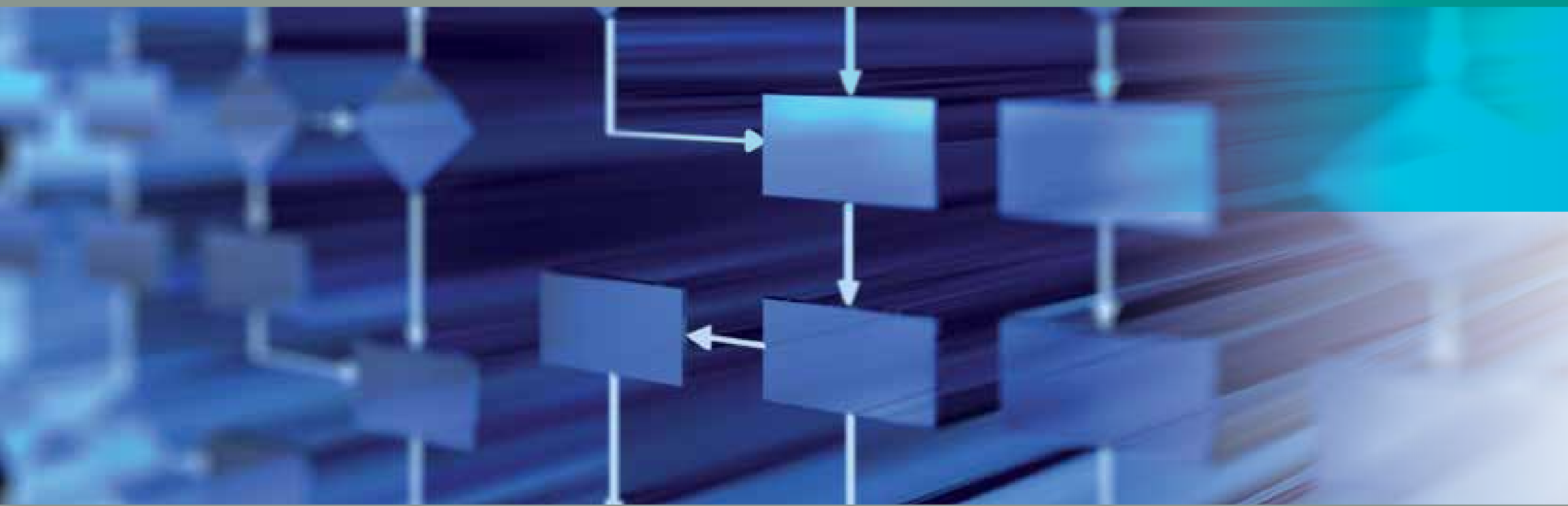

WIPO Economics & Statistics Series

April

2013

Working Paper No. 7

How does geographical mobility of inventors influence network formation?

Ernest Miguelez

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

How does geographical mobility of inventors influence network formation?

Ernest Miguelez[1]

Abstract

The goal of this paper is to assess the influence of spatial mobility of knowledge workers on the formation of ties of scientific and industrial collaboration across European regions. Co-location has been traditionally invoked to ease formal collaboration between individuals and firms, since tie formation costs increase with physical distance between partners. In some instances, highly-skilled actors might become mobile and bridge regional networks across separate locations. This paper estimates a fixed effects logit model to ascertain precisely whether there exists a 'previous co-location premium' in the formation of networks across European regions.

Key words: inventors' mobility, technological collaborations, co-location, European regions, panel data

JEL: C8, J61, O31, O33, R0

Disclaimer

The views expressed in this article are those of the author and do not necessarily reflect those of the World Intellectual Property Organization or its Member States.

[1] World Intellectual Property Organization, Geneva, Switzerland; e-mail: ernest.miguelez@wipo.int; Tel.: (+41) 223388344. Fax: (+41) 223388350

Contact information: Economics and Statistics Division, World Intellectual Property Organization, 34 Chemin des Colombettes, P.O. Box 18, CH-1211 Geneva 20, Switzerland, & AQR-IREA, University of Barcelona

Acknowledgements

Part of this work was carried out while I was visiting the Rotman School of Management, at the University of Toronto (Toronto, Canada). The use of the School's facilities is gratefully acknowledged. I would also like to share my appreciation of helpful comments received from Ajay Agrawal, Christian Catalini, Francesco Lissoni, Rosina Moreno, Julio Raffo, and participants to the Geography of Innovation EuroLio Seminar (Saint Etienne, France, 26th-28th January 2012), the XV Applied Economics Meeting (La Coruña, Spain, 6th-7th June 2012) and the ESD-WIPO Special Summer Seminar Series (Geneva, Switzerland, 3rd September 2012). This paper was awarded the ALdE's Young Researchers Prize 2012 in the Applied Economics Meeting and the award "Càtedra Universitat-Empresa per al Foment de la Innovació Emrpesarial", from University Rovira i Virgili. I also acknowledge financial support from the Ministerio de Ciencia e Innovación, ECO2008-05314 and ECO2011-30260-C03-03, and from the Ministerio de Educación, AP2007-00792. However, any mistake or omission remains mine.

1. Introduction

It has become commonplace in the literature that innovation and technological advances fuel the pace of the economic development of countries (Aghion and Howitt, 1998; Grossman and Helpman, 1991, 1994; Jones, 1995). In parallel, networks are increasingly important for innovation, mainly due to the growing complexity of current knowledge production processes. Cross-pollination of ideas, barters of tacit knowledge or the division of labour, have been regarded to be the underlying forces heading to network formation (Katz and Martin, 1997). Yet, what drives the selection of one particular partner among all possible partners?

In the present paper, I focus on one particular issue that, surprisingly, has been largely under-investigated to date, that is, the role of mobile employees on the formation of linkages across the space. To this end, I construct and estimate a Knowledge Linkage Production Function (KLPF). Its underlying logic states that the likelihood to observe a tie between innovators located in different European regions can be explained by the individuals' characteristics, as well as by whether or not they were co-located in the past. Thus, I conjecture that the benefits of co-location in building up formal interactions and networks do survive after the individuals' separation and are conducive to tie formation across the space, above and beyond other individuals' features. Thus, trust, mutual understanding, and hence information diffusion, are more likely to exist between separated actors if they shared a common spatial context in the past.

Indeed, the setting up of research collaboration ties is costly. There are many potential partners to choose, but their ability and their complementarity with one's knowledge skills are unknown. The costs of searching potential partners are likely to be high. Other costs such as those derived from negotiation between the partners, formation of contracts, agreement on the amount of knowledge and information that have to be exchanged, managing and administration of the common project, as well as monitoring of partners' fulfilments, are also likely to be important and condition actors on whether or not to collaborate and, above all, with who they want to collaborate. In such a setting, spatial co-location may smooth these frictions and therefore formal networks are more likely to arise between individuals located in close physical proximity.

If co-located agents are more willing to build up ties, what happens when they move? Recent research stresses the importance of mobile inventors in setting up relations with their former colleagues and flowing knowledge back to their prior location (Agrawal et al., 2006; Oettl and Agrawal, 2008). Indeed, the benefits of co-location have been shown to manifest amongst people who then move away but continue in contact (Storper and Venables, 2004). My main hypothesis in this paper states that knowledge workers invest in developing social capital in the spatial context in where they reside, made up of trust and mutual understanding as well as a dense network of friends and acquaintances, and at least partially, these features endure after the innovator has left this specific context. If these informal relationships are maintained after separation, they are likely to be conducive to network formation between the individuals involved, even if they do not share geographic, social, cognitive, institutional, or organizational similarities.

The study of this phenomenon is important from a policy perspective and motivates its analysis. Broadly speaking, there exists a reasonable agreement on the fact that knowledge flows tend to be local (Audretsch and Feldman, 1996; Jaffe et al., 1993). This is because knowledge is better transmitted through frequent interactions and face-to-face meetings, rather than through long-distance communication technologies.

