NBER WORKING PAPER SERIES

GONE BUT NOT FORGOTTEN:
LABOR FLOWS, KNOWLEDGE SPILLOVERS,
AND ENDURING SOCIAL CAPITAL

Ajay K. Agrawal
Iain M. Cockburn
John McHale

Working Paper 9950
http://www.nber.org/papers/w9950

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2003

Gone But Not Forgotten: Labor Flows, Knowledge Spillovers, and Enduring Social Capital
Ajay K. Agrawal, Iain M. Cockburn, and John McHale
NBER Working Paper No. 9950
September 2003
JEL No. F22, O31, R12, R23

ABSTRACT

It is well known that patent citations occur disproportionately between patents issued to inventors living in the same location, which has been taken as evidence of geographically localized knowledge spillovers. In this study, we find that patent citations also occur disproportionately often in locations where the cited inventor was living prior to being issued the patent in question, which we interpret as evidence of a significant role played by social capital in promoting knowledge spillovers. We first develop a model of purposeful investments in social capital by co-located inventors that incorporates the effect of expected mobility. Using patent and citation data, we then test two hypotheses motivated by the model. First, we find strong evidence in support of the enduring social capital hypothesis; social ties that facilitate knowledge transfer persist even after formerly co-located individuals are separated. Consistent with the model, we find that individuals with higher ex ante mobility are somewhat less likely to invest in location-specific social relationships, but the pattern of spillovers implied by patent citations is consistent with them investing in those social relationships that survive subsequent geographic separation. Second, we find strong evidence that the social ties associated with co-location are particularly important for facilitating knowledge spillovers across technology fields or communities of practice where alternative mechanisms for transferring knowledge are more costly.

Ajay Agrawal
Rotman School of Management
University of Toronto
105 St. George Street
Toronto, ON M5S 3E6
Canada
ajay.agrawal@rotman.utoronto.ca

Iain M. Cockburn
School of Management
Boston University
595 Commonwealth Ave
Boston, MA 02215
and NBER
cockburn@bu.edu

John McHale
Queen's School of Business
Goodes Hall
Queen's University
143 Union Street
Kingston, Ontario, K7L 3N6
Canada
jmchale@business.queensu.ca

I. Introduction

An often-discussed source of competitive advantage for a national or regional economy is its favored access to knowledge spillovers within its network of highly skilled workers.[1] Social ties between skilled workers are thought to be especially important for the effective transmission of non-codified (but codifiable) and tacit knowledge, which tends to diffuse through the population by direct communication. Recent work on the economic approach to social capital has stressed the importance of such social networks for economic interaction more generally, and the strong tendency for the effects of social capital to weaken with geographic distance (e.g., Glaeser *et al*., 2002).[2]

Although recognition of the importance of localized knowledge spillovers goes back at least to Alfred Marshall (Marshall, 1920; Krugman 1991), the difficulties of measuring such tacit knowledge flows impeded their study. As pointed out by Krugman[3], "they leave no paper trail by which they can be measured and tracked." The work of Jaffe *et al*. (1993, hereafter referred to as JTH) pointed, however, to one important exception. They argued that "[K]nowledge flows do sometimes leave a paper trail in the form of patent citations," which can be followed to "test the extent of spillover localization".[4] Taking patent citations as a proxy for knowledge spillovers, they found strong evidence of geographic localization even after controlling for the tendency of inventive activities to be geographically clustered by technological field.[5]

If co-location facilitates greater access to knowledge spillovers due to stronger social ties, what happens when an inventor moves? In this paper, we explore the possibility that citations to a patent, and thus knowledge spillovers, also occur disproportionately at locations where inventors were living *prior* to their current inventive activity. Our hypothesis is that individuals invest in the development of social ties with others with whom they are co-located, and that at least a portion of those ties endure even after the

---

[1] Modern endogenous growth theory casts knowledge spillovers from investments in human capital and research and development as a central character in generating the increasing returns that sustain long-term growth (e.g., Romer, 1986 and 1990).

