The geography of knowledge spillovers: the role of inventors' mobility across firms and in space

Stefano Breschi[*] Camilla Lenzi[*] Francesco Lissoni[*,⊙] Andrea Vezzulli[*]

[*] CESPRI, Università "L. Bocconi", Milan (Italy)

[⊙] Dept. of Mechanical Engineering, Università di Brescia (Italy)

Abstract

Spatial concentration of knowledge flows is often brought up as a key source of increasing returns to location, and a major explanation for the clusterization of innovation activities. While mainstream economics most often sums up all sources of such concentration under the notion of "localized knowledge spillovers", evolutionary economic geography addresses the spatial and social dimensions of proximity as separate geographical dimensions. In this spirit, we revisit both the JTH test of the localization of knowledge spillovers (Jaffe et al. 1993) and its extension by Agrawal et al. (2006). We find that inventors who patent across different companies and geographical locations contribute extensively to the observed citation patterns, both directly (through personal self-citations) and indirectly, by linking the various companies via a social network conducive to more citations. We also find that social networks convey knowledge both to the mobile inventors' current locations, and to the prior ones. We conclude that spatial distance is just a proxy for social distance, of which the professional ties between inventors are an important component.

1. Introduction

In the past 20 years, research on the geography of innovation has revolved largely around the concept of "localized knowledge spillovers" (hereafter LKSs). LKSs are "pure externalities" (Griliches, 1992): they exist insofar scientific and technological knowledge may escape its producer's control, and yet diffuse only locally. LKSs may explain why innovation activities are often found to be spatially clustered (Feldman, 1999).

For long supported only by circumstantial evidence, the LKS hypotheses was first tested by Jaffe, Trajtenberg and Henderson (1993; hereafter JTH). The three authors argued that knowledge spillovers may be measured by the "citations to prior art" contained in most patent documents, and produced a statistical experiment showing that such citations come disproportionately from the same geographical area of the cited patents. The experiment requires matching each citing patent to a control one, with the same application date and technological classification, in order to compare their location in space.

The JTH experiment has become a classical reference for most empirical work on the geography of innovation, both within mainstream economics and for unorthodox approaches, such as evolutionary and institutional ones.. However, its interpretation as proof of the existence of LKSs relies on the twin assumptions that scientific and technological knowledge is largely tacit, so that face-to-face contacts are the necessary vehicle for its diffusion, and that geographical proximity is a necessary condition for those contacts to take place. As a result, JTH's work treats *geographical proximity* as a proxy for *social proximity*, which inventors may derive from professional collaboration or common affiliation to companies, technical and scientific societies, or former institution of higher education. A limitation of this strategy is that it makes it difficult to distinguish between different social ties, according either to their nature (for example, professional vs. friendly) or strength (such as older vs. more recent ties).

Various attempts have been made to overcome this limitation. Agrawal et al. (2006) have tested whether inventors who move from one company to another and across different locations still pass on knowledge to former colleagues active in the cities they have left. Breschi and Lissoni (2006a,b) and Singh (2005) have resorted to a more direct approach, which consists in measuring professional ties between inventors resulting from co-invention data, and in applying to the resulting data some standard tools of social network analysis.

In this paper we put together the two strands of research. In particular, we explore the social network of inventors who are mobile in space, and test whether its geographical spread may help explaining the spatial distribution of the knowledge flows generated by inventors who move both across organizations and geographical locations.

This research effort is relevant for evolutionary economic geography for at least two reasons. First, it does not merely pay lip service to the conceptualization of knowledge as tacit and situated, but it explores in depth the implications of such conceptualization. In order to do so, it avoids taking the logical shortcut of assuming that tacitness necessarily imply the localization of knowledge flows, as it often happens in the applied literature based upon both the institutionalist and mainstream economic approaches (Breschi and Lissoni, 2001; Boschma and Frenken, 2006).

Second, it promotes the social network as the main unit of analysis for knowledge diffusion, instead of the city or region. By doing so, it shares one of the distinctive features of evolutionary economic geography outlined by Boschma and Frenken (2006), which consists in studying whether the spatial relations between economic agents matter for technological change, and how those relations and their importance may change over time.

In section 2 we sum up the key details of the JTH experiment, and of the related experiment by Agrawal et al. (2006). In section 3 we show that patents contain enough information to measure social distance quite accurately, recall a few notions of social network analysis, and apply them to a large set of patent applications in Organic Chemistry, Pharmaceuticals, and Biotechnology, signed by US inventors. In the same section, we provide descriptive evidence on the extent of inventors' mobility across firms and space, and on the resulting shape and geographical features of the social network of inventors.

In section 4 we use our patent sample to reproduce the Agrawal et al.'s version of JTH experiment and to show that inventors' mobility across firms and social ties between inventors largely explain both the original JTH and Agrawal et al.'s results.

In the Conclusions we stress that our evidence casts some doubts on the common interpretation of citation-measured knowledge flows as pure externalities, or spillovers, and outline our future research plans.

2. The JTH s experiment: methodology and interpretation

2.1 Methodology

The JTH experiment starts with the selection of a sample of *originating* (cited) patents. For each originating patent, all subsequent patents that cite it as prior art are then collected, previous exclusion of company self-citations, i.e. pairs of citing-originating patents assigned to the same company[1]. The address of inventors recorded in patent documents is then used to assign patents to a geographical area, in order to compare the locations of citing and originating patents[2].

A *control* sample of patents is also built. Each citing patent is matched to a randomly drawn patent, with the same technology class and application date, but no citation link to the corresponding originating patent.