Among other reasons, co-location enables the formation of local formal networks, which are main conduits of knowledge barters and ideas diffusion. Recently, however, scholars have started to claim that excessively close actors may have little to exchange after a certain number of interactions (Boschma and Frenken, 2010). Indeed, the production of ideas requires the combination of different – though related, complementary pieces of knowledge to be most effective. However, at some point, co-located agents may start to combine and recombine local knowledge that eventually becomes redundant and less valuable. As a result, lock-in (Arthur, 1989; David, 1985) and subsequent economic stagnation may occur. In contrast, truly dynamic regions in the era of the knowledge economy will be those whose firms are able to identify and establish interregional and international connections to outside sources of ideas (Gertler and Levitte, 2005; Maskell et al., 2006). I speculate that one main mechanism to identify and access distant pools of knowledge is through mobile high-human-capital employees who left the region but did not break their ties with their former social contexts. By means of such a mechanism, mobility introduces variation into the local economy, which can prevent the region from entering non-dynamic development paths.

To better comprehend the determinants of cross-regional knowledge linkages between European regions, as well as the influence exerted by the mobility of labour, I make use of micro-data on European inventors who have applied for EPO[2] patents in the biotechnology industry[3], over the period 1978-2005. A fixed-effect logit model will be estimated to ascertain whether there exists a 'previous co-location premium' on the likelihood to build up formal ties across regions. As in Fafchamps et al. (2010), I deal with the endogenous nature of my focal variable by exploiting the fact that when two inventors have already co-authored together, they have enough information about each other and about the match quality. Hence, features such as informal relationships, trust, mutual understanding, and so on, inherent to the spatial context in which they were co-located, are unlikely to affect tie formation aside from through their prior co-authorship – as it will be explained in detail afterwards.

In brief, the contributions of the present paper are manifold. First, in broad terms, it provides additional and consistent evidence on the determinants of knowledge linkages formation between physically separated actors, putting a special emphasis on the role played by different types of similarities between the pair. Second, it provides the first empirical test on the role of individuals' geographical mobility on the formation of networks throughout the space, which, in turn, are conducive to spread knowledge. To the best of my knowledge, any study has empirically tested its role as a mean of knowledge ties formation.

---

[2] EPO stands for European Patent Office.

[3] According to the OECD, biotechnology refers to the "application of science and technology to living organisms, as well as parts, products and models thereof, to alter living or non-living materials for the production of knowledge, goods and services" (van Beuzekom and Arundel, 2009, pp. 9). Biotech is indeed an ideal candidate for studying the interplay between innovation, networks and geography. Over time, the share of multi-inventor biotech EPO patents in Europe has sharply increased, which is also reflected by team size: the average number of inventors per patent goes from 2.64 during the first part of the eighties until 3.55 around 2005. On top of this, biotech patents are growingly being co-authored with outside-to-the-region peers. This suggests that in spite of the anchored spatial clustering of the biotech industry, individuals, firms and institutions increasingly rely on external-to-the-region partners with whom jointly patent – figures can be requested upon request from the author.

In addition, it also provides indirect evidence on the role of spatial proximity and co-location, by estimating the 'previous co-location premium', whilst controlling for a number of time-variant features as well as time-invariant pair-wise fixed effects.

The remaining of the paper is organized as follows: Section 2 reviews previous studies, bringing together dispersed, but related, literature, and outlines the conceptual framework. Section 3 describes the empirical approach taken here and the data sources. Section 4 summarizes some remarkable findings and Section 5 presents conclusions and policy implications.

2. Background Framework and Contributions of the Present Analysis

The study of social networks formation has long attracted a great deal of interest from various research streams, spanning the limits across disciplines and sub-disciplines. In part, this is due to the pervasiveness of organizations' and individuals' cooperative practices in knowledge creation, which has become a salient feature of innovation management and is regarded to be a source of outstanding economic performance of firms.

This has given rise to a flourishing number of scholarly research topics, such as the study of cooperation determinants (Cassiman and Veugelers, 2002) or the study of partnering choices. Among the later, two strands of literature stand out: the network structural effects perspective and the proximity perspective (Cassi and Plunket, 2010). The former emphasises the importance of the amount of knowledge that each partner can access from the others in the network – their network position (Autant-Bernard et al., 2007). The second strand of literature argues that partnering decisions are often based on the logic of 'homophily' (McPherson et al., 2001; Ter Waal and Boschma, 2009). 'Homophily' refers to the homogeneity of individuals' personal relations in a range of socio-demographic and personal characteristics. Tie formation between peers is crucially determined by this similarity. Among others, 'homophily' may refer to physical proximity between partners. Indeed, geographic propinquity creates contexts in which homophile relations form and knowledge linkages may arise. Trust, mutual understanding, informal relations or serendipitous encounters, group identification, socialization, and, in general, social capital formation, which are enhanced in close geographical proximity, has been pinpointed to be main facilitators to surmount the barriers to start collaborating. As a result, knowledge interactions are more likely to occur between individuals who are closely located.

Yet, geographic proximity is one of many forms of 'homophily' that may boost knowledge interactions and network formation. Other non-geographical similarities have been highlighted as producing the same type of outcomes: that is, social proximity, cognitive proximity, institutional proximity, or organizational proximity (Boschma, 2005), producing a lively debate on the topic. All these proximities have some fundamental things in common: they reduce uncertainty, help solving coordination problems and, on top of this, lower the cost of identifying partners. Accordingly, all of them are likely to influence network formation across regions.

In this framework, several empirical exercises have attempted to identify the determinants of linkages formation in scientists' co-authorships, firms embarking in R&D alliances, or inventors' co-patents. Fafchamps et al. (2010) estimate network effects in co-authorship formation among economists over a twenty-year period.

Their findings consistently show that collaborations between pairs of economists emerge faster if they are closer to each other in the network of co-authors. Time-variant characteristics such as the individuals' productivity or their propensity to collaborate, as well as the cognitive proximity between the pair, are equally found to influence team formation.

In parallel, network effects *vis-à-vis* geographic proximity and other meaningful similarities is the leitmotif of a growing number of studies, such as Mariani (2004), Ter Wal (2011), Cassi and Plunket (2010), for the case of European inventors of the chemical industry, biotech inventors in Germany, and genomics inventors in France, respectively, and Autant-Bernard et al. (2007) and Paier and Scherngell (2011), for the case of European firms' R&D collaborations as captured by joint participation within the European Framework Programmes. Their findings can be summarized as follows: social, organizational, institutional, and cognitive proximities between agents are found to influence network formation. Notwithstanding, no empirical analysis has succeeded in explaining the role of geographical distance away. Furthermore, network effects matter more in the later stages of an industry life cycle, when the industry moves to an exploitation stage (Ter Wal, 2011). At the early stages of the industry, geographic proximity between actors is, however, more conducive to tie formation. It is also found that when firms lack the competences and size to manage themselves within global R&D networks, geography becomes crucial to induce people collaborate (Mariani, 2004). In addition, geography plays a critical role when collaboration involves very different organizations (like industry-university interactions). Geography is also found to be highly complementary, rather than substitute, with social proximity as conduit to form social ties.

The present inquiry largely builds upon these later contributions, and estimates a knowledge linkage production function to disentangle the different effect of social, cognitive, organizational, and institutional proximities, on the probability to observe a tie. Different from these studies, however, I enlarge the empirical analysis to the whole Europe, on the one hand, and I will control for pair-wise unobserved time-invariant heterogeneity, on the other hand.

A main tenet of the present paper is that geographical proximity remains essential for knowledge interaction and hence network formation, as sustained by most of the studies sketched above. Bradner and Mark (2002) undertook an interesting experiment on collaboration patterns. They invited a number of people to choose a collaboration partner through computer-mediated mechanisms. The subjects were only told about the city in which potential partners were located. Intriguingly, the authors found that individuals had a striking tendency to start collaborating with those they believed were located in the same or nearby cities, rather than those located in cities far apart. Their results, they argue, can be explained by *social impact* and *social identity* effects. Latané (1981) claims that the time spent interacting, paying attention, recollecting, and attempting to persuade others depend on physical proximity and co-location. These variables constitute the *social impact* of a given agent over the others, and are strongly conducive to network formation. Similarly, Tajfel (1978) argues that *social identity* effects lead people to view their cohorts in a more positive light than the others simply because of their own desire to be viewed as superior to outsiders. People living and working close by are more likely to belong to the same cohort than those individuals living far apart (op. cit.). In a similar vein, as Storper and Venables (2004) posit, screening of potential partners is pivotal to enhance network formation processes.