[2] An important piece of circumstantial evidence for the importance of localized knowledge spillovers is that industries for which new knowledge plays an important role tend to be more spatially concentrated (Audretsch and Feldman, 1996). Another intriguing piece of evidence is that clusters of biotechnology firms developed around academic scientists who published genetic sequencing discoveries in academic journals (Zucker *et al*., 1998).

[3] P. 53

[4] P. 578

[5] One concern with this result is whether it reflects communication between the inventors and thus true knowledge spillovers. The survey evidence reported in Jaffe *et al.* (2002, Chapter 12) partially allays this concern, since they find that citations are a signal of communication, albeit a noisy one.

individual has moved. Thus, their *past* neighbors also have some degree of favored access to the new knowledge generated at their *new* locations.

To better understand the mechanisms at work, we first develop a simple model of purposeful investments in social relationships that facilitate the exchange of non-rival, non-contractible knowledge. In the model, the opportunities for such investments are limited to co-located inventors. We are particularly interested in how the prospect of mobility, and thus the possible future geographic separation of individuals, alters the incentive to invest in relationships with currently co-located colleagues, and thus has implications for the presence, in equilibrium, of enduring social capital.

Two effects are at work in this model. First, high prospective mobility diminishes the incentive to invest in local relationships when the value of communication is adversely affected by separation. Second, however, high prospective mobility also biases investments towards relationships whose value to the individuals involved is relatively insensitive to their degree of geographic separation. Which of these effects dominates is an empirical question, and in this paper we present evidence consistent with the presence of investments in enduring social capital that facilitates communication between previously co-located individuals even after they become geographically separated.

The essence of our empirical methodology for testing the enduring social capital hypothesis is to seek evidence of disproportionate cites to inventions at locations where the individual lived prior to their invention. Following the pioneering methodology developed in JTH, we compare the extent to which actual citations are disproportionately located in a particular location relative to a distribution of control citations that have the same temporal and technological characteristics. Importantly, this comparison allows us to control for any technology-based clustering of inventive activity, which may otherwise confound any inference drawn from co-location of citations.

Given our interest in inventor mobility, we use citations to inventor-patent pairs as our unit of analysis. For example, a 1990 patent that has three inventors will generate three inventor-patent pairs. Each inventor-patent pair will be associated with a unique 1990 location (Metropolitan Statistical Area, or MSA) that is based on the inventor's address as recorded in the 1990 patent. If that inventor had previously patented somewhere else (i.e., is a "mover"), we also record their most recent *prior* location. We then examine the proportion of the subsequent citations (1990-2002) that occurred at the inventor's 1990 location, and, for movers, the proportion of subsequent citations that occurred at the inventor's prior location.

Our results support the geographically localized knowledge spillover finding of JTH. Like JTH, we find strong evidence of a disproportionate number of cites that are co-located with the inventor. More interestingly, in the context of the present paper, we also find evidence of a disproportionate number of cites in locations where the inventor had previously lived (and patented)—that is, we find evidence of a *prior location premium*. As discussed above, one plausible interpretation of this finding is that individuals invested in social ties with others at their prior location during their residency there, and at least part of that social capital endured to support above average knowledge flows back to their prior location. In effect, prior co-location allows for investments in social relationships that condition the subsequent distribution of knowledge spillovers from mobile inventors.

Finally, we find evidence that co-location is particularly important for cross-field knowledge spillovers. We hypothesize that knowledge spillovers within communities of practice (Lave and Wenger, 1993; Brown and Duguid, 1991) or invisible colleges (Crane, 1965 and 1969) are less likely to be geographically mediated. Groups of researchers interested in similar problem areas are likely to communicate with each other via mechanisms such as conferences, publications, and trade shows such that being co-located is less important for facilitating knowledge spillovers. We find evidence that both current and prior co-location increase the likelihood of cross-field spillovers even more, proportionately, than within-field spillovers.