A test follows, which consists in comparing the frequency with which citing-originating patent pairs match geographically (in our experiment, at the city level) to the corresponding frequency for control-originating patent pair. If the former turns out to be significantly greater than the latter, this should be interpreted as evidence of localisation effects of spillovers *over and above* the agglomeration effects arising from other sources[3].

The evidence reported by JTH shows indeed that citations are highly localised. Citing patents are up to two times more likely than the control patents to come from the same state, and up to six times more likely to come from the same metropolitan area.

Agrawal et al. (2006) propose a variation over the original JTH experiment which sheds some light on the nature and durability in time of social ties. In particular, they focus on the citations received by patents signed by "movers", the latter being inventors who appear to have changed location over time (that is, they have signed at least two patents, reporting at least once a different address). To identify such patents, each citing-cited (control-cited) patent pair is "unbundled" in order to obtain as many observations as the number of co-inventors listed on the originating patent; that is, from citing(control)-cited pairs, one obtains several so-called citing(control)-cited-inventor "triples". This allows identifying the relevant movers and their patents, whose "triples" are then retained for running the co-location test. The latter is performed considering both the movers' "current" location (the city where the mover resided at the time of the patent application) and the "prior" ones (the city/cities where the mover used to reside at an earlier time, as witnessed by her earlier patent applications).

Agrawal et al. find that movers' patents are more heavily cited than expected not only by inventors resident in the movers' current locations, but also by those who reside in the prior ones. Among the latter, inventors with a past working spell in the same companies as the movers appear to be the majority. These results suggest that movers keep in touch with their former colleagues, or at least that they leave a lasting knowledge legacy behind them.

2.2 Interpretation and the role of social networks

The mobility of R&D scientists and engineers within a localized labour market and the existence of localized markets for technologies have both been reported by various authors as potential explanations of JTH results[4].

As for labour mobility, Almeida and Kogut (1999) have replicated the JTH exercise for each US state. They find evidence of localised knowledge flows only in those few regions (most notably, the Silicon Valley) where the intra-regional mobility of inventors across companies is high.

Markets for technologies may also explain the JTH results to the extent that co-location is encouraged by technology users' need to consult frequently with their suppliers. Research contracts signed by the same independent inventor with different companies may produce patents that appear to be unrelated in terms of ownership, but very close in terms of technological contents and geographical distance (Mowery and Ziedonis, 2001).

The above-mentioned studies suggest that patents linked by a citation may also be personally or socially linked. A personal tie occurs whenever the same inventor is responsible for two patents from two different companies, either because she moved from one to another, or because she is an independent inventor who sold ideas to both. A social tie exists whenever two inventors *i* and *j* working for two different companies have a common professional acquaintance, in this case a fellow inventor who has worked jointly on a patent both with *i* and *j* (he/she may be either an employee who has moved across the two companies, or an independent professional who has consulted or done research for both of them). Other social ties between two patent may involve longer chain of acquaintances, such as when inventors *i* and *j* are connected by two other inventors *z* and *w* who first worked together, and then moved on to work separately with *i* and *j*, and so on.

More indirect social ties may also exist, such as serendipitous encounters between any two inventors who meet at workshops or through friends and other non-professional acquaintances. We fail to capture these ties with patent data, which record only formal collaboration instances.

In principle, social ties may or may not be concentrated in the geographical space: robust social links between inventors may convey tacit information even when the inventors have just a few chances to meet personally (witness many fruitful academic cooperation experiences), or well after the inventors have last met.

Agrawal et al.'s (2006) extension of JTH experiment improves our understanding of what types of social ties really matter for knowledge diffusion. Still, it tells us little of who knows whom in the inventors' community. We know that patents are disproportionally concentrated in the hands of relatively few large companies, wherein inventors may or may not have a chance to meet and exchange knowledge, or get in touch with each other's ideas via chains of mutual acquaintances; at the same time, these companies may provide for the codification of their inventors' knowledge in order to help its circulation among all technical employees.

Both the original JTH experiment and its extension by Agrawal et al. (2006) make use of a database with limited information on inventors' social capital. None of the databases contains information on the inventors' social ties; as a consequence, both geographical distance and company affiliations end up being two summary proxies for all kinds of ties, and for physical distance as such.

Having reclassified patents according to inventors, we will be able to show the importance of social ties as distinct from that of spatial localization. In particular, we will stress the role of a specific professional community (that of inventors) as channels of knowledge diffusion, and show that the localization of knowledge flows is largely explained by the limited mobility in space of such community. Thus, we will "open the black box" of localized social ties, often evoked as the conduit of knowledge, but seldom told apart from one another and subject to measurement.

We regard our exercise as the first step towards a more comprehensive direct measurement of all social ties between inventors and, more generally, knowledge producers. It is only by detecting what types of social ties matter (in our case, we focus on collaboration ties between inventors, recorded by patent documents) that researchers will be able to explain why spatial distance is often found to limit knowledge diffusion, and possibly reach strategic or policy prescriptions.

3. Methodology and data

3.1 Social networks: definition and methodology

Our methodology exploits patent-recorded information on inventors' names, surnames, addresses, and company affiliation (Breschi and Lissoni, 2004 and 2006). The following hypothetical example illustrates the main idea (see Figure 1). Let's consider five patents (1 to 5) and four assignees

(α, β, γ, δ). Assignee α owns two patents (1 and 2), while assignees β, γ and δ one each. Patents have been produced by thirteen distinct inventors (A to M). For example, patent 1, assigned to company α, has been produced by a team comprising inventors A, B, C, D and E. It is reasonable to assume that, due to the collaboration in a common research project, those five inventors are *socially linked* by some kind of knowledge sharing. The existence of such a linkage can be graphically represented by drawing an undirected edge between each pair of inventors, as in the bottom part of Figure 1.

