-2.06]
Geographical proximity	0.601***	-0.020	-0.006	-0.000	0.591***	-0.002
	[35.34]	[-0.30]	[-0.14]	[-0.01]	[38.08]	[-0.05]
Closure	1.446***	0.137				
	[7.04]	[0.37]				
Closure x geographical proximity	-0.491***	0.023				
	[-10.40]	[0.26]				
Bridge			0.063	2.246*		
			[0.22]	[2.26]		
Bridge x Technological proximity				-2.982*		
				[-2.32]		
Same applicant	1.640***	0.724*	0.736*	0.771*	2.583***	-2.405*
	[11.79]	[2.32]	[2.38]	[2.51]	[6.16]	[-2.06]
Same type	-0.042	1.115***	1.116***	1.147***	-0.351	-0.144
	[-0.58]	[3.80]	[3.79]	[4.08]	[-1.42]	[-0.13]
Same type x Techno. proximity					-0.181	1.788
					[-0.32]	[1.23]
Same applicant x Techno. proximity					0.438	4.262**
					[1.28]	[2.81]
Degrees - Avrg	0.251***	0.848**	0.919**	0.840**	0.072	0.929***
	[4.39]	[2.71]	[2.92]	[2.82]	[1.47]	[3.46]
Degrees - Abs.diff.	-0.563***	-0.373*	-0.392*	-0.334*	-0.023	-0.368*
	[-6.18]	[-2.36]	[-2.45]	[-2.14]	[-0.29]	[-2.47]
Border	-1.198***	-0.823+	-0.807+	-0.746+	-1.321***	-0.781+
	[-10.69]	[-1.80]	[-1.77]	[-1.66]	[-13.36]	[-1.70]
Originality		1.603*	1.615*	1.268		1.437+
		[1.97]	[1.98]	[1.48]		[1.76]
# inventors per patent		-0.592	-0.569	-0.542		-0.594
		[-1.59]	[-1.50]	[-1.43]		[-1.57]
Experience - Avrg	-0.153**				-0.152**	
	[-2.60]				[-2.82]	
Experience - Abs.diff	0.459***				0.220**	
	[5.01]				[2.81]	
Stock - Abs.diff	-0.176**				-0.127*	
	[-2.64]				[-2.11]	
Stock - Avrg	-0.142				-0.098	
	[-1.04]				[-0.83]	
Constant	-4.041***	-7.205***	-7.347***	-8.679***	-4.292***	-5.706**
	[-17.98]	[-4.14]	[-4.18]	[-4.34]	[-17.12]	[-2.80]
Observations	22,854	1,988	1,988	1,988	22,854	1,988
Log Likelihood	-3306.28	-855.4	-855.5	-852.5	-3830.1532	-851.8
D.F.	23	22	21	22	23	22
Chi2	.	65.44	63.95	72.98	.	87.21
Alpha (overdispersion test)		2.575***	2.577***	2.543***		2.558***
		[12.37]	[12.43]	[12.26]		[12.47]

Model 2 and 3 consider network mechanisms by introducing *closure,* and test whether the position of actors in the network has an impact on the quality of patents. The question raised here is whether redundancy in local networks harms patents as opposed to more distant ties that could bring more novelty and creativity. Although strongly significant in the logistic estimation (model 2a), it does not support our expectations, since closure is not significant (model 2b), which suggests that the actor's network position *per se* does not impact performance.

In order to investigate further the role of networks, model 3b interacts social proximity and 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 their close social network. In contrast, the interaction term is positive and significant meaning that the impact of technological proximity positively depends on social proximity.

Model 4a and 4b test the interaction between closure and geographical proximity. It is negative and significant in the logistic model, which confirms that social and geographical proximity are substitutes for explaining the formation of collaborations. However, we do not find a similar relationship when it comes to inventive performance.