We think these results are interesting in the context of three literatures. First, they provide additional insight into the processes through which economic knowledge diffuses. The results are consistent with the conjecture that social ties facilitate knowledge spillovers, and they show how the geographic distribution of relevant social capital is determined in sometimes subtle ways.

Second, the results are relevant in the context of measuring what an economic location loses when a portion of its skilled workforce leaves. The migration literature has suggested the importance of knowledge spillovers to the gains and losses of locations from mobile labor (e.g., Borjas, 1995). But these effects generally have been viewed as un-measurable. The JTH findings show that knowledge spillovers are geographically localized, which suggests an important source of location-specific loss when inventors leave. Our results suggest, however, that the losing location can nonetheless retain some degree of favored access to the knowledge generated by the departed inventor from their new location.[6]

Finally, our results may be of interest to those studying the links between labor mobility and social capital accumulation (e.g., Glaeser *et al*., 2002). Using patent citation data as a proxy for knowledge flows and modeling knowledge flows as being facilitated by social relationships, our work shows how the rich data that is available on the geographic locations of patenting and citing inventors can be used to empirically examine how prospective mobility affects social capital accumulation.

The rest of the paper is structured as follows. In the next section, we develop a simple model of purposeful investments in social relationships between (prospectively) mobile inventors. The model yields two hypotheses that we test in the remainder of the paper. Section III outlines our general methodology for using patent and citation location data to test these hypotheses. Section IV describes our data and Section V our results. Section VI concludes with a summary of our main findings and some suggestions for future work.

II. Social Relationships, Knowledge Spillovers, and Inventor Mobility: A Simple Model

In this section, we develop a simple model of tacit knowledge communications that are facilitated by social relationships between presently and formerly co-located inventors. Inventor co-location plays two key roles in the model. First, it creates opportunities to develop social relationships. Thus co-location may be thought of as a sort of social treatment, whereby the inventor in exposed to potential social acquaintances.[7] And second, conditional on a social relationship existing, the value of the knowledge flowing between the related individuals depends on whether they are currently co-located or separated. For the purposes of the model, we assume that the knowledge in question is both non-rivalrous and non-contractible. The non-rivalry assumption implies that the knowledge has the public-good characteristic of not losing value when it is communicated to other inventors. Of course, such knowledge is easily excludable; the knowledgeable inventor can simply refuse to communicate the knowledge to others. We

[6] Interest in estimating the losses from the out-migration of skilled workers (a.k.a. the "brain drain") has been growing as the competition for talent between and within national (or regional) economies has increased (Desai *et al*., 2002).

[7] Although we focus on relationships developed as a result of co-location, we allow for the possibility that social relationships can be developed without co-location.

assume, however, that the knowledgeable inventor is unable to write enforceable contracts to sell the knowledge.[8]

We treat social relationships as a technology for communicating non-rivalrous, non-contractible knowledge between inventors, and define the total set of an inventor's relationships as that inventor's social capital. These relationships are assumed to be the result of purposeful costly investments by forward looking inventors—potential social acquaintances are sought out, meetings arranged, small talk engaged in, etc. We remain agnostic on how exactly social relationships work to facilitate communication, but note three possible mechanisms. First, once a relationship is established, it may actually be pleasurable for the parties to exchange information about their work. Second, even where the information exchange is costly, the establishment of a long-term relationship may allow for the development of trust that facilitates reciprocal knowledge transfer.[9] Third, where inventors care about the opinions their colleagues hold about their work and their willingness to cooperate, the development of social relationships may contribute to social pressures to reveal (at least) the non-rivalrous part of what they know.[10]

We assume that inventors will only make investments for which the expected value of the resulting communication (net of any ancillary communication costs or benefits) is greater than the cost of the upfront investment in developing the relationship. Opportunities for any given inventor i to invest in social relationships with presently co-located inventors are assumed to arrive randomly over time. Suppose, for example, that an opportunity has arisen to invest in a relationship with inventor j. If the investment is made, it results in valuable information flows between the inventors. The expected value of this flow (again, net of any ancillary costs and benefits) is assumed to be a constant, k_{ij}, as long as the two inventors remain co-located.