[Figure 1 here]

Repeating the same exercise for each team of inventors, we end up with a graph representing the network of all inventors. This allows us to measure how tight is the connection between each pair of patents. In order to see how, we first give a few definitions:

i) For any pair of inventors, one can measure the distance between the two by calculating the so-called *geodesic distance*. The geodesic distance is defined as the minimum number of edges that separate two distinct inventors in the network[5]. In Figure 1, for example, the geodesic distance between inventors A and C is equal to 1, whereas the same distance for inventors A and H is 3. While A and C shared directly their knowledge while working on patent 1, A and H are more likely to have exchanged some word-of-mouth technical information through the mediation of other actors (such as B and F).

ii) Any two inventors may belong to the same *social component* or to *socially disconnected components*. A component of a graph can be defined as a subset of the entire graph, such that all nodes included in the subset are connected through some path. In Figure 1, for example, inventors A to K belong to the same component, whereas inventors L and M belong to a different component. A pair of inventors belonging to two distinct components have distance equal to infinity (i.e. there is no path connecting them).

iii) By *cross-firm* inventor we mean any inventor whose name has been reported in patent documents assigned to different organizations. This kind of inventors plays a fundamental role in connecting teams of inventors belonging to different organizations. For example, in Figure 1, inventor F worked for both company α and β, thus connecting the team of inventors (B,D) with the team of inventors (H,I). Similarly, inventor G worked both for company α and γ, thus connecting the team (B,D,F) with the team (I,J,K).

Using these definitions, we may now turn to illustrate how the existence of a linkage between patents can be ascertained. Three possible relations exist between any pair of patents from different firms:

1) The two patents exhibit *no social connection*, such as when the inventors behind them belong to socially disconnected components[6].

2) The two patents are linked by a *social connection*, such as when their inventors belong to the same social component. We can calculate the social distance *between patents* as the *minimum* geodesic distance between the two closest individuals from the two teams of inventors (geodesic distance)[7]. As such, the social distance between two socially connected patents may vary from 1 to any positive discrete value.

3) The two patents are linked by a *personal connection*, such as when at least one inventor belongs to both patents' teams. The social distance between two personally connected patents is zero.[8]

A limitation of this approach relates to the absence of rules to establish the *decay* of social links. In fact, we know for sure when two inventors come into contact, namely when they work together on the same patent for the first time. But we cannot be sure they keep in touch (and exchange information) after that common experience, unless we find them working on more joint patents in the following years. Some contacts established through co-inventorship may be dropped by one or both parts, but we do not know which ones.

We have addressed this problem by setting a 5-year maximum life span for all social ties: the tie binding two inventors who worked on the same patent at time *t* is cancelled at time *t+5*, and the network is re-calculated accordingly, unless the same inventors are found to work again together at a time comprised between *t* and *t+5*. This means that for every year we calculate a different social network[9].

3.2 Data

To implement the methodology just described, we rely on a biographical dataset of 63188 US inventors and their 66349 patent applications at the European Patent Office (EPO), filed between 1978 and 2002, in the following fields: Organic Chemistry, Pharmaceuticals, and Biotechnology (table 1)[10].

The choice of the three technical fields is explained by the high degree of inter-relation among them. In a previous paper we show how it is often the case that patents whose main technological class falls in one field have secondary classes in one or both of the others (Breschi, Lissoni and Malerba, 2003). Here we stress that many inventors active in one field also sign patents in one or both of the other fields. This explains why the total number of inventors in our sample (last line of table 1) is smaller than the sum of all inventors in each technological class, whatever time period we observe. In particular, in between 1978 and 2002, 13439 inventors (21% of the total sample) have patented in more than one of the three fields.

[Table 1 here]

The selected inventors represent the universe of all US inventors listed on EPO documents in the selected fields and years. The resulting affiliation network of patents, applicants and inventors, is therefore comprehensive of all social ties established through joint inventive activity, as it is the one-mode projection of the same network onto just inventors. From such network we have derived measures of social distance between all patents in the sample.

A major problem in measuring the geographic dispersion of patents and patent citations relates to the way patents are assigned to locations. Patent documents report the postal address of each inventor, from which we derive the MSA (Metropolitan Statistical Area) and State location of each inventor in each year of activity.

However, patents can have multiple inventors, each one with a different address. Therefore, the location of patents in geographic space cannot be resolved in an unequivocal way. In case of multiple inventors, JTH assigned each patent to the country/state in which pluralities of inventors resided, with ties assigned arbitrarily. Here, we take a slightly different approach and argue that two patents match geographically to the extent that they share *at least* one inventor's location.

3.3 Descriptive statistics

The vast majority (73%) of all inventors in Organic Chemistry, Pharmaceuticals, and Biotechnology are found to sign patents for only one company throughout their inventive career (which, in many cases, is limited to only one patent). Most "cross-firm" inventors sign patents for just 2 different assignees (17% of all inventors; 64% of cross-firm inventors); only a very few of them sign patents for more than 5 assignees (0,6% of all inventors; 4% of cross-firm inventors; table 2, second column).

[Table 2 here]

If inventors' cross-firm activity looks limited, mobility in space look even more limited. Only 28,4% of all cross-firm inventors (9,2% of all inventors) have been active in more than one MSA. Figures grow with the number of assignees the inventors have signed patents for, and go over 40% for cross-firm inventors with 4 assignees or more. The latter, however, are very few. Almost no inventor has been active in more than 2 MSAs[11].

Table 3 reports some descriptive statistics for the one-mode network of inventors. By construction, the size of the network changes over time (see section 3.1 above), with a tendency to expand due to the "patent explosion" of the 1990s (Hall, 2004). Table 3 shows how the network of inventors exhibits a characteristic typical of "small world" network of scientists, namely the asymmetry in size between the principal component (which in 1999 collects 46% of the inventors) and all the other components (in 1999, the second largest component collects only 0,9% of the inventors)[12].

