Papers in Evolutionary Economic Geography

10.15

The determinants of co-inventor tie formation: proximity and network dynamics

Lorenzo Cassi and Anne Plunket

Utrecht University
Urban & Regional research centre Utrecht

http://econ.geo.uu.nl/peeg/peeg.html

The determinants of co-inventor tie formation: proximity and network dynamics[•]

Lorenzo Cassi

lorenzo.Cassi@univ-paris1.fr

Université Paris 1, CES - Centre d'Economie de la Sorbonne and

OST - Observatoire des Sciences et des Techniques, Paris

Anne Plunket

anne.plunket@u-psud.fr

Université Paris Sud 11, ADIS

November 2010

[•] The order of authorship is alphabetical. This research is funded by a grant from the ANR – Agence Nationale pour la Recherche – *Corpus Genomique* project. We are grateful to the seminar and conference participants of ARSDLF Clermont 2009, Colloque proximité, Poitier 2009, ADIS internal seminar, Fudan-Paris 1 joint seminar June 2010, Toulouse Eurolio, 2010, Dime workshop Utrecht, September 2010 and especially to Koen Frenken for their comments. We wish to thank Endri Nocka, Sounia Chanfi and especially Olivier Antelo for research assistance on this project.

The determinants of co-inventor tie formation: proximity and network dynamics

Abstract

This paper investigates the determinants of co-inventor tie formation using micro-data on genomic patents from 1990 to 2006 in France. We consider in a single analysis the relational and proximity perspectives that are usually treated separately. In order to do so, we analyse the determinants of network ties that occur within existing components *and* between two distinct components (i.e. bridging ties). We test the argument that formation of these two different types of ties results from distinct strategies in accessing resources. Doing so, we contrast network and proximity determinants of network formation and we investigate if social network allows economic actors to cross over geographical, technological and organizational boundaries.

Keywords: Social networks, relational perspective, proximity, co-patenting, network formation.

JEL codes: D85, O31, R12, Z13

1. Introduction

The fact that social networks matter for innovation is now widely acknowledged, and even considered as a truism (Lobo and Strumsky, 2008). A growing body of literature convincingly argues that knowledge is far from being in the air and accessible to all actors but rather follows specific channels between socially and personally linked individuals (Breschi and Lissoni, 2005, 2009; Knoben, 2009). These "social proximity" arguments strongly contrast with previous studies on geographical proximity that investigate agglomeration economies and argue that knowledge circulates more or less freely among co-located industrial and academic actors suggesting that they benefit from a premium depending on their location (Jaffe, 1989; Jaffe et al. 1993; Audretsch and Feldman, 1996; Aharonson et al. 2008; Boufaden and Plunket, 2008; Knoben, 2009).

Although social networks suggest that innovation and diffusion of knowledge do not simply depend on the location, it cannot be ignored that it strongly interacts with it (Boschma, 2005; Torre et Rallet, 2005). Networks and proximity appear as highly interrelated phenomena since the formation of networks is highly spatially localized at least in its earliest stages (Ponds et al. 2010) and mainly found within organizational, institutional and cognitive boundaries (Singh, 2007). Surprisingly enough, the dynamics of network formation remains an under-investigated question that has been, recently, highlighted as a major research objective for the geographical analysis of innovative networks (Boshma and Frenken, 2009). These debates raise a number of questions: first, to what extent geographical and social proximity are overlapping phenomena? Second, to what extent networks enable to reduce geographical, organizational and cognitive boundaries and offer the opportunity to access non-local knowledge (Gluckler, 2007).

The aim of this paper is to investigate these questions by analysing the determinants of scientific and technological network collaborations, namely inter-individual co-inventions. We address this issue through the formation of network ties using a longitudinal analysis of French co-patenting data in the field of genomics between 1990 and 2006.

The number of individuals directly or indirectly linked, namely the network connectivity, is a pre-requisite for knowledge diffusion; it is the result of three non-mutually exclusive mechanisms. First, connectivity increases as single individuals attach themselves to already existing groups of individuals, that is, network components. Second, it increases as indirectly connected people form a direct link within an existing component. Third, connectivity increases as actors manage to connect existing but separate components, thus establishing a bridge between disconnected groups of individuals. Our study focuses on these two latter types of ties in order to highlight the strategies underling their formation.

In order to do so, we consider the determinants of network ties that occur within existing components (namely intra-component ties) and between two components (namely bridging ties). Bridging and intra-component ties have very different consequences on connectivity and component size. The former allows linking (at least indirectly) different groups of inventors and establishing channels that facilitate access to resources or other assets. The latter type allows establishing a direct link between actors already (indirectly) connected and increasing the cohesion of a group of individuals, favouring trust and enabling to share resources.

Our main argument is that the formation of bridging and intra-component ties results from these distinct strategies in accessing resources. We test our argument by contrasting the emerging literatures on network and proximity dynamics. The network perspective explains tie formation through preferential attachment and closure whereas the proximity perspective distinguishes various driving forces behind network formation (Boschma, 2005) based on similarity. People become connected because they share similar spaces, geographical, organizational, and cognitive or similar institutional incentives (being an academic or working for a private company).

Our main contribution is twofold. First, we show that the network formation is the result of specific and distinct strategies that can only be shown by explicating the types of ties. Second, we contrast network and proximity determinants of network formation. Doing so, we investigate if social connections allow economic actors to cross over geographical, technological and organizational boundaries.

The paper is organized as follows: section 2 presents the theoretical framework and stresses the element of novelty in our work relative to the existing literature. Section 3 provides a description of data and an explanation of how networks have been built up. Section 4 describes the estimation design and discusses the results of the econometric analysis. Section 5 concludes.

2. Theoretical framework and background literature

An increasing body of literature investigates innovation networks considering clusters of firms within regions and their impact on performance. Since networks are crucial for innovation, it seems important to consider the conditions under which these networks are formed and the relative importance of factors acting as major network drivers. However, the dynamics of network formation is still an under-investigated question. Network formation has only recently started to be empirically analysed and most existing studies run some form of *pairwise regression* (Bramoullé and Fortin, 2009), in which case the variable to be explained is represented by the links themselves.

2.1. The determinants of tie formation

Within existing studies, the formation of network ties are explained by different bodies of literature that offer two distinct perspectives: (a) the relational perspective assumes that trust, knowledge access and control of information are conferred through the actors' positions within networks; (b) the proximity perspective focuses on the relative position of economic actors in space, however defined.[1]

These two perspectives rely on different mechanisms, however they highly interact in shaping the evolution of observable social networks. The proximity determinants explain the contexts in which people meet and may become connected: for instance, two individuals are located in the same region. Once connected, they are part of a network that offers opportunities to form new ties and, doing so, to cross over organisational, institutional and geographical boundaries. However, since the different streams of literature rely on two distinct perspectives, they have been developed more or less independently.

[1] Sociologists identify a further perspective related to compatibilities and complementarities between actors' attributes (e.g. race) – the so-called *assortative* perspective (Rivera et al., 2010).

2.1.1. The relational perspective

The relational perspective focuses on direct and indirect connections among individuals; it is sometimes referred to as a 'within-the-network' approach, since the "focal predictor of network change is hypothesized to be the shape and structure of the network in a prior time period" (Rivera et al., 2010, p. 97).

Two main explanations have been identified: closure and preferential attachment. The former concerns the tendency of actors to form clusters; the latter deals with the actors' propensity to link to the most connected actors.