[8] One obvious problem that will impede standard market transactions is that it typically will be impossible to allow the buyer to "inspect" the knowledge product before sale.

[9] Repeated interaction over time, coupled with investment in establishing relationships, can enhance the efficiency of information exchange, in the sense of Williamson's "relational contracting" in the face of transactions costs.

[10] Work exploiting the concept of social capital has increased dramatically in sociology, organizational theory, and political science. Much of the recent political science work has focused on the value of social capital as an asset to an organization, community, or nation (e.g., Putnam, 2000, Fukuyama, 1995.) Economists are naturally drawn to a view of social capital as something that is purposefully accumulated by self-interested individuals for its perceived value to them (Glaeser *et al*., 2002). It is interesting, however, that this approach has quite a lot in common with the work of sociologists on social capital that followed the pioneering studies of Pierre Bourideau and James Coleman (see the survey by Portes, 1998). Burt (1992), for example, offers a view of network building that is explicitly set in terms of profit-seeking investments in relationships. We thank Bill Cooper for drawing our attention to work in the sociological tradition, though he is in no way responsible for our misinterpretations.

We are particularly interested in how the prospect of future mobility affects the extent and composition of investments in social relationships. The per-period probability of separation between a pair of inventors is denoted by m_{ij} and the value of the communication flow conditional on separation is denoted by k_{ij}^*. We take it that separation generally increases the costs of communication and impedes knowledge flows, though we allow for the possibility that access to certain kinds of knowledge might become even more valuable once the inventors are separated.[11] In judging the value of a potential social relationship, an inventor must then consider the net communication value while co-located, the probability of separation, and net communication value while separated, all the while discounting the value of future communication flows by an appropriate interest rate.[12]

Putting the pieces together, the value of an investment by i in j is given by the following value equation:

$$(1) \qquad V_{ij} = \frac{1}{1+r}\left[k_{ij} + (1-m_{ij})V_{ij} + m_{ij}V_{ij}^*\right].$$

The relationship with the currently co-located inventor is treated as an asset for inventor i. The value of the asset at the beginning of the period is equal to the discounted value of the communication k_{ij} (which is assumed to take place at the end of the period) and the discounted expected value of the asset at the end of the period. The expected asset value at the end of the period is a separation-probability-weighted average of the value of the relationship conditional on co-location, V_{ij}, and the value of the relationship

[11] The sociological literature points to some reasons to expect valuable enduring communications. First, the work of Granovetter (1973) on the strength of weak ties points to how the extent of overlap between two individuals' networks is correlated with the strength of ties between them. Given that one's friend's friends are also likely to be their friends, the friend may actually provide them with little information that they cannot get from the rest of their friendship network. In contrast, an acquaintance to which they are only weakly tied may provide them with truly novel information. Similarly, an additional co-located inventor in a dense network of local inventors may provide little additional information. Suppose, however, that one moves, so that their ties with all members of the network in their former home weaken. In this case, it is easy to imagine that having multiple ties to the former network—some which previously seemed redundant—is important to making sure that one retains access to critical information flows. In this case, a high prospect of mobility may actually increase an inventor's incentive to invest in a relationship with co-located individuals. More recently, Burt (1992) emphasized opportunities for value-creating brokerage that accrue to individuals who can fill "structural holes" in networks—that is, relationships with people who don't have relationships with one another. Inventor migration may create such brokerage opportunities, thus increasing the value of post-separation communication, and again increase the incentive for pre-separation investments. In the economics literature, Rauch (2001) and Saxenian (2002) offer interesting evidence on how emigrant diasporas serve as intermediaries in matching business partners and facilitating contracting between their present and former homes.