[Table 3 here]

Inventors from different companies are linked one to another by cross-firm inventors. To the extent that the latter do not move much across cities the resulting social network will be concentrated in space. Table 4 confirms this intuition for the principal component: C5 and C4 concentration indexes always result above 50%, when calculated for active inventors (namely, those inventors in the network who sign patents in the current year). These figures are relatively stable over time, despite an increase in the number of MSAs reached by the expanding network, which suggests that peripheral areas host just a very few inventors.

[Table 4 here]

4. The role of "movers", and their social ties

4.1 Sampling

Following as closely as possible the JTH methodology, as adapted by Agrawal et. al (2006), we select 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 [13].

For each cohort of originating patents, we eliminate all applications that either received citations only from foreign organisations, or whose applicant was a US organisation, *but did not* report any US inventor[14]. The choice of excluding citations from foreign companies implies that our study does not investigate the extent of inter-national localisation of patent citations. This choice has been mainly dictated by data constraints, as the inclusion of citations coming from foreign organisations would have implied the construction of the worldwide network of inventors. We also remove all observations in which citing and originating patents have been assigned to the same organisation (i.e. company self-citations).

We additionally exclude those patent pairs (citing-cited) in which the same individual is listed in both documents (personal self-citations). Given that our focus is especially on movers, we follow Agrawal et al. (2006) and exclude these cases, to the extent that they cannot represent any knowledge flow between two distinct individuals. This leaves us with a total of 1933 originating (i.e. cited) patents, 2868 citing patents and 3517 citing-originating patent pairs.

In order to create a sample of control patents we randomly extract, for each citing patent, another patent with the following characteristics: no citation to the originating patent and same IPC 4-digit class and priority year of the citing patent. This procedure yields a sample consisting of 3283 control patents for a total of 3517 control-originating patent pairs (the same control patent can control for more than one citing patent).

Next, we "unbundle" individual inventors of the originating patents: from each patent we extract as many observations as the number of inventors listed on it, in order to obtain two "patent triples" (citing-cited-inventor and control-cited-inventor) for each inventor. The resulting sample contains 10988 observations.

We then extract information on the spatial and social distance between inventors within each triple in the sample.

In the first place, we ask whether at least one inventor from the citing (control) patent is located in the same current MSA of the inventor of the originating patent, and create a dummy variable accordingly. We also ask the same question relatively to the prior location of the inventor, if the latter is a mover. Thus, for movers, we end up with two dummy variables for each triple, which take value one in case of geographical matching of the citing (control) patent with the mover's current and prior location, respectively.

In the second place, for each patent triple, we measure the social distance between the mover and the inventors of the citing (control) patents. Using information derived from the network of inventors, we classify all patent triple into two (mutually exclusive) groups according to the linkages connecting the individual inventors that have produced them:

- ✓ *connected*: these are patent triples in which there is at least one inventor from the citing (respectively, control) patent that is connected to the inventor of the originating patent through a finite path in the co-invention network; we also calculate the social distance within each triple as the geodesic distance between the mover and the closest among the co-inventors of the citing (control) patent[15].
- ✓ *unconnected*: these are patent triple whose respective teams of inventors are not connected to each other in the co-invention network (although we cannot exclude the existence of other types of informal social ties).

4.2 Social ties in citing vs. control patents

Social connections between inventors are conducive to knowledge diffusion. It has been shown that the probability to observe a citation link between any pair of patents with different dates is a positive function of the existence of a personal or social tie between the two patents (Singh, 2005). In particular, social links matter when they are very close, that is when they exhibit low (≤5) geodesic distance (Breschi and Lissoni, 2004).

As a consequence, we expect that the probability to observe a personal or social link within a cited-citing patent pair is higher than the probability to observe a similar link in a cited-control pair. Table 5 shows that 35% of citing patents are linked to originating ones through a social chain of finite length; the same percentage for control patents is only 24%.

[Table 5 here]

Table 6 shows that the mean geodesic distance for connected patent pairs is significantly lower for citing patents (8,72) than for control ones (10,01). While over 24% of social ties between citing and originating patents is below six degrees of separation, the same figure for the control-cited pairs is under 16%.

[Table 6 here]

Finally, we have calculated the geographical co-location of both the citing and the control patent samples, with respect to the originating patents. As expected, closely connected patent pairs, both in the citing and the control sample, are highly co-localized with the cited one. For inventors at less than six degrees of separation, the percentage of co-location is around 29% both between citing-cited patents and for control-cited patents (table 7). The only major difference emerges for connected patents at more than 20 degrees of separation. Overall, these results suggest that, once we control for social ties, there is no reason to expect an association between geographical proximity and citations.

[Table 7 here]

4.3 Analysis: the JTH experiment

As a first step in our analysis, in the "all inventors" row of table 8 we replicate the JTH exercise on our data. The second and third columns of the table report the percentage of, respectively, citing and control patents that are co-located with the originating ones, at the MSA level, while the fourth column reports the z statistic for the difference between the two and (in brackets) the result of a 1-tail test on the hypothesis that $p_c > p_{nc}$ (where p_c is the co-location probability of citing patents, and p_{nc} the co-location probability of controls)[16].

[Table 8 here]

The geographic matching rates we find both for the citing and the control patents are almost double those reported by JTH (at the city level). To the extent that localization effects are expected to fade with time, these differences may be explained by the shorter period of observation we use (9 years max, from 1991 to 1999, in our sample, against a maximum of 14 years in JTH original exercise).

