XVth Meeting of Applied Economics
A Coruña, 7-8 June 2012

HOW DO GEOGRAPHICALLY MOBILE INNOVATORS INFLUENCE NETWORK FORMATION?[1]

Very preliminary version

Ernest Miguélez
Economics and Statistics Division, World Intellectual Property Organization (WIPO),
34, Chemin des Colombettes, CH-1121 Geneva 11, Switzerland.
Tel.: (+41) 223388344. Fax: (+41) 223388350
& AQR-IREA, University of Barcelona.
E-mail: ernestmiguelez@gmail.com

Abstract
In this paper, I aim 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 have been traditionally invoked to easy both knowledge diffusion through repeated face-to-face interactions and to facilitate formal collaboration between individuals and firms. Tie formation is costly and decreases as distance between the partners involved increases, making ties between co-located individuals more likely than between spatially separated peers. However, too much inward-looking regions face the risk of lock-in, as local knowledge interaction becomes redundant and less valuable. In some instances, highly-skilled actors might become mobile and bridge regional networks across long physical distances. That is to say, the effect of trust and mutual understanding between members of a co-located community may well survive the end of their co-localisation, and therefore communication and knowledge diffusion between them may overcome long distances, and so, the formation of networks across the space. In this paper I estimate a fixed effects logit model to ascertain whether there exist a 'previous co-location premium' in the formation of networks across European regions. The role of mobility and non-permanent proximity has been lately discussed elsewhere, but, to my knowledge, barely empirically tested.

Key words: inventors' mobility, technological collaborations, co-location, brain drain, panel data

JEL: C8, J61, O31, O33, R0

[1] The views expressed in this article are those of the author and do not necessarily reflect those of the World Intellectual Property Organization. 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, and participants to the Geography of Innovation EuroLio Seminar (Saint Etienne, France, 26th-28th January 2012). 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). Besides, the growing complexity and quality requirements of current innovations have been confronted through an increase of collaborative practices between peers. Cross-pollination of ideas, barters of tacit knowledge or the division of labour, has been regarded to be the underlying forces heading to network formation (Katz and Martin, 1997). Yet, what drives the selection of one particular partner rather than the other?

In the present inquiry I focus my attention 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 is that the likelihood to observe a tie for the first time 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 and similarities between the two. 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 complementary with one's knowledge skills are unknown. Partners' searching costs are likely to be high. Other costs such as those derived from negotiation between the partners, contracts formation, 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 form ties, what happens when they move? Recent research stresses the importance of mobile inventors in setting up relations with their former colleagues and flow back knowledge to their origin 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 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 that 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 other 'similarities', such as geographic, social, cognitive, institutional, or organizational.

The study of this phenomenon is important from a policy perspective and motivates its analysis. Thus, 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 conduit 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. Under this setting, 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 mechanisms to identify and access distant pools of knowledge is through mobile high-human-capital employees who left the region but do 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 patents in the biotechnology industry to the European Patent Office (EPO hereafter), over the period 1978-2005. A fixed-effect logit model will be estimated to ascertain whether there exist a 'previous co-location premium' on the likelihood to build up formal ties across regions. As in Fafchamps et al. (2010), we deal with the endogenous nature of our foci variables 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.

In brief, the contributions of the present analysis 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. 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 provides some stylized facts as regards the geography of the biotech industry in Europe, helping to better motivate the present analysis. Section 3 reviews some literature, bringing together dispersed, but related, literature, and outlines the theoretical framework. Section 4 describes the empirical approach taken here and the data sources. Section 5 summarizes some remarkable findings and Section 6 presents conclusions and policy implications.

2. Stylized facts on technological collaborations in biotechnology

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). As much of other high-tech industries, biotech in Europe shows a striking tendency to cluster in space. This makes this industry an ideal candidate for the study of the geography of innovation and the specific role of networks and mobile workers within this strand (Ter Waal, 2011).

In 2003, for instance, Denmark was leading country in Europe in terms of biotechnology patent application per capita at the EPO, producing nearly six times the EU-27 average (Félix, 2007). Up to 7 European countries (Denmark, Iceland, Switzerland, Sweden, Germany, the Netherlands, and Belgium) were listed in the top ten worldwide ranking of biotech patent applications per capita, whereas Germany was leading country in Europe in absolute terms. In addition, Iceland, Denmark and Switzerland lead the ranking of biotech R&D intensity in 2003 (biotech R&D expenditure over country value added), far above the majority of Eastern and Southern European countries (op. cit.).