[12] We allow for the possibility that the value of an investment by i in j depends on the investment j is making in i. For example, i might find it worthwhile to invest in a relationship with j if j is also making an investment in i, or, conversely, i might only find it worthwhile to invest in a relationship with j if j is not investing in a relationship with i. In the case where investing in j is not a dominant strategy, we assume that investment by i and the required action by j for the investment by i to be worthwhile is a unique Nash equilibrium.

conditional on separation, V_{ij}^*. r is the instantaneous real interest rate. At the end of the period, one or both of the inventors can move. The instantaneous probability that one or both move conditional on not having moved already is given by m_{ij}. For simplicity, we assume that if both inventors move, the probability that they end up in the same place is zero. Thus, m_{ij} is the probability that the two inventors will be separated.[13] The value of the relationship conditional on separation, V_{ij}^*, is itself determined by a value equation,

$$(2) \qquad V_{ij}^* = \frac{1}{1+r}\left[k_{ij}^* + V_{ij}^*\right] = \frac{k_{ij}^*}{r},$$

where k_{ij}^* is the post-separation value to i of the long-distance communication between i and j.[14]

Substituting (2) into (1) allows us to solve for the expected value to i of an investment in a social relationship with j.[15]

$$(3) \qquad V_{ij} = \frac{k_{ij} + \left(\frac{m_{ij}}{r}\right)k_{ij}^*}{(r + m_{ij})}.$$

From the perspective of an inventor considering whether to make an investment in a relationship with another presently co-located inventor, the relative value of a dollar's worth of long-distance communication in terms of dollars of local communication is given by $\frac{m_{ij}}{r}$. It will also be useful for the

[13] We further assume that moves are permanent. Equation (2) can be extended to allow for the possibility of return by allowing the value of the end-of-period asset to be per-period return probability weighted average of the end of period value of the separated relationship, V_{ij}^*, and the value of reunited relationship V_{ij}. This formulation makes the strong assumption, however, that the communication returns to its initial value upon being reunited.

[14] We make the strong assumption that the value of the post-separation value of communication is invariant to who moves. The model can easily be extended to allow for different communication values depending on which party moves.

[15] Note that when k_{ij}^* is zero, the value of the investment in the co-located inventor is given simply by the perpetuity $\frac{k_{ij}}{r + m_{ij}}$.

results below to note that effect of an increase in the probability of separation on the expected value of a relationship is given by

$$(4)\qquad \frac{\partial V_{ij}}{\partial m_{ij}} = -\frac{k_{ij} - k_{ij}^{*}}{(r + m_{ij})^{2}}.$$

Each social relationship investment opportunity is associated with a unique cost, C_{ij}. This cost captures the value of the time and energy that must go into the development of the relationship, though we do not rule out the possibility that the cost is zero or even negative—developing a relationship can be fun as well as economically valuable. We assume that the inventor will invest in all relationships for which the value is greater than the cost. Figure 1 shows the set of profitable social relationships for a given level of upfront investment cost and a given per-period probability of separation, where each social relationship is described by a pair of local communication and long-distance communication values (k_{ij}, k_{ij}^{*}). If some proportion of inventors take advantage of opportunities to invest in relationships with other co-located inventors (i.e., the set of profitable investment opportunities for relationships with presently co-located inventors is not empty), and some positive fraction of those inventors have not been separated, then we should observe that knowledge spillovers from inventors are disproportionately localized.[16] This finding was reported in JTH and is confirmed with our data (see Appendix).

However, will we also see disproportionate spillovers to a mobile inventor's *prior* locations? That is, will we also see a *prior location premium* due to enduring social capital? Consider an inventor who previously resided in location *x* and now resides in location *y*. To what extent will the knowledge that the inventor generates in their new location *y* also spill disproportionately back to location *x*? To better understand the mechanisms at work, it is useful to shift attention back to the inventor's decisions while residing in location *x* relating to investments in social relationships with other *x*-located inventors. Given that the inventor actually did move, it is reasonable to suppose that there was a high *ex ante* probability of separation. From equation (4), we can see that a high probability of separation will have a large adverse effect on the value of relationships when k_{ij} is large relative to k_{ij}^{*}. When this is true, the inventor is unlikely to develop many relationships, and those relationships that are developed are unlikely to retain much value upon separation (i.e. k_{ij}^{*}/r is low). In this case, we will not see a significant enduring social capital effect. However, it is possible that inventors do have opportunities to develop relationships that