However, much of what is valuable from potential partners is tacit, and therefore can only be communicated as a highly contextual metaphor. A good knowledge of potential partners is therefore required, which can only be achieved through socialization. Socialization refers to the mean by which individuals signal the others that they belong to the same social group. Socialization, they argue, is mostly achieved through frequent face-to-face interactions enhanced by shared spatial contexts (op. cit.).

However, the benefits of physical proximity for the formation of linkages between inventors established through long periods of co-location are durable and manifest among people after they become separated in the space. That is to say, the effect of mutual understanding between members of a co-located community may well survive the end of their co-localisation, and therefore communication and the formation of networks across the space may overcome long distances. In this respect, an increasing number of scholars have recently unearthed the role of mobile skilled workers that, by not breaking their ties with their former colleagues, favour the diffusion of knowledge and ideas across firms, regions and even countries. Kaiser et al. (2011) identify positive effects on firm's innovation of enterprises losing an employee hired by a competing firm, for the case of Denmark. Similarly, Corredoira and Rosenkopf (2010) show disproportionately larger number of citations from the sending to the receiving firm after an employee has left the former for the later, for the case of US innovators. The 'outbound mobility' effect is even stronger when mobility occurs between geographically dispersed firms, since co-located organizations usually exploit other cross-firms interactions channels (op. cit.). According to their views, the leaving employee probably stays in contact with their former colleagues, constituting in this way a source of knowledge diffusion from the hiring to the sending firm. This same issue was also devised in a study by Agrawal et al. (2006). Exploring inventors' mobility across different MSAs, the authors find that knowledge flows are around 50% more likely to go to the innovator's prior location than if he had never lived there. Thus, social ties created during inventors' co-localisation, which facilitate knowledge diffusion, persist even after the inventors' separation and are conducive to knowledge flows. Oettl and Agrawal's (2008) study builds upon the same idea. The authors estimate a fixed-effects negative binomial model to analyse backward knowledge flows between countries from the leaving innovator to their former co-located colleagues. Indeed, mobile knowledge workers provide access to distant knowledge pools that neither the receiving firm and country nor the source firm and country might otherwise enjoy.

These and related studies (see also Agrawal et al., 2008; Agrawal et al., 2011; Kerr, 2008) rest on the logic of the 'enduring social capital hypothesis' (Agrawal et al., 2006). That is, informal ties between individuals, shared trust and mutual understanding, built after years of co-location and shared spatial context, may well survive the spatial separation of the individuals and be a source of knowledge diffusion, as these studies have consistently shown. My tenet is that the enduring social capital between previously co-located peers is also conducive to knowledge linkages formation across different locations, which in turn is a way to access distant pools of knowledge and ideas' diffusion across the space.

In sum, as it will be discussed subsequently in detail, the present analysis tries to find evidence on the role of previous co-location on the formation of knowledge linkages across the space. To the best of my knowledge, few papers have dealt with this issue, despite the importance of research collaborations and skilled labour mobility from the academic and policy perspectives. Only recently, Jöns (2009) provides case study evidence on the role of foreign academic visiting to Germany during the second half of the XXth century as a source of subsequent academic mobility and collaborations that significantly contributed to the country's reintegration into the international scientific sphere.

3. Research design

Estimation Framework

This section describes the way in which I chose to assess the influence of the focal variable – the 'previous co-location', in the likelihood to build cross-regional knowledge ties. As explained before, a fixed-effects conditional logit model is estimated, which enables controlling for important time-invariant confounders that might have biased previous econometric analyses.

Recall from previous sections that my general framework is the study of individuals' linkages formation between separate European regions – basically NUTS3, though robustness analysis includes NUTS2 estimations. Hence, amongst all the potential partners to be chosen from other regions, I am particularly interested to know what drives the selection of one particular collaborator rather than the other, conditional upon not residing in the same region and not having co-patented before.

For each pair of inventors, a link is formed if and only if the associated payoffs are expected to be positive, $\pi_t^{ij} > 0$. This in turn depends upon i's and j's observable time-variant and non-observable time-invariant characteristics, X_t and γ^{ij} respectively, as well as a well-behaved error term, ε:

$$\Pr_t^{ij} = \Pr(\pi_t^{ij} > 0) = \beta_n \cdot X_t + \gamma^{ij} + \varepsilon_t^{ij}, \tag{1}$$

where n stands for the number of regressors included in the model. The i's and j's observable features refer to i's individual characteristics, j's individual characteristics, as well as a set of proximities between the two – social, institutional, cognitive, and organizational. In addition, a dummy variable reflecting whether the two individuals were spatially co-located in the past (valued 1) or not (valued 0) is introduced to test the main hypothesis of the paper, that is, the existence of the 'previous co-location effect'. Thus, the latent payoffs of collaborating are described by the following expression:

$$\pi_t^{ij} = \beta_t^i \cdot X_t^i + \beta_t^j \cdot X_t^j + \beta_t^{ij,\text{proximities}} \cdot X_t^{ij,\text{proximities}} + \beta_t^{ij,\text{co-location}} \cdot X_t^{ij,\text{co-location}} + \gamma^{ij} + \varepsilon_t^{ij}. \tag{2}$$

The coefficient of interest, $\beta_t^{ij,\text{co-location}}$, will reflect networking practices' changes attributed to mobility. As it is customary in the related literature, a logit model is used to estimate the latent payoff.

Denote y_t^{ij} as the observed dependent variable, defined as a dummy taking the value 1 if a given pair of inventors collaborate at time t and 0 otherwise, conditional upon not having collaborated before, $t-s$. More formally, the specific data-generating process is expressed as follows:

$$\Pr(y_t^{ij} = 1 \mid y_{t-s}^{ij} = 0) = \frac{\exp(\beta_{t,n} \cdot X_{t,n} + \gamma^{ij})}{1 + \exp(\beta_{t,n} \cdot X_{t,n} + \gamma^{ij})}, \tag{3}$$

where $y^{ij}_{t-s} = 0$ stands for the fact that the pair has never collaborated before. The r.h.s. variables are lagged to avoid simultaneity bias. Thus, the probability of forming a tie in time t will be a function of a number of regressors computed within a time window of five years, from $t-5$ to $t-1$. In equations (1) to (3), γ^{ij} is a pair-wise fixed effect that takes on board all time-invariant unobservable characteristics that a cross-sectional setting cannot account for. I refer here to variables such as age, sex, race, educational and cultural backgrounds, current location, time-invariant research interests, and other features of the inventors' character, as well as the country of residence, physical distance to his partners, and the like. The introduction of pair-wise fixed effects is highly valuable, since allows a better identification of the influence of time-variant variables on the likelihood to observe a tie between regions. However, the introduction of fixed effects precludes testing other interesting variables, such as geographic proximity, which is actually one of the leitmotifs of large part of the related literature. I claim, however, that the 'previous co-location' variable may provide indirect evidence on the role of geography on network formation, whilst controlling for pair-wise fixed-effects at the same time. The way in which the variables are built is explained in detail in the following subsection.

Data Sources and Variables Construction

I start by retrieving all EPO patent applications from 1978 to 2005 having at least one technology class code corresponding to biotechnology. The REGPAT OECD database, January 2010 edition, is used (Maraut et al., 2008). Among the numerous information contained in patent data, it is included the technology or technologies into which the patent is classified. Thus, the front page of an EPO patent contains a number of codes corresponding to the International Patent Classification (IPC) allowing the classification of patents onto different broad technologies. I follow Schmoch's (2008) technological classification to select and retrieve biotechnology patents.[4] Afterwards, I retrieve all the information regarding the inventors having at least one biotechnology patent and contained in the database. Only inventors reporting a European postal address are considered. If an inventor has patented from Europe and also while residing abroad, all the information concerning his years in a non-European country is disregarded. Note that a single ID for each inventor and anyone else is missing in the database. However, in order to draw the spatial mobility and networking history of inventors, it is necessary to identify them individually. I use their name and surname, as well as other useful details contained in the patent document, for singling out individual inventors using patent documents. In brief, I first clean, harmonize and code all the inventors' names and surnames. Afterwards, I test whether each pair of names belong to the same individual, using a wide range of characteristics, such as their address, the applicants and groups of applicants of their patents, their self-citations, or the technological classes to which their patents belong – up to 15 different tests were run.[5]

[4] This means retrieving all patents which IPC codes start with one of the following 4-digit strings: C07G, C12M, C12N, C12P, C12Q, C12R, or C12S.
[5] See, among others, Miguélez and Gómez-Miguélez (2011) for a detailed description of the methods I used to identify single inventors from patent data, as well as for a review of related literature.