To illustrate this point further, I also look at the spatial distribution of patent applications across 287 European NUTS2 regions.[2] The Gini index –provided upon request- shows substantially large figures throughout the whole period under analysis. Thus, from values around 85-90 in the initial eighties until a value of 75 around 2005, the index has remained steadily high over time, pointing at a persistent spatial concentration of this industry throughout the period.

Next, the first panel of Figure 1 shows the share of patents produced by two or more inventors in this industry. Biotech inventions have been a matter of increasing collaborative practices over time, as the steady increase of the share of co-authored patents indicates. This rise is also reflected by team size: the average number of

[2] As explained in section 4, biotechnology patent data is retrieved from the REGPAT OECD database, January 2010 edition.

inventors per patent goes from 2.64 during the first part of the eighties until 3.55 around 2005.

[Insert Figure 1 about here]

This rise in collaborative practices, which might partially have helped sustaining the strong spatial concentration of this industry, goes however hand by hand with a deepening geographical spread of teams' composition. Panel (b) in Figure 1 shows the annual share of multi-inventor patents that includes co-authors of different NUTS2 regions. As observed, 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 to jointly patent. Subtracting one set of shares from the other yields a non-parametric measure of the likelihood of a cross-regional tie, conditional upon co-authored patents. This is presented in the third panel of Figure 1. The slight declining trend of the resulting line points at the fact that cross-regional collaborations augment their role more than proportionally compared to the evolution of co-authored patents in general. This gives further evidence that knowledge networks become progressively more interconnected across different locations. Given these two phenomena –persistent clustering and augmented a-spatial collaborations, my interest lies in ascertaining what fosters linkages formation between individuals located at a distance, above and beyond physical proximity.

Figure 2 plots the annual share of multi-patent inventors reporting, for each time-window, more than one address of residence –in different NUTS2 regions of Europe. That is, the figures show the share of spatially mobile inventors over time. Admittedly, geographic mobility is more the exception rather than the rule. However, again, the sharp increasing trend of mobile inventors is noticeable. Thus, in the present paper I aim to test whether, amongst other features influencing partnering selection across regions, there is room for the role played by geographic mobility of innovators and their informal linkages with those left behind.

[Insert Figure 2 about here]

3. 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. Particular attention has been paid from innovation economics and organization science, since the pervasiveness of organizations' and individuals' cooperative practices in knowledge creation have became a salient feature of innovation management and is regarded to be a source of outstanding firms' outcomes.

This has given rise to a flourishing number of hot 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, i.e. 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. Ties formation between peers is crucially determined by this similarity. Among others, 'homophily' may refer to physical proximity between partners. Indeed, geographic propinquity creates context in which homophile relations form and knowledge linkages 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. Some things have all these proximities 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 recent years, 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 Waal (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 Waal, 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 geographic, 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, in order to provide more reliable estimations.

A main tenet of the present paper is that geographic 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 invite 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 who they believe 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 all 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) advocates for *social identity* effects, which 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, for innovators interactions to arise, screening of potential partners is pivotal. 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.).

My focal argument in the present paper states that the benefits of physical proximity for the formation of linkages between inventors established through long period 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. Hence, they provide evidence of the 'enduring social capital hypothesis', that is, social ties created during inventors' co-localisation, which facilitated knowledge diffusion, persist even after the inventors' separation and are conducive to knowledge flows between them. 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. Insights from social capital research offer a theoretical framework for their empirical analysis. 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. That is, mobile skilled employees broker across structural holes (Burt, 1992) building up some sort of weak ties (Granovetter, 1973) that are especially useful to access distant knowledge networks. Thus, the authors provide evidence on the existence of knowledge flows in the form of knowledge externalities –the market does not price the flows back- when innovators move between countries.

These and related studies (see also Agrawal et al., 2008; Agrawal et al., 2011; Kerr, 2008) support 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.

Admittedly, Agrawal and co-authors' analyses do not take on board the influence of formal social ties between partners, which are likely to drive their results on disproportionate knowledge flows to prior locations (Breschi et al., 2010). I control for this possibility by including network effects among the explanatory variables.

In sum, as it will be discussed subsequently in detail, the present analysis tries to find evidence on the role of previous co-location –therefore long-term, but temporary proximity- 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.