[16] This also requires that k_i is strictly positive on at least one of those investments.

have large values of k_{ij}^{*}, both absolutely and relative to k_{ij}. In this case, a higher *ex ante* probability of separation can actually increase investments in social capital relationships (see equation (4)), and those investments that do take place will retain their value upon separation (i.e. k_{ij}^{*}/r is high). Thus, even with forward looking and highly mobile inventors, it is possible that we will observe an enduring social capital effect. This leads to the *enduring social capital hypothesis*.

H1: Knowledge spillovers also go disproportionately to the inventor's <u>prior</u> location.

Up to this point we have made no distinctions between the types of relationships an inventor might form. To generate our second hypothesis, we focus on a single distinction—relationships between inventors working in the same technological field (within-field relationships) and relationships between inventors working in different technological fields (across-field relationships). It is very likely that intra-technology relationships will yield the most value to an inventor, since they are more likely to generate invention-relevant communications. On the other hand, the inventor may have access to much of this knowledge through professionally based (as distinct from co-location based) social relationships, which are less likely to be location-specific, as well as through such non-relationship based channels as journals, conferences, and the internet. Knowledge from outside their "community of practice" may be harder to come by outside of social relationships facilitated by co-location.

Consider the effects of such social relationships on both codified and tacit types of knowledge. The ability to utilize and build on codified knowledge depends on access and awareness. It seems plausible that all inventors have reasonably equal *access* to most codified knowledge, such as that published in journals, whether or not they are in the same field and whether or not they are co-located. It also seems plausible that inventors in the same field are equally *aware* of new codified knowledge, whether or not they are co-located. Looking across fields, however, it is less likely that inventors will become aware of codified knowledge outside their field. Therefore, social relationships facilitated by co-location may be relatively more important as a means of generating awareness of new knowledge across fields than within fields.

For example, consider new knowledge in the field of evolutionary biology that is published as an article in the journal *Evolution*. We assume that all scientists, regardless of field or location, have equal access to the journal. We also assume that all scientists in the field of evolutionary biology are equally likely to be aware of this new, published knowledge, regardless of their location, since they read the journal and attend related conferences. In other words, the awareness of scientists in this field is largely independent of geography. However, scientists outside of this field do not regularly read *Evolution* and do not regularly attend related conferences. The likelihood that these scientists are aware of this new knowledge is a function of their likelihood of having a social relationship with the knowledge creator, which is, as we have argued, a function of geographic proximity. So, we assume that the awareness of scientists outside the field is geographically mediated.

Similarly, we consider the effects of geographic proximity on tacit knowledge. Access to new, tacit knowledge requires direct communication with the knowledge creator and therefore is not equal for all. We assume that direct access to the knowledge creator is both more likely to occur (and at lower cost) for those who have a relationship with that person. This is true even amongst scientists in the same field. However, since scientists in the same field have opportunities to interact at conferences and other gatherings of the "invisible college," co-location is likely to lead to a larger proportionate increase in the probability of a knowledge spillover for inventors who are not in the same field.

These considerations lead us to our second hypothesis that co-location (current and prior) has a greater proportionate effect on inter- than on intra-technological knowledge flows:

H2: The proportionate increase in knowledge spillovers due to co-location is greater for spillovers across technology fields than for spillovers within technology fields.

