Journal of Economic Geography Advance Access published March 4, 2009

Journal of Economic Geography (2009) pp. 1–30 doi:10.1093/jeg/lbp008

Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows

Stefano Breschi[*,‡] *and Francesco Lissoni*[*,†]

Abstract

This article illustrates the contribution of mobile inventors and networks of inventors to the diffusion of knowledge across firms and within cities or states. It is based upon an original data set on US inventors' patent applications at the European Patent Office, in the fields of drugs, biotechnology and organic chemistry. The study combines the methodology originally proposed by Jaffe et al. (1993, *Quarterly Journal of Economics*, 108: 577–598) with tools from social network analysis, in order to evaluate extent of the localization of knowledge flows, as measured by patent citations. After controlling for inventors' mobility and for the resulting co-invention network, the residual effect of spatial proximity on knowledge diffusion is found to be greatly reduced. We argue that the most fundamental reason why geography matters in constraining the diffusion of knowledge is that mobile researchers are not likely to relocate in space, so that their co-invention network is also localized. In the light of these results, we revisit common interpretations of localized knowledge flows as externalities.

Keywords: knowledge diffusion, localized spillovers, mobility, networks
JEL classifications: O33, R12, Z13
Date submitted: 20 July 2008 **Date accepted:** 30 December 2008

1. Introduction

Research on the geography of innovation in the past 20 years or so has largely revolved around two basic questions. Does knowledge tend to flow more easily across spatially proximate agents than across agents localized far apart in space? And, to what extent localized knowledge flows may be also characterized as spillovers? Both questions and the answers given to them have been extremely influential in shaping the approach of policy makers to regional development. On one hand, evidence of spatially bounded knowledge flows has contributed to shift the attention away from traditional policies targeted to sustain less-advanced regions and towards a new set of enabling policies aimed at nurturing the birth and growth of new high-tech clusters (Audretsch, 1998; Feldman, 1999; Breschi and Malerba, 2005). On the other hand, suggestions that such localized knowledge flows take the special characteristic of knowledge externalities (spillovers) have prompted the intervention of regional and local policy makers in support of R&D activities at local universities and public research centres, thought to be the most important sources of knowledge spillovers for local firms (Jaffe, 1989).

*KITES-CESPRI—Università Commerciale L. Bocconi, Milan, Italy.
†DIMI, Department of Mechanical and Industrial Engineering, Università di Brescia, Brescia, Italy.
‡Corresponding author: Stefano Breschi, Kites-Cespri, Department of Economics, Università L. Bocconi, Via Sarfatti 25, 20136 Milan, Italy. *email* <stefano.breschi@unibocconi.it>

A large part of the qualitative literature on agglomerations such as high-tech clusters and industrial districts has been devoted to systematically explore and tell apart the mechanisms through which knowledge flows locally (for a survey: Breschi and Lissoni, 2001b). Such studies have contributed to social scientists' awareness that codified media are less effective means of spreading 'tacit' knowledge than personal, face-to-face contacts. And yet, scientists and technologists nurture a variety of personal contacts: Which, among them, are so sensitive to distance that an increasing spatial separation will impair the transmission of knowledge? And which do contribute most to the dissemination of knowledge immediately relevant for the innovation process? Two possible candidates for such knowledge mediation role stand out. First, many authors have stressed the role of non-market-based social ties, such as those one can have with former fellow students and teachers, or with fellow affiliates to technical associations. According to this argument, geography matters in driving the diffusion of knowledge, because proximity facilitates the development of social relationships, by exposing agents to relevant social events and by reducing the costs of forming and sustaining such ties (Agrawal et al., 2006).

Second, it has also been suggested that spatial proximity and co-location may facilitate the transfer of knowledge through contractual and market-based channels, such as the labour market (Almeida and Kogut, 1999) or licensing and formal collaboration networks (Mowery and Ziedonis, 2004). From this perspective, geography operates either because knowledge-intensive workers, such as R&D employees and consultants, move from one organization to the other but are relatively immobile in space, or because economic transactions related to knowledge assets require frequent interactions and trust-building activities, and are therefore more effectively conducted if agents are co-located.

Whereas explanations based on the idea of non-market based social interactions place special emphasis on interpreting knowledge flows as spillovers (pure externalities), the same does not hold true for market-based channels. To the extent that a knowledge exchange is the main object of an economic transaction, the contracting parties will make an effort to price it correctly, thus limiting the extent of pure spillovers. Pecuniary externalities may also arise, but these cannot be readily described as originating a local public good.

Very few quantitative studies have so far attempted to sort out these two channels of knowledge transmission (for a survey, see Breschi and Lissoni, 2001a). The purpose of this work is to propose a methodology that may be of help in this direction, one that combines the use of patent citations, as originally proposed by Jaffe et al. (1993), with the geographical analysis of inventors' mobility and co-invention networks. The results we obtain by applying such methodology to selected technological fields in the United States suggest that inventors' activity across firms (mobility, consulting, contract research, etc.) explain a large part of localized knowledge flows and reduce the logical room left for explanations based upon non-market based social ties.

The article is organized as follows: Section 2 discusses how patent citations data have been used in past research to infer the existence of localized knowledge spillovers and the controversial nature of results based upon them. In Section 3, we show how the wealth of information contained in patent data can actually be exploited to track the existence of co-invention linkages between citing and cited patents and to filter out different channels of knowledge transmission. Section 4 provides a description of the data set used in this article, and Section 5 reports the main empirical findings. Section 6 concludes.

2. Using patent citations to track localized knowledge spillovers

In what has become a classical reference, Jaffe et al. (1993) (hereafter, JTH) argue that knowledge flows can be measured by patent citations and that the latter provide material for a natural experiment on geography.[1] The experiment starts with the selection of a sample of cited patents, that is, patents that have originated a measurable knowledge flow. For each cited patent, all subsequent citing patents are collected: they represent the destination of the knowledge flow originating from the cited ones. Since JTH are interested in capturing knowledge spillovers, pairs of citing-cited patents that are assigned to the same company are excluded from the analysis precisely because such ‘self-citations’ cannot represent an externality. Each citing-cited patent pair is coded to record whether or not the two patents are spatially co-located in the same geographic area; the resulting fraction of spatially co-located pairs is then computed. Yet, this fraction is unlikely to tell anything significant about the actual localization of knowledge flows, as it depends on the existing spatial distribution of patenting activity. As long as the production of patents in a given industry is highly concentrated in a few regions, one would expect a large fraction of citing-cited patents to be also spatially co-located. In order to control for this, JTH propose to build a patent sample that mimics the distribution of patent production and thereby may be used as a benchmark against which one can evaluate the fraction of spatially co-located citing-cited patents. For each citing patent, a control patent is selected such that it does not cite the originating (cited) patent, but has the same technological class and approximately the same application date of the citing one. The JTH experiment thus consists in comparing the fraction of citing-cited patents that match geographically with the corresponding fraction of control-cited patents. More formally, this exercise amounts to comparing the probability of a geographic match between two patents, conditional on there being a citation link, with the probability of a match not conditioned on the existence of a citation link (JTH, 1993: 581). The evidence reported by JTH shows that citing patents are up to three times more likely than control patents to come from the same state as the cited ones, and up to six times from the same metropolitan area.

2.1. Knowledge spillovers and social networks

JTH results have been widely cited and interpreted as a proof that knowledge externalities (spillovers) are both important and to a large extent local in nature.[2] To the extent that patent citations represent uncompensated flows of knowledge, this interpretation seems to be warranted. Yet, in order for this interpretation to be valid, one has to provide an answer to an ancillary question: what are the mechanisms through which knowledge is transmitted from the origin (the cited patent) to the destination (the citing patent)? While JTH are relatively agnostic with respect to this

1 For a discussion of the merits of patent citations as diffusion indicators, see Jaffe and Trajtenberg (2002) and Breschi and Lissoni (2004).

2 Elsewhere, we have provided a detailed critical review of the large body of literature that has developed around the concept of localized knowledge spillovers (Breschi and Lissoni, 2001a,b).

question, a view has come to prevail in the literature that identifies in the network of informal social relationships the communication channel supporting the localization of knowledge spillovers [for a discussion of this view and a list of references, see Breschi and Lissoni (2001a,b)]. Examples such as those of people getting good ideas at conferences or technical meetings abound, as well as references to Marshallian metaphors of 'knowledge in the air'. Both the examples and the metaphors come in handy to stress that patents represent a piece of codified information, but that the knowledge stock they draw from is to a large extent tacit. In order to use that knowledge productively, one needs to have access to and interact with the individuals that have generated and still master it, that is, the patent inventors. Here is where geography and social ties come in. On one hand, the transfer of tacit knowledge requires direct communication among individuals, such as frequent face-to-face contacts and interpersonal meetings. On the other hand, the effectiveness with which such direct communication takes place sharply declines with spatial distance. In fact, co-location of knowledge workers within the same geographical area not only exposes them to many potential social acquaintances, but it also reduces the costs of forming and maintaining social links.

The popularity that these arguments have achieved in the literature is probably due to their simple and intuitive nature and to the fact that they fit nicely with the empirical evidence presented by JTH. A few studies, however, have expressed a few doubts that challenge such received wisdom. By and large, two types of critiques have been raised. First, it has been argued that although knowledge may well flow locally, that does not necessarily imply any externality. Whether or not spillovers are occurring depends on the type of relationship between knowledge-exchanging agents: besides social ties of an informal type, one should also consider knowledge transactions in which the two parts are careful not to transmit any knowledge for free. A second line of critique is much more specific and concerns the JTH methodology.

2.2. The controversial nature of localized knowledge flows

In the quantitative literature on the geography of innovation following JTH, the role of social ties as carriers of localized knowledge spillovers has been more often assumed than demonstrated. Scientists, technologists, students, and other knowledge workers in a given area may get in touch through a variety of mechanisms, some of which bring along the possibility of gratuitous knowledge exchanges (as in communications and discussions at scientific and technical societies' meetings), while others promote market-related knowledge transactions (as in contract research or scientists' and engineers' recruitment over the labour market). While the former type of knowledge flows falls squarely in the category of pure externalities, the latter does not, although it may comprise them to the extent that some of the knowledge transacted is not entirely paid for.