As for differences between citing and control patents, however, our results are very close to the original JTH ones. The proportion of citing patents co-located with originating ones is significantly greater than the proportion of control patents.

The descriptive statistics we presented in section 3 suggest that the different composition of the citing and control sample may be responsible for this result. We know that the citing patents are much more likely than the control ones to be socially connected to the originating ones, and that the cross-firm inventors responsible for the connection are most often immobile in space. Hence we expect that controlling for personal connections may reduce the observed differences in co-location rates between citing and control patents.

As the citing patent sample contains a higher number of socially connected patents with geodesic distance equal to or lower than 5 (which we know to be highly co-located with originating ones), we also expect to observe a similar result when controlling for social connections.

The "Only connected" line of table 8 show that by excluding all the unconnected patent triples from both the citing and the control samples[17], the co-location percentages of both the citing and control samples increase .On the contrary, when considering only the unconnected patents (third line) the co-location percentages go down, although the difference between citing and control patents remain significant. When considering connected patents, the difference in co-location percentages between citing and control samples is significantly higher than when considering unconnected ones. This result suggests that professional ties, as measured in our network of inventors, may explain a sizeable part of JTH findings. However, the persistence of some locational differences between citing and control triples leaves open the possibility that social ties other than those captured by our network may also affect inventors' knowledge exchanges, and result in patent citations.

4.4 Analysis: the role of movers

In order to focus on movers, we restrict our sample to those triples wherein the inventor of the originating patent has signed at least one patent before the originating one(s), and he/she has done so when residing in a different MSA[18]. Since movers are a very small subset of all inventors (see section 3) this restriction causes a dramatic drop in our sample size, from over 10000 observations to just 594, resulting from 287 originating patents, 279 inventors, 477 citing patens and 499 controls. This figure approximately represents the 5,5% of the initial sample, a percentage which is consistent with the one found by Agrawal et al. (2006) that is approximately 6%.

The distribution of patent triples between connected and unconnected ones does not change with respect of the original sample (36% connected triples among citing ones; 26% connected among controls). The path length distribution of connected patents, however, changes slightly, with more citing triples connected with paths shorter than six degrees of separation than in the general sample (34% vs. 24%; distribution for control sample does not change much; see table 9)

The analysis of co-location patterns reveals some interesting results. Citing patents do not appear to be more likely to come from the same current locations of movers, no matter whether we consider the connected or the unconnected ones, or all (table 10).

On the contrary, when we consider movers' prior location, we find that citing patents still tend to be more likely to be signed by inventors from those locations (11% vs. 8%; see table 10); however, when considering social ties, this result holds (indeed is reinforced) only when patents are socially connected (14% vs. 8,5%), and not when there is no social connection (9% vs. 7%, but the z-test suggests the difference is not significant).

[Table 10 here]

These results look sharper than those achieved by Agrawal et al. (2006), both for purely statistical reasons (the smaller sample) and for substantial ones.

Our smaller sample may explain why we see the citing patents' "co-location" premium for movers' current locations to become insignificant, while Agrawal et al. find it to diminish, but still resist.

As for prior locations, we do not only find confirmation of Agrawal et al.'s result of higher co-location of citing patents; we also find that this holds only for connected patents. This suggests that the social ties that bind movers to their prior locations originate from within their professional networks, that is can be traced back to their former co-inventors, and to those co-inventors' collaborators up to the sixth degrees of separation or so.

5. Conclusions

Our analysis has shown that the results obtained by both the original JTH experiment and its variation by Agrawal et al. (2006) can be largely explained by the importance of social ties from within the network of inventors.

Inventors who work for different companies are responsible for a large number of citations, but are scarcely mobile in space: they move or diffuse their knowledge across different firms, but not so much across different localities. They also contribute to create social networks, which also spread knowledge across firms, but not in space. These results are confirmed by another exercise, based on the same methodology, on a sample of Italian data (Breschi and Lissoni, 2006).

Those few inventors who move not only across companies, but also across locations (the "movers") maintain their ties with former co-inventors and their networks who are still located in the cities they have left. It is those cities, not those where the movers reside, that seem to enjoy an advantage over others in terms of access to the movers' knowledge.

Our results also raise a few substantive issues that deserve to be further discussed and investigated.

In the first place, our results qualify the original intuition of those economists and sociologists that first stressed the tacit content of technological knowledge: knowledge always travel along with people who master it. If those people move away from where they originally learnt, researched, and delivered their inventions, knowledge will diffuse in space. Otherwise, access to it will remain constrained in bounded locations. That is, knowledge flows are localized to the extent that cross-firm activity and the resulting social networks also are localized. Why US cross-firm inventors exhibit the observed (quite limited) mobility patterns is of course an important question, but one that goes beyond the scope of the present study.

In addition, our results suggest that, social ties derived from co-invention activity do not disappear when inventors move across space; on the contrary they seem to convey more knowledge than those built by movers in their new locations.

Network of inventors, of course, capture only a tiny subset of all the relevant social contacts enabling an individual to achieve an invention. However, most inventors listed on EPO patents are professionals, and their population is much more than a tiny and unchecked sample of all the individuals who can influence them; rather, it is the most immediate and influential social environment from which inventors draw ideas and information, at least for the technical contents of their patents.

The fact, that by controlling for the role of the network of inventors we manage to belittle the apparent role of spatial proximity, but not to eliminate it, suggest that other social networks, different from the professional one we considered here, may matter. In the near future, we will move in this direction by considering both patents and scientific publications, so that our network will include both inventors and pure academic scientists.