One of the characteristics that distinguish social from biological or technological networks is clustering (Newman and Park, 2003). Coleman (1988), and many others after him, have argued that being embedded in a very dense, interconnected, "cliquish" network brings benefits by enhancing the trust among individuals and thereby encouraging joint activities and the sharing of tacit and complex knowledge. Consequently, the effect of sharing a mutual acquaintance increases the likelihood of forming a dyad between unconnected actors: open triads tend to close over time. Moreover, clustering occurs even between individuals who are separated by more than a single intermediary, yet "[r]esearch has found that shorter network distances are correlated with increased tendencies toward connections" (Rivera et al., 2010, p. 101)

However, being embedded in very dense and strongly cohesive networks may also harm individuals in their search of new knowledge and their learning process. In fact, Burt (1992) argues that knowledge accessing is more efficient when individuals occupy structural holes that enable to link unconnected actors. Individuals positioned in structural holes are able to broker knowledge flows across unconnected groups (e.g. Gargiulo and Benassi, 2000). In sum, if clustering seems to be quite a general tendency, some strategic reasons may lead actors to avoid these configurations and, instead, seek out structural holes.

Skewed degree (i.e. number of links per node) distribution is one of the recurring features of most networks. The main explanation initially proposed by Barabási is the preferential attachment model (Barabási and Albert, 1999): the rate at which actors acquire new ties is a function of the number of ties they already have. This is explained by the fact that actors looking for new partners consider the other agents' number of existing ties as a proxy of, for instance, productivity.

However, in some cases establishing and maintaining a partnership could require a not negligible (opportunity) cost and this can limit the number of partners an actor can efficiently collaborate with. Thus the relationship between degree centrality and tie accumulation could be weaker in all those networks where some actor's constraints (e.g. time or resource) are important. Moreover, Newman and Park (2003) have noticed that social networks, differently from other types of network such as biological or technological ones, display a specific characteristic: a tendency for the most connected actors to connect among themselves. Popular actors tend to attach to popular actors; likewise, low degree actors do with their peers.

2.1.2. The proximity mechanisms

Geographical proximity is at the heart of the network formation issue and often appears as one of the main drivers since many ties take place between actors located within a short distance (Boschma and Frenken, 2009). However, it has also been pointed out that face-to-face and frequent contacts do not need permanent proximity, suggesting that agents do not need to be located in the same region. Thus geographical proximity is not a necessary condition for collaboration and networking (Boschma, 2005; Torre and Rallet, 2005). This issue is still on the agenda since many studies show high clustering of networks within regions and large metropolitan areas but no empirical studies have yet been carried out that would allow disentangling the respective geographical and density or network effects (Lobo and Strumsky, 2008).

Besides geography, the proximity literature highlights other forms of proximity such as cognitive, organisational and institutional proximity (Boschma, 2005). Cognitive proximity, for instance, means that economic actors share the same knowledge base or technology; as a result of cumulativeness, knowledge spillovers should be stronger when inventors and organizations share a strong technological proximity. Network tie formation may also result from a technological brokerage strategy which aim is to connect previously separated technological communities (Stuart and Podolny, 1999; Burt, 2004) thus leading to cross-disciplinary fertilization (Fleming et al., 2007). In this respect, two different stories may explain tie formation. In early stages, when basic science needs to be first investigated, inventors tend to become connected to partners working in the same technological field because they seek primarily specialization effects. This effect may even be stronger when actors share a geographical proximity (Jaffe, 1989; Aharonson et al., 2008; Boufaden and Plunket, 2008). In later stages, especially in exploitation phases, agents may rather seek complementary and distant knowledge bases.

Organisational proximity refers to the fact that two individuals share the same affiliation or industrial group at the time of tie formation or in prior periods. These ties are believed to be beneficial for innovation collaborations because they reduce uncertainty and opportunism. They are also more manageable when individuals share similar routines and processes, as well as they ease confidentiality requirements (Singh, 2005). However, the role of patent applicants is most often ignored although co-patenting may reveal some applicants' strategies in terms of knowledge production, diffusion and exploitation. Fleming et al. (2007), for example, highlight IBM Almaden Valley Labs' structural role in giant component formation as IBM highly invested in research and offered a doctoral program for Stanford University students thus favouring the connection between IBM and their doctoral students' future appointments. Similarly, in his study on patent citations, Singh (2005) shows that citations are three times larger when they happen within the same firm, whereas they are only 66% more likely when there is spatial propinquity, that is, when they emanate from the same region.

Finally, institutional proximity refers to the fact that collaboration is facilitated if it takes place under the same institutional setting, either between firms or between academics. In this case, individuals share similar incentives and coordination routines. Collaborations between academic and firms may encounter a number of difficulties due to conflicting interests, and difficulties in coordinating labour and accessing funds (Ponds et al. 2007). As explained by these authors, geographical proximity can compensate the difficulties induced by institutional and organizational distance.

2.2. Selective mechanisms at work

According to Gluckler (2007), the evolution of networks results from two separate sets of mechanisms. The first is cumulative and related to the historical process, where both initial conditions and the observed sequence of events matter. The second is selective and deals with the strategies that individual actors implement in order to gain benefit from being part of the network, given the constraints represented by the current structure of the network itself.

On the one hand, cumulative mechanisms occur through the formation of a cluster and more precisely of a clique, namely a group of actors where everyone is connected to everyone. The simple idea is that a group of individuals tends to cluster themselves in order to fulfil social obligations. An individual actor can meet some difficulties as she fails to collaborate with someone else who in turn collaborates with the majority of her collaborators. This is exactly one of the costs of redundancy mentioned by Burt, that is, the loss of autonomy.

On the other hand, Selective mechanisms can be interpreted more explicitly in terms of economic rationality: actors tend to allocate their resources efficiently, given some constraints. According to this logic, actors choose to form a link with actors endowed with specific assets (e.g. relevant knowledge) or to fill structural holes in order to profit from their strategic position (e.g. control the information flow).

What is the relationship between this distinction, *cumulative versus selective* mechanisms, and the distinction between *relational versus proximity* mechanisms illustrated in the previous section?

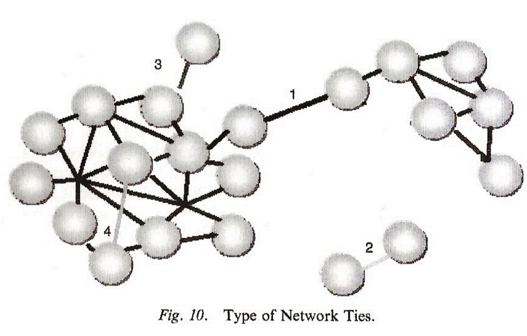
Concerning the relational mechanisms, while the tendency to cluster (i.e. closure) is considered to be a cumulative mechanism, it is more difficult to classify preferential attachment. As emphasised in the previous section, social networks seem to differ from other kind of networks in terms of degree correlation: individuals tend to attach themselves to other individuals with a similar number of partners (i.e. degree). According to Newman and Park (2003), this empirical regularity would be a consequence of a tendency to cluster *and* a tendency to belong to groups and communities of different size. Since individuals tend to establish links within their group, the consequence is that individuals end up with similar degree centrality. Following this reasoning, preferential attachment should be considered as a cumulative mechanism. However, a different argument can be put forward. Individuals could be interested in establishing links with individuals with a different degree for strategic reasons. A low degree individual may be interested to become attached to someone with a greater degree in order to gain indirect access to resource and to have greater visibility. A high degree individual could also be interested in establishing a link with someone with lower degree, because the latter is less constrained and could consecrate more time and make greater effort for an effective collaboration. Therefore, if individuals are similar (respectively dissimilar) in terms of degree, then preferential attachment should be viewed as a cumulative (respectively selective) mechanism (*Hypothesis 1)*.