Actually, this possibility was already adumbrated by JTH in their original paper. While discussing the validity of drawing inferences about knowledge spillovers from patent citations, they note that

> [..], we can classify the links that might exist between two inventions into one of three groups: spillovers accompanied by citations, citations that occur where there was no spillover, and spillovers that occur without generating a citation. Our experiment uses the first set, but clearly

> the other two are non-empty. [..] A deeper problem is created by ‘‘real’’ citations that are not spillovers. [. . .] To the extent that the flow of rents between the parties is governed by a complete contract, there could conceivably be no externality running from the original inventor to the contractor. If we now add to this hypothetical contract the assumption that such contracted development is relatively likely to be localized, we have the potential for the observed localisation of citations to be greater than the true localisation of knowledge spillovers (JTH, 1993: 583–584, italics added)

This crucial passage has remained virtually unnoticed in the subsequent literature, perhaps because the authors themselves rushed to observe that ‘in general the contract between the two parties will be quite incomplete, making it more likely than not that the citing organization could capture some rents from the original invention and hence benefit from at least a partial spillover’ and that ‘though there are a number of considerations, all difficult to quantify, we believe that on balance it is reasonable to draw inferences about spillovers from citations’ (JTH, 1993: 583). Yet, the point remains. If one has to conclude that the observed localization of patent citations captures the true localization of knowledge spillovers, one should embark upon the difficult task to classify the types of relationships between knowledge-exchanging agents and to weigh the importance of cases that cannot be readily interpreted as spillovers [this point is well discussed by Moen (2000); see also Griliches (1992)]. To date, in the literature, only few attempts in this direction have been made.

In two important contributions, Almeida and Kogut (1999) and Zucker et al. (1998) argue that the reason why knowledge flows are spatially bounded has to be found in the working of labour market for scientists and engineers, rather than in the direct communication taking place over more informal social networks. Knowledge is transferred by individuals that move from one organization to the other, but do not relocate in space.[3] However, Almeida and Kogut (1999) make their point by simply showing that JTH-like evidence is more readily found in regions with a more lively labour market for scientists and engineers, while Zucker et al. (1998) base their claim only on a number of case studies.

The existence of localized markets for technologies has also been reported as a potential explanation for JTH results (Lamoreaux and Sokoloff, 1999). Mowery and Ziedonis (2004) replicate the JTH methodology by distinguishing between citations that are linked through a licensing contract and those that are not. Their results show that market-based citations are far more spatially localized than non-market (i.e. spillover) citations. Although, their sample is quite limited and refers only to university-licensed patents, their insight is a relevant one. Precisely because tacit knowledge necessary to develop the licensed invention is embodied in individuals, access to it requires providing knowledgeable workers with the appropriate incentives to engage in collaborative development. This objective may be achieved by offering those individuals some kind of compensation, such as equity stakes or research contracts (Jensen and Thursby, 2001). This in turn requires that frequent interactions are established, which are most

3 There may be several reasons why individuals moving across organizations do not relocate in space. In the case of academic scientists, it has been argued that spatial immobility is likely to derive from the fact that they prefer to enter into contractual agreements with firms located within commuting distance of their university, where they tend to retain affiliation for reasons of reputation. More generally, various social and economic forces (e.g. family and friendships) may induce people to remain within the same geographical area.

effectively conducted at a short spatial distance. Once again, one might observe a localized knowledge flow, which is not a spillover given that its carrier is a standard market transaction.

The quantitative exercise we carry out in this work differ from the literature cited here, because it aims explicitly at identifying the role of a key category of knowledge workers, the inventors, and it addresses chiefly the results of market-based mechanisms, such as job mobility, consultancy and research collaboration.

2.3. Is there a flaw in the JTH methodology?

Some of the scepticism towards the notion of localized knowledge spillovers has been directed not to substantive interpretation issues, but to the empirical methodology used to prove their importance. In particular, Thompson and Fox-Kean (2005) argue that the evidence reported by JTH is spurious, due to a fatal flaw in the experiment. Their point is a rather technical one, but no re-examination of the JTH experiment can escape discussing it.

The critique relates to the sample selection process used by JTH to match citing and control patents. JTH select control patents within the same USPTO 3-digit technological fields of the citing patents. However, the degree of within-class heterogeneity at this level of classification may be very large, resulting in a sample of control patents bearing little resemblance with the citing ones. As a consequence, JTH results might be simply explained by the fact that control and citing patents come from different industries. Thompson and Fox-Kean show that, by matching citing and control patents at a finer level than the three-digit main technology field, replicas of the JTH experiment fail to provide any evidence of localization effects.[4]

3. Tracing the links between inventions

As argued earlier, understanding the role of geography in mediating knowledge flows and the nature of such flows requires tracing the links between inventions and the channels through which knowledge is transmitted. Our methodology to trace such links is based on the use of information on inventors listed in patent documents. The basic idea may be illustrated with reference to Figure 1. The figure illustrates a snapshot of a hypothetical network of inventors at some time *t*, resulting from formal collaborations (co-inventorship). Each node represents an individual inventor, whereas ties linking pairs of nodes indicate that the corresponding inventors have been part of the same team on one or more patents (i.e. they have co-invented at least one patented invention).[5] Dashed lines enclosing nodes, finally, correspond to the different

4 In their reply to Thompson and Fox-Kean, Henderson et al. (2005) observe that too tight a technological match has the undesirable side effect of limiting too much the size of the patent samples. They also observe that knowledge spillovers captured by patents necessarily imply an act of invention, since pure imitation cannot produce any patent (for which novelty is required). Therefore, one must allow for some technological distance between patents: if spillovers generate inventions, these will be original enough to be classified not in the same main and secondary more-than-three-digit classes of the originating patent.

5 In the language of graph theory, patent data may be represented as a three-mode (or affiliation) network. The first mode is given by the patent documents, the second mode is given by the inventors and the third mode is given by the patent applicants. The corresponding graph is called a tripartite graph as lines connect only vertices from one mode to vertices from another mode and there are no connections inside

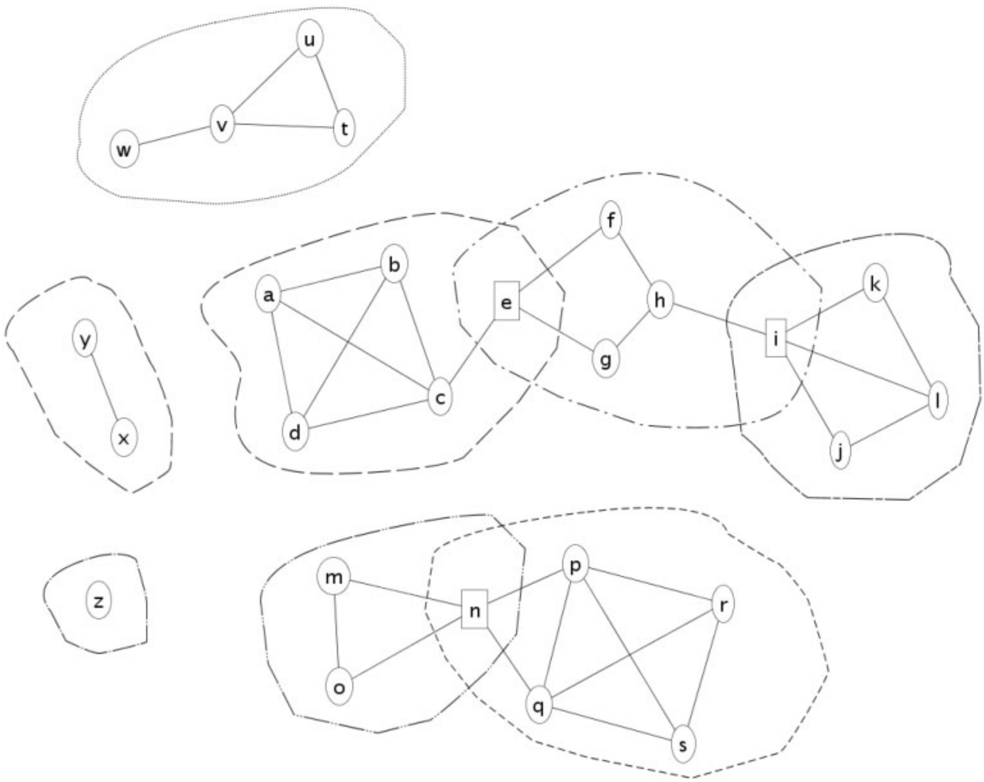

Figure 1. A snapshot of the network of inventors at time t.

organizations (i.e. patent applicants) for which the inventors have produced their patents. We may refer to the overall network of inventors as the co-invention network to stress the fact that it captures a specific set of social ties derived from formal collaborations.

A few important points have to be noted. First, there are a few inventors who appear to have been working for more than one organization (the squared nodes enclosed in overlapping lines). This happens when the same individual has been listed as inventor in patents assigned to different patent applicants. We define them as *mobile inventors* to point out that such individuals have moved across companies offering their services to different organizations. Mobile inventors may be either employees that switch from one employer to another (that is, mobile workers in a strict sense), or consultants or

modes. For example, each inventor is linked (i.e. affiliated) to one or more patents, but there are no direct connections among inventors. Three-mode networks can be transformed into 'ordinary' networks, where vertices are only from one mode. In particular, one can transform the patent-inventor ties into inventor-inventor ties. Two inventors are in relation (i.e. in the corresponding graph there exists an undirected line) if they took part in at least one common patent. The line value tells the number of patents in which both of them have been part of the same team. The network reported in the figure is exactly the one-mode projection onto just inventors of the tripartite graph of patents, inventors and applicants (Wasserman and Faust, 1994).

even academic scientists that offer their services to different companies.[6] All these types of mobile inventors transfer their knowledge at a price, not for free. As a consequence, they can originate an externality only if part of the knowledge they transfer comes from their former employees or customers, the latter not being entirely compensated for their loss (Fosfuri and Ronde, 2004). Producing evidence on this goes beyond the scope of our work. However, it is important to stress that knowledge travelling with mobile inventors looks much less like a 'spillover' than the exchanges occurring between attendants at conferences, alumni's meetings and social gatherings of all sorts: while the latter can be safely presumed to fall in the sphere of gratuities, the recruitment of inventors as employees or consultants does not.

The second point to be noted is that inventors working for different patent applicants may be connected by a path, namely an unordered sequence of lines and nodes. For example, inventor *a* is connected to inventor *k*, despite the fact that the two individuals are enclosed within different circles (i.e. they have been working for different organizations). It is quite important to remark that the connection between *a* and *k* is made possible by a chain of co-invention ties (i.e. *a* co-invented with *c*, *c* co-invented with *e*, etc.), which necessarily passes through a mobile inventor. In other terms, mobile inventors act as bridges across teams of inventors working for different organizations, making communication among them potentially possible. For any connected pair of inventors *a* and *k*, we define the distance $d(a,k)$ between the two inventors as the shortest path connecting them. In the example above, $d(a,k)$ is equal to six.