Dependent Variable

I look first at all the realized ties during the whole period of analysis, building up all the possible pairs, that is, all the couples of inventors that have a co-patent. I remove all ties occurring within the same region. I also disregard the pairs in which at least one of the inventors has only one patent. Recall that I am interested in knowing whether there exist a collaboration premium due to being co-located (residing in the same region) in the past. To that end, I need to exploit the information concerning the inventors' past location. Similarly, I drop all the pairs in which the focal co-patent is the first patent for at least one of the inventors of the pair, even if he has additional subsequent patents. Again, this is done because I need to observe patenting history before the date of the focal co-patent.

Each pair of inventors is considered active from the first year in which both inventors have a patent to the last year in which both of them have a patent as well. Note, however, that for now I am only interested in the determinants of the inventors' first collaboration, so I remove the years after their first collaboration. Suppose that they have a co-patent at year t, t_{ij}. Therefore, I create a variable y_{ij} that takes value 1 at $t = t_{ij}$, and 0 at $t < t_{ij}$. That is, for each pair, I end up having a sequence from the first time they patent independently until their common co-patent, resulting in an unbalanced panel. All in all, I end up having 7,376 pairs of inventors forming linkages across NUTS3 regions. On average, the pairs take 4.5 years from their independent patenting to their common co-patent, ranging from a minimum of 2 years to a maximum of 21 years.

Explanatory Variables

All the explanatory variables are built within time-windows of five years.[6] Recall that the r.h.s. variables are lagged one year to avoid biases due to system feedbacks. I discuss the appropriateness of this approach later on. Thus, ties in year t are explained by a set of explanatory variables computed from year $t-5$ to $t-1$. In consequence, I remove all years of the dependent variables corresponding to the period 1978-1982, since a 5-year window lag for the explanatory variables cannot be computed from the raw data. All the explanatory variables are built using information from the REGPAT OECD database, January 2010 edition, unless otherwise noted.

Previous co-location: the main hypothesis of the present paper is tested by introducing a dummy variable valued 1 if the two inventors resided in the same NUTS3 region in the period $t-5$ to $t-1$, and 0 otherwise. Since this variable is re-built for each year, it shows time variation and can be included in the estimations alongside the fixed-effects.

Social proximity: to compute this variable, I start by defining the co-inventorship network, from $t-5$ to $t-1$, where inventors are nodes and co-patents are the links between these nodes. Afterwards, I compute the shortest path between every pair of inventors of my sample for each time window, p_t^{ij}, that is, the shortest geodesic distance between the two. Consider the following example: if inventors i and j have both co-invented with z, but not between them, their shortest path is 2.

[6] Different time windows do not alter significantly the qualitative results.

Recall that my focus is on the determinants of first co-patenting, so the minimum shortest path possible between pairs of inventors is always 2. If two inventors do not have any common co-author, at any geodesic distance, their shortest path is infinite. For this reason, it is better to work with the inverse of the geodesic distance, that is, social proximity, defined as

$$s_t^{ij} = \frac{1}{p_t^{ij}} \tag{4}$$

which varies between 0 and 0.5. Social proximity equals 0.5 if the two inventors share at least one common co-author, and equals 0 when they are not connected at all.

Cognitive proximity: to proxy cognitive proximity I use an index of technological similarity, or research overlap, as suggested in Jaffe (1986). Thus, I compute the uncentered correlation between individuals' vector of technological classes in the form of:

$$t_{ij} = \frac{\sum f_{ih} f_{jh}}{\left(\sum f_{ih}^2 \sum f_{jh}^2\right)^{1/2}} . \tag{5}$$

In (5), f_{ih} stands for the share of patents of one technological class h according to the IPC classification (out of 300 technological classes in the subdivision chosen) of the inventor i, and f_{jh} for the share of patents of one technological class h of the inventor j. Values of the index close to the unity would indicate that a given pair of inventors share almost the same fields of research, and values close to 0 means that they do not share research expertise at all.

Institutional proximity: proxied with a dummy variable valued 1 if the couplet of inventors used to work for the same type of applicant (company, university, non-profit organization, or hospital) according to their patent portfolio within the period $t-5$ to $t-1$, and 0 otherwise. Information on applicants' classification is retrieved from the EEE-PPAT database (Du Plessis et al., 2009) and merged with my sample.

Organizational proximity: when the inventors of the pair have worked for the same organization in the past, they are a priori more willing to collaborate; that is to say, knowledge workers are more likely to form ties within organizational boundaries. I proxy this variable with a dummy taking the value 1 if the pair of inventors share at least one common applicant according to their patent portfolio within the period $t-5$ to $t-1$, and 0 otherwise. Harmonized and coded applicants' data are retrieved from the KITES-PatStat database (Bocconi University – Milan), and merged with my sample.

As my estimations could be compromised if time-varying features of the individual inventors have an impact on the likelihood to observe a tie, I include additional variables derived from the raw database.

Productivity: More productive innovators tend to attract other inventors to work with them. Omitting individuals' ability to produce patents may lead to inconsistent results. To proxy individuals' ability, q_t^i, I count the number of patents of each inventor through the time-window $t-5$ to $t-1$, weighted by the number of citations each patent has received, to account for heterogeneity in patent quality and relevance – citations data are retrieved from the OECD Citations database, January 2010. Note that the dependent variable is undirected, so I need to have the same regressors, irrespective of the order of indexation. I chose to enter the regressors in a symmetrical way as in Fafchamps et al. (2010), that is, the average productivity

$$\overline{q}_t^{ij} = \frac{q_t^i + q_t^j}{2}, \tag{6}$$

and the absolute difference in productivity,

$$\Delta q_t^{ij} = \left| q_t^i - q_t^j \right|. \tag{7}$$

Degree centrality : I also need to control for observed time-varying individuals' propensity to collaborate. Likewise, the concept of *preferential attachment* (Barabási and Albert, 1999) states that highly connected actors are more likely to attract additional connections. To that end, I compute the innovators' degree centrality, dc_t^i, within each time-window $t-5$ to $t-1$. Degree centrality stands for the number of co-authors a given inventor has in a given time period. Again, I introduce symmetrically this variable as the average degree centrality,

$$\overline{dc}_t^{ij} = \frac{dc_t^i + dc_t^j}{2}, \tag{8}$$

and the absolute difference in degree centrality,

$$\Delta dc_t^{ij} = \left| dc_t^i - dc_t^j \right|. \tag{9}$$

4. Results

Descriptive Figures

This section presents summary figures of the phenomena under study. First of all, table 1 provides an overview of the biotechnology sector in Europe and some figures of my final dataset. From that table we learn the following main findings: first, the biotech industry accounts for 6.77% of all European inventors throughout the whole period (1978-2005), but only for 3.71% of the patents, which seems to indicate the importance of research teams in inventive activity – making the present analysis worthwhile. Only 37.36% of inventors (19,459) are multi-patent – and therefore constitute my focal group of analysis – of which only 9.15% are mobile across the space – report more than one NUTS3 region of residence. The number of observed cross-regional pair-wise linkages is, respectively, for NUTS3 and NUTS2, 70,852 and 49,351. However, after the necessary restrictions imposed described above, the focal group of analysis reduces to 7,376 and 4,902 pairs (respectively, 10.41% and 9.94%), which represents the 10.53% of all biotech inventors. This percentage is apparently low, indeed. Note, however, that these 5,484 inventors have, on average, larger number of patents per inventor, larger number of co-authors, and accumulate more citations to their work, witnessing the importance and economic impact of this subgroup for inventive activity and knowledge diffusion.