Concerning proximity explanations, the general argument is that being close in some sense (e.g. geographical, technological, organizational and institutional) increases the likelihood of tie formation. On the one hand, being close facilitates exchange and communication with reduced cost. On the other hand, the benefit in terms of variety is limited, by definition, because of the similarity between involved individuals. Thus, establishing a link between actors that are close or *similar* corresponds to a cumulative rather than a selective mechanism. While, establishing a link between individuals that are different should underline a selective mechanism: it costs more but allows getting greater benefits (*Hypothesis 2)*.

2.3. Structural consequence of tie formation

Following Amburgey et al. (2008), it is possible to classify each new link according to the connectivity to the overall network. Taking two individual inventors as our focal point, they may become connected through four categories of links, 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.

Fig. 1 Type of network ties



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

The formation of each type of links has different implications for the overall network structure, as summarised in Table 1.

Tab. 1 Consequence of tie formation

Type of link	Size of the network	Number of components	Size of components
1. Bridging links	↔	↓	↑
2. New components link	↑	↑	↑
3. Pendant links	↑	↔	↑
4. Intra-components links	↔	↔	↔

Bridging and intra-component ties have very different consequences on network structure. The former allows linking (at least indirectly) different groups of inventors and establishing channels that facilitate the access to resources or other assets. The latter type allows establishing a direct link between actors already (indirectly) connected and increasing the cohesion of a group, favouring trust and enabling to share resources.

It is reasonable to assume that two different strategies underlie the formation of these two types of ties. A bridging tie is most likely explained by selective rather than cumulative mechanisms; the opposite may be true for intra-component ties.

Our empirical exercise will allow identifying the determinants of the independent formation of bridging and intra-component ties. Analysing the determinants of each type of tie and comparing the results should allow to disentangle the roles played by relational and proximity mechanisms in network tie formation and test our main hypotheses: similarity between individuals determines intra-component tie formation in case of cumulative mechanisms, on the contrary, diversity and variety determine bridging tie formation when selective mechanisms are at work.

3. Patent networks in genomics

3.1. Description of the data and network formation

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 French postal address and their co-inventors whatever their location within or outside France.

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) 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 2104 patents filed by 496 applicants and 4456 inventors, 7976 couple patent-inventor among which 6034 reporting a French postal address and 1942 with a foreign address.

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 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.

In order to build the network,[2] 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.

The aim of our paper is to understand how networks evolve over time through the formation of dyads between co-inventors. 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 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[3] and technological field (IPC codes). For this reason, this information cannot be used to highlight organizational or technological determinants.

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 periods.

3.2. Networks structural and dynamic properties

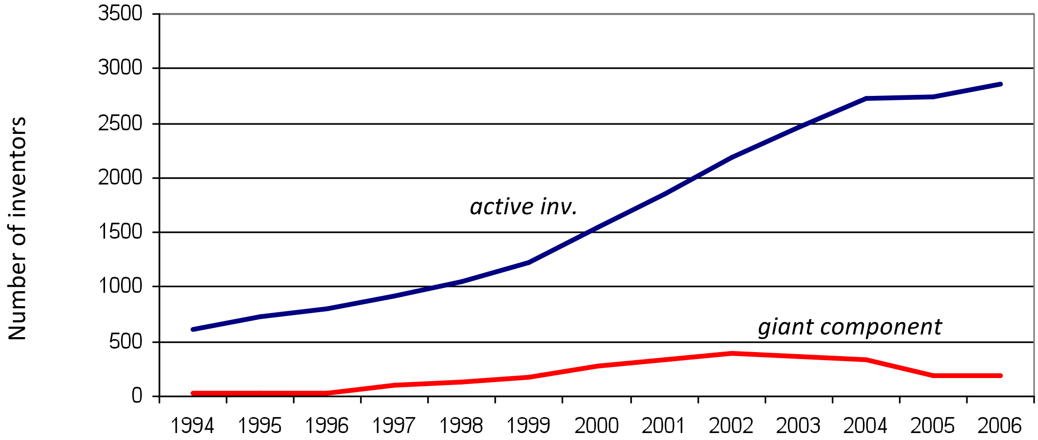
Figure 2 displays the number of active networks over time. At the beginning of 2000, the number of inventors clearly grows and then stabilizes around 2004.

---

[2] 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.

[3] 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.

Fig. 2 Size of the network (active inventors) and of the giant component



More striking is the time-varying pattern depicted by the giant component: first, it appears to be relatively small throughout the period compared to the size in similar studies (e.g. Fleming and Frenken, 2007). Second, it reaches its maximum in the year 2002, and starts decreasing before reaching a plateau.

While previous analyses rather focus on the giant component, our paper tracks the network dynamic by considering all sub-components (Baum et al., 2003; Lee Fleming et al. 2007; Fleming and Frenken, 2007). It is interesting to consider the formation of the giant component over time and understand why some network subparts become connected and grow over time whereas others do not. The formation of the largest component may be the result of two scenarios that are not necessarily mutually exclusive. In the first, the largest component may result from the connection of relatively large existing components that increase over time, have their own dynamics and finally become connected in a larger one. In the second scenario, the largest component may result from an incremental process wherein small components become connected, within a short time period, to a single relatively large component. In the former scenario, bridging ties would play a pivotal role for network connectivity, while it would not be the case in the latter.

Fig. 3 Evolution of the first four 1998 components

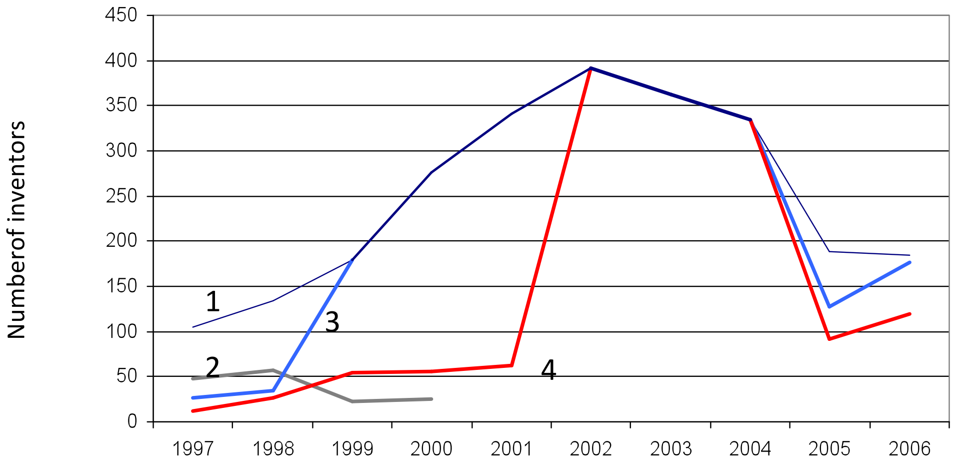


Figure 3 reports the evolution of the first four largest components in the 1998 network. The first component (137 inventors in 1998, around 13% of the active inventors) is mainly composed of inventors located in the Paris region, Ile-de-France (the same holds for the second and partially for the fourth component), while the bulk of the third component is located in the Rhône-Alpes region. The components also differ in terms of patent applicants. The first component includes some big pharmaceutical corporations (e.g. Aventis and Centillion) and some foreign universities; the second mainly includes public actors as CNRS, INSERM and some Parisian universities as well as biotechnological firms (e.g. Neurotech SA). Finally, the third component revolves around one main applicant, Bio Merieux, while the fourth one is mainly composed of inventors working for a spin-off of CSIRO, the Australian government research agency and for a French firm located in the central region of Auvergne. The most striking is that the 'public' component, i.e. the second one, breaks up during first years, while the other components converge to a giant component. Finally, in the most recent years, the size of the giant component drops down with its members splitting into three subgroups. In sum, looking at the giant component formation confirms the usefulness of analysing the specific role of bridging ties and their determinants.