The third point to observe is that, besides connected pairs of inventors, there are also pairs of inventors (such as *a* and *w*), who are not connected, as there is not any finite sequence of lines and nodes between them.[7] The absence of a co-invention chain does not imply that these inventors are socially unrelated, as it may be the case that they are connected through some other type of informal linkages, such as kinship or friendship, which are not captured in patent data (a typical example may be the relationship between two relatives in the same line of business or between two alumni of the same university).

It may also be that inventors in our networks are both connected through a co-invention chain *and* other types of social ties. However, it is reasonable to assume that the shorter the co-invention chain between two inventors, the more intense their professional relationship, and the higher the probability that knowledge exchanges are entirely or partially mediated by contractual arrangements (such as an employment contract or a consultancy contract). On the contrary, when it comes to explain knowledge flows between two inventors at a long or infinite distance in the co-invention network, we are at a loss, in the sense that we do not have information on how the two inventors may be related, so we may not exclude social ties of any kind, and are left free to imagine all sorts of spillovers from one to another.

As a consequence, a question can be asked, whose answer may cast light on the nature of localized knowledge flows: what remains of localization effects once we take into account and control for mobility and the specific ties captured by the co-invention network? Or, to put it more bluntly, what room is left to informal social ties in

6 On academic inventors as consultants, see Thursby et al. (2009), who report data for the United States. For European evidence, see Lissoni et al. (2008).

7 For pairs of disconnected inventors the distance $d(a,w)$ is equal to infinite. At the other extreme, the distance of an inventor from himself, i.e. $d(a,a)$, is equal to 0.

explaining localized knowledge spillovers, once mobility of inventors and co-invention ties have been isolated and factored out of the explanation?

In order to answer this question, we need to measure the links between patents on the basis of links between their inventors. To this purpose, let us imagine to take a patent P filed at time $t+1$ and citing a patent M filed at some time $s \leq t$. Let us also assume that the two patents have been assigned to two different organizations (i.e. we do not consider self-citations). The possible links between the two patents can be classified into one of three groups, according to the relationship between their inventors:

1. Among the inventors listed in the citing patent, there is at least one individual that has been also listed among the inventors of the cited patent. The two patents are thus linked by an inventor that has moved from the assignee of the cited to the assignee of the citing patent, either as an employee or consultant. We will refer to this type of linkage between citing and cited patents as 'mobility';
2. Among the inventors listed in the citing patent, there is at least one inventor connected to at least one of the inventors listed in the citing patent. In this case, we may define the distance between the two patents as the shortest path between the two closest inventors from the two teams of inventors.[8] We will refer to this type of linkage as co-invention connectedness to denote the fact that patents are linked by ties among individuals established through chains of formal collaboration;
3. Among the inventors listed in the citing patent, there is no inventor connected to any of the inventors listed in the citing patent. For this reason, we will label this third group as non-connected patents or patents at infinite co-invention distance (which does not exclude social connections of other kind).

In the next section, we describe the data we used to implement the methodology just discussed.

4. Data and descriptive statistics

Our patent data come from the EP-CESPRI data set, which contains complete information on all patent applications to the European Patent Office (EPO) from 1978 to 2002. We have selected from this data set all patent applications reporting at least one US inventor and applied for by a US organization in one of the following technological fields: Organic Chemistry, Pharmaceuticals and Biotechnology.[9]

8 The following example may help the reader to understand this case. Let us assume that citing patent P applied for at time $t+1$ has been produced by a team of inventors comprising individuals [*a, c* and *z*]. Individuals *a* and *c* were already present in the network of inventors at time *t*, whereas individual *z* was not, as he appeared listed in P for the first time. On the other hand, let us assume that the cited patent M was produced by a team of inventors comprising individuals [*i* and *k*]. If we apply the definitions given in the text, we observe that inventors [*a* and *c*] and inventors [*i* and *k*] were already connected in the network of inventors at time *t*, as reported in Figure 1. Moreover, we also observe that the two closest individuals, among the connected ones, are *c* and *i*, whose distance is equal to four. Therefore, we may conclude that citing and cited patents are connected to each other through the inventors that have generated them and that the distance between the two patents is equal to four.

9 Technologies have been identified on the basis of the primary International Patent Classification (IPC) codes. Pharmaceuticals correspond to IPC code A61K, Organic Chemistry to IPC code C07 (with the exclusion of C07B), and Biotechnology to IPC codes from C12M to C12S.

The sample thus generated consists of 66,349 patent applications, which have been applied for by 5820 different organizations (mostly business companies) and have been signed by 63,188 distinct inventors. Since the validity of our analysis crucially depends on the correct identification of individual inventors, we have carried out a thorough work of cleaning and standardizing inventor's names and addresses, in order to tell homonymous inventors apart from mobile ones.[10] Using the sample just described, we have proceeded in two ways. On one hand, we have used data on inventors and teams of inventors to derive the corresponding co-invention network as described in the previous section. On the other hand, we have replicated the original experiment of JTH by selecting a sample of citing, cited and control patents and accounting for their network properties.

Two aspects of our methodology need a more detailed discussion. First, the choice of restricting attention to only three technological fields; second, the use of EPO data rather than USPTO data.

As far as the first issue is concerned, our choice is first and foremost explained by time and computational limits. Although machine-assisted, our procedure for cleaning names produces results that need to be checked and corrected manually, when necessary. In addition, the need to construct, handle and store very large data matrices (to build and measure the co-invention network) puts a constraint on the maximum number of nodes that current software programmes, and the computers we have access to can process. Notice that our sample includes over 60,000 inventors, which implies that the matrix recording the presence or absence of a co-invention tie between any two of them consists of over 3.6 billions of cells.

Given the necessity to select only a limited number of technological classes, we went for those corresponding to industries best known for relying on patents as appropriation tools, so that patents can be considered a good indicator of innovation activity.[11] In addition, the three selected fields exhibit a high degree of technical inter-relatedness: patents in one field tend to cite patents in one of the other two fields and do not cite patents in other fields; and many inventors active in one field also sign patents in one or both of the others. Overall, our selected fields cover 13.5% of all EPO patent applications (13.0% of applications by US companies).[12]

10 The following procedure has been adopted. Inventors have been identified first by assigning a unique code to all inventors with the same name, surname, and address; and then by running Massacrator©, a programme that assigns scores to any pair of inventors with the same name and surname but different address, on the basis of information suggesting the two inventors may be the same person (such as the technological class of their patents, the identity of their patent applicants, their location in space and the identity of their co-inventors). The threshold score for accepting identity between two individuals with same name and surname but different address has been set to a relatively high value to ensure a rather conservative approach in merging individuals. For details, see Lissoni et al. (2006). In a similar way, patent applicants' names have been standardized and subsidiaries have been consolidated using various editions of Who Owns Whom.

11 On the importance of patents for Pharmaceutical, Biotech and Chemical industry see Cohen et al. (2000) and Levin et al. (1987).

12 Biotech, Chemical and Pharmaceutical patents cover respectively 2.7%, 6.2% and 3.7% of all EPO patent applications, respectively (figures for applications by US companies are 4.1%, 6.7% and 5.3% respectively). Calculations over the entire EP-CESPRI database suggest that Organic Chemistry patents receive 44% citations from Pharmaceutical patents and 13% from Biotechnologies, plus 17% from Basic Chemistry; Pharmaceuticals receive citations mainly from Biotechnology (29%) and Organic Chemistry (34%); Biotechnology receive 42% of its citations from Pharmaceuticals and 18% from Organic Chemistry, plus 19% from Control Technologies. As for forward citations, Organic Chemistry sends to Pharmaceuticals 34% of its cross-field citations and 18% to Biotechnology, plus 12% and 16%

Regarding the second issue, USPTO patent data would have appeared at first sight a more natural choice for a study, such as ours, that addresses the geography of innovation in the United States. Yet, the use of EPO data can be defended on a methodological ground for various reasons. In the first place, by using patent applications of US companies at the EPO we are dropping from the analysis low-quality patents, which are not worth extending to Europe (through a costly procedure such as the EPO one). Second, EPO patent records provide more information than USPTO ones, such as the inventors' home address (street and zipcode).[13] These data are crucially important for a correct identification of individual inventors and to avoid problems related to homonymy (see footnote 10). Finally, the EPO examination procedure is such that patent citations are less numerous, more directly relevant to prior-art, and much less 'noisy' than those one can find on USPTO documents. In short, EPO citations are of better quality than USPTO ones [for a more detailed discussion along these lines, see Breschi and Lissoni (2004) and Michel and Bettels (2001)].[14] At the same time, the use of EPO data may impose some limitations on our results; we come back to them in Section 5.1.

4.1. Network of inventors and inventors' mobility

Table 1 illustrates the extent of the phenomenon of inventors' mobility by reporting the distribution of all inventors in our sample according to the number of different organizations for which they have signed patents. In addition to this, the table reports information on the extent of geographical mobility of inventors.

We observe that the vast majority (73%) of all inventors have signed patents for just one organization throughout their inventive career. Most mobile inventors signed patents for just two different assignees (17% of all inventors; 64% of mobile inventors);

to Macromolecular Chemistry and Basic Chemistry; 24% and 35% of citations from Biotechnology go to Organic Chemistry and Pharmaceuticals, respectively (plus 19% to Control Technologies), while 44% and 27% citations from Pharmaceuticals go respectively to Organic Chemistry and Pharmaceuticals. No other fields besides those mentioned here contribute more or receive more than 10% citations to the fields selected for this study. Data are available on request. An alternative way to measure interrelatedness between technological fields, which returns similar results, is based upon the examination of secondary classifications of EPO patents (Breschi et al., 2003).

13 More precisely, USPTO data report the home street address and zipcode of inventors only for non-assigned patents, i.e., patents which are owned by inventors themselves, and not companies. This means that most patents (and the most valuable ones) do not carry such information. For example, the NBER database used by JTH provides information on the street address for only 11.5% of the observations in the 'Inventors' file (7.7% on zipcodes). Our own calculations on raw UPSTO data suggest a figure for street address of around 18.5%, which we obtain by some pre-cleaning of inventors' names. On the contrary, almost all EPO patents report the inventors' home address, with a few exceptions due to some companies' policy to provide only their headquarters' address.