The social network approach to knowledge diffusion proposed in this paper may be further extended to comparative analysis. Ideally, one could compare the extent of knowledge localization in different regions or nations, and explain it with the different degree of mobility and resulting network dispersion of inventors, scientists, and knowledge producers in general.

References

Agrawal A., Cockburn I., McHale J. 2006, "Gone But Not Forgotten: Knowledge Flows, Labour Mobility, and Enduring Social Relationships", *Journal of Economic Geography* 6(5), 571-59

Almeida P., Kogut B. 1999, "Localisation of knowledge and the mobility of engineers in regional networks", *Management Science*, 45(7), 905-917.

Breschi S., Lissoni F. 2001, "Knowledge Spillovers and Local Innovation Systems: A Critical Survey", *Industrial and Corporate Change*, vol. 10 n. 4, 975-1005.

Breschi S., Lissoni F. 2004, "Knowledge networks from patent data: Methodological issues and research targets", in: Glänzel W., Moed H., Schmoch U. (eds), *Handbook of Quantitative S&T Research*, Kluwer Academic Publishers.

Breschi S., Lissoni F. 2006, "Mobility and Social Networks: Localised Knowledge Spillovers Revisited", *Annales d'Economie et de Statistique* (forthcoming)

Breschi S., Lissoni F., Malerba F. 2003, "Knowledge Relatedness in Firm Technological Diversification", *Research Policy* 32/1, pp. 69-87

Callon M.(1994), "Is Science a Public Good?", *Science, Technology, and Human Values* 19 (4), pp.395-424

Feldman M.P. 1999, "The New Economics of Innovation, Spillovers and Agglomeration: A Review of Empirical Studies", *Economics of Innovation and New Technology* 8, pp.5-25.

Boschma R.A., Frenken K. (2006), "Why is economic geography not an evolutionary science? Towards an evolutionary economic geography", *Journal of Economic Geography* 6/3, pp. 273-302

Griliches Z. 1992, "The Search for R&D Spillovers", *Scandinavian Journal of Economics* 94, Supplement, 29-47

Hall B. (2004), "Exploring the Patent Explosion", *Journal of Technology Transfer* 30, pp.35-48

Jaffe A.B., Trajtenberg M., Henderson R. 1993, "Geographic localisation of knowledge spillovers as evidenced by patent citations", *Quarterly Journal of Economics* 10, 577-598.

Lamoreaux N., Sokoloff K.L. (1999), "The Geography of the Market for Technology in the Late-Nineteenth and Early-Twentieth Century United States," in Gary D. Libecap, ed., *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth*, Stamford, CT: JAI Press

Lissoni F., Sanditov B., Tarasconi G. (2006), "The Keins Database on Academic Inventors: Methodology and Contents", CESPRI Working Paper n. …

Moen J. 2000, "Is mobility of technical personnel a source of R&D spillovers?", NBER Working Paper n. 7834.

Mowery D.C., Ziedonis A.A. 2001, "The geographic reach of market and non-market channels of technology transfer: comparing citations and licenses of university patents", NBER Working Papers n. 8568

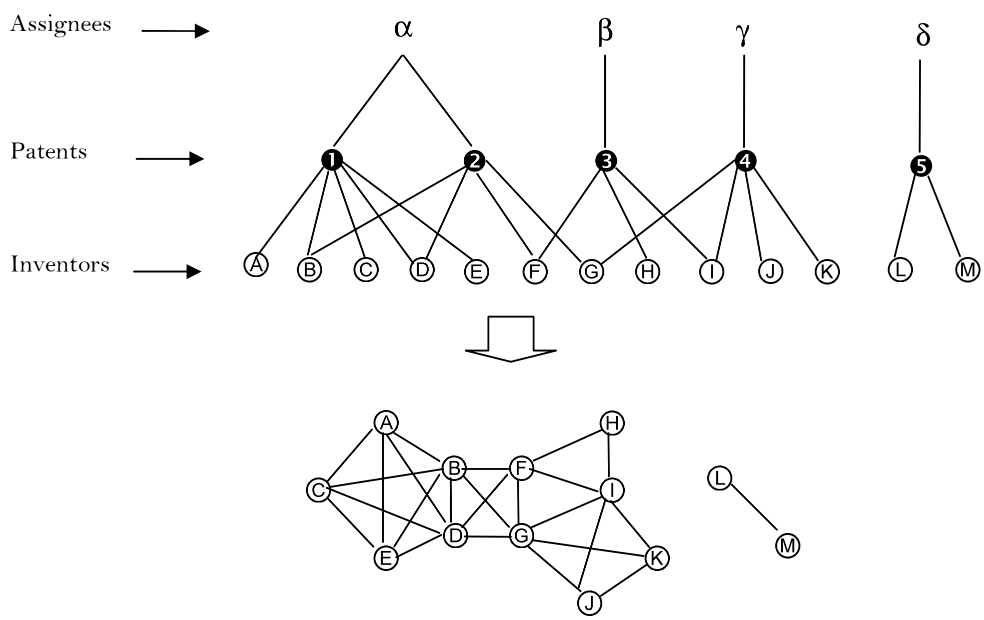
Newman M.E.J. 2001, "The structure of scientific collaboration networks", *Proceedings of the National Academy of Science USA* 98, 404-409.

Singh J.H. (2005), "Collaborative Networks as Determinants of Knowledge Diffusion Patterns", *Management Science* 51 (5), pp. 756-770

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

FIGURES

Figure 1 – Bipartite graph of patents and inventors



Top: Bipartite graph of assignees (α, β, γ, δ), patents (1 to 5) and inventors (A to M), with lines linking each patent to the respective inventors and assignees.