Table 2 reports the number and share of new links relative to the period we intend to explain, i.e. 1995-2006.

Tab. 2 New link: type of networks ties

Links	Total number	%
1. Bridging links	244	1,88
2. New components link	8723	67,03
3. Pendant links	3853	29,61
4. Intra-components links	193	1,48
Total	13013	100

Most ties happen to involve new inventors either through the formation of new components or through pendant links. Indeed the most adopted strategy to enter into a network is forming a new component. The corollary is that one should already have patented (i.e. sent a signal), before attaching to some active inventors.

A fortiori, this implies a more central role for bridging ties. If the majority of inventors enter into a network establishing a new component, the overall network's connectivity depends mainly on actors able to link already existing components (i.e. bridging link case) rather than inventors able to attach themselves *directly* to already active inventors (i.e. pendant link case).

Moreover descriptive statistics (Table 3) suggest that intra-component ties are to a large proportion formed within the same applicant or with subsidiaries, whereas bridging ties are formed by different types of applicants, namely between academia and private companies.

Tab 3. Organizational and institutional relationships among types of ties

	Bridging ties		Intra-component ties	
	Total	%	Total	%
Between academia[1]	12	4,92	1	0,52
Between firms and academia[1]	112	45,90	23	11,92
Between firms[1]	43	17,62	21	10,88
Same applicant (academia or firms)	77	31,56	148	76,68
Total	244	100,00	193	100,00

[1] Different applicants

4. Estimation design and econometric results

In order to test our hypotheses, we use two different estimation procedures. We first implement a conditional logit model in order to estimate the likelihood of forming a network tie. Since we distinguish intra-component from bridging ties, we have to estimate a first set of models to estimate the respective impact of relational and proximity factors; we consider the differences between two types of ties as compared to the probability of not forming any tie, but it does not allow considering the likelihood of forming a bridging rather than an intra-component tie. For this reason, we estimate a second set of models based on a multinomial logit in order to predict the likelihood of forming a tie across separate components as compared to within one's component. The differences between the two types of ties may thus be considered regarding their specific network dynamic and configuration. This latter framework highlights the specificities of both configurations: intra-component ties are rather based on similarity and proximity whereas bridging ties enable to cross over technological, organizational and geographical boundaries.

4.1. The dependent variables and estimation procedures

4.1.1. The conditional logistic approach

For two inventors i and j, the probability of forming a tie p_{ij} follows a conditional logit distribution given by (Cameron and Trivedi, 2005):

$$p_{ij} = \frac{\Pr(y_{ij} = m \mid \mathbf{x})}{\Pr(y_{ij} = \text{No Tie} \mid \mathbf{x})} = \frac{\exp(\mathbf{x}'\beta)}{\sum_m \exp(\mathbf{x}'\beta)} \text{ with } m = N, B, I$$

N = Network tie, I = intra-component tie, B = bridging tie

$\mathbf{x}$ represents a vector of covariates whereas β is a vector of parameters to be estimated. If the tie is observed, the dependent variable takes the value of 1 and it is 0 otherwise. Three cases are considered whether we distinguish between the ties. So, the estimations consider subsequently the likelihood of forming a bridging tie (B) and an intra-component (I) versus no tie.

In order to estimate this model, we first compute all existing and potential ties between any two pairs of inventors. All the possible and realized dyads generate around four millions observations and the realized links represent only a marginal portion of all possible ties. Since this gap raises important difficulties of estimation, we adopt a case-control approach (Sorenson, Rivkin and Fleming, 2006). For any realized tie and their related co-inventors, we randomly select five 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. In sum, for each realized tie, we have ten controls. Each realized tie and its controls represent a group and the estimation is realized within this group; we use a cluster robust procedure to adjust standard errors for intra-group (matched case-control) correlation. The corollary is that variables characterized by constant within-group effects, such as year dummies, cannot be estimated. The same sample is then used for the multinomial estimations.

4.1.2. The multinomial logit approach

The multinomial logit model is equivalent to a series of pairwise logit regressions, except that the whole sample is used in order to reduce the potential biases that may arise from dropping part of the observations. In this framework, it is supposed that inventors do not simply choose between two outcomes (forming versus not forming a tie) but that they face three choices, forming a bridging tie (B), an intra-component tie (I) or not forming any tie (No tie). One of the outcomes J is chosen as the "base category" or comparison group. In our case, it is the intra-component tie formation. This means that we estimate the likelihood of forming a bridging tie (B) as compared to an intra-component (I) as well as the likelihood of not forming a tie as compared to an intra-component tie.

Let y_{ij} be the dependent variable with J nominal outcomes that are not ordered. $\Pr(y_{ij} = B \mid \mathbf{x})$ is the probability of observing outcome B given explanatory variables vector $\mathbf{x}$.

The probability may be written as follows (Cameron and Trivedi, 2005):

$$p_{ij} = \frac{\Pr(y_{ij} = J \mid \mathbf{x})}{\Pr(y_{ij} = \mathrm{I} \mid \mathbf{x})} = \frac{\exp(\mathbf{x}'\beta)}{\sum_m \exp(\mathbf{x}'\beta)} \text{ with } J = B, \text{No Tie}$$

4.2. The independent variables

Two sets of variables are considered, the relational and the proximity variables. *The relational perspective* is tested using *social proximity and degree centrality* measures in order to grasp the closure effects and the related accumulative advantages. The *social proximity* is computed as the inverse of the geodesic distance d_{ij} between two inventors i and j, that is, the shortest path connecting them in the network. This measure is only appropriate for intra-component ties since the geodesic distance between unconnected nodes is infinity, which is the case for all bridging ties by definition. *Social proximity* increases the likelihood of forming a tie since inventors may collaborate more easily with their partners' partners because "knowing" them facilitates trust and collaboration. Instead of considering social proximity per se, we prefer to limit proximity to geodesic distance equal to 2, by focusing on the *number of common partners*. This way, we account explicitly for the impact of triadic closure, that is, the fact that inventors form ties with their direct co-inventor's partners[4]. We expect an inverted U-curve relation between collaboration and the number of common partners. We expect the likelihood of forming a tie to increase with the number of common partner's until a certain level. After this level, we expect the probability to decrease because an inventor cannot manage an increasing number of collaborations.

In order to take into account the preferential attachment effect, we consider the *degree centrality measure*. Since we study the likelihood of two inventors to form 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' degree (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 ties, we expect a different sign. In particular, we expect the average measure to be positive and the difference to be negative for intra-component ties and *vice*

[4] We have tested our models using social distance *per se*, and we obtain qualitatively similar results as when we introduce "the number of common partners".

versa for bridging ties. When cumulative mechanism is at work, 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.