Table 1 Summary Figures

Absolute number of inventors in biotech (1978-2005)	52,081
Share of inventors in biotech	6.77%
Number of patents in the biotech industry	38,624
Share of patents in the biotech industry	3.71%
Average number of patents per inventor	2.19
Average co-authors per inventor	5.11
Average number of citations received per inventor	0.83
Number of multi-patent inventors	19,459
Geographically mobile inventors (NUTS3)	1,781
Share mobile inventors over multi-patent inventors	9.15%
Total number of potential ties	1,356,189,240
Total number of realized ties	124,681
Realized ties across different NUTS3 regions	70,852
Realized ties across different NUTS2 regions	49,351
Observed ties under analysis (NUTS3)	7,376
Observed ties under analysis (NUTS2)	4,902
Final set of inventors under study	5,484
Share of biotech inventors under study	10.53%
Average number of patents per inventor final dataset	6.78
Average number of co-authors per inventor final dataset	6.10
Average number of citations received per inventor final dataset	3.22

Note: Recall that the final dataset refers to the final number of inventors used in the empirical analysis, retrieved after the necessary restrictions imposed described before.

Table 2 goes one step further in the analysis of this subgroup. In there, summary figures of the number of patents per inventor, number of co-authors and citations received are shown broken down into two groups: geographically mobile inventors (those with more than one NUTS3 region of residence) and non-mobile inventors. Noticeably, mobile inventors are more productive, have more co-authors, and their work is more valuable, according to the number of citations received. The figures indicate that mobile innovators differ systematically in their observable characteristics from those who do not move across regions. Clearly, controlling for such features in the econometric analysis is pivotal.

Table 2. Two-Group Mean Comparison. Mobile vs. Non-Mobile Innovators

	Mobile inventors	**Non-mobile inventors**	**Absolute difference**
Observations	1,383	4,101	
Average # of patents per inventor	8.79	6.11	2.68***
Average # of co-authors per inventor	7.25	5.71	1.55***
Average # of citations per inventor	3.99	2.97	1.02***

Notes: *** p<0.01, ** p<0.05, * p<0.1.

Table 3 provides summary statistics of the variables used in the present analysis for the case of linkages across NUTS3 regions (NUTS2 linkages figures can be provided upon request). Finally, table 4 displays the correlation matrix.[7] Other than the high correlations between both productivity measures and between both degree centrality measures, the correlation among the focal independent variables is, in general, sufficiently small and collinearity does not pose a significant problem in the estimations. I do not find those high correlations a serious concern to the extent that these four variables are only used to control for confounding individuals' features that might bias the point estimates of the focal variables of the present analysis.[8]

Table 3 Summary Statistics, Unbalanced Panel, Linkages Across NUTS3

	# obser.	**Mean**	**Coef. Var.**	**Min**	**Max**
Cross-regional co-patents	33,005	0.22	1.86	0	1
Social proximity	33,005	0.09	1.96	0	0.50
Cognitive proximity	33,005	0.38	1.04	0	1
Institutional proximity	33,005	0.50	0.99	0	1
Organizational proximity	33,005	0.29	1.58	0	1
Previous co-location	33,005	0.06	3.90	0	1
Average productivity	33,005	1.15	1.32	0	21.55
Abs. diff. productivity	33,005	1.49	1.64	0	41.17
Average centrality	33,005	6.74	1.36	0	128
Abs. diff. centrality	33,005	9.12	1.65	0	236

Note: Descriptive figures do not included variables in first differences.

[7] Potential spurious correlation between r.h.s. variables and the dependent one due to non-stationary panels may arise. This may happen because the dependent variable is by construction a sequence of zeros followed by a single 1. Any regressor exhibiting a trend will mechanically create a correlation with the dependent variable (see Fafchamps et al., 2010). Unit root tests for panel data are performed to identify regressors that exhibit a trend. Unfortunately, I am unaware of unit root tests for unbalanced panels including very short series. To solve this pitfall, I first drop out all the panels with 10 or less periods and perform Im-Pesaran-Shin tests (Im et al., 2003), which allows for unbalanced panels. Afterwards, I keep separately the panels with 5, 10 and 15 periods and perform unit root tests for balanced short panels (Harris and Tzavalis, 1999). The null hypothesis of these tests is that the panel contains unit roots, whilst the alternative is that the panels are stationary. Those variables for which most of these tests do not reject the null are said to exhibit trend. I only find some evidence of trend for the case of the productivity variables, both the average and the absolute difference, and the degree centrality variables, again both the average and the absolute difference. Results of these tests can be provided upon request from the author. To address this issue, these four variables are included in first differences in all the estimations.

[8] To ensure that this is not an issue, I repeated all the estimations by including either one or the other highly correlated variables each time. No remarkable change is worth to be reported.

Table 4 Correlation Matrix, Unbalanced Panel, Linkages across NUTS3

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1. Cross-regional co-patents	1									
2. Social proximity	0.19	1								
3. Cognitive proximity	0.19	0.42	1							
4. Institutional proximity	0.15	0.60	0.45	1						
5. Organizational proximity	0.18	0.46	0.66	0.62	1					
6. Previous co-location	0.02	0.12	0.07	0.15	0.10	1				
7. Average productivity	0.11	0.15	0.19	0.16	0.20	0.04	1			
8. Abs. diff. productivity	0.04	0.05	0.04	0.04	0.04	0.01	0.83	1		
9. Average centrality	0.16	0.19	0.19	0.17	0.21	0.03	0.60	0.48	1	
10. Abs. diff. centrality	0.09	0.06	0.06	0.06	0.06	0.01	0.49	0.51	0.88	1

Note: Correlations involving variables 7 to 10 are computed using their first differences transformation.

Fixed Effects conditional Logit Estimation

I now turn to examining the estimation results. Recall that I estimate an unbalanced panel, from 1983 to 2005. Conditional logit methods are used to drop out the fixed effect (Chamberlain, 1992). Note, however, that the inclusion of pair-wise fixed effects prevents me to directly test the role of geographical proximity. Table 5 reports the fixed-effects logit estimations for the linkages formed across different NUTS3 regions in Europe. Note that all the proximities considered (social, cognitive, institutional, and organizational) are significant and with the expected sign, confirming prior evidence on the role of different, more meaningful types of proximities to explain agents' knowledge interactions and linkages formation. As it is shown below, these results are robust to the choice of the spatial scale (NUTS3, NUTS2), different specifications and time windows, and the inclusion of fixed-effects. Results concerning productivity and collaborative propensity of innovators (their degree centrality) accord with the theory. Thus, both the average productivity and the average connectivity enhance knowledge linkages formation.[9] That is, the more productive or connected, on average, are two given inventors, the more willing to collaborate they are. The absolute difference of both variables is, however, negative and significant. That is to say, the likelihood of collaborating falls when authors are dissimilar in terms of their productivity and their propensity to collaborate.

'Previous co-location' is the main variable under scrutiny in the present inquiry. The associated coefficient is positive and significant throughout all the estimations of table 5. This finding holds even when controlling for a large number of potential time-varying confounders as well as for pair-wise time-invariant fixed-effects. Thus, there exists a premium derived from being co-located in the past on the likelihood to form ties between currently non-co-located individuals, all else equal. This result confirms my main hypothesis. Put differently, informal ties between individuals, shared trust and mutual understanding, built after years of co-location and shared spatial context, may well survive the spatial separation of the individuals and be a source of knowledge interaction among peers. This result provides further evidence on the role of 'pure geography' as well. The spatial, highly contextual, conditions in which interactions take place and social capital is built up are important for economic outcomes. Its effects manifest through agents that shared this same context but are not currently co-located, even controlling for a wide range of time-variant and time-invariant features.

[9] The average productivity is only significant when 10-year windows are used to compute the r.h.s. variables, as it is shown later on.

Further, to see not only the statistical, but also the economic significance of these results, the marginal effects were also calculated and evaluated at the means – except for the case of dummy variables, evaluated at the change from 0 to 1.

Thus, I find that having shared a common spatial context in the past 5 years increases the probability to build up cross-regional linkages by around 5.3%, holding other covariates at the reference points. This result may seem certainly unimportant in economic terms. In order to make these figures comparable, note that the marginal effect of institutional proximity – they worked for the same type of institution in the recent past – is around 5.2%, whilst having worked for the same organization increases the probability to observe a tie, with respect to those pairs that did not work for the same firm in the past, around 4.5%. The Appendix section contains the marginal effects for the analogous models estimated in table 5.