Source: Breschi and Lissoni (2004, 2006)

TABLES

Table 1 – US inventors' patenting activity at EPO, 1978-2002 (selected IPC classes)

	1978-1985	1986-1995	1996-2002	*1978-2002*
Patents				
Pharmaceuticals (1)	1784	8739	9360	*19883*
Organic Chemistry (2)	6152	13433	9845	*29430*
Biotechnologies (3)	1525	7013	8498	*17036*
Total	*9461*	*29185*	*27703*	*66349*
Inventors				
Pharmaceuticals (1)	2409	11897	14494	*24762* °
Organic Chemistry (2)	5679	15289	15640	*29799* °
Biotechnologies (3)	2145	10549	12787	*22066* °
Total	*9430**	*32215**	*36661**	*63188**°

(1) IPC: A61K = preparations for medical/dental/toilet purposes
(2) IPC: C07, excl.C07B
(3) IPC: C12M-S = biochemistry; microbiology; enzymology; genetic engineering
* Total n. of inventors < sum of tech. classes, as one inventor may patents across classes
° Total n. of inventors < sum of time intervals, as one inventor may patents at different times

Table 2 – Inventors' activity across assignees, MSAs *

N. of assignee joined	N. of inventors (% of all inv.)	% of inventors active in >1 MSA
1	46458 (73,52)	2,3
>1	16730 (26,48)	28,4
of which		
2	*10645 (16,85)*	22,8
3	*3679 (5,82)*	34,1
4	*1439 (2,28)*	40,3
5	*562 (0,89)*	46,4
6-10	*392 (0,62)*	46,7
>10	*13 (0,02)*	53,8
All inventors	63188 (100)	9,2

* The table reports the distribution of all inventors in the sample, according to the n. of different assignees for which they have recorded patents over the period examined. The calculation includes patents registered by individual inventors, i.e. not assigned to organisations. The calculation includes also patents that have been co-assigned to different organisations.

Table 3 – Network of inventors: size of selected components; selected years*

	N. of Inventors	% of inventors (excl. isolates)
1991 (n. of components=2525)		
Principal component	3664	23,39
2nd largest "	287	1,83
Smallest "	2	0,01
1995 (n. of components=2969)		
Principal component	7038	33,29
2nd largest "	250	1,18
Smallest "	2	0,01
1999 (n. of components=3595)		
Principal component	14077	45,97
2nd largest "	277	0,90
Smallest "	2	0,01

* Each year's network is calculate taking into account all patent applications for the selected technological classes over the previous 5 years

Table 4 – Concentration of inventors, by MSA; principal network component, selected years

Year	Nr. of inventors*	Nr. of active inventors §	Nr. of MSAs with active inventors	% of active inv. in top 5 MSAs π
1991	3664	1509	59	59,2
1995	7038	2934	80	52,0
1999	14077	5656	103	53,8

* N. of inventors in the principal component of the inventors' network. Each year's network is calculate taking into account all patent application for the selected technological classes over the previous 5 years

§ N. of inventors in the network having signed at least one patent application in the given year, for the selected technological classes

π Top 5 MSAs: San Francisco, New York, Baltimore, Philadelphia, Boston

Table 5 – Sample of cited, citing and control patents: summary statistics

	N. of patents	N. of pat-ent triples	N. of inv-entors	N. of inv. per patent	% of connected triples
Cited	1933	-	4375	2,96	-
Citing	2868	10988	7436	3,45	35,0
Control	3283	10988	8690	3,24	24,2

Table 6 - Socially connected citing (control) patents: frequency of geodesic distances from the originating patent

Geodesic distance	Citing	Control
1-5	24,40	15,44
6-10	49,6	45,29
11-20	23,3	36,14
>20	2,7	3,13
Mean	*8,72*	*10,01*
Median	*8*	*9*
std deviation	*4,94*	*4,93*

Obs=3894 citing (2656 control)

Table 7 - Geographic matching: % frequency for socially connected patents at the MSA level (citing and controls)

Geodesic distance	Citing (*n. of triples*)	Control (*n. of triples*)
1-5	29,4	29,5
6-10	13,5	9,6
11-20	7,0	7,0
>20	28,8	4,8
unconnected	9,7	8,2
all	12,0	9,0

Table 8 – Geographical matching at MSA level: % frequency and test of proportions*

	n. of obs	Citing: % match	Control: % match	z-statistic (*P>z*)
All inventors (JTH experiment)	10988	12,01	9,05	7,16***
Only connected citing-cited patents, and related controls	4779	14,69	10,40	6,33***
Only unconnected citing-cited patents, and related controls	6209	9,95	8,00	3,80***

§ One-tail z-test on difference between co-location percentages: 2,81 (0.0234)***
* Statistically significant at 1%***, 5%**, 10%*. One tail z-test P(Z>z)

Table 9 - Socially connected citing (control) patents: frequency of geodesic distances from the originating patent (movers)

Geodesic distance	Citing	Control
1-5	34,42	16,13
6-10	36,28	38,71
11-20	27,44	43,87
>20	1,86	1,29
Mean	*8*	*10,19*
Median	*7*	*10*
std deviation	*5,18*	*4,9*

Obs=215 citing (155 control)

Table 10 – Geographical matching at MSA level, for movers' current and prior location: % frequency and test of proportions*

	n. of obs	Citing: % match	Control: % match	z-statistic (*P>z*)
Movers' current location				
All	594	7,74	8,42	-0.43
Connected	272	10,66	8,82	0,72
Disconnected	322	5,28	8,07	-1,42
Movers' prior location				
All	594	11,28	7,58	2.18***
Connected	272	13,97	8,46	2,04**
Disconnected	322	9,01	6,83	1,02

* Statistically significant at 1%***, 5%**, 10%*. One tail z-test P(Z>z)

Notes

[1] The exclusion of company self-citations is motivated by the fact that these citations do not represent spillovers. Originating patents with no citations are similarly excluded from the analysis, because they are supposed not to have generated any knowledge flow.