The proximity perspective is assessed through geographical, technological, organizational/institutional distance. In order to calculate the *geographical distance* in kilometres, we locate each inventor at the NUTS 3 level based on the 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 centroïd.[5] We calculate the distance in kilometres divided by 100. Geographical distance is supposed to have a negative impact on the likelihood of forming a tie since distance increases transaction costs.

Collaboration is easier among inventors that share similar technological interest and specialization. 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.

[5] 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 than consider the impact of organizational and institutional proximity. Organizational proximity occurs when two inventors file a patent for the same applicant. Institutional proximity characterizes two inventors working for similar types of organizations, academia and public research centres or private companies. We suppose that inventors are more likely to form ties within their organizational boundaries and with inventors belonging to similar organizational types. Since we interact these variables with geographical distance, and in order to ease interpretations, we choose to avoid interacting variables with different signs. For this reason, we consider organizational distance, namely *different applicants*, which takes the value 1 when two inventors file patents for different applicants, and institutional distance, namely *different organizational types*, which takes the value 1 when it is a link between academia and a private company.

We interact these variables with geographical distance in order to test how interacting institutional/organizational and spatial distance may impact the likelihood of forming a tie. Our hypothesis is that inventors will choose intra-component links when they need similar competences that may be found in a close neighbourhood. Instead, they will choose bridging links when they need distinct competences that may not be found in their own environment.

We introduce two types of control. We first control for the distinction between French located inventors and foreigners. 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 one of the French border countries, that is, Spain, Germany, Italy, Switzerland, and Belgium. We expect the impact to be positive.

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

4.3. Dyads formation: a conditional logit approach

Table 4 presents the results from a series of conditional logit models with cluster robust standard errors. Models 1-7 demonstrate the impact of organizational distance on the likelihood of forming intra-component and bridging ties. Models 3, 5 and 7 include interactions between organizational and geographical distance. Across models, variables and controls remain overall consistent in sign and magnitude suggesting that they are rather robust to the introduction of additional variables.

Models 1-3 consider the *relational perspective* through social proximity proxied by the number of common partners. Direct network links have a strong impact on forming intra-component ties and the effect seems quadratic as expected. Increasing the number of common partners enhances the opportunities to meet new inventors and so the probability to form ties. This supports the assumption of triadic closure and accumulative advantages provided by networks. Inventors tend to form ties with their partner's partners within given network components.

Since social proximity is infinite by definition in the case of separate components, and in order to enable comparison between intra-component and bridging ties, we test the impact of networks though the effect of degree centrality through the Models 4-7. The results for the absolute difference and average for the inventors' prior degrees show distinct patterns of dissimilarity between both types of ties.

As expected, the inventors' relative position within the network explains intra-component tie formation. The difference in degrees has a negative impact whereas average degree has a positive impact in this case. This confirms that the likelihood of forming such ties decreases when inventors are more dissimilar and it increases when they have high degrees namely when they are more visible and attractive within the network. Yet, these impacts are only slightly significant as opposed to the bridging ties for which the size is opposite but highly significant suggesting that bridging ties are driven by selective mechanisms and diversity. The corollary is found in the negative sign for the averages. The attractiveness is not a question of visibility for bridging ties; inventors are looking for other characteristics and resources.

The number of years since the first patent does not seem to play an important role in the formation of network ties. This impact is probably already grasped in the number of common partners and degree centrality, which are dependent on accumulative tendencies.

Regarding the *proximity mechanisms*, all the source of similarity increases the likelihood of forming network ties, as expected. Regarding geographical and technological distance, the results are similar among the ties suggesting that the likelihood of forming any type of tie is larger when co-inventors share similar technological fields and work in close geographical distance.

Yet, the impact is even stronger for the intra-component ties in which the impact is nearly twice as large suggesting that the higher the technological distance and the lower the likelihood of forming an intra-component tie. On the other hand, the geographical distance has a less negative impact for intra-component ties than for bridging ties. This could indicate that intra-component ties ease the collaborations that occur across larger distances but within organizational boundaries, since we control for different applicants (i.e. organizational distance).

Organizational proximity is strongly significant and positive; the likelihood of forming a tie increases when inventors patent for the same applicant, even in the case of bridging ties. This confirms the fact that inventors patent first of all with individuals that belong to their own organization (Singh, 2005).

In sum, collaborations mainly occur when inventors are located in close geographically distance to each other, they work in similar technological areas and presumably belong to the same organization. If this is not the case, the impact is negative, and it is even more negative when two inventors belong to separate organizations and are geographically distant. The interaction term *Geographical distance x different applicant* is strongly negative and significant in both models, whether they include social proximity or degree centrality. This suggests that combining geographical and organizational distance has a very significant and deleterious effect on intra-component ties. Unlike intra-component ties, the interaction term appears positive although not significant in the case of bridging ties.

Tab. 4 Conditional logit – Determinants of network ties with organizational distance							
	Model 1 Intra-comp	Model 2 Intra-comp	Model 3 Intra-comp	Model 4 Intra-comp	Model 5 Intra-comp	Model 6 Bridge	Model 7 Bridge
# of partners in common	2.596***	7.379***	7.164***				
	(0.374)	(1.595)	(1.504)				
# of partners in common - sq		-3.460**	-3.435***				
		(1.115)	(1.015)				
Degrees - Abs.diff.				-0.247+	-0.274+	0.292*	0.293*
				(0.144)	(0.141)	(0.130)	(0.130)
Degrees - Avrg				0.351	0.394	-0.529**	-0.527**
				(0.253)	(0.247)	(0.192)	(0.193)
Technological distance	-2.288*	-2.334*	-2.209*	-3.420***	-3.260***	-1.996***	-1.995***
	(1.021)	(1.114)	(1.114)	(0.958)	(0.950)	(0.574)	(0.576)
Geographical distance	-0.795***	-0.775***	0.072	-0.913***	-0.027	-1.243***	-1.337***
	(0.156)	(0.177)	(0.253)	(0.154)	(0.191)	(0.126)	(0.232)
Border	-1.482*	-0.751	-1.578*	-0.397	-1.295	0.806**	0.812**
	(0.639)	(0.658)	(0.717)	(0.724)	(0.788)	(0.252)	(0.249)
Different applicant	-1.955***	-1.907***	-1.042***	-2.657***	-1.603***	-1.593***	-1.657***
	(0.264)	(0.270)	(0.305)	(0.271)	(0.267)	(0.199)	(0.274)
Geographical distance x different applicant			-1.625***		-1.841***		0.110
			(0.363)		(0.306)		(0.258)
Experience - Abs.diff	-0.210	-0.027	0.180	0.190	0.283	0.038	0.040
	(0.349)	(0.360)	(0.367)	(0.286)	(0.292)	(0.186)	(0.188)
Experience - Avrg	-0.246	-0.345+	-0.437*	-0.383*	-0.420*	-0.190+	-0.187
	(0.204)	(0.190)	(0.194)	(0.170)	(0.179)	(0.114)	(0.114)
Observations	1604.000	1604.000	1604.000	1604.000	1604.000	2421.000	2421.000
Log Likelihood	-145.434	-136.956	-128.700	-190.237	-174.092	-399.677	-399.600
Pseudo R-Square	0.611	0.634	0.656	0.492	0.535	0.263	0.263