4. Research design

Empirical model

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

Recall from previous sections that my general framework is the study of individuals' linkages formation between separate European regions –both NUTS3 and NUTS2 level. 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, conditioned 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, $\pi_t^{ij} > 0$. Hence, the probability that the link between inventor i and j is formed at time t is a function of positive payoffs, which in turn depend upon i's and j's observable time-variant and non-observable time-invariant characteristics, X and γ^{ij} respectively, as well as a well-behaved error term, ε:

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

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 –geographic, 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:

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

The coefficient of interest, $\beta_t^{ij,\mathrm{co-location}}$, 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, where the dependent variable, y_t^{ij}, is defined as a binary variable 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 $y_{t-1}^{ij}=0$ stands for the fact that I only consider in time t if the pair has never collaborated before; and n stands for the number of regressors included in the model. 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 the (1)-(3) equations, γ^{ij} is a pair-wise fixed effect that takes on board all time-invariant unobservable variables 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 I construct the variables is explained in detail in the following subsection.

Data sources and variables construction

I start by retrieving all patents applied to the EPO 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 to which the inventors themselves assign their patent. Thus, the front page of an EPO patent includes 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.[3] 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, I disregard all the information concerning his years in a non-European country. 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.[4]

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 –separately, NUTS3 and NUTS2. I also disregard the pairs in which at least one of the inventors has only one patent. Recall that I am interested to know 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 interested only in the determinants of the inventors' first collaboration, so we remove the years after their first collaboration. Suppose that they have a co-patent at year t_{ij}. Therefore, I create a variable y_{ij} that takes value 1 at $t = t_{ij}$,

[3] 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.
[4] See Miguélez and Gómez-Miguélez (2011) for a description of the methods I used to identify single inventors from patent data, as well as for a review of related literature.

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 7,377 pairs of inventors for the case of linkages across NUTS3 regions (4,903 for linkages across NUTS2 region). On average, the pair takes 4.5 (4.4) 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.[5] 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 1-year lagged 5-year time-window 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 (NUTS2) 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 fixed-effects estimation.

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. 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-

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

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

$$s_t^{ij} = \frac{1}{p_t^{ij}} \tag{6}$$

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:

$$t_{ij} = \frac{\sum f_{ih} f_{jh}}{\left(\sum f_{ih}^2 \sum f_{jh}^2\right)^{1/2}} . \tag{7}$$

In (7), 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.

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).

Institutional proximity: Institutional proximity is 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 Pleassis et al., 2009) and merged with our sample.

Productivity: 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. 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, we 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 we 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

$$\bar{q}_t^{ij} = \frac{q_t^i + q_t^j}{2}, \tag{9}$$

and the absolute difference in productivity,

$$\Delta q_t^{ij} = \left| q_t^i - q_t^j \right|. \tag{10}$$

Degree centrality: I also need to control for observed time-varying individuals' propensity to collaborate. Besides, 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 follows: average degree centrality,

$$\overline{dc}_t^{ij} = \frac{dc_t^i + dc_t^j}{2}, \qquad (11)$$

and the absolute difference in degree centrality,

$$\Delta dc_t^{ij} = \left| dc_t^i - dc_t^j \right|. \qquad (12)$$

5. 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 our 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 our 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 are, respectively, for NUTS3 and NUTS2, 70,852 and 49,351. However, after the necessary restrictions imposed described above, our focal group of analysis reduces to 7,377 and 4,903 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 numbers 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.

[Insert Table 1 about here]

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. The figures indicate the importance of controlling for such features in the econometric analysis in order not to bias our estimates.

[Insert Table 2 about here]

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. 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 our estimation. 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.[6]

[Insert Tables 3 and 4 about here]

[6] For the case of the fixed-effects conditional logit estimations, 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 the null is not rejected 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. To address this issue, I de-trend these variables by regressing them on a pair-wise fixed effect and a liner time trend, and using the residuals of these estimations instead of the original variables in the main estimations (as in Fafchamps et al., 2010). Results of the tests are not presented here to save space but can be provided upon request.

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 again, however, that the inclusion of pair-wise fixed effects prevents me to directly test the role of geographic 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. 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) are according to the theory. Thus, both the average productivity and the average connectivity enhance knowledge linkages formation. 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 none co-located individuals, all else equal. This result confirms my main hypothesis: European innovators are more willing to form a knowledge linkage across different NUTS3 regions, in the form of co-patents, if they coexisted in the same physical territory in the recent past. 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. In addition, its effects manifest through agents that shared this same context but are currently not co-located, even controlling for a wide range of time-variant and time-invariant features.