Figure 4. Predicted number of forward citations – Closure versus non Closure

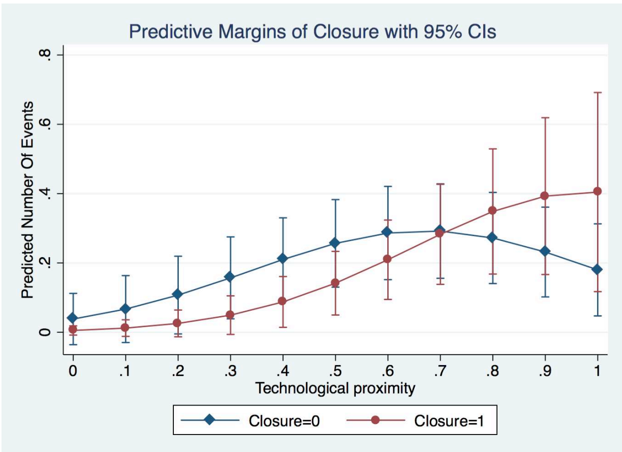
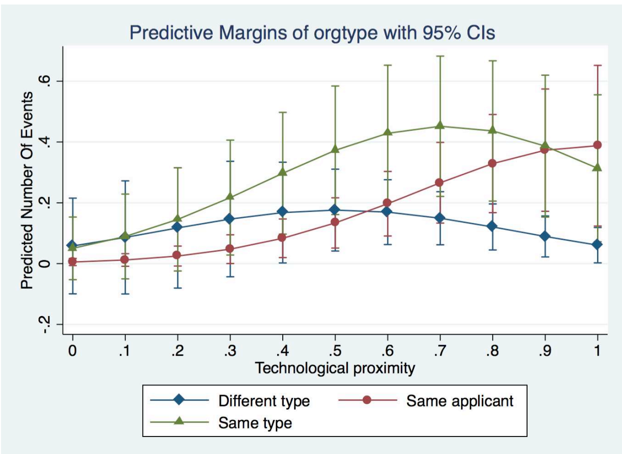


Figure 4 displays how the network position (i.e. closure vs. non closure tie) has a different impact on forward citations whether inventors are technologically close or rather distant. Looking at the curve for socially connected individuals, it emerges that the two types of proximities are complements 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[14].

Regarding bridging ties, the impact on technological performance is opposite as shown by models 5 and 6. The coefficient for Bridge is positive when technological proximity is equal to zero. This means that socially distant links yield higher patent quality for technological distance up to an optimal level; after that, as technological proximity increases the contribution of bridging ties decreases performance. 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 we consider organizational instead of social proximity (models 7a and 7b) as illustrated graphically. Figure 5 displays the predicted number of citations, estimated for different organizational types involving: (1) *Same type*, that is, mainly collaborations between private companies; (2) *Different type*, that is, collaborations between firms and public research and (3) Same applicant. If the first two curves display a similar U-inverted shape, the latter shows a very different pattern. *Same applicant* curve shows that performance increases, as individuals become closer in terms of technology. Collaborations established within the same organization allow individuals to get better results when they exploit their common specialization as opposed to collaborations taking place across organizational boundaries. 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 it does for social proximity.

Figure 5. Predicted number of forward citations – Organizational proximity



[14] The marginal effect of closure with interaction is tested with a Wald test; the null hypothesis that closure=0 & closure*technological proximity=0, is rejected with a p-value of 0.0394 (chi2(2)=8.34).

6. Conclusions

The aim of the paper is to investigate how the network and proximity characteristics of co-inventors ease access to different sources of knowledge in different social, geographical and organizational contexts, and explain the formation and the quality of their collaboration. The main conclusion is that these 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) and Broekel and Boschma (2011). First, we find that proximity variables, in every declination tested – social, geographical, technological and organizational – have a positive impact on establishing collaborations. 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. We also confirm that social, geographical and organizational variables are substitutes, as they behave in a similar way in explaining the formation of network ties (Cassi and Plunket, 2012). Second, when performance is considered, geographical distance is not significant. Third, technological proximity displays diminishing returns. Moreover, 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, which is not the case for bridging ties that are able to manage effective collaborations at an optimal technological distance. The same is true when organizational proximity is considered instead of social proximity. Organizational and social proximity act as moderators in the process of knowledge creation, as more proximity yields higher technological performance when more specialization is needed. Instead, when actors seek more distant competences, bridging ties become prominent.