Cluster Robust standard errors in parentheses + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Tab. 5 Conditional logit – Determinants of network ties with institutional distance							
	Model 1 Intra-comp	Model 2 Intra-comp	Model 3 Intra-comp	Model 4 Intra-comp	Model 5 Intra-comp	Model 6 Bridge	Model 7 Bridge
# of partners in common	3.292***	8.290***	8.271***				
	(0.403)	(1.334)	(1.334)				
# of partners in common - sq		-3.634***	-3.638***				
		(0.875)	(0.871)				
Degrees - Abs.diff.				-0.090	-0.113	0.282*	0.270*
				(0.127)	(0.125)	(0.126)	(0.126)
Degrees - Avrg				0.015	0.055	-0.493**	-0.471**
				(0.211)	(0.208)	(0.180)	(0.180)
Technological distance	-2.928**	-2.924**	-2.988**	-4.723***	-4.872***	-2.348***	-2.348***
	(0.937)	(1.005)	(1.009)	(0.901)	(0.920)	(0.583)	(0.587)
Geographical distance	-1.112***	-1.098***	-0.991***	-1.497***	-1.393***	-1.418***	-1.613***
	(0.160)	(0.165)	(0.180)	(0.165)	(0.172)	(0.122)	(0.148)
Border	-1.085+	-0.445	-0.452	-0.433	-0.480	0.809***	0.779**
	(0.645)	(0.657)	(0.664)	(0.822)	(0.837)	(0.236)	(0.239)
Different organisational type	-0.917**	-0.923**	-0.484	-1.172***	-0.744*	-0.088	-0.418+
	(0.303)	(0.306)	(0.430)	(0.247)	(0.305)	(0.159)	(0.237)
Geographical distance x different org. type			-0.951*		-1.086*		0.436*
			(0.478)		(0.531)		(0.205)
Experience - Abs.diff	-0.049	0.103	0.130	0.317	0.294	-0.061	-0.055
	(0.291)	(0.304)	(0.310)	(0.251)	(0.249)	(0.173)	(0.175)
Experience - Avrg	-0.305+	-0.400*	-0.402*	-0.432**	-0.420**	-0.165	-0.168
	(0.171)	(0.162)	(0.162)	(0.146)	(0.144)	(0.110)	(0.110)
Observations	1604.000	1604.000	1604.000	1604.000	1604.000	2421.000	2421.000
Log Likelihood	-168.028	-157.308	-155.789	-255.392	-253.049	-429.721	-427.234
Pseudo R-Square	0.551	0.580	0.584	0.318	0.324	0.207	0.212

Cluster Robust standard errors in parentheses + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

	Network tie		Network tie		Network tie		Network tie		Network tie	
	Bridge	Notie	Bridge	Notie	Bridge	Notie	Bridge	Notie	Bridge	Notie
Degrees - Abs.diff.	0.336*	0.080	0.378*	0.132	0.383*	0.138	0.323*	0.068	0.317*	0.066
	(0.154)	(0.100)	(0.156)	(0.111)	(0.157)	(0.110)	(0.151)	(0.099)	(0.151)	(0.099)
Degrees - Avrg	-0.968***	-0.278	-0.957***	-0.264	-0.939***	-0.241	-0.937***	-0.247	-0.919***	-0.236
	(0.251)	(0.169)	(0.255)	(0.187)	(0.256)	(0.187)	(0.245)	(0.167)	(0.244)	(0.167)
Technological distance	2.565**	4.067***	1.600+	2.725***	1.549+	2.691***	2.554**	4.075***	2.557**	4.078***
	(0.892)	(0.743)	(0.908)	(0.765)	(0.888)	(0.736)	(0.897)	(0.750)	(0.894)	(0.748)
Geographical distance	0.196	1.400***	-0.210	0.828***	-0.507*	0.265+	0.218	1.422***	0.004	1.353***
	(0.174)	(0.146)	(0.187)	(0.156)	(0.224)	(0.159)	(0.171)	(0.142)	(0.190)	(0.149)
Border	1.659*	0.519	1.694*	0.606	1.716*	0.629	1.398+	0.286	1.298+	0.231
	(0.766)	(0.707)	(0.765)	(0.706)	(0.797)	(0.743)	(0.761)	(0.701)	(0.765)	(0.703)
Different applicant			1.761***	2.941***	1.126***	2.117***				
			(0.275)	(0.232)	(0.288)	(0.233)				
Geographical distance x different applicant					1.197***	1.513***				
					(0.337)	(0.275)				
Different organizational type							1.684***	1.497***	1.176***	1.214***
							(0.272)	(0.226)	(0.328)	(0.267)
Geographical distance x different org. type									1.197*	0.868+
									(0.537)	(0.506)
Experience - Abs.diff	-0.034	-0.088	-0.314	-0.473**	-0.315	-0.472*	-0.280	-0.301+	-0.270	-0.298+
	(0.241)	(0.165)	(0.240)	(0.183)	(0.244)	(0.184)	(0.241)	(0.166)	(0.240)	(0.166)
Experience - Avrg	0.107	0.327*	0.151	0.382**	0.170	0.406**	0.141	0.357**	0.136	0.352**
	(0.165)	(0.127)	(0.170)	(0.143)	(0.171)	(0.142)	(0.168)	(0.132)	(0.167)	(0.131)
Constant	1.029*	1.323***	0.898+	0.579	0.964+	0.762*	0.878+	1.201***	0.968*	1.211***
	(0.491)	(0.328)	(0.504)	(0.383)	(0.503)	(0.376)	(0.484)	(0.325)	(0.484)	(0.326)
Observations	4184.00		4184.00		4184.00		4184.00		4184.00	
Log Likelihood	-1364.81		-1233.52		-1218.59		-1336.15		-1332.73	
LR Chi Square	358.28		624.71		471.54		413.25		371.43	
Pseudo R-Square	0.14		0.22		0.23		0.15		0.16	

Tab. 6 Multinomial logit – Bridging and no tie versus intra-component ties

Robust standard errors are in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001 – Comparison group : intra-component ties

Table 5 presents similar results, but unlike Table 4, it focuses instead on institutional distance, and on its interaction with geographical distance. Belonging to different organizational types, academia versus private companies decreases the likelihood of forming a network tie. This is explained by the difficulties induced when two different organisations collaborate and especially when their incentives and organizational modes differ. As for organizational distance, the interaction with geographical distance even decreases the likelihood of forming an intra-component tie. The interaction term is positive in the case of bridging ties and significant. This means that geographically and institutionally distant ties occur between unrelated inventors and components. Since intra-component ties are rather geographically bounded and take place between individuals that are at least indirectly linked and especially with the partner's partners, they link similar inventors from an institutional and a technological perspective. For this reason, diversity may only be found by reaching separate components.

Finally, forming a tie with a foreigner located in countries bordering France has a negative impact on intra-component ties whereas it is positive on bridging ties.

4.3. Bridging versus intra-component ties: a multinomial logit approach

Until now we have considered the determinants of bridging and intra-component ties as compared to not forming any tie. We have seen from the previous regression tables that the behaviours are rather similar for what regards geographical, technological and organizational/institutional distance. The difference between intra-component and bridging ties appears when we combine geographical and organizational or institutional distances. Said differently, it appears that bridging ties occur when different competences and resources are needed and this is realized outside organizational and geographical boundaries.