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. Thus, I find that having shared a common spatial context in the 5-year past increases the probability to build up cross-regional linkages by around 3.2%, holding other covariates at the reference points. Put differently, mobility of skilled labour increases the probability of linkages to distant pools of knowledge by 3.2%. This result may seem certainly unimportant in economic terms. In order to make these figures comparable, note that the marginal effect of social proximity is around 13%, that is, a 1% increase of this variable increases by 0.13% the probability to observe a tie. Similarly, marginal effects are 7% for cognitive proximity, 2.2% for institutional proximity, and 3% for organizational proximity.

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 forms 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 significance. This is an important result for identification. Hence, I can 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 similar present distance. However, the importance of having shared a common spatial context in the past is independent of this present physical proximity and therefore it turns to be important in itself.

[Insert Table 5 about here]

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 my results I repeat the former analysis by only considering those pair-wise linkages across different NUTS2 regions. Fortunately, most of the results and qualitative conclusions remain unaltered with respect to the former estimations. One important remark is in order: the only significant change with respect the former table corresponds to the 'previous co-location' coefficient, which increases by more than 3 times with respect to the former results (marginal effect of 7%). That is to say, the importance of having shared a common spatial and social context in the past is especially valuable when the chances to meet and interact are substantially reduced. In line with Corredoira and Rosenkopf (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. Equally, geographic and cognitive proximities show now positive and significant complementarity effects.

[Insert Table 6 about here]

Causal interpretation

The challenge in analysing 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 we have used time-lagged explanatory variables, avoiding in this way system feedbacks and minimizing as much as possible resulting endogeneity problems. Presumably, however, omitted relevant variables may also be a source of endogeneity and affect my results. Thus, for instance, even if I have ensured that the analysis is performed on the likelihood to observe a knowledge linkage for the first time, 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.

It is also reasonable to think that 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). At the same time, 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. In order to minimize this effect, I included a measure of the individuals' productivity in the form of the weighted number of patents. Despite this, 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 my results. This is reinforced by the fact that technology fields are actually very narrow, even narrower than what IPC classes may take on board (for a discussion on spurious correlations due to narrow 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. Again, this is likely to affect the results to the extent that specific technology fields might also be spatially concentrated, and therefore two researchers that used to live together may share a common very specific cognitive background even if no social capital, trust, and mutual understanding effects took place.

For all these reasons, it is reasonable to think that the findings encountered so far might be the result of an omission of relevant variables, and therefore our estimates are inconsistent. In order to take on board these drawbacks, my identification strategy mimics Fafchamps' et al. (2010) one and exploits information concerning subsequent collaborations between the pairs of inventors of my sample. The idea of this approach is to estimate the same models but for the years after the first collaboration. 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. 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 give support to my previous findings. Thus, 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, or research overlap, that might confound with this premium.

Table 7 below 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 on the likelihood to observe a knowledge linkage between inventors that have already collaborated. Admittedly, the sample size of the second specification is considerably lowered, and therefore the results should be treated with caution. However, the number of observations is still large enough and therefore I can be reasonable confident on the causal effect of the 'enduring social capital hypothesis' on knowledge linkages across NUTS3 European regions.

[Insert Table 7 about here]

6. 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 was 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 followed Fafchamps’ et al. (2010) methodology and perform a counterfactual experiment using subsequent collaborations. Hence, I showed 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 peripherally 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.

Hence, my results shed also new light on the lively debate around the ‘brain drain’ vs. ‘brain circulation’ paradigms (Saxenian, 2006). Thus, national and regional policies aimed to send skilled individuals abroad for long, though temporary, periods of time should not be seen as a ‘brain loss’ anymore, but as a ‘brain investment’. Hence, 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. In general, countries are reluctant to encourage outward mobility despite the 'brain circulation' arguments. 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 an asset loss.

However, 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 which knowledge flows and knowledge linkages may go back easily than if they 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. and Howitt, P. (1998) *Endogenous Growth Theory*, MIT Press.

Agrawal, A.; Kapur, D. and 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. and 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.

Agrawal A., Cockburn I. and McHale J. (2006) "Gone but not forgotten: labour flows, knowledge spillovers, and enduring social capital", *Journal of Economic Geography* 6, 571-91.