14 Note that a bias might be induced by the fact that EPO patents report only very few, highly relevant citations compared to USPTO ones. This would lead us to place in the control sample some patents that, according to USPTO data, would belong to the citing sample. If this was a major problem, however, our results on the geographical location of the two samples would sensibly differ from the original JTH ones, which is not the case (see below, Section 5). In addition, it is also worth pointing out that USPTO patents (if not examined through the PCT procedure) contain a very large number of preposterous citations that have nothing to do with true knowledge flows, having been inserted by assignees' lawyers eager to show compliance with the 'duty of candour' rule or to mislead the examiners (on this point, see again Breschi and Lissoni, 2004). It this holds, it could be that by using USPTO data one ends up doing the opposite mistake, that of including in the citing sample several patents that ought to be used as controls.

Table 1. Mobility of inventors

Number of ≠ assignees	Number of inventors (% of all inv.)	Percentage of inventors active in >1 MSA	Percentage of inventors active in >1 State
1	46,458 (73.5)	2.3	2
>1	16,730 (26.5)	28.4	25.7
Of which			
2	10,645 (16.9)	22.8	20.5
3	3,679 (5.8)	34.1	31.5
4	1,439 (2.3)	40.3	38
5	562 (0.9)	46.4	43.6
6	392 (0.6)	46.7	43.9
>10	13 (0.0)	53.8	46.2
Total	63,188 (100)	9.2	8.3

very few of them signed patents for more than five assignees (0.6% of all inventors; 4% of mobile inventors). If inventors' mobility across organizations looks relatively limited, mobility in geographical space appears to be even more so. Although the propensity to change location is higher for inventors that move across assignees, only 28.4% of all mobile inventors (9.2% of all inventors) have been active in more than one Metropolitan Statistical Area (MSA). Similar figures apply to inter-state mobility.[15] As one would expect, the extent of geographical mobility increases with the number of different assignees for which the inventors have signed patents. Whereas only 23% of all inventors that have been working for just two assignees have changed MSA, the same fraction goes up to 40% for inventors that have signed patents for four different assignees (see Table 1, third column).

As far as the network of inventors is concerned, we have built three basic sets of matrices on the basis of information contained in patent documents: a matrix whose generic cell [*i*,*j*] reports the presence or absence of a co-invention tie between inventors *i* and *j*, a matrix reporting the presence or absence of a path connecting inventors *i* and *j*, and a matrix reporting the length (i.e. geodesic distance) of the shortest path connecting the two inventors. In order to build these matrices, we have adopted the following rule. Each matrix capturing the co-invention network at time t includes only co-invention ties formed between t and $t-5$. Thus, for example, for two inventors that have co-authored a patent document applied for at time $t-6$ and that have never worked together afterwards, the matrix recording the presence or absence of a co-invention tie at time $t-1$ will report a 1 in the corresponding cell, whereas the same matrix at time t will report a 0. The three sets of matrices are thus updated each year, by dropping older than 5-year ties and adding up new ones. The fundamental reason for applying this rule is that the effectiveness with which a co-invention tie transmits knowledge between inventors is likely to decay with the age of the link. While it is reasonable to assume that

15 As far as inter-state mobility, this is somehow overestimated. Separate tabulations, not reported here, break down inter-state mobility by pairs of states and reveal that a large fraction of such mobility occurs within MSAs, such as Philadelphia and New York, whose boundaries span across two neighbouring states. More generally, the evidence shows that inter-state mobility is by and large limited to five states (California, Pennsylvania, New York, New Jersey and Massachusetts) or to movements between these states and the others.

Table 2. Network of inventors

	1991	1995	1999
Largest connected component			
Number of inventors	3664	7038	14077
As percentage of all inventors	23.4	33.3	46.0
Average distance	16.3	13.6	12.0
Second largest component			
Number of inventors	287	250	277
As percentage of all inventors	2.8	1.2	0.9

recently established ties convey knowledge and information, one may doubt that a co-invention tie established long ago and never renewed still performs the same function today. In the absence of any clear indication from the literature, we believe that the adopted time-window captures reasonably well the period of time over which co-invention ties may act as channels of knowledge transmission.

Table 2 reports some descriptive statistics on the resulting network of inventors for three different years. In particular, the table reports the absolute and percentage size of the so-called largest connected component. This may be defined as the largest subset of mutually reachable inventors, which is the largest subset wherein all inventors are connected to each other. We note from the table that the size of the largest component increases over time, reflecting presumably both a higher propensity to collaborate and an increase in the number of patent applications. However, the most relevant aspect emerging from the data is that the largest connected component comprises a very large fraction of all inventors, particularly as compared to the second most numerous one. In 1999, for example, around 46% of all inventors in the selected technological fields were connected to each other through some path. The evidence thus suggests that the extent of connectivity among inventors is quite large. In addition, the table reports the average length of paths connecting individuals in the largest connected component. Quite interestingly, the average separation among inventors tends to decrease as the network gets more connected. In 1991, the average distance (i.e. the shortest path) among two inventors randomly drawn from the largest component was about 16, while the corresponding figure in 1999 was just equal to 12.[16] Thus, over time, more and more inventors get connected, while at the same time their average distance declines.

4.2. Sample selection of citing, cited and control patents

Following as closely as possible the methodology developed by JTH, we have selected for this study three cohorts of originating patents, consisting respectively of the 1991, 1992 and 1993 patent applications that received at least one subsequent citation by the end of 1999.[17] For each cohort of originating patents, we dropped all patents that

16 The combination of a large connected component and short average distance among vertices is typical of so-called ‘small-world’ networks, which have been recently investigated in the physics literature (Newman, 2001).

17 The priority year has been used to date patents.

either received citations only from foreign organizations, or whose applicant was an US organization, but did not report any US inventor. It must be pointed out that the choice of excluding citations from foreign companies implies that our study does not investigate the extent of inter-national localization of patent citations, focusing only on intra-national localization effects. This choice has been mainly dictated by data constraints, as the inclusion of citations coming from foreign organizations would have implied the construction of the worldwide network of inventors. Following JTH, we also removed all observations in which citing and cited patents have been assigned to the same organization (i.e. self-citations). This leaves us with a total of 2014 originating (cited) and 3012 citing patents. However, the total number of citing-cited patent pairs is 3700, since some originating patents have been cited by more than one citing patent. In order to create a sample of control patents mimicking the spatial distribution of patent production, we followed once again the methodology used by JTH. For each citing patent, we randomly extracted another patent having the following characteristics: the control patent does not cite the same originating patent, it is classified in the same IPC four-digit class as the citing patent, and it has a priority date as close as possible to the priority date of the citing patent. This procedure yielded a sample consisting of 3419 control patents for a total of 3700 control-cited patent pairs.[18] Overall, our sample consists of 3700 citing-cited patent pairs, which are matched to 3700 control-cited patent pairs.

For each patent pair (i.e. citing-cited and control-cited), we then computed two fundamental variables. In the first place, we defined the extent of geographical co-localization between patents by using the inventors' addresses as reported in patent documents. In particular, we coded two patents as matching geographically if at least one inventor from one patent was located in the same MSA (or State, depending on the level of analysis) of at least one inventor from the other patent.[19] In the second place, for each patent pair, we also measured the type of linkage (if any) connecting them. Following the methodology described in Section 3 above, we classified all patent pairs into three (mutually exclusive) groups according to the linkages connecting the individual inventors that have produced them:

- mobility: patent pairs signed (among others) by the same inventor; given that we are excluding company self-citations, this means that the two patents are linked by an inventor who has moved from the cited to the citing organization or at least worked for both;

18 Note that a patent may cite more than one prior art patent. For each citing patent, we have randomly drawn as many control patents as the number of times it cites previous patents. For example, for a patent citing two previous patents, we have randomly extracted two control patents. This explains why the number of controls (3419) is larger than the numbr of citing patents (3012).

19 A major problem in measuring the geographic dispersion of patents and patent citations relates to the way patents are assigned to locations. Our approach in this respect is slightly different from JTH. In the case of multiple inventors, they assigned unequivocally each patent to the MSA/state in which pluralities of inventors resided, with ties assigned arbitrarily, and then compared the assigned locations across patent pairs. We do not assign a patent to a given location, but simply argue that two patents match geographically if they have in common at least one location among those reported in the inventors' addresses.

Table 3. Sample of citing, cited and control patents

Patent pairs	Number	Type of links (% of all pairs)		
		Connected		Not connected
		Mobility (social distance = 0)	Co-invention network (social distance > 0)	(social distance → ∞)
Citing-cited	3700	5	25.9	69.1
Control-cited	3700	0.2	19.6	80.2

Table 4. Distance between connected patent pairs (co-invention network)

Distance	Citing-cited	Control-cited
1	5.2	1.1
2–5	17.1	13.4
6–10	46.3	32.6
10–20	28.3	40.5
>20	3	4.3
Average	9.2	10.5
Median	8.5	9.5
Std deviation	10.5	11.7

- connected (*via* co-invention network): patent pairs for which there is at least one inventor from the citing patent that is connected to an inventor from the cited patent through a finite path in the co-invention network; as argued above (see Section 3), for these patent pairs, we may define the distance between the two patents as the shortest path between the two closest inventors from the two teams.[20]
- not connected: patent pairs whose respective teams of inventors are not connected to each other in the co-invention network; of course, we cannot exclude the existence of other types of social connections, possibly of a more informal type.

Table 3 reports the distribution of patent pairs in our sample by type of links connecting them. The table shows that citing-cited patent pairs are significantly more likely to be linked either by mobile inventors (5% versus 0.2%) or via the co-invention network (25.9% versus 19.6%) than control-cited pairs. Moreover, for patent pairs linked via the co-invention network, we have also computed the distance between them. This is reported in Table 4.

While 5.2% of all citing-cited pairs linked via the co-invention network are at a distance 1 (i.e. at least two inventors from their respective teams had previously

20 Formally, the distance between the two patents is given by *min d(i,j)* for each *(i,j)* where *i* labels inventors of the citing patent and *j* labels inventors of the cited patent.

co-invented together), the corresponding figure for control-cited pairs is only 1.1%. Similarly, 17.1% of all citing-cited pairs are at a distance comprised between two and five as compared to just 13.4% for control-cited pairs. More generally, we observe that the mean distance for connected patent pairs is significantly lower for citing-cited patents (9.2) than for control-cited ones (10.5). Overall this evidence suggests two important things. First, mobile inventors are a fundamental carrier of the knowledge flows captured by patent citations. Second, inventors responsible for citing and cited patents are linked through relatively short chains of acquaintances in the co-invention network.