In order to further investigate differences between both types of ties, we estimate a multinomial logit model in order to assess the probability of forming bridging rather than intra-component ties. These estimations are presented in Table 6. This estimation procedure provides direct evidence for the selective mechanisms involved in choosing bridging rather than intra-component ties. Whatever the proximity issues, when there is a distance in terms of geography, technology, organization or institutional type, the bridging tie will always be preferred. Bridging ties occur when inventors seek different competences, technologies, and this requires going beyond once network component and geographical area.

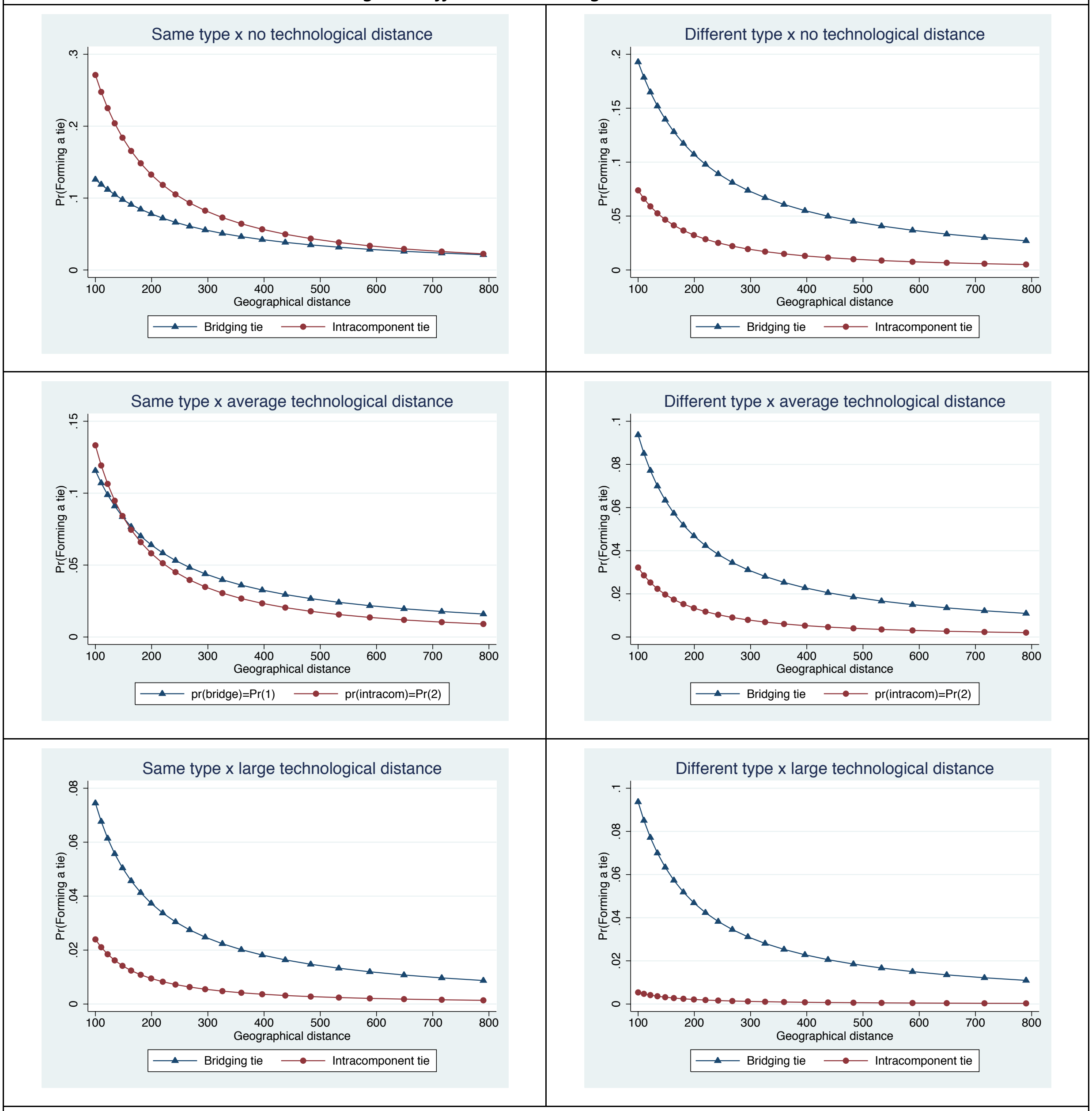
Finally, it is worth considering more precisely the interaction between three forms of proximity in order to fully understand how bridging ties allow individuals and firms to cross over different types of boundaries. Figure 4 displays the probabilities of forming bridging and intra-component ties for three levels of technological distance, that is, none, average and large technological distances given the co-inventors geographical and organisational or institutional distances.

It appears that intra-component ties are preferred when inventors belong to the same organization and share exactly the same research area. Within organizational boundaries and with no technological distance, geographical distance can be overcome (figure 4, upper left). When technological distance reaches an average level, intra-component ties will be preferred for short distances (under 150 km). For higher geographical distances, even within organizational boundaries, inventors will prefer bridging ties, but the differences in probability are marginal. The picture becomes really sharper when technological distance becomes larger as well.

Bridging ties appear to be preferred when there is organizational and institutional distance, namely for academia-firm linkages, whatever the level of technological distance. The probability of forming intra-component ties is, in this case, decreasing as technological distance increases, and it nearly becomes null when there is no technological overlap between inventors. These results are somewhat counterintuitive because we would expect that social proximity would ease to cross over geographical boundaries, but this does not seem to be the case. On the contrary, social proximity seems very much correlated to geographical but also technological and organizational boundaries. The likelihood of inter-regional bridging ties increases with technological distance and different applicants. These ties are formed outside one's component and in other regions in order to find different technological competences that are not easy to find in the close technological, geographical and organizational neighbourhoods.

If the likelihood of forming a tie is increased within one's organization for bridging as well as intra-component ties, interregional collaboration offer the opportunity to find new partners outside the organizational boundaries.

Fig. 4 Relative probabilities of forming Bridging versus intra-component ties given different technological distances



These probabilities correspond to the second and fourth multinomial logit estimation of table 6 with all the variables set at their mean except for geographical distance which ranges from 0 to 800 km and the technological distance which is set to zero, its average and its extreme value depending on whether we consider no, average or large technological distances. Please note, *type* means *organizational type*.

5. Conclusion

This paper investigates the dynamics of network formation in genomics patenting in France over the last two decades and it tests the respective role of cumulative mechanisms and actor strategies in determining the evolution of inventor networks.

In performing this study, we have a double motivation. First, we seek to consider in a single analysis the relational and proximity perspectives that are usually treated separately while they rely on highly overlapping mechanisms. Following Gluckler's (2007) conceptual framework of cumulative versus selective mechanisms, we contribute to explain how both perspectives lead to the formation and the evolution of networks. Second, instead of analysing the determinants of network ties in general, we distinguish among ties according their relative impact on the overall network connectivity. Doing so, it is possible to analyse the determinants of different strategies underlying tie formation.

Our results confirm the distinct characteristics of bridging ties (Powell et al. 2005, Baum et al., 2007). While intra-component ties rely on accumulative mechanisms, partly redundant and mainly based on triadic closure, bridging ties result from selective mechanisms that enable the access to different resources and competences. Our main contribution relies on the combination of various forms of proximity while controlling for preferential attachment and closure.