III. Empirical Methodology

The hypotheses we wish to test both relate to the geographic distribution of tacit knowledge flows. Such knowledge flows are notoriously difficult to measure. Following the work of Adam Jaffe, Manuel Trajtenberg, and co-authors (see the collected papers in Jaffe and Trajtenberg, 2002), we use patent citation data as an indicator of communication and ultimately knowledge spillovers between inventors.[17]

[17] Patent citations are not straightforward to interpret in terms of communication between inventors, and the signal to noise ratio for this measure is therefore likely to be rather low. Patents cite other patents as "prior art," with citations serving to delineating the property rights conferred. Some citations are supplied by the applicant, others by the patent examiner, and some patents may be cited more frequently than others because they are more salient in terms of satisfying legal definitions of prior art rather than because they have greater technological significance. Cockburn, Kortum, and Stern (2002) report, for example, that some examiners have "favorite" patents that they cite preferentially because they "teach the art" particularly well. Nonetheless, Jaffe *et al*. (2002) surveyed cited and citing inventors to explore the "meaning of patent citations" and found that "communication is important, and that patent citations do provide an indication of communication, albeit one that also carries a fair amount of noise" (p. 380).

We adapt the methodology of one of the seminal papers in this literature (JTH) in ways that allow us to examine the effects of inventor mobility on the geography of knowledge flows.

The essence of the JTH methodology is the comparison of citing patents with control patents in terms of the frequency with which each is located in the same region as the original patent. A finding of a disproportionate number of co-located citations relative to co-located control patents is interpreted as evidence of localized knowledge spillovers. The reason for using controls is that patent citations will tend to be co-located with the original inventions even in the absence of knowledge spillovers when inventive activity in particular technological areas is clustered geographically. Thus the spillover effect is identified as the extent of co-location that exists over and above what we would expect given the geographic concentration of inventive activity by technological area.

More formally, we define the probability of co-location in our sample of control patents as the unconditional probability of co-location, $P(Co\text{-}location)$, and the probability of co-co-location given that an actual citation has occurred as the conditional probability of co-location, $P(Co\text{-}location \mid Citation)$. Our basic hypothesis test is that the difference between the conditional and unconditional probabilities—i.e., the co-location premium—is positive and statistically significant. In economic terms, however, it will often make more sense to think about how inventor co-location affects the probability of a citation rather than how the occurrence of a citation affects the probability of the inventors being co-located. In other words, the interesting causal relationship is from co-location to the likelihood of a knowledge spillover (as proxied by a citation). Of course, the two probabilities are related by Bayes Rule.

$$(5) \qquad \frac{P(Citation \mid Co-location)}{P(Citation)} = \frac{P(Co-location \mid Citation)}{P(Co-location)}.$$

Subtracting 1 from both sides, we see that the proportionate increase in the probability of a citation conditional on co-location is equal to the proportionate increase in the probability of co-location conditional on a citation. Thus in presenting our results we also report the ratio of [$P(Co\text{-}location \mid Citation)$ - $P(Co\text{-}location)$] to $P(Co\text{-}location)$. This ratio measures the proportionate increase in the probability of a citation due to co-location.

The unit of analysis in JTH is a citation to an originating patent. Given our focus on the ties between individual inventors, our unit of analysis is a citation to an *inventor* on an originating patent—what we call an inventor-patent-citation. Thus, a single patent that has two inventors and is cited by five subsequent patents will generate ten unique observations.[18] Each observation is assigned to a location (MSA) based on the city and state information associated with their home address as reported on the front page of the patent.

As in JTH, however, a simple measurement of location matches would not account for any geographic clustering of innovative activity within particular technological areas. For example, an inventor on a patent for a particular type of medical device might be located in Boston, and the patent might receive a large fraction of citations from patents which include at least one inventor located in Boston. This might reflect knowledge spillovers through social ties, or it could simply reflect the large fraction of overall patenting for medical devices that occurs in Boston.

We use the following procedure to construct the set of control patents. A control patent is selected for each inventor-patent-citation observation that matches the citing patent on the following dimensions: 1) application year and 2) technology classification. Having generated the set of patents with the same application year and same original three-digit U.S. classification as the citing patent, we identify the patent in the set that has the closest grant date to the citing patent. Next, we confirm that the control patent does not cite the original patent. If it does, we remove the patent from the set of potential control patents and select the next best control patent. Finally, if there are no patents that match the citing patent in at least application year and original classification without citing the original patent, the observation (original patent) is removed from the data set.