A number of limitations must be raised. First, 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. Second, our definition of social proximity, only captures a subset of relevant interpersonal relations related to the patenting activity. An extension could be to supplement social proximity with additional data such as collaborations through publications to have a broader picture of network connections. The third limitation is related to the motivation of individuals. Our framework does not allow accounting explicitly for the motivation nor for the strategies of individuals in establishing their collaboration. For this reason, our analysis proposes to disentangle the effect of different dimensions of proximity in establishing one type of tie rather than the other. Finally, the performance of an invention is not only due to the inventors' 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.

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 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. (2012), Research collaboration in co-inventor networks: combining closure, bridging and proximities, *Regional Studies*, under review

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

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.

Gilsing V., Nooteboom B., Vanhaverbeke W., Duysters G. and A. van den Oord, (2008), Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density, *Research Policy*, 37(10), 1717-1731

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

Trajtenberg, M., R. Henderson and A. Jaffe (1997), "University *versus* Corporate Patents: A Window on the Basicness of Invention", *Economics of Innovation and New Technology*, Vol. 5 (1), pp. 19-50.

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

Appendix

Table 1a. Descriptive statistics												
	All ties				Closure ties				Bridge ties			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
# citations (family)	0.25	1.18	0.00	18.00	0.31	1.38	0.00	18.00	0.23	1.07	0.00	18.00
Closure	0.41	0.49	0.00	1.00								
Bridge	0.49	0.50	0.00	1.00								
Geographical proximity	0.37	0.47	0.00	1.00	0.44	0.48	0.00	1.00	0.30	0.45	0.00	1.00
Geog. Proximity in km	154.6	273.6	0.00	2944.2	120.3	238.2	0.00	1478.3	191.0	303.8	0.00	2944.2
Border	0.05	0.23	0.00	1.00	0.02	0.15	0.00	1.00	0.08	0.28	0.00	1.00
Technological proximity	0.72	0.21	0.00	1.00	0.74	0.20	0.00	1.00	0.71	0.21	0.00	1.00
Same applicant	0.42	0.49	0.00	1.00	0.60	0.49	0.00	1.00	0.26	0.44	0.00	1.00
Same type	0.35	0.48	0.00	1.00	0.28	0.45	0.00	1.00	0.42	0.49	0.00	1.00
# inventors	6.29	3.62	2.00	40.00	6.41	3.06	2.00	18.00	6.24	4.21	2.00	40.00
Degree - Avrg	7.22	4.74	1.00	35.50	8.82	5.19	1.00	32.50	5.84	3.95	1.00	35.50
Degree - Abs. Diff	6.25	7.01	0.00	51.00	6.99	7.34	0.00	51.00	5.49	6.66	0.00	50.00
Experience - Avrg	4.34	2.58	1.00	14.50	4.16	2.65	1.00	14.00	4.44	2.44	1.00	14.50
Experience - Abs. Diff	3.17	3.04	0.00	15.00	3.07	3.09	0.00	14.00	3.23	2.97	0.00	15.00
Stock - Abs.diff	5.05	7.18	0.00	78.00	0.72	5.34	-34.50	39.00	0.55	3.42	-18.50	30.50
Stock - Avrg	0.61	4.35	-34.50	39.00	6.46	8.64	0.00	78.00	3.88	5.73	0.00	61.00
Pool - Avrg	0.61	4.35	-34.50	39.00	0.72	5.34	-34.50	39.00	0.55	3.42	-18.50	30.50
Pool - Abs.diff	5.05	7.18	0.00	78.00	6.46	8.64	0.00	78.00	3.88	5.73	0.00	61.00
Originality	0.65	0.19	0.00	0.90	0.63	0.22	0.00	0.90	0.66	0.17	0.00	0.90