5. Empirical findings

As a first step in our empirical analysis, we simply replicate the original JTH experiment by calculating the fraction of all citing-cited and control-cited patent pairs that match geographically. Results are reported in the first rows of Table 5 at the MSA and State levels, respectively. The second column of the table reports the percentage of citing patents that are co-located with the cited ones; the third column reports the same percentage for the control sample; the fourth column reports the value of the z statistic for the difference between the two proportions, and (in brackets) the result of a one-tail test on the null hypothesis that $pc = pnc$ (where pc is the co-location probability of citing-cited patents, and pnc the co-location probability of control-cited patents). Finally, the last column reports the observed odds ratio.[21]

Our results confirm, by and large, the original JTH findings. Citing patents are significantly more likely than control patents to come from the same geographical area of cited patents. At the MSA level, 17% of citations are localized, compared to 11% of controls; likewise, at the state level, 21% of citations are localized, compared to 14% of controls. Looking at the odds ratio, citing patents are 73% more likely than

21 The z test is defined as:

$$z = (\bar{p}_c - \bar{p}_{nc})/\sqrt{\bar{p}(1-\bar{p})(1/n_c + 1/n_{nc})}$$

where $\bar{p}_c$ and $\bar{p}_{nc}$ are the sample proportion estimates for the citing-cited and the control-cited patents that match geographically; n_c and n_{nc} are the size of the citing-cited and control-cited samples (in our case: $n_c = n_{nc}$); and

$$\bar{p} = (m_c + m_{nc})/(n_c + n_{nc})$$

where m_c and m_{nc} are, respectively, the number of citing-cited patents and the number of control-cited patents that match geographically. The one-tail test of our interest calculates the probability attached to values higher than z from a standard normal distribution. The odds of co-location for citing-cited patent pairs are defined as:

$$O_c = p_c/(1 - p_c).$$

Let O_{nc} be the corresponding odds of co-location for control-cited patent pairs. The odds ratio is defined as $\mathrm{OR} = O_c/O_{nc}$. OR values higher than one suggest a higher co-location probability of citing-cited patents, as opposed to control-cited ones. JTH used a t-distributed statistic to test the geographical localisation of patent citations. However, since the test is meant to compare two population proportions a z-test seems to us to be more appropriate.

Table 5. Geographical matching at the MSA and state levels

	Number of observations	Citing cited	Control cited	z-test ($P > z$)	Odds ratio
MSA level					
All patent pairs (JTH experiment)	3700	17.4	10.8	8.1 (0.00)	1.73
All pairs except those linked by mobility (social distance = 0)	3512	13.5	10.8	3.5 (0.00)	1.29
All pairs except those connected at social distance ≤ 5	3217	11.8	9.9	2.5 (0.01)	1.22
Only not connected pairs	2299	11.7	9.4	2.5 (0.01)	1.21
State level					
All patent pairs (JTH experiment)	3700	20.8	14	7.8 (0.00)	1.62
All pairs except those linked by mobility (social distance = 0)	3512	17.2	13.8	3.9 (0.00)	1.3
All pairs except those connected at social distance ≤ 5	3217	15.5	13	3.0 (0.00)	1.24
Only not connected pairs	2299	15.1	12.5	2.5 (0.01)	1.24

control ones to come from the same MSA of the cited patents; the same value at the state level is 62%.[22]

As argued earlier, however, patent pairs differ according to the type of social links between patents, which in turn implies different means of knowledge transmission. The relevant question therefore is whether and to what extent the same localization pattern found for all patent pairs holds true once we look at patent pairs linked by specific types of relationships. In order to address this crucial question within the framework of the JTH experiment we remove from the whole sample all patent pairs (i.e. both citing-cited and control-cited) linked by a specific mechanism and re-calculate the rate of geographical matching. In the second rows of Table 5, we report the percentage of citing and control patents that match geographically with the cited ones, after removing from the sample all patent pairs linked by mobile inventors, respectively at the MSA and State levels. The proportion of citing patents that match geographically with the cited ones drops dramatically, while the proportion of co-localized controls remains virtually unchanged. At the MSA level, around 13% of citations are now localized, compared to 11% for controls. Likewise, at the state level, 17% of citations are localized, compared to almost 14% for controls. Although the difference between the two proportions is still statistically significant, citing patents are now only 29% more likely to come from the same MSA as the cited patents. What these results

22 Although our results on the geographical localization of citations are qualitatively comparable to those reported by JTH, the absolute values of geographical matching are different. In particular, our geographic matching rates are much higher than JTH's: nearly double at the state level, more than double at the MSA level. There may be several explanations for these differences. First, we use EPO patent applications, whereas they use patents granted at the USPTO. Second, we focus on three specific technology fields, while they use a sample of patents from any technological class. Third, the citation window we use (9 years, from 1991 to 1999) is shorter than the one used by JTH (14 years, from 1975 to 1989). To the extent that localization effects tend to fade with time, this aspect is likely to explain a great deal of the differences with the original experiment. Finally, our period of observation refers to the 1990s, whereas the period of observation in JTH refers to the 1980s. Once again, to the extent that the production of patents has become more spatially concentrated over time, this is also likely to explain the differences mentioned above.

suggest is that mobility of inventors across organizations account for a very large fraction of the observed localization of patent citations. Once we remove patents linked through mobile inventors evidence of localization of patent citations considerably weakens.

To move further this intuition, we remove, in addition to patent pairs linked by mobile inventors, also the patent pairs linked via the co-invention network at a distance lower than six. The reason for setting this distance threshold is that there are several insights from the literature suggesting that the effectiveness of co-invention ties in transmitting knowledge sharply declines with the length of the path connecting two agents.[23] From Table 4, we observe that localization effects drop even further and that the drop is remarkably larger for citations than for controls. As a consequence, the difference between the two proportions narrows down further: at the MSA level, 11.8% of citations are localized as compared to 9.9% of controls. As suggested by the odds ratio, citing patents are now only 22% more likely than controls to come from the same MSA as the cited patents.

As a final step, we remove from the sample all pairs connected either by mobility or by co-invention ties at any distance. Quite interestingly, the difference between the proportions of geographically matching patents does not change remarkably compared to the previous case. This confirms that networks of inventors convey knowledge only through relatively short co-invention chains. However, it also confirms that the bulk of the observed localization of patent citations is due to mobile inventors and to the core of the co-invention network they create. Any social tie formed outside this network explains a much smaller fraction of localization effects.

Besides comparing the co-location proportion of citing and control patents (with cited ones), we can regress the probability of citing-cited patents' co-location as a function of social distance, controlling for the co-location on control-cited patent pairs.

Notice that the evidence we have produced so far suggests that the effect of social distance on the citing-cited co-location probability is non-linear: we expect patents at zero distance (i.e. connected by mobile inventors) or at less than six steps, in fact, to be disproportionately more likely to be co-located than others connected at higher distances; in turn, the latter may not be much more likely to be co-located than the unconnected ones (infinite distance).

In order to account for such non-linearity we experiment with three different measures of social distance. The first measure consists of a set of dummy variables, each representing a different distance, from zero (as when a mobile inventor is present in both patents) to 20, with distances higher than 20 as the reference case. The second measure combines a dummy variable for patents at zero distance with a continuous

23 This insight has been confirmed both in theoretical and empirical studies. On one hand, simulation models have shown that the ability for one agent to benefit from knowledge possessed by others rapidly declines with the geodesic distance separating the origin and the destination of knowledge flows (Cowan and Jonard, 2004). On the other hand, empirical studies examining the probability to observe a citation link between two patents have shown that this is a function of how close the corresponding inventors are in the network of co-invention (Singh, 2005). In particular, indirect ties matter only to the extent that the path connecting individuals is relatively short (Breschi and Lissoni, 2004). In our case, we ended up selecting out the links at less than 6° of separation in a stepwise fashion; that is, we noticed that by excluding from our sample the patent pairs at less than six degrees of distance the co-location rates of citing patent fell noticeably, but that by excluding also the patent pairs at more than five degrees of distance it did not change much any longer.

Table 6. Summary statistics for regression variables

	Min	Median	Max	Mean	Std Dev
Citing-cited co-location (state level)	0	0	1	0.208	0.406
Control-cited co-location (state level)	0	0	1	0.14	0.347
Citing-cited co-location (MSA level)	0	0	1	0.174	0.379
Control-cited co-location (MSA level)	0	0	1	0.11	0.311
Dummies for social distance					
Citing-cited social distance = 0	0	0	1	0.05	0.217
Citing-cited social distance = 1	0	0	1	0.014	0.115
Citing-cited social distance = 2	0	0	1	0.01	0.101
Citing-cited social distance = 3	0	0	1	0.006	0.077
Citing-cited social distance = 4	0	0	1	0.012	0.107
Citing-cited social distance = 5	0	0	1	0.016	0.127
Citing-cited social distance = 6	0	0	1	0.023	0.15
(More dummies for social distance, not reported)[a]					
Citing-cited inverted social distance	0	0	1	0.097	0.246
Control-cited inverted social distance	0	0	1	0.028	0.087
Mobile inventor between citing-cited (dummy)	0	0	1	0.049	0.217
Mobile inventor between control-cited (dummy)	0	0	1	0.002	0.04
Citing-cited social proximity	0	0	1	0.057	0.22
Control-cited social proximity	0	0	1	0.004	0.045

[a]Citing-cited social distance dummies from 7 to 20 and control-cited social distance dummies from 1 to 20.

measure of ‘inverted social distance’, ranging from 0 to 1. The latter takes the following values:

$$\begin{cases} 1/\text{social distance} & \text{if social distance} > 0 \\ 1 & \text{if social distance} = 0 \end{cases}$$

so that the dummy variable for patents at zero distance allows to distinguish the latter from those at distance one.

The third measure is also a continuous measure inversely proportional to distance, namely *social proximity* $= e^{-\text{social distance}}$, which also ranges from 0 to 1 and takes value one for patents at zero distance and lower than one for all the others (zero for not connected patents). The exponential transformation is meant to stress even further the non-linearity effect discussed above. We apply such distance measures also to control-cited patent pairs and control for them in all our regressions.

Table 6 reports summary statistics for both the dependent and the independent variables in our regression. Table 7 reports the results of logit regressions of the probability of co-location of citing-cited patents either at the MSA or at the State level, each with a different measure of social distance. Regressions are numbered from (1) to (6) for ease of reference. Estimated odds ratios at 95% confidence levels are calculated to quantify the impact of a unit variation of the independent variable; these are most easily interpreted for regressions (1) and (4), where all independent variables are dummies.