[2] The rules followed by JTH to locate patents in space are indeed too complex to be summed up here. Two full paragraphs of their article are devoted to explain them (p. 585). Our rules will be slightly different and simpler, as explained below.

[3] "[We estimate] the probability of a patent matching the originating patent by geographic area, *conditional* on its citing the originating patent, with the probability of a match *not conditioned on the existence of a citation link*. This noncitation-conditioned probability gives a baseline or reference value against which to compare the proportions of citations that match" (JTH, 1993, p. 581).

[4] None of these explanations is entirely compatible with the "spillover" interpretation of such results. We come back to this point in the Conclusions

[5] For technical terms from social network analysis, see Wasserman and Faust (1994).

[6] With reference to Figure 1, this is the case, for example, of patent 5 and patent 1.

[7] When two patents are socially connected, all of their inventors belong necessarily to the same social component, but not all them are at the same geodesic distance. For example, in Figure 1 patents 4 and 1 are socially connected, but inventor K (from patent 4) and inventor A (from patent 1) exhibit a geodesic distance of 3, while inventors G (patent 4) and B (patent 1) have a geodesic distance of just 1. G and B are the closest inventors, and it is the geodesic distance between them that we pick up as the social distance between patents 4 and 1. In other words, the social distance between the two patents is the minimum geodesic distance between their inventors. See Breschi and Lissoni (2004) for further details, and a discussion of this choice.

[8] For example, inventors G, K and J (from patent 4) and H and F (from patent 3) belong to the same social component; in addition, inventor I appears both in patent 4 and 3. In the absence of I, patents 4 and 3 would be socially connect at distance 1 (the geodesic distance between G and F). The presence of I reduces this distance; we capture this reduction by setting the distance to zero.

[9] We have experimented with different time windows, and the resulting networks turn out to be quite similar to those based upon the 5-year time frame.

[10] Our data are a subset of the KEINS database, a much larger database that classifies all EPO patent applications from 1978 (EPO's first year of activity) to the current year, both by company and inventor. Fields are defined as 3- or 4-digit IPC (International Patent Classification) classes, or groups of classes. Pharmaceuticals correspond to class A61K, Organic Chemistry to class C07 (with the exclusion of C07B), and Biotechnology to classes from C12M to C12S. Companies in the KEINS database are identified by name and address. Group subsidiaries have been identified with the help of different editions of *Who Owns Whom*. As for inventors, these are 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+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 patents' applicants, their location in space, and the identity of their co-inventors). For details, see Lissoni et al. (2006).

[11] Data available on request

[12] On networks of scientists and their "small world" properties see Newman (2001). Differently from those networks, however, ours does not exhibit low average geodesic distance: two randomly chosen inventors from the principal component are separated by more than six degrees.

[13] The priority year has been used to date patent applications.

[14] The nationality of inventors has been derived by the address reported in patent documents.

[15] This implies that the geodesic distance assigned to each triple citing-cited-inventor or control-cited-inventor is the same for all the inventors in the originating patent team while, in principle, this distance could vary across triples (i.e. across inventors in the originating patent team). Nevertheless, the geodesic distance between each inventor of the originating patent and the relative citing (respectively, control) patent is equal either to the minimum geodesic distance between the citing-cited pair (respectively, the control-cited pair) if this inventor allows for the shortest path between the two or the minimum geodesic distance between the citing-cited pair (respectively, the control-cited pair) plus 1 (i.e. the geodesic distance of the closest inventor plus the geodesic distance of an inventor in the patent team to the closest inventor, that is by definition 1). Thus, there is a

potential risk, though limited, of underestimating the geodesic distance between an inventor of the originating patent and the citing (respectively, control) patent.

16 We define: $z = (\hat{p}_c - \hat{p}_{nc}) \Big/ \sqrt{\hat{p}(1-\hat{p})(1/n_c + 1/n_{nc})}$, where: $\hat{p}_c$ and $\hat{p}_{nc}$ are the sample proportion estimates for the citing and the control patents; n_c and n_{nc} are the size of the citing and control samples (in our case: $n_c = n_{nc}$); and $\hat{p} = (colocated_c + colocated_{nc})/(n_c + n_{nc})$, where $colocated_c$ and $colocated_{nc}$ are the number of co-located citing patents and controls, respectively.

The 1-tail test of our interest calculates the probability attached to values higher than z from a standard normal distribution.

JTH's test of proportion is slightly different, as it is based on a t-distributed statistic (JTH, p. 589): $t = (\hat{p}_c - \hat{p}_{nc}) \Big/ \sqrt{\left[\hat{p}_c(1-\hat{p}_c) + \hat{p}_{nc}(1-\hat{p}_{nc})\right]/n}$, where $\hat{p}_c$ and $\hat{p}_{nc}$ have the same meaning as above, and n is the size of both the citing and the cited samples ($n = n_c = n_{nc}$).

17 More precisely, we exclude all the citing patents that turn out to be personally connected, and the related controls; we also exclude all the control patents that turn out to be personally connected, and the citing patents they were meant to match.

18 In particular, we looked at the inventors of the originating patents and tracked backward their patent history. For each originating patent, we identified the first previous patent filed (or the set of previous patents filed in case of multiple patents in the same year), if any. We next compared the inventor's location (MSA) at the time of the originating patent and that of the first previous patent filed. If these locations differ, then we consider the inventor as mover. In case of multiple previous patents, we relax this criterion and consider an inventor as mover if at least one of the previous locations differs from the one at the time of the originating patent.