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

Audretsch, D.B. and 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. and Massard, N. (2007) "Social distance versus spatial distance in R&D cooperation: Empirical evidence from European collaboration choices in micro and nanotechnologies" *Papers in Regional Science* 86(3): 495-519.

Balconi, M.; Pozzali, A. and Viale, R. (2007) "The "codification debate" revisited: a conceptual framework to analyze the role of tacit knowledge in economics" *Industrial and Corporate Change*, 16(5): 823-849.

Barabási, A.-L. and 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. and 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. and Mark, G. (2002) *Why distance matters: Effects on Cooperation, Persuasion and Deception*. In Proceedings of Computer Supported Cooperative Work (CSCW), New Orleans, Louisiana, 226-235.
Breschi, S., Lenzi, C., Lissoni, F. and Vezzulli, A. (2010) The geography of knowledge spillovers: the role of inventors' mobility across firms and in space, in: R. Boschma and R. Martin (eds.) The Handbook of Evolutionary Economic Geography, Cheltenham: Edward Elgar, pp. 353- 369.
Burt, R.S. (1992) *Structural Holes: The Social Structure of Competition*. Cambridge, Massachusetts: Harvard University Press.
Cassi, L. and 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. and 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. and 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.
Ejermo O. and Karlsson C. (2006) “Interregional inventor networks as studied by patent coinventorships”, *Research Policy* 35: 412-430.
European Commission (2007) Green Paper. The European research area: new perspectives. COM(2007) 161.
European Commission (2010) A vision for strengthening world-class research infrastructures in the ERA. Report of the Expert Group on Research Infrastructures.
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.
Feldman, M.P., Aharonson B.S. and Baum J.A.C. (2005) “The importance of proximity and location” *Rotman School of Management, University of Toronto*, unpublished manuscript.
Félix, B. (2007) “Biotechnology in Europe. Patents and R&D investments”, Statistics in Focus, Science and Technology 100/2007, Eurostat.
Florida, R. (2002) *The Rise of the Creative Class, and how it's transforming work, leisure, community and everyday life*. New York, Basic Books.
Gallié E.P. and Guichard, R. (2005) “Do collaboratories mean the end of face-to-face interactions? An evidence from the ISEE project”, *Economics of Innovation and New Technologies* 14: 517–532.
Gertler, M.S. and Levite, Y.M. (2005) “Local nodes in global networks: the geography of knowledge flows in biotechnology innovation” *Industry and Innovation* 12(4): 487-507.
Granovetter, M.S. (1973) The Strength of Weak Ties *The American Journal of Sociology* 78(6): 1360-1380.

Grossman, G.M. and Helpman, E. (1991) *Innovation and Growth in the Global Economy*. Cambridge, Massachusetts: MIT Press.
Grossman, G. and Helpman, E. (1994) "Endogenous innovation in the theory of growth" *Journal of Economic Perspectives* 8: 23-44.
Harris, R.D.F. and Tzavalis, E. (1999) "Inference for unit roots in dynamic panels where the time dimension is fixed" *Journal of Econometrics*, 91(2): 201-226.
Hoekman, J.; Frenken, K. and van Oort, F. (2009) "The geography of collaborative knowledge production in Europe" *The Annals of Regional Science*, 43(3): 721-738.
Im, K.S.; Pesaran, M.H. and Shin, Y. (2003) "Testing for unit roots in heterogeneous panels" *Journal of Econometrics*, 115(1): 53-74.
Jaffe, A.B. (1989) Real effects of academic research, *American Economic Review* 79(5): 957-970.
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. and 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. and 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. and 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
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*, Kluwer Academic publishers, Dordrecht.
Miguélez, E. and Gómez-Miguélez, I.G. (2011) Singling out individual inventors from patent data, IREA Working Paper 2011/05.
Oettl, A. and Agrawal, A. (2008) "International labor mobility and knowledge flow externalities", *Journal of International Business Studies*, vol. 39(8): 1242-1260.
Olsson, O. (2000) "Knowledge as a Set in Idea Space: An Epistemological View on Growth" *Journal of Economic Growth* 5(3): 253-275.
Paier, M. and Scherngell, T. (2011) "Determinants of Collaboration in European R&D Networks: Empirical Evidence from a Discrete Choice Model" *Industry & Innovation* 18(1): 89-104.
Picci, L. (2010) "The Internationalization of Inventive Activity: A Gravity Model using Patent Data", *Research Policy*, 39(8): 1070-1081.