Table 5. Fixed-Effects Conditional Logit Estimations. Linkages across NUTS3

	(i)	(ii)	(iii)
Social proximity	1.922***	1.828***	1.922***
	(0.169)	(0.208)	(0.169)
Cognitive proximity	0.802***	0.811***	0.802***
	(0.076)	(0.092)	(0.076)
Institutional proximity	0.343***	0.388***	0.343***
	(0.064)	(0.076)	(0.064)
Organizational proximity	0.296***	0.431***	0.296***
	(0.071)	(0.089)	(0.071)
Previous co-location	0.346***	0.369***	0.347**
	(0.119)	(0.119)	(0.142)
Average productivity	0.034	0.034	0.034
	(0.025)	(0.025)	(0.025)
Abs. diff. productivity	-0.029**	-0.029**	-0.029**
	(0.014)	(0.014)	(0.014)
Average centrality	0.051***	0.050***	0.051***
	(0.005)	(0.005)	(0.005)
Abs. diff. centrality	-0.010***	-0.010***	-0.010***
	(0.003)	(0.003)	(0.003)
Geographic*Social		0.037	
		(0.048)	
Geographic*Cognitive		-0.004	
		(0.028)	
Geographic*Institutional		-0.023	
		(0.024)	
Geographic*Organizational		-0.041*	
		(0.024)	
Previous co-location*Geographic			-0.000
			(0.023)
Pairwise fixed-effects	yes	yes	yes
Observations	33,005	33,005	33,005
Pairs of inventors	7,376	7,376	7,376
McFadden's Adjusted R-squared	0.141	0.141	0.141
Log-likelihood	-8423.156	-8413.434	-8423.156

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

Specification (ii) introduces interactions between geography and some selected proximities. The logic behind these interactions is to test the idea that physical proximity plays a critical role as a platform to enhance the effects of the other more meaningful similarities (Boschma, 2005; Rodríguez-Pose and Crescenzi, 2008). If this condition was met, I would expect to find complementarities between geography and the other types of proximity in the form of positive and significant coefficients when their interactions are introduced. Results seem to indicate a substitutive, rather than complementary, relationship between geography and the other proximities considered. That is, a wide range of meaningful proximities enables knowledge interactions between separated individuals independently from their current physical distance.

Column (iii) introduces the interaction between the 'previous co-location' variable and current geographic proximity between the inventors' regions centroids. The interaction term measures if the marginal effect of being previously co-located depends on the current geographical distance between the two partners. The coefficient is not significant. This is an important result for identification. Hence, I interpret this finding as evidence that the 'previous co-location premium' is not a result of not being in the same NUTS3 region but close enough to maintain frequent physical interactions. Clearly, two non-co-located individuals are more likely to interact if they are at a nearby present distance. However, the importance of having shared a common spatial context in the past is independent of this present physical proximity and therefore previous co-location turns out to be important in itself.

Causal Interpretation

The challenge in interpreting mobile high-human-capital employees' effects on network formation, however, is that mobility itself is not exogenously determined. Endogeneity issues are discussed here. Recall that in the foregoing I have used time-lagged explanatory variables, so as to minimize system feedbacks. Despite this, omitted variables could also be a source of endogeneity and biased estimates. Although the data set is rich in observed characteristics of the inventors, many dimensions which are likely to affect the network formation decision remain unobserved. If these unobserved factors are correlated with the outcome, the estimated mobility-networking relationship would be biased.

First, inventors' motivation or some features of their talent may remain unobserved. More talented and productive individuals are more willing to move across regions (think about their chances to be hired by out-of-the-region firms or their chances to get a work permit before the Schengen agreement became effective).

Similarly, more talented individuals are also more prone to be required by other innovators to start collaborating. Mobility and productivity are likely to be correlated and drive the results. Observed productivity measures were included among the regressors. Yet, other measures of productivity observed by the inventor's peers but not observed using patent data, such as his scientific publication record, might be as important as their patent portfolio and importantly bias the results. Second, unobserved similarity in research interest between the pair may also drive the results. Current technologies are actually very narrow, even narrower than what IPC classes may take on board (for a discussion on spurious correlations due to broad patent technology fields, see Thompson and Fox-Kean, 2005). Therefore, at the end of the day the potential inventors with whom to collaborate are very few. Finally, two inventors might have worked together before in a scientific paper, in a national patent, or in an EPO patent which was not finally successful. Indeed, having worked together may increase the likelihood to form a tie because the individuals involved have enough information about each other, as well as enough mutual trust and understanding. However, this relationship has nothing to do with the fact that they shared a common space in the past and they built up social capital and informal relationships that endure after their physical separation. To the extent that collaboration in scientific papers or patents for national offices are more willing to occur between co-located individuals, the 'previous co-location' variable may take these effects on board if they are not controlled for. For all these reasons, it is reasonable to think that the findings encountered so far could be the result of an omission of relevant variables, and therefore my estimates would be inconsistent.

My identification strategy mimics the one in Fafchamps et al. (2010) and exploits information on subsequent collaborations between pairs. The underlying logic is that, if the listed omitted variables are relevant and drives the results concerning the 'previous co-location' premium, there is no reason to think that they do not drive the results for subsequent collaborations. That is to say, the argument states that after patenting together, two inventors have enough information about the match quality and their likelihood to start collaborating should not depend on the benefits of having shared the same spatial context in the past. Contrariwise, time-varying confounders such as their unobserved talent or motivation, for instance, are likely to be correlated with this match quality and therefore drive network formation before and after their first collaboration. As in Fafchamps et al. (2010), I perform a counterfactual-type experiment, by testing the role of 'previous co-location' on subsequent collaboration, conditional upon having collaborated before. While this type of experiment does not completely resolve for the omission of relevant variables, the potential qualitative results of this exercise may give support to my previous findings. T hus, I would expect my focal variable no longer matter unless the 'previous co-location' premium is correlated with time-varying unobserved previous work, individuals' talent, motivation, or research overlap, that might confound with this premium.

Table 6 replicates the main results of table 5 in column (i), and re-estimates the model for the subsample of subsequent collaborations in column (ii). The point estimates of the 'previous co-location' variable decreases dramatically, whilst the standard error increases, making strongly non-significant the effect of this variable. Admittedly, the sample size of the second specification is considerably lowered, and therefore the results should be viewed as a robustness check.

Table 6. Fixed-Effects Conditional Logit Estimations: First and Subsequent Collaborations

	(i)	(ii)
Social proximity	1.922***	0.311
	(0.169)	(0.366)
Cognitive proximity	0.802***	-0.123
	(0.076)	(0.384)
Institutional proximity	0.343***	-0.709
	(0.064)	(0.495)
Organizational proximity	0.296***	0.326
	(0.071)	(0.541)
Previous co-location	0.346***	0.018
	(0.119)	(0.188)
Average productivity	0.034	0.021
	(0.025)	(0.029)
Abs. diff. productivity	-0.029**	-0.017
	(0.014)	(0.020)
Average centrality	0.051***	-0.024***
	(0.005)	(0.004)
Abs. diff. centrality	-0.010***	0.009***
	(0.003)	(0.003)
Pairwise fixed-effects	yes	yes
Observations	33,005	3,846
Pairs of inventors	7,376	762
McFadden's R-squared	0.141	0.022
Log-likelihood	-8423.156	-1297.242

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

Robustness Analysis

This section summarizes complementary estimations I performed in order to ensure the robustness of my main results. First of all, we acknowledge that, for the case of some countries of Europe, the NUTS3 administrative borders do not correspond to meaningful regions where economic interactions take place within relatively confined boundaries, but to arbitrary parts of them. In order to see whether the choice of the spatial scale bias the results I repeat the former analysis but only considering those pair-wise linkages across different NUTS2 regions. Fortunately, as illustrated in column (i) of table 7, most of the results and qualitative conclusions remain unaltered with respect to the former estimations. Note, however, that most of the computed marginal effects (see Appendix, table A.2) decrease the size of the coefficient, being 'previous co-location' the exception. Thus, it seems that the importance of having shared a common spatial and social context in the past is especially beneficial when the chances to meet and interact are substantially reduced. In line with Corredoira and Rosenkopf's (2010) interpretation, proximate agents may exploit other interaction channels, and therefore the 'previous co-location' premium becomes more valuable when these channels are less likely to be available.