Our study contributes to specify the conditions under which inventors choose their collaborations given the various levels of proximity. Our results confirm Ponds et al.'s (2007) findings that research collaboration involving different kinds of organisations are more geographically localised than collaboration between similar organisations. But this result only holds for intra-component ties. Ponds et al. (2007) argue that geographical proximity is a way of overcoming the institutional differences between organisations. Our study goes one step further and shows that geographical proximity is highly complementary to social proximity in determining collaboration even when partners belong to different (types of) organisations. Trust and reputation seem therefore to play a prominent role. However this happens only when technological distance is rather reduced: the advantages of closure disappear as technological distance increases. When different competences are needed, bridging ties are formed and this is realized outside organizational and geographical boundaries. In other words, as soon as we combine different forms of distances, collaborations take place outside the actors' close local network and they involve different groups and communities of inventors. These collaborations are preferred even if they are likely to be more risky and uncertain.

Finally, our study enables to advance some explanations regarding industrial clustering and specialization effects. It appears that local clustering is mainly based on intra-component, that is, closure ties that facilitates collaborations between academic and non-academic organisations within similar technological fields, thus contributing to increase local specialisation effects. While the cluster increases over time, different technological resources are needed, and these are brought to the network through bridging ties, that enable to bring together communities that are geographically and technologically separate.

References

Aharonson, B. S., Baum, J. A. and Plunket, A. (2008), Inventive and uninventive clusters: The case of Canadian biotechnology, *Research Policy* **37**(6-7), 1108 - 1131.

Amburgey, T., Al-Lahan, A., Tzabbar, D. and Aharonson, B. (2008), The structural evolution of multiplex organizational networks: research and commerce in biotechnology, *Advances in Strategic Management,* **25**, 171-209.

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

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

Baum, J. A. C., Shipilov, A. V. and Rowley, T. J. (2003), Where do small worlds come from?, *Ind Corp Change* **12**(4), 697-725.

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

Boschma, R. and Frenken, K. (2009), The spatial evolution of innovation networks. A proximity perspective, *in* Ron A. Boschma & Ron Martin, (ed.), Handbook of Evolutionary Economic Geography, Edward Elgar

Boufaden, N. and Plunket, A. (2008), Proximity and innovation: Do biotechnology firms located in the Paris Region benefit from localized technological externalities?, *Les Annales d'Economie et Statistique* **87/88**, 198-220.

Bramoullé, Y. and Fortin, B. (2009), The Econometrics of Social Networks, CIRPEE.

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.

Breschi, S. and Lissoni, F. (2005), Cross-firm and Social Networks: Localized Knowledge Spillovers Revisited, *Les Annales d'Economie et de Statistique*, **79/80**, 189--209.

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

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

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. and Frenken, K. (2007), The Evolution Of Inventor Networks In The Silicon Valley And Boston Regions, *Advances in Complex Systems (ACS)* **10**(01), 53-71.

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.

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

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

Jaffe, A. B., Trajtenberg, M. and Henderson, R. (1993), Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, *The Quarterly Journal of Economics* **108**(3), 577-598.

Knoben, J. (2009), Localized inter-organizational linkages, agglomeration effects, and the innovative performance of firms, *The Annals of Regional Science* **43**(3), 757-779.

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.

Lobo, J. and Strumsky, D. (2008), Metropolitan patenting, inventor agglomeration and social networks: A tale of two effects, *Journal of Urban Economics* **63**(3), 871-884.

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

Newman, M.E. J. and Park, J. (2003), Why social networks are different from other types of networks, *Phys. Rev. E* **68**(3).

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

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

Rivera, M. T., Soderstrom, S. B. and Uzzi, B. (2010), Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms, *Annual Review of Sociology* **36**(1), 91-115.

Singh, J. (2007), External Collaboration, Social Networks and Knowledge Creation: Evidence from scientific publications, DRUID Summer Conference 2007.

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 flow, *Research Policy* **35**(7), 994-1017.

Stuart, T. and Podolny, J. (1999), Positional Causes and Correlates of Strategic Alliances in the Semiconductor Industry, *Research in the Sociology of Organizations*, 161-182.

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

Watts D.J. & S.H. Strogatz (1998), Collective dynamics of 'small-world' networks, *Nature*, **393**, 440-442.

Appendix :

A1. Variables: definitions

Variables	**Definitions**
Dependant variables	
Intra-component tie	Takes value 1 if two inventors already in the network form an intra-component tie
Bridging tie	Takes value 1 if two inventors already in the network form a bridging tie
Network variables	
Number of common partners	The number of partners for which the geodesic distance is equal to 2 in the prior period (in logs)
Number of common partners sq	The quadratic term of the number of common partners (in logs)
Absolute difference in degree	Absolute value of the differences between the co-inventors' respective degree centrality
Average degree	The average value of the co-inventors' respective degree centrality
Proximity variables	
Geographical distance	The distance in km / 100 between NUTS3 regions prior to attachment (in logs)
Technological distance	The complement of Jaffe's index using IPC codes for each co-inventor's patents prior to attachment (in logs)
Different applicant	Takes value 1 if co-inventors patent for different applicants prior to attachment, 0 otherwise. It is a proxy for organisational distance
Different organisational type	Takes value 1 if co-inventors patent for different organizational types (firm, university or individual) prior to attachment, 0 otherwise. It is a proxy for institutional distance.
Geographical distance x different applicant	The interaction term between geographical distance and different applicant
Geographical distance x different org. type	The interaction term between geographical distance and different organizational type
Other Controls	
Absolute difference in experience	Absolute value of the differences between each co-inventors' number of years since first patent
Average experience	The average value
Border	Takes value 1 if one of the co-inventors belong to a border country to France, 0 otherwise

A.2. Variables: descriptive statistics

Variable	**Observations**	**Mean**	**Std. Dev.**	**Min**	**Max**
1. Number of common partners	4184	0.039	0.192	0	1.792
2. Number of common partners sq	4184	0.05	0.272	0	3.258
3. Abs. difference in degrees	4184	1.659	0.885	0	4.025
4. Average degrees	4184	1.978	0.538	0.693	3.676
5. Technological distance	4184	0.228	0.147	0	0.693
6. Geographical distance	4184	1.261	0.779	0	2.854
7. Different applicant	4184	0.886	0.318	0	1
8. Geographical distance x Different applicant	4184	1.193	0.815	0	2.854
9. Different organizational type	4184	0.394	0.489	0	1
10. Geographical distance x Different organizational type	4184	0.49	0.774	0	2.854
11. Experience - Avrg	4184	1.624	0.471	0.693	2.773
12 Experience - Abs.diff	4184	1.291	0.733	0	2.773

Correlation table

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	1.000											
2	0.976*	1.000										
3	0.029	0.031*	1.000									
4	0.101*	0.103*	0.691*	1.000								
5	-0.148*	-0.140*	-0.081*	-0.218*	1.000							
6	-0.212*	-0.190*	-0.029	-0.061*	0.162*	1.000						
7	-0.386*	-0.362*	-0.028	-0.045*	0.142*	0.307*	1.000					
8	-0.273*	-0.251*	-0.041*	-0.071*	0.171*	0.919*	0.524*	1.000				
9	-0.115*	-0.106*	0.001	-0.032*	0.017	-0.017	0.288*	0.050*	1.000			
10	-0.122*	-0.111*	-0.012	-0.056*	0.077*	0.367*	0.226*	0.403*	0.785*	1.000		
11	-0.004	0.002	0.207*	0.265*	-0.141*	-0.069*	0.031*	-0.061*	0.121*	0.058*	1.000	
12	-0.021	-0.018	0.115*	0.117*	-0.034*	-0.028	0.025	-0.027	0.065*	0.045*	0.531*	1.000