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.
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.
Rychen, F. and Zimmermann, J.B. (2008) Clusters in the Global Knowledge-based Economy: Knowledge Gatekeepers and Temporary Proximity. *Regional Studies* 42(6): 767-776.
Saxenian, A. (2006) *The New Argonauts: Regional Advantage in a Global Economy*, 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. and 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. and Fox-Kean, M. (2005) "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment", *American Economic Review* 95(1): 450-460.
Torre A. (2008) "On the Role Played by Temporary Geographical Proximity in Knowledge Transmission" *Regional Studies* 42(6): 869-889.
van Beuzekom, B. and Arundel, A. (2009) *OECD Biotechnology Statistics 2009*, OECD, Paris.

Figure 1. Shares of multi-inventor and multi-region biotech patents

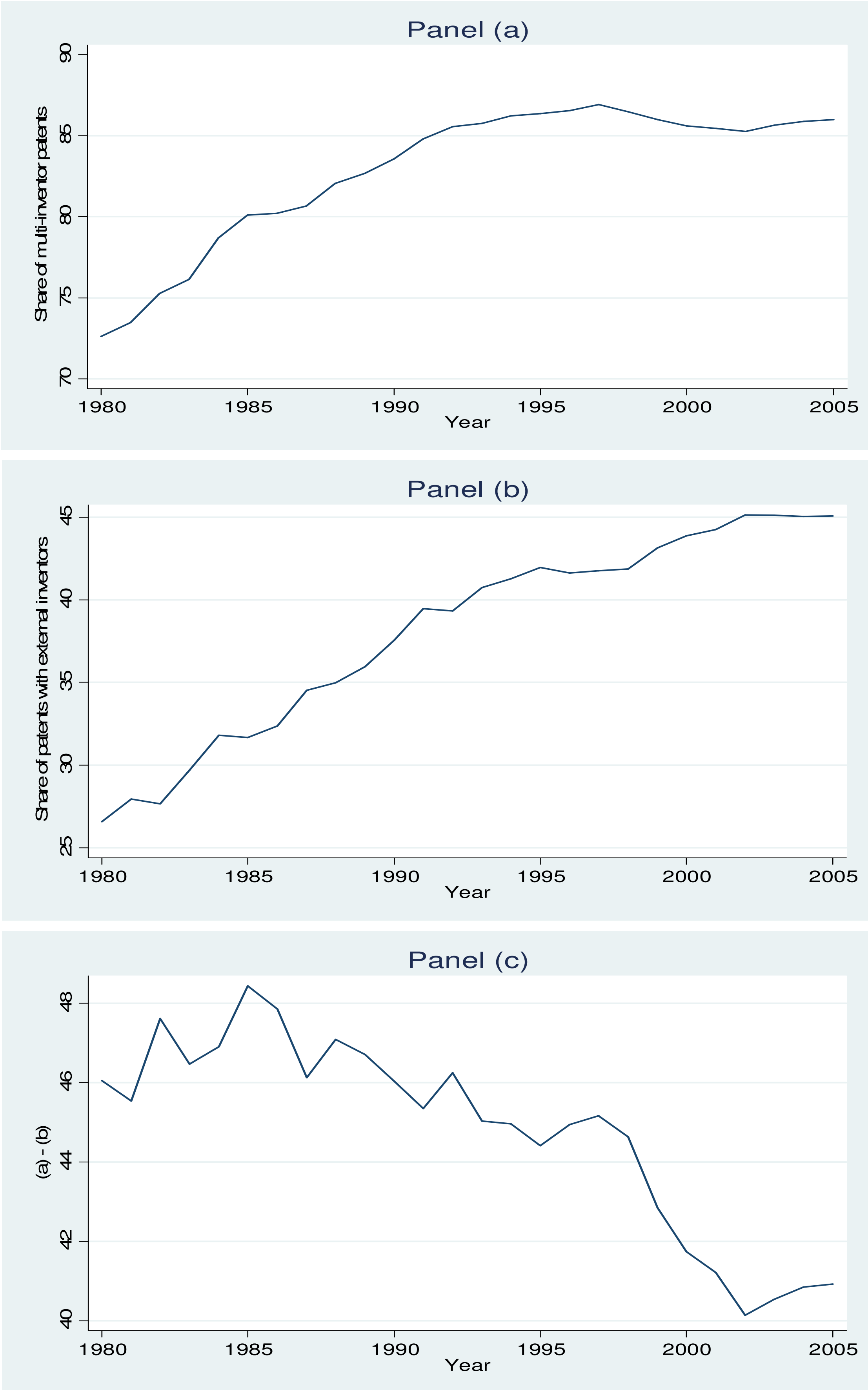

Figure 2. Share of mobile inventors across NUTS2 regions

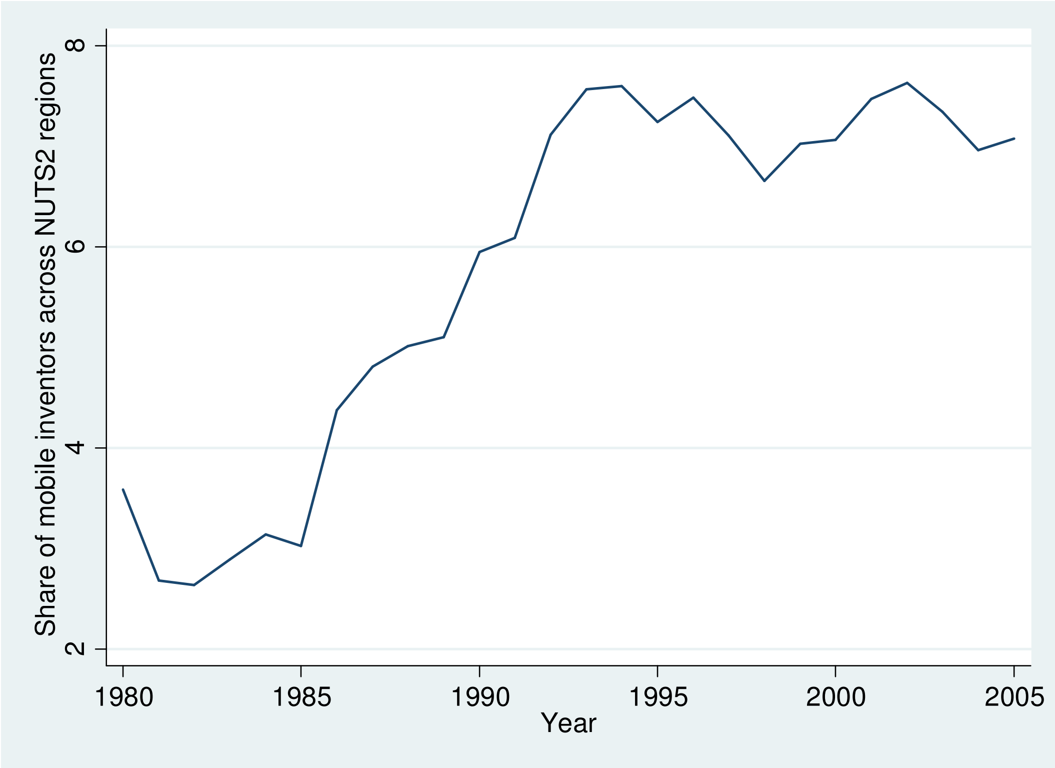

Table 1. Summary figures

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)	7,377
Observed ties under analysis (NUTS2)	4,903