Regressions (1) and (4) confirm that the citation probability falls quite sharply with social distance, starting from very high levels at low social distances. The estimated odds ratios (reported in squared brackets) indicate that citing-cited patents at social

Table 7. Citing patents' probability of co-location with cited ones, at the MSA and state level: Logit regressions

	MSA			State		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	−2.06***	−2.12***	−2.01***	−1.80***	−1.84***	−1.75***
	(0.07)	(0.060)	(0.055)	(0.060)	(0.060)	(0.0512)
Control-cited co-location	0.67***	0.68***	0.72***	0.73***	0.76***	0.7707***
	(0.142)	(0.138)	(0.135)	(0.118)	(0.114)	(0.113)
	[1.5–2.5]	[1.5–2.6]	[1.6–2.7]	[1.6–2.6]	[1.7–2.7]	[1.7–2.7]
Citing-cited social distance = 0	4.40***			4.08***		
	(0.271)			(0.270)		
	[48.1–139.3]			[34.9–100.5]		
Citing-cited social distance = 1	2.11***			1.77***		
	(0.295)			(0.294)		
	[4.6–14.7]			[3.3–10.4]		
Citing-cited social distance = 2	1.56***			1.48***		
	(0.35)			(0.344)		
	[2.4–9.4]			[2.2–8.6]		
Citing-cited social distance = 3	1.63***			1.38***		
	(0.449)			(0.447)		
	[2.1–12.3]			[1.7–9.6]		
Citing-cited social distance = 4	1.21***			1.20***		
	(0.354)			(0.338)		
	[1.7–6.7]			[1.7–6.5]		
Citing-cited social distance = 5	0.91***			0.59*		
	(0.308)			(0.306)		
	[1.4–4.5]			[0.99–3.3]		
Citing-cited social distance = 6	1.11***			0.98***		
	(0.260)			(0.250)		
	[1.8–5.0]			[1.6–4.3]		
More social distance dummies[a]						
Citing-cited inverted soc. distance		2.35***			2.06***	
		(0.269)			(0.265)	
		[6.2–17.8]			[4.7–13.4]	
Control-cited inverted soc. distance		0.63			0.456	
		(0.557)			(0.527)	
		[0.6–5.6]			[0.6–4.4]	
Mobile inventor btw. citing-cited		2.07***			2.04***	
		(0.366)			(0.364)	
		[3.9–16.3]			[3.8–15.7]	
Mobile inventor btw. control-cited		−2.31			−2.34	
		(1.727)			(1.639)	
		[0.003–2.9]			[0.01–2.4]	
Citing-cited social proximity			4.53***			4.17***
			(0.276)			(0.270)
			[53.9–159.2]			[38.3–110.4]
Control-cited social proximity			−0.02			−0.49
			(1.018)			(1.034)
			[0.1–7.2]			[0.08–4.7]
L-ratio test	693.99***	625.15***	602.43***	630.81***	559.70***	541.76***
Aikake information criterion	2807.7	2800.5	2819.3	3241.8	3236.9	3250.9
Bayesian information criterion	3081.2	2837.8	2844.1	3515.3	3274.2	3275.8
Nagelkerke R-square	0.284	0.258	0.249	0.245	0.219	0.213
Obs	3700	3700	3700	3700	3700	3700

Note: Std error in brackets; 95% confidence odds ratios in square brackets.[a]Citing-cited soc. distance dummies from 7 to 20 and control-cited from 1 to 20 (none significant). Significant at ***99%, **95%, *90%.

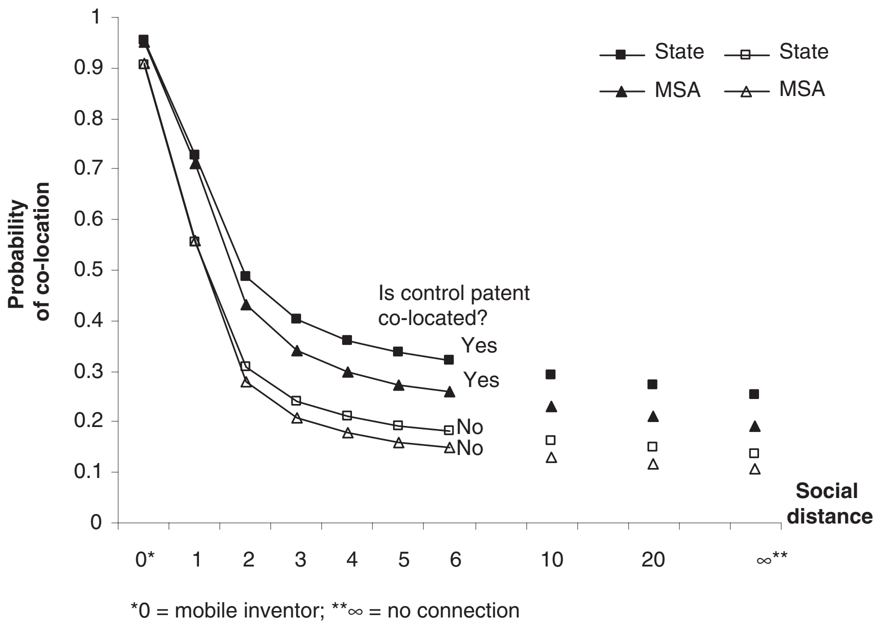

Figure 2. Predicted probability of co-location of citing-cited patents (MSA and state level) as function of social distance.
Note: From regressions (2) and (5) in Table 7 [*Control-cited co-location* and *Control-cited social proximity* both set equal to zero (median value).

distance 0 are more than 48 times more likely to be co-located in the same MSA and more than 34 times more likely to be co-located in the same state than patent pairs separated by twenty or more steps. On the other hand, patent pairs at social distance 1 are *only* five times more likely to be co-located in the same MSA and three times more likely to be co-located in the same state than the baseline; while for social distances higher than six the co-location odds are not significantly higher than one.

Regressions (2), (3), (5) and (6) confirm these results, with the second measure of distance (inverted distance, plus dummy for zero distance) doing a slightly better job according to both Aikake and Schwarz (Bayesian) information criteria. In all regressions, the effect of the co-location of control-cited patent pairs has a positive and highly significant value.

Figure 2 reports estimates for the probability of co-location of cited and citing patents based upon regressions (2) and (5), respectively, as a function of social distance.[24] From each regression, two lines are calculated, one that reports the

24 Although distance enters the two regressions as an inverted measure (plus the dummy for patents at zero distance) the horizontal axis of the figure reports social distance as conventionally measured. This means that social distance >1 in the graph corresponds to inverted social distance <1 and zero-distance dummy equal to zero in the regression; distance equal to one in the graph correspond to inverted distance =1 and zero-distance dummy also equal to zero in the regression; finally, distance equal to zero in the graph (mobile inventor) correspond to inverted distance =1 and zero-distance dummy equal to one in the regression.

probability of citing-cited co-location, when the control patents are also co-located with the citing ones, the other for the opposite case. All four lines in the figure are calculated for median values of all the other independent variables.

As expected, the probability of co-location at the State level is higher than that at the MSA level. The co-location of control-cited patents, which is a proxy of the geographical concentration of industry, also increases the probability of co-location of citing and cited patents. The figure also illustrates nicely the sharp decline of such probability with the increase of social distance: patents at infinite distance are only slightly less likely to be co-located than those at distance 10 or 20, while the latter have a co-location probability which is less than half of those at distance 1. Patents at zero social distance have a probability of co-location close to one.

5.1. Discussion and limitations

Our results show that mobility of inventors across organizations accounts for the largest fraction of the observed localization of patent citations. The fundamental reason why we observe geographical localization of patent citations is the low propensity of a special category of knowledge workers and providers of knowledge-intensive services (the inventors) to relocate in space.

As for word-of-mouth exchanges of tacit knowledge, a diffusion factor often cited by the literature, they seem to take place chiefly among co-inventors, and, in particular, among those connected through short social chains. These short chains are very much likely to arise through and be contained within formal collaboration activities between inventors, or common employment contracts with the same company.

How does geography enter this picture? The explanation is a subtle one and involves once again mobile inventors. As observed when commenting Figure 1, teams of inventors from different organizations are linked to each other by inventors that move across organizations and act as bridges across them. To the extent that such moving individuals do not relocate in space and remain within the same region, the resulting co-invention network will be also spatially localized. In other words, the most closely connected inventors will also tend to be spatially close to each other and this explains why knowledge flows measured by patent citations exhibit localization effects. This explanation does not square entirely with the conventional wisdom on the importance of informal social ties in diffusing tacit knowledge. In the absence of localized movements of inventors and the ensuing creation of closed networks of formally collaborating agents (co-inventors), informal linkages (such as those we leave out from our co-invention networks) are likely to explain only a minor fraction of the observed phenomenon.

To give further support to this interpretation, we compute the average geographical distance between all pairs of inventors located at a given social distance in the co-invention network. Results are illustrated in Figure 3. The x-axis reports the distance among inventors in the co-invention network, whereas the y-axis reports the average geographical distance. As the network is constantly updated, we report for illustrative purposes only the results calculated for 1999, which comprises co-invention ties formed between 1995 and 1999. The figure shows that the average spatial distance among inventors sharply increases with the distance in the co-invention network. In other words, inventors connected by short paths in the co-invention network tend also to be located at a relatively short geographical distance from each other. Thus, for example,

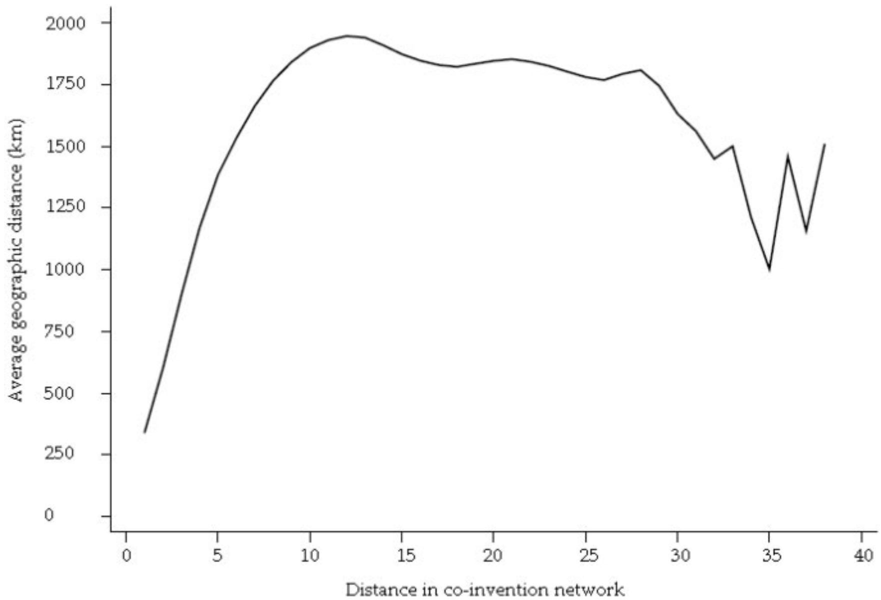

Figure 3. Average geographical distance in the co-invention network (1999).

pairs of inventors at one step in the co-invention network are on average 340 kilometres away from each other. As a consequence, the likelihood that they are geographically co-located within the same state or MSA is quite high. On the contrary, inventors separated by six steps in the co-invention network are on average located at a spatial distance around 1500 kilometres. The likelihood that they come from same state or MSA is therefore correspondingly much lower.