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

Correlation table

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1
1	**# citations (family)**	1.0000																
2	**Closure**	0.0378	1.0000															
3	**Bridge**	-0.0222	-0.8262*	1.0000														
4	**Geographical proximity**	-0.0005	0.1252*	-0.1367*	1.0000													
5	**Border**	-0.0217	-0.1165*	0.1294*	-0.1884*	1.0000												
6	**Technological proximity**	0.0395	0.0789*	-0.0442*	0.0166	-0.0148	1.0000											
7	**Same applicant**	0.0110	0.3050*	-0.3236*	0.2044*	-0.0787*	0.0268	1.0000										
8	**Same type**	0.0408	-0.1384*	0.1314*	-0.0735*	0.0062	-0.0213	-0.6282*	1.0000									
9	**# inventors**	0.0081	0.0280	-0.0140	-0.0658*	-0.0456*	0.0005	-0.1156*	0.1010*	1.0000								
10	**Degree - Avrg**	0.0154	0.2835*	-0.2873*	0.0086	-0.0067	-0.0355	0.0400	-0.0078	0.1169*	1.0000							
11	**Degree - Abs. Diff**	-0.0330	0.0891*	-0.1059*	-0.0017	0.0328	-0.0927*	-0.0082	-0.0375	-0.0020	0.7193*	1.0000						
12	**Experience - Avrg**	-0.0409	-0.0593*	0.0364	0.0284	0.0000	0.0059	0.0342	-0.0228	-0.0470*	0.2233*	0.1469*	1.0000					
13	**Experience - Abs. Diff**	-0.0170	-0.0259	0.0215	-0.0439	0.0128	-0.0251	0.0319	0.0032	-0.0093	0.1888*	0.1674*	0.4778*	1.0000				
14	**Stock - Abs.diff**	0.0447*	0.0222	-0.0133	0.1108*	0.0151	-0.0196	0.0033	0.0135	-0.0125	0.1038*	0.1776*	0.0473*	0.0449*	1.0000			
15	**Stock - Avrg**	0.0269	0.1645*	-0.1600*	-0.0192	-0.0050	-0.0545*	0.1201*	-0.0508*	-0.0631*	0.4656*	0.4763*	0.2350*	0.3092*	0.1723*	1.0000		
16	**Pool - Avrg**	0.0447*	0.0222	-0.0133	0.1108*	0.0151	-0.0196	0.0033	0.0135	-0.0125	0.1038*	0.1776*	0.0473*	0.0449*	1.0000*	0.1723*	1.0000	
17	**Pool - Abs.diff**	0.0269	0.1645*	-0.1600*	-0.0192	-0.0050	-0.0545*	0.1201*	-0.0508*	-0.0631*	0.4656*	0.4763*	0.2350*	0.3092*	0.1723*	1.0000*	0.1723*	1.0000
18	**Originality**	0.0361	-0.0768*	0.0548*	0.0164	0.0128	-0.0108	0.0194	-0.0606*	-0.0746*	-0.0479*	0.0246	0.0218	-0.0060	0.0233	0.0324	0.0233	0.0324

*** p<0.05**

Supplement 1 – Robustness check – Handling autocorrelation

Robustness check

Dyadic data are typically not independent since residual involving the same individual are likely to be correlated, that is Cov(ε_{ij},ε_{ik})≠0. In consequence, standard errors must correct for cross-observation in the error terms involving the same inventors. The quadratic assignment procedure enables to handle this non-independence using a permutation procedure. Given the number of possible dyads, the procedure is difficult to apply on the whole sample. As a robustness check, we have applied the Netlogit procedure of the SNA R package on the final sample. The quadratic assignment procedure used is the Dekker's semi-partialling procedure, namely Netlogit in the following table. Coefficient estimates are the same; the p-values differ since the QAP corrects standard errors.

We apply the procedure on a subset of dependent variables and their interactions. Results are similar to those displayed in the paper. In the first regressions, all proximity measures are positive and significant. The interaction term is negative and very significant in the second set of variables, confirming that social and geographical proximity are substitute. Although the interaction term is significant, Closure when geographical proximity is equal to 0 is not significant anymore, meaning that social proximity does not increase the likelihood of forming ties when individuals are too distant geographically.

Semi-Partialling Quadratic Assignment Procedure

	Estimate	Logit Pr(>=|b|)	Netlogit Pr(>=|b|)	Estimate	Logit Pr(>=|b|)	Netlogit Pr(>=|b|)
(intercept)	-4.038	0.000	0.000	-3.929	0.000	0.000
Closure	1.693	0.000	0.027	0.959	0.000	0.229
Geographical proximity	0.676	0.000	0.000	0.702	0.000	0.000
Technological proximity	1.682	0.000	0.000	1.656	0.000	0.000
Closure x Geographical prox.				-0.536	0.000	0.001
AIC		24177.36			23920.75	
BIC		24226.33			23981.96	

Reference

Dekker D., Krackhardt D. and Snijders T. (2003), Multicollinearity Robust QAP for Multiple-Regression