Final set of inventors under study (links across NUTS3)	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

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***

Table 3. Summary statistics, unbalanced panel, linkages across NUTS3

	# obser.	Mean	Co. Var.	Min	Max
Cross-regional co-patents	33,007	0.22	1.86	0	1
Social proximity	33,007	0.09	1.96	0	0.50
Cognitive proximity	33,007	0.38	1.04	0	1
Institutional proximity	33,007	0.50	0.99	0	1
Organizational proximity	33,007	0.29	1.58	0	1
Previous co-location	33,007	0.06	3.90	0	1
Average productivity	33,007	1.15	1.32	0	21.55
Abs. diff. productivity	33,007	1.49	1.64	0	41.17
Average centrality	33,007	6.74	1.36	0	128
Abs. diff. centrality	33,007	9.12	1.65	0	236

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	0.18	0.46	0.66	1						
5. Organizational proximity	0.15	0.60	0.45	0.62	1					
6. Previous co-location	0.02	0.12	0.07	0.10	0.14	1				
7. Average productivity	0.06	0.23	0.25	0.30	0.26	0.04	1			
8. Abs. diff. productivity	0.02	0.12	0.09	0.10	0.11	0.01	0.88	1		
9. Average centrality	0.12	0.34	0.26	0.30	0.28	0.02	0.58	0.50	1	
10. Abs. diff. centrality	0.07	0.17	0.10	0.13	0.14	0.00	0.49	0.51	0.91	1