However, we must point out that the importance attached by our results to inventors' mobility and connectedness may suffer of a selection bias effect, due to our choice of using EPO patents (over US inventions) instead of USPTO ones. As explained in Section 4, EPO patents do not include low-quality patents, which are not worth extending to Europe. It is possible that low-quality patents come from relatively isolated inventors, who may nonetheless cite the originating patents when co-located with the latter's inventors, thanks to contacts different from the co-invention network (e.g. conferences and local press). As a result, by using EPO patents, we would end up underestimating the co-location proportion of citing-cited patents and also over-estimating the importance of the co-invention network in explaining the geographical co-location of citing-cited patents, relatively to the control sample.[25]

To make sure this problem is real, one needs first to test that connectedness via the co-inventor network and patent quality are negatively correlated, and then that the proportion of high-quality patents, conditional to co-location, is higher in the citing sample than in the control sample. Both things seem to receive some support with our data.

25 We thank two anonymous referees for drawing our attention to this point.

We can measure the quality of a patent by the number of citations it receives, controlling for the patent's age (the longer a patent has been around, the higher the probability to receive citations). We find that, on average, connected patents in our two samples (citing and control) have received more citations than non-connected ones, that is, they are of higher quality, although we cannot find a strong relationship between quality and social distance.[26] We also find that, considering only patents co-located with cited ones, the proportion of connected (i.e. high quality) patents in the citing sample is higher than in the control sample (48% versus 28%).

It is difficult to say how much including low-quality patents (that is, using USPTO rather than EPO data) could alter our results. We leave to others the (daunting) task of building a dataset similar to ours with USPTO data and check.[27] However, our results certainly hold for the most important knowledge flows, namely those that contribute to high quality, highly valuable inventions.

5.2. Is there a flaw in JTH methodology? A reprise

In Section 3.3 above, we mentioned how the original JTH methodology has been recently criticized for providing merely spurious evidence of localization effects (Thompson and Fox-Kean, 2005). The flaw would consist in sampling controls from too broad technological fields, which bear little resemblance to the citing patents' technological contents and results in relative over-dispersion in the physical space. We believe this critique to be only half-right: what is flawed is not the methodology itself, but the interpretation of the results it produces. In order to elaborate on this point, we first replicate the revised methodology proposed by Thompson and Fox-Kean (2005) and then assess to what extent the outcome is compatible with our findings on the role of mobility and co-invention networks.

Following the revisions proposed by the two authors, we re-sample the set of control patents in our experiment to force a narrower technological matching between citing and control patents. In particular, we extract three further samples of control patents: first we require the matching to occur not at the four-digit IPC primary classification level, but at the 12-digit one; then, we impose the additional condition that patents also match at the level of four-digit secondary classes; finally, we restrict even this additional condition, and impose matching also at the level of 12-digit secondary classes. We then recalculate the percentage of co-located citing-cited and control-cited patents for all the new samples. Results are reported in Table 8. For a better comparison, the table also reports results based on the original experiment (from Table 5).

Our results seem to confirm the critique raised by Thompson and Fox-Kean. For control-matching at the 12-digit primary classification (second row), the share of co-located control patents increases sensibly, while remains virtually unchanged for citing ones. As a consequence, the difference between the two proportions narrows down, as suggested also by the odds ratio. Restricting even further the sampling criteria

26 Considering jointly the patents in the citing and control samples, we calculate that, from their date of filing to 2005, connected patents have received 2.58 citations each, while non connected ones have received only 2.21 (the difference is significant at 99%). In terms of citations per year, the figures are 0.29 and 0.23, respectively, whose difference is still significant. Patents connected through short chains, however, are not more cited than those connected over longer chains.

27 See Trajtenberg et al. (2006) for a data building effort that may help in this direction.

Table 8. Geographical matching: the effect of narrowing technological fields

	Number of obs.	Citing cited	Control cited	z-test ($P > z$)	Odds ratio
MSA level					
Matching at primary 4-digit (JTH experiment)	3700	17.4	10.8	8.1 (0.0)	1.73
Matching at primary 12-digit	3168	17.4	13.7	4.1 (0.0)	1.33
Matching at primary 12-digit and secondary 4-digit	2458	17.5	15	2.4 (0.0)	1.2
Matching at primary 12-digit and secondary 12-digit	1882	16.7	16.6	0.1 (0.9)	1
State level					
Matching at primary 4-digit (JTH experiment)	3700	20.8	14	7.8 (0.0)	1.62
Matching at primary 12-digit	3168	20.8	18.4	2.4 (0.0)	1.16
Matching at primary 12-digit and secondary 4-digit	2458	21	18.6	2.1 (0.0)	1.17
Matching at primary 12-digit and secondary 12-digit	1882	20.2	20.1	0.0 (1.0)	1

leads to a further decrease of the difference between the two proportions, down to the point where no significant differences in the geographical matching of citing and control patents can be found (fourth row).

Should we interpret this evidence as proof that knowledge flows are not geographical localized? Differently from Thompson and Fox-Kean, we believe this conclusion is not warranted. At a closer look, in fact, another explanation, entirely consistent with the framework proposed in this article, is possible. As long as inventors and teams of inventors are necessarily specialized, extracting patents from within narrowly defined technological fields implies that the corresponding inventors will belong to a relatively closed and self-contained technological community. As a consequence, when one draws control patents in such a way, it is quite likely that inventors of control patents and inventors of cited patents either correspond to the same individuals or are connected to each other by very short paths in the co-invention network. As shown earlier, however, the fundamental point is that closely connected inventors tend also to be spatially co-located. It does not come as a surprise, therefore, to observe that the narrower the technological focus, the larger the fraction of control patents that match geographically with the cited ones.

In order to confirm this intuition, we compute the fraction of citing-cited and control-cited patents that are connected either by mobile inventors or via the co-invention network, according to the method used to select controls. Results are reported in the first two columns of Table 9. The table shows that the narrower the technological criteria, the smaller the differences between control and citing patents in terms of connectivity to the cited patents. In other terms, the probability that a (randomly extracted) control patent is connected through mobile inventors or co-invention ties to the cited patent increases with the strictness of the technological matching criteria. In addition to this, for pairs of connected patents we have also computed the average distance in the co-invention network. Results are reported in the last two columns of Table 9.

By and large, the evidence seems to give further confirmation to our line of reasoning. By imposing a narrower technological matching, the average path connecting inventors gets shorter and the difference between citing and control patents in terms of network distance to the cited patents almost vanishes. Once again, to the extent that inventors

Table 9. Connectedness of citing and control patent samples, by control selection criteria

	Percentage of connected patents by (mobility or co-invention nw.)		Average path length between connected patents	
	Citing-cited	Control-cited	Citing-cited	Control-cited
Matching at primary 4-digit (JTH experiment)	30.9	19.8	9.2	10.5
Matching at primary 12-digit	28.5	22.1	8.3	9.8
Matching at primary 12-digit and secondary 4-digit	31.1	25.2	8	8.5
Matching at primary 12-digit and secondary 12-digit	31.4	27.1	7.1	7.8

connected through short paths in the co-invention network are also spatially co-located, this represents the most fundamental explanation for the results observed by Thompson and Fox-Kean. Knowledge flows are indeed spatially localized, but their carriers are mobile inventors and the resulting co-invention network.

Regression analysis further proves our point. Table 10 reports, in column (1), the logit regression for the probability of citing-cited co-location at the state level of Table 7 (column 5), and three more regressions for increasingly restrictive samples. We notice that moving from regression (1) to (4) the importance of controlling for the co-location of industry increases: odds-ratios for this variable are in the range of 1.7–2.7 in regression (1), and in the 3.6–6.3 for regression (4). This suggests that by increasing the technological proximity of the control and citing sample we also increase the probability to observe similar geographical patterns, as suggested by Thompson and Fox-Kean.

We also notice that moving from regression (1) to (4) the social distance between control and cited patents becomes significant, with a negative sign. This suggests that, other things equal, the more dense the co-inventor network is, the less likely it is that citing-cited patents are co-located. We explain this result by considering that the more restrictive our definition of a technological field is, the higher its social density, irrespectively of spatial proximity. So, it may well be that knowledge travels at a longer distance, since some co-invention social chains will span outside the location of cited patents.

Figure 4 reports the estimates of the probability of citing-cited co-location from regressions (1) and (4) in Table 10, respectively. When we consider the restricted sample, the probability of co-location of citing patents follows more closely that of controls; in other words, looking at the geographical distribution of an industry tells us already a lot on the distribution of spillovers. In fact, while the technological boundaries of our sample make no difference in terms of predicted probability when control patents are not co-located, they make a big deal of difference in the opposite case: in this case, the regression on the restricted sample suggests a much higher probability of citing-cited co-location than the regression on the original JTH sample, for any level of social distance.

The predictive power of social distance, however, remains high and unaffected by changes in the sample: socially close patents remain much more likely to be co-located than socially distant ones.