Column (ii) repeats the main estimation of table 5 but using 10-year time windows to compute the explanatory variables. Some of the coefficients are changed with respect to the former tables. In particular, the point estimates corresponding to social proximity diminishes dramatically, whilst increases for the case of cognitive, institutional, and organizational proximities, as well as for my focal variable, the 'previous co-location' effect. However, the main conclusions remain unchanged.

Finally, column (iii) mimics the estimation of column (ii) slightly changing the computation of the 'previous co-location' variable. In particular, this dummy variable is now valued 1 if the two inventors resided in the same NUTS3 region in the period $t-10$ to $t-6$, and 0 otherwise. The logic of doing that is to ensure that the decision to collaborate was not taken before the decision to move and get separated in the space. As can be seen, the focal variable remains significant and increases sharply its point estimates, pointing at the absence of any remaining simultaneity effect.

Table 7. Fixed-Effects Conditional Logit Estimations. Robustness Analysis

	(i)	(ii)	(iii)
Social proximity	1.733***	0.225**	0.221**
	(0.217)	(0.091)	(0.091)
Cognitive proximity	0.857***	1.771***	1.775***
	(0.091)	(0.114)	(0.114)
Institutional proximity	0.356***	1.013***	1.054***
	(0.075)	(0.088)	(0.089)
Organizational proximity	0.417***	1.105***	1.113***
	(0.093)	(0.105)	(0.105)
Previous co-location	0.950***	0.644***	2.903***
	(0.194)	(0.159)	(0.285)
Average productivity	0.040	0.062**	0.057**
	(0.031)	(0.027)	(0.027)
Abs. diff. productivity	-0.037**	-0.045***	-0.042***
	(0.017)	(0.015)	(0.015)
Average centrality	0.048***	0.066***	0.066***
	(0.006)	(0.005)	(0.005)
Abs. diff. centrality	-0.006*	-0.012***	-0.012***
	(0.003)	(0.003)	(0.003)
Pairwise fixed-effects	yes	yes	yes
Observations	21,683	30,336	30,336
Pairs of inventors	4,902	6,828	6,828
McFadden's Adjusted R-squared	0.143	0.248	0.256
Log-likelihood	-5542.073	-6814.430	-6739.450

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. Column (i) shows the results corresponding to linkages across NUTS2 regions. Column (ii) corresponds to linkages across NUTS3 regions, but computing the explanatory variables over 10-year time-windows. In column (iii), the potential prior co-location between inventors in the immediate 5 past years is disregarded. The observations corresponding to the years 1983 to 1987 are not included in the estimations (ii) and (iii).

5. Concluding Remarks

Throughout the previous pages I attempt to appraise the role played by skilled individuals that move across the space, bridging in this way physically distant pools of knowledge. I defend that these actors play a critical role in the formation of an integrated and coherent European Research Area, whereas at the same time they are pivotal means by which knowledge is entered into the territory in order to introduce variation and avoid regional lock-in problems. My main tenet is that the way in which they are beneficial is through the formation of knowledge linkages (in my case, co-patents in the European biotech industry) across regions more disproportionately with their former colleagues than if they had never lived there – the 'previous co-location' premium.

The results confirm, by and large, that indeed there exists a ‘previous co-location’ premium in the likelihood to observe a knowledge linkage across the space, even when controlling for a large number of time-varying variables as well as pair-wise time-invariant fixed-effects. I also claim that this relationship is likely to be causal. Thus, I follow Fafchamps’ et al. (2010) methodology and perform a counterfactual experiment using subsequent collaborations. Hence, I show that the ‘previous co-location’ variable affects only the likelihood to observe a tie for the first time, arguing that features such as informal relationships, trust, mutual understanding, and so on, inherent to the spatial context in which the two inventors were co-located, are unlikely to affect tie formation aside from through their prior co-authorship.

The implications of my results are manifold as regards to the way in which knowledge diffuses across the space as well as the formation of the European Research Area. Particular implications can be derived for the case of European peripheral regions. The related literature has largely shown evidence of the physical stickiness of knowledge flows, especially in the form of spillovers. This fact helps to explain how peripherality persistently hampers regional innovation of these regions: the sticker the knowledge, the lower the access to this asset by peripheral territories (Rodríguez-Pose and Crescenzi, 2008). A direct way to access this otherwise unreachable distant knowledge pools is through mobile skilled employees ‘migrating’ from peripheral to core regions, possibly positioned in the technological frontier, who do not break their ties with their former colleagues, and enables knowledge interactions back with their past location.

My results shed also new light on the lively debate around the ‘brain drain’ vs. ‘brain circulation’ paradigms (Saxenian, 2006). In general, countries are reluctant to encourage outward mobility. This is in part because of the belief that local economic development heavily relies on attracting and retaining talent (Florida, 2002), thereby outward mobility is usually seen as a loss of local endowments. Few policy initiatives move in the other direction, though. For instance, the Spanish Ministry of Education offers Integrated Programmes to finance long-term research stays abroad for pre and post-doctoral students, in order to establish connections with distant R&D centres. Certainly, this is the exception rather than the rule among the European countries (OECD, 2008) and, more importantly, have experienced severe budgetary constrains from the beginning of the financial crisis – which may compromise accessing relevant out-of-the-region knowledge and ensuing economic growth in the long run.

If the local economic tissue is in position to reinforce the local identity as well as the sense of belonging to it by those who left, the region will be able to encourage mobile talent to come back after some years of working abroad in probably more technologically advanced regions or, at least, maintaining linkages with their home colleagues with whom knowledge flows and knowledge linkages may go back more easily than if they had never lived there. To that end, policies targeted to maintain and reinforce this sense of belonging to a given place, as well as, and more important, policies aimed to keep creating talent (strengthening the education levels of the indigenous population), are strongly recommended.

References

Aghion, P., Howitt, P., (1998), Endogenous Growth Theory. Cambridge (MA): The MIT Press.

Agrawal, A., Cockburn, I., McHale J., (2006), Gone but not Forgotten: Labour Flows, Knowledge Spillovers, and Enduring Social Capital, *Journal of Economic Geography*, 6, 571-591.

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.

Agrawal, A., Kapur, D., McHale, J., Oettl, A., (2011), Brain Drain or Brain Bank? The Impact of Skilled Emigration on Poor-Country Innovation, *Journal of Urban Economics*, 69(1), 43-55.

Arthur, W.B., (1989), Competing Technologies, Increasing Returns, and Lock-in by Historical Events, *Economic Journal* 99, 116-131.

Audretsch, D.B., Feldman, M.P., (1996), R&D Spillovers and the Geography of Innovation and Production, *American Economic Review* 86(3), 630-640.

Autant-Bernard, C., Billand, P., Frachisse, D., 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), 495-519.

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

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

Boschma, R., Frenken, K., (2010), The Spatial Evolution of Innovation Networks. A Proximity Perspective, in R. Boschma and R. Martin (eds.). *The Handbook of Evolutionary Economic Geography.* Cheltenham, Edward Elgar, pp. 120-135.

Bradner, E., Mark, G., (2002), Why Distance Matters: Effects on Cooperation, Persuasion and Deception, Proceedings of Computer Supported Cooperative Work (CSCW), New Orleans, Louisiana, pp. 226-235.

Cassi, L., Plunket, A., (2010), The Determinants of Co-Inventor tie formation: proximity and network dynamics. Papers in Evolutionary Economic Geography (PEEG) 1015, Utrecht University, Section of Economic Geography.

Cassiman, B., Veugelers, R., (2002), R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium, *American Economic Review*, 92(4), 1169-1184.

Chamberlain, G., (1992), Binary Response Models for Panel Data: Identification and Information, Harvard University, Unpublished Manuscript.