Table 5. Fixed-effects conditional logit estimations. Linkages across NUTS3

	(i)	(ii)	(iii)
Social proximity	1.290***	1.178***	1.291***
	(0.175)	(0.214)	(0.175)
Cognitive proximity	0.684***	0.691***	0.685***
	(0.079)	(0.095)	(0.079)
Institutional proximity	0.197***	0.235***	0.196***
	(0.067)	(0.079)	(0.067)
Organizational proximity	0.267***	0.405***	0.267***
	(0.073)	(0.091)	(0.073)
Previous co-location	0.279**	0.299**	0.257*
	(0.120)	(0.120)	(0.146)
Average productivity	0.080*	0.077*	0.080*
	(0.046)	(0.046)	(0.046)
Abs. diff. productivity	-0.084***	-0.082***	-0.084***
	(0.025)	(0.025)	(0.025)
Average centrality	0.181***	0.180***	0.181***
	(0.010)	(0.010)	(0.010)
Abs. diff. centrality	-0.030***	-0.030***	-0.030***
	(0.005)	(0.005)	(0.005)
Geographic*Social		0.044	
		(0.049)	
Geographic*Cognitive		-0.003	
		(0.029)	
Geographic*Institutional		-0.018	
		(0.025)	
Geographic*Organizational		-0.044*	
		(0.025)	
Previous co-location*Geographic			0.006
			(0.024)
Observations	33,007	33,007	33,007
Pairs of inventors	7,377	7,377	7,377
McFadden’s Adjusted R-squared	0.169	0.170	0.169
Log-likelihood	-8143.750	-8134.999	-8143.715

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

Table 6. Fixed-effects conditional logit estimations. Linkages across NUTS2

	(i)	(ii)	(iii)
Social proximity	1.119***	1.426***	1.122***
	(0.225)	(0.283)	(0.225)
Cognitive proximity	0.737***	0.609***	0.736***
	(0.094)	(0.112)	(0.094)
Institutional proximity	0.199**	0.263***	0.199**
	(0.079)	(0.094)	(0.079)
Organizational proximity	0.403***	0.503***	0.401***
	(0.095)	(0.122)	(0.095)
Previous co-location	0.879***	0.873***	1.093***
	(0.205)	(0.205)	(0.258)
Average productivity	0.075	0.080	0.075
	(0.059)	(0.059)	(0.059)
Abs. diff. productivity	-0.088***	-0.091***	-0.088***
	(0.033)	(0.033)	(0.033)
Average centrality	0.177***	0.176***	0.177***
	(0.013)	(0.013)	(0.013)
Abs. diff. centrality	-0.017***	-0.017***	-0.017**
	(0.007)	(0.007)	(0.007)
Geographic*Social		-0.237*	
		(0.134)	
Geographic*Cognitive		0.151**	
		(0.077)	
Geographic*Institutional		-0.076	
		(0.062)	
Geographic*Organizational		-0.069	
		(0.066)	
Previous co-location*Geographic			-0.140
			(0.095)
Observations	21,685	21,685	21,685
Pairs of inventors	4,903	4,903	4,903
McFadden's Adjusted R-squared	0.174	-5332.245	-5337.362
Log-likelihood	-5338.346	0.175	0.174

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

Table 7. Fixed-effects conditional logit estimations: first and subsequent collaborations

	(i)	(ii)
Social proximity	1.290***	0.669*
	(0.175)	(0.378)
Cognitive proximity	0.684***	0.137
	(0.079)	(0.385)
Institutional proximity	0.197***	-0.719
	(0.067)	(0.492)
Organizational proximity	0.267***	0.113
	(0.073)	(0.541)
Previous co-location	0.279**	0.209
	(0.120)	(0.197)
Average productivity	0.080*	0.020
	(0.046)	(0.056)
Abs. diff. productivity	-0.084***	-0.048
	(0.025)	(0.040)
Average centrality	0.181***	-0.063***
	(0.010)	(0.008)
Abs. diff. centrality	-0.030***	0.025***
	(0.005)	(0.007)
Observations	33,007	3,846
Pairs of inventors	7,377	762
McFadden's R-squared	0.169	0.042
Log-likelihood	-8143.750	-1269.651

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