Table 10. Probability of co-location of citing-cited patents at the state level: Logit regressions for increasingly restricted samples

	Four-digit (JTH experiment) (1)	Primary class, 12-digit (2)	Primary and secondary class (12 and 4-digit) (3)	Primary and secondary class (both 12-digit) (4)
Intercept	−1.84***	−1.92***	−1.94***	−2.06***
	(0.060)	(0.061)	(0.071)	(0.085)
Control-cited co-location	0.76***	1.15***	1.26***	1.55***
	(0.114)	(0.113)	(0.128)	(0.144)
	[1.7–2.7]	[2.5–3.9]	[2.8–4.5]	[3.6–6.3]
Citing-cited inverted social distance	2.06***	2.53***	2.58***	2.62***
	(0.265)	(0.317)	(0.350)	(0.410)
	[4.7–13.4]	[6.8–23.5]	[6.6–26.2]	[6.1–30.6]
Control-cited inverted soc. distance	0.456	−1.01**	−0.78*	−1.46***
	(0.527)	(0.413)	(0.437)	(0.564)
	[0.6–4.4]	[0.2–0.8]	[0.2–1.1]	[0.1–0.7]
Mobile inventor between citing-cited	2.04***	1.40***	1.43***	1.06**
	(0.364)	(0.414)	(0.463)	(0.514)
	[3.8–15.7]	[1.8–9.1]	[1.7–10.4]	[1.1–7.9]
Mobile inventor btw. control-cited	−2.34	0.64	−0.70	0.33
	(1.639)	(0.545)	(0.644)	(0.687)
	[0.01–2.4]	[0.7–5.5]	[0.1–1.8]	[0.4–5.3]
L-ratio test	559.70***	499.58***	408.47***	326.01***
Aikake information criterion	3236.9	2749.5	2132.3	1579.4
Bayesian information criterion	3274.2	2785.8	2167.2	1612.7
Nagelkerke R-square	0.219	0.228	0.238	0.251
Observed	3700	3168	2458	1882

Note: Std err. in brackets; 95% odds r.s in square brackets. Significant at ***99%, **95%, *90%.

6. Conclusions

This article has provided an analysis of localized knowledge flows as captured by patent citations, on the basis of a methodology that allows to identify the role of mobile inventors and co-invention networks. When applied to US inventors' patent data, in the fields of Biotechnology, Organic Chemistry and Pharmaceuticals, such methodology suggests that mobile inventors and short social chains of co-inventors are largely responsible for the localization of knowledge flows. In an application of the same methodology to Italian data, we found very similar results (Breschi and Lissoni, 2006). In the light of them, the logical room left to more informal social ties, conventionally thought to be responsible for the localized diffusion of tacit knowledge, appears to be greatly reduced. In this perspective, the main reason why geography really matters is that mobility of technologists across organizations, either as employees or consultants, is bounded in space. Consequently, the organizations that are more likely to benefit from knowledge held by such individuals and embedded in their network of co-invention ties are only those located at a close geographical distance from them.

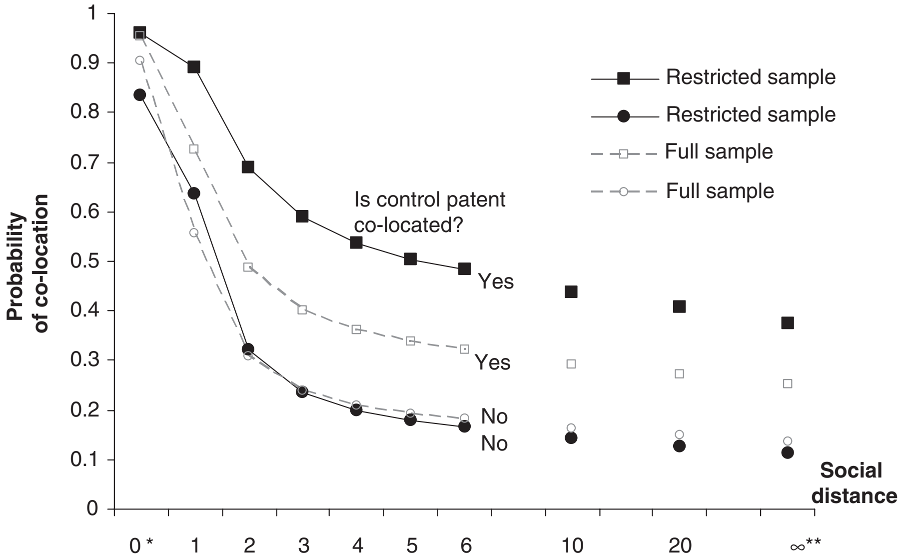

Figure 4. Probability of co-location of citing-cited patents (state level) as function of social distance (restricted versus full sample).

Mobility and co-invention are often attached to market transactions, whether in the labour market or in markets for knowledge-intensive services. Therefore, the knowledge flows they bring about may be classified as externality or spillovers only to the extent that they escape a complete contract definition. Whether this is really the case pertains future research. However, our results cast some doubts on conventional wisdom that assigns great importance to more informal, non-market related knowledge exchanges such as those originating from kinship, friendship and social gatherings. For these doubts to be confirmed, one will need to extend our evidence to more technologies and industries besides those considered here as well as to knowledge flows pertaining lower-quality inventions than those represented by EPO patents by US inventors.

Future research will have also to explore the motives behind inventors' low propensity to relocate in space. This will be of help in making (regional) policy-makers aware of the difference between measures aimed at supporting the creation of knowledge externalities through co-location of activities, and measures targeted to the labour market for knowledge workers, such as policies to smooth the frictions obstructing mobility or to attract and retain the best talents in the region (Florida, 2003; Saxenian, 2005).

In all of this research, new generations of patent databases, such as ours, which contain or may be attached to information on individual inventors will certainly play a key role.

Acknowledgements

Valerio Sterzi provided skilful research assistance. Nicola Lacetera, Koen Frenken, Robin Cowan, Diego Puga and three anonymous referees provided very useful comments. Usual disclaimers apply.

Funding

Financial support from Bocconi University is gratefully acknowledged. Francesco Lissoni also benefited from a Fulbright Visiting Scholarship, and the kind hospitality of the Sloan School of Management, MIT.

References

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.

Almeida, P. and Kogut, B. (1999) Localisation of knowledge and the mobility of engineers in regional networks. *Management Science*, 45: 905–917.

Audretsch, D. (1998) Agglomeration and the location of innovative activity. *Oxford Review of Economic Policy*, 14: 18–29.

Breschi, S. and Lissoni, F. (2001a) Knowledge spillovers and local innovation systems: a critical survey. *Industrial and Corporate Change*, 10: 975–1005.

Breschi, S. and Lissoni, F. (2001b) Localised knowledge spillovers vs. innovative milieux: knowledge "tacitness" reconsidered. *Papers in Regional Science*, 80: 255–273.

Breschi, S. and Lissoni, F. (2004) Knowledge networks from patent data: methodological issues and research targets. In H. Moed, W. Glänzel, U. Schmoch (eds) *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*, pp. 613–643. Berlin: Springer Verlag.

Breschi, S. and Lissoni, F. (2006) Cross-firm inventors and social networks: localised knowledge spillovers revisited. *Annales d'Economie et de Statistique*, 79–80.

Breschi, S., Lissoni, F., Malerba, F. (2003) Knowledge relatedness in firm technological diversification. *Research Policy*, 32: 69–87.

Breschi, S. and Malerba, F. (2005) *Clusters, Networks and Innovation*. Oxford: Oxford University Press.

Cohen, W. M., Nelson, R. R., Walsh, J. P. (2000) Protecting their intellectual assets: appropriability conditions and why U.S. manufacturing firms patent (or not). NBER Working Paper No. 7552.

Cowan, R. and Jonard, N. (2004) Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control*, 28: 1557–1575.

Feldman, M. P. (1999) The new economics of innovation, spillovers and agglomeration: a review of empirical studies. *Economics of Innovation and New Technology*, 8: 5–25.

Florida, R. (1993) The economic geography of talent. *Annals of the Association of American Geographers*, 92: 743–755.

Fosfuri, A. and Ronde, T. (2004) High-tech clusters, technology spillovers, and trade secret laws. *International Journal of Industrial Organization*, 22: 45–65.

Griliches, Z. (1992) The search for R&D spillovers. *Scandinavian Journal of Economics*, 94: 29–47.

Henderson, R., Jaffe, A., Trajtenberg (2005) Patent citations and the geography of knowledge spillovers: a reassessment: comment. *American Economic Review*, 95: 461–464.

Jaffe, A. (1989) Real effects of academic research. *American Economic Review*, 79: 957–970.

Jaffe, A. and Trajtenberg, M. (2002) *Patents, Citations and Innovations*. Cambridge, MA: Cambridge University Press.

Jaffe, A. B., Trajtenberg, M., Henderson, R. (1993) Geographic localisation of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108: 577–598.

Jensen, R. and Thursby, M. (2001) Proofs and prototypes for sale: the licensing of university inventions. *American Economic Review*, 91: 240–259.

Lamoreaux, N. and Sokoloff, K. L. (1999) The geography of the market for technology in the late-nineteenth and early-twentieth century United States. In G. D. Libecap (ed.) *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth*. Stanford, CT: JAI Press.

Levin, R. C., Klevorick, A. K., Nelson, R. R., Winter, S. G. (1987) Appropriating the returns from industrial research and development. *Brookings Papers on Economic Activity*, 18(3): 783–831.

Lissoni, F., Llerena, P., McKelvey, M., Sanditov, B. (2008) Academic patenting in Europe: new evidence from the KEINS database. *Research Evaluation*, 16: 87–102.

Lissoni, F., Sanditov, B., Tarasconi, G. (2006) The Keins database on academic inventors: methodology and contents. *CESPRI Working Paper*, 181.

Michel, J. and Bettels, B. (2001) European patent office, patent citation analysis: a closer look at the basic input data from patent search reports. *Scientometrics*, 51: 185–201.

Moen, J. (2000) Is mobility of technical personnel a source of R&D spillovers? *Working Paper 7834*. Cambridge, MA: National Bureau of Economic Research.

Mowery, D. C. and Ziedonis, A. A. (2004) The geographic reach of market and nonmarket channels of technology transfer: comparing citations and licenses of University Patents. In J. Cantwell (ed.) *Globalization and the Location of Firms*. Northampton, MA: Edward Elgar.

Newman, M. E. J. (2001) The structure of scientific collaboration networks. *Proceedings of the National Academy of Science USA*, 98: 404–409.

Saxenian, A. (2005) From brain drain to brain circulation: transnational communities and regional upgrading in India and China. *Studies in Comparative International Development*, 40: 35–61.

Singh, J. H. (2005) Collaborative networks as determinants of knowledge diffusion patterns. *Management Science*, 51: 756–770.

Thompson, P. and Fox-Kean, M. (2005) Patent citations and the geography of knowledge spillovers: a reassessment. *American Economic Review*, 95: 450–460.

Thursby, J., Fuller, A., Thursby, M. (2009) US faculty patenting: inside and outside the university. *Research Policy*, 38(1): 14–25.

Trajtenberg M., Shiff G., Melamed R. (2006), The ''names game'': harnessing inventor' patent data for economic research. *NBER working paper 12479*.

Wasserman, S., Faust, C. (1994) *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Zucker, L. G., Darby, M. R., Armstrong, J. (1998) Geographically localized knowledge: spillovers or markets? *Economic Inquiry*, 36: 65–86.