Corredoira R. A., Rosenkopf L., (2010), Should Auld Acquaintance Be Forgot? The Reverse Transfer of Knowledge through Mobility Ties, *Strategic Management Journal*, 31 (2),159-181.

David, P.A., (1985), Clio and the Economics of QWERTY, *American Economic Review*, 75(2), 332-337.

Du Plessis, M., Van Looy, B., Song, X., Magerman, T., (2009), Data Production Methods for Harmonized Patent Indicators: Assignee sector allocation. EUROSTAT Working Paper and Studies, Luxembourg.

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

Florida, R., (2002), *The Rise of the Creative Class, and How it's Transforming Work, Leisure, Community and Everyday Life*. New York: Basic Books.

Gertler, M.S., Levite, Y.M., (2005), Local Nodes in Global Networks: the Geography of Knowledge Flows in Biotechnology Innovation, *Industry and Innovation*, 12(4), 487-507.

Grossman, G.M., Helpman, E., (1991), Innovation and Growth in the Global Economy. Cambridge (MA), The MIT Press.

Grossman, G., Helpman, E., (1994), Endogenous Innovation in the Theory of Growth, *Journal of Economic Perspectives*, 8, 23-44.

Harris, R.D.F., Tzavalis, E., (1999), Inference for Unit Roots in Dynamic Panels where the Time Dimension is Fixed, *Journal of Econometrics*, 91(2), 201-226.

Im, K.S., Pesaran, M.H., Shin, Y., (2003). Testing for Unit Roots in Heterogeneous Panels, *Journal of Econometrics*, 115(1), 53-74.

Jaffe, A.B., (1986). Technological Opportunity and Spillovers of R&D: Evidence from Firms patents, Profits and Market Value, *American Economic Review* 76, 984-1001.

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

Jones, C., (1995), R&D-Based Models of Economic Growth, *Journal of Political Economy*, 103, 759-784.

Jöns, H., (2009), 'Brain Circulation' and Transnational Knowledge Networks: Studying Long-Term Effects of Academic Mobility to Germany, 1954-2000, *Global Networks*, 9(3), 315-338.

Kaiser, U., Kongsted, H.C., Ronde, T., (2011), Labor Mobility, Social Network Effects, and Innovative Activity. IZA DP No. 5654, Institute for the Study of Labor, Bonn.

Katz, J.S., Martin B.R., (1997), What is research collaboration?, *Research Policy*, 26, 1–18.

Kerr, W.R., (2008), Ethnic Scientific Communities and International Technology Diffusion, *The Review of Economics and Statistics*, 90(3): 518-537.

Latané, B., (1981), The Psychology of Social Impact, *American Psychologist,* 36(4), 343-356.

Maskell, P., Bathelt, H., Malmberg, A., (2006), Building Global Knowledge Pipelines: The Role of Temporary Clusters, *European Planning Studies*, 14(8): 997-1013.

Maraut, S., Dernis, H., Webb, C., Spiezia, V., Guellec, D., (2008), The OECD REGPAT Database: A presentation. STI Working Paper 2008/2, OECD.

Mariani, M., (2004), Location of Research and Collaboration among Inventors in the Chemical Industry, in: F. Cesaroni, A. Gambardella, W. Garcia-Fontes (Eds.) *R&D, Innovation and Competitiveness in the European Chemical Industry*. Dordrecht: Kluwer Academic publishers.

McPherson, M., Smith-Lovin, L., Cook, J.M., (2001), Birds of a Feather: Homophily in Social Networks, *Annual Review of Sociology*, 27: 415-444.

Miguélez, E., Gómez-Miguélez, I., (2011), Singling out Individual Inventors from Patent Data, IREA Working Paper 2011/05.

OECD (2008), *The Global Competition for Talent: Mobility of the Highly Skilled*.Paris: OECD.

Oettl, A., Agrawal, A., (2008), International Labor Mobility and Knowledge Flow Externalities, *Journal of International Business Studies*, 39(8), 1242-1260.

Paier, M., Scherngell, T., (2011), Determinants of Collaboration in European R&D Networks: Empirical Evidence from a Discrete Choice Model, *Industry & Innovation* 18(1), 89-104.

Rodríguez-Pose, A., Crescenzi, R. ,(2008), Mountains in a Flat World: Why Proximity Still Matters for the Location of Economic Activity, *Cambridge Journal of Regions*, 1(3), 371-388.

Saxenian, A., (2006), *The New Argonauts: Regional Advantage in a Global Economy*. Cambridge (MA), Harvard University Press.

Schmoch, U. (2008) Concept of a Technology classification for Country Comparisons. Final Report to the World Intellectual Property Organisation (WIPO). Fraunhofer Institute for Systems and Innovation Research, Karlsruhe http://www.wipo.int/ipstats/en/statistics/patents/pdf/wipo_ipc_technology.pdf

Storper, M., Venables, A., (2004), Buzz: Face-to-Face Contact and the Urban Economy, *Journal of Economic Geography*, 4(4), 351-370.

Tajfel, H., (1978), *Differentiation between Social Groups: Studies in the Social Psychology of Intergroup Relations*. London, Academic Press.

Ter Wal, A.L.J., (2011), The Dynamics of the Inventor Network in German Biotechnology: Geographical Proximity versus Triadic Closure. Papers in Evolutionary Economic Geography (PEEG) 1102, Utrecht University, Section of Economic Geography.

Ter Wal, A.L.J. and Boschma, R., (2009), Applying Social Network Analysis in Economic Geography: Framing Some Key Analytic Issues, *The Annals of Regional Science*, 43(3): 739-756.

Thompson, P., Fox-Kean, M., (2005), Patent Citations and the Geography of Knowledge Spillovers: A Reassessment, *American Economic Review*, 95(1), 450-460.
van Beuzekom, B., Arundel, A. (2009) *OECD Biotechnology Statistics 2009*, Paris, OECD.

Appendix A. Marginal Effects

Table A. 1. Fixed-Effects Conditional Logit Estimations. Marginal Effects

	(i)	(ii)	(iii)
Social proximity	0.294***	0.279***	0.294***
	(0.036)	(0.036)	(0.037)
Cognitive proximity	0.123***	0.118***	0.123***
	(0 .016)	(0.016)	(0.016)
Institutional proximity	0.052***	0.050***	0.052***
	(0.010)	(0.010)	(0.010)
Organizational proximity	0.045***	0.050***	0.045***
	(0.010)	(0.010)	(0.011)
Previous co-location	0.053***	0.054***	0.053***
	(0.015)	(0.014)	(0.015)
Average productivity	0.005	0.005	0.005
	(0.004)	(0.004)	(0.004)
Abs. diff. productivity	-0.004**	-0.004**	-0.004**
	(0.002)	(0.002)	(0.002)
Average centrality	0.008***	0.007***	0.008***
	(0.001)	(0.001)	(0.001)
Abs. diff. centrality	-0.002***	-0.001***	-0.002***
	(0.000)	(0.000)	(0.000)

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

Table A.2. Fixed-Effects Conditional Logit Estimations, Robustness Analysis. Marginal Effects

	(i)	(ii)	(iii)
Social proximity	0.158***	0.025**	0.024**
	(0.034)	(0.010)	(0.010)
Cognitive proximity	0.078***	0.194***	0.189***
	(0.016)	(0.012)	(0.011)
Institutional proximity	0.037***	0.124***	0.126***
	(0.010)	(0.011)	(0.011)
Organizational proximity	0.045***	0.108***	0.106***
	(0.012)	(0.009)	(0.009)
Previous co-location	0.125***	0.057***	0.119***
	(0.020)	(0.011)	(0.006)
Average productivity	0.004	0.007**	0.006**
	(0.003)	(0.003)	(0.003)
Abs. diff. productivity	-0.003*	-0.005**	-0.005**
	(0.002)	(0.002)	(0.002)
Average centrality	0.004***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)
Abs. diff. centrality	-0.001*	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. Column (i) shows the results corresponding to linkages across NUTS2 regions. Column (ii) corresponds to linkages across NUTS3 regions, but computing the explanatory variables over 10-year time-windows. In column (iii), the potential prior co-location between inventors in the immediate 5 past years is disregarded. The observations corresponding to the years 1983 to 1987 are not included in the estimations (ii) and (iii).