IV. Data

We use the "front page" bibliographic data for patents published by the United States Patent and Trademark Office (USPTO) as the basis for most of the empirical work. These data contain the application date and issue date of each patent, the names and locations of the inventor(s), a technology classification, and a list of other patents cited. We augment these data with the NBER Patent-Citations data file for additional fields including the 1-digit technology category code, the 2-digit subcategory code, and the assignee code.

[18] Such a patent would only generate 5 observations using the JTH method since their unit of analysis is the patent-citation, rather than the inventor-patent-citation.

We begin with the full set of issued patents that have their application year as 1990.[19] There are 108,672 such patents. From these, we select the set of patents that are from North America.[20] There are 60,974 such patents. We then discard all patents that have not received any citations, since our study is based on examining citations as a proxy for knowledge spillovers and social ties. We don't believe this elimination results in selection bias since we are interested in comparing the fraction of citations that are from the same location as the original patent, a measure that is conditional on there being citations. Consequently, approximately 8.7% of the remaining patents are discarded, leaving 55,664 as the set of "originating patents" that form the basis of the empirical analysis.

Each of the originating patents has an average of approximately 10.2 citations, resulting in 568,960 unique inventor-patent-citation observations. A small fraction of these observations are removed because their citing patents do not map to an MSA or because our process for generating control patents is not able to find an adequate control for the citing patent. This process reduces the number of observations to 564,590. Next, we discard the 11.6% of observations for which the citing patent is a self-citation by one or more of the inventors,[21] leaving us with 499,341 observations. Finally, we "unbundle" individual inventors, of which there are an average of approximately two per patent, resulting in a final sample size of 992,362 observations.

We follow an identical procedure for generating the 1989 dataset. There are slightly fewer North American patents in 1989 (56,896 rather than 60,974). Ultimately, we generate 938,419 observations for the 1989 cohort. Finally, we also generate 555,962 and 528,148 observations for the 1980 and 1975 cohorts, respectively.

[19] We replicate the entire study with 1989 patents, which is a completely distinct set from 1990, and also report these results throughout the paper. In all cases, the results are similar across the two years. In addition, we conduct some analyses with 1975 and 1980 data in order to offer a direct comparison with the JTH study, which uses data from those years.

[20] We use the geographic assignment procedure developed by JTH to determine whether patents are from North America. This procedure works as follows. Where there is a single inventor, the patent is assigned to the location of that inventor. Where there are multiple inventors, the patent is assigned to the location of the majority of inventors. In other words, if there are two inventors from the Boston and one inventor from Paris, the patent is assigned to Boston. If there is a tie across inventor locations (e.g., one inventor in Boston and the other in Paris), the patent is randomly assigned to one of these locations. Finally, for the section where we compare our data directly to that of JTH, it is important to note that some North American inventors are located in regions that are not mapped to an MSA. In these cases, we assign the patent to a "phantom MSA." Phantom MSAs are created for each US state and Canadian province.

[21] We consider assignee name matches or inventor name matches as self-cites. This is perhaps a stricter definition than often used in citation-based empirical research, which often only considers assignee name matches as self-citations. Since we are particularly concerned with "movers," we want to eliminate the possibility of an individual citing their own prior work while at a new firm and thus filing under a new assignee name, since this does not represent a spillover.

Every observation is assigned to an "originating location" based on the home address of the inventor. Inventors are assigned to an MSA based on their city and state information.[22] There are 268 US MSAs and consolidated metropolitan statistical areas (CMSAs) and 25 Canadian census metropolitan areas (CMAs) – hereinafter collectively referred to as the "MSAs."[23] We have also created 63 "phantom MSAs" for individuals located in one of the 50 states or 13 provinces or territories that are in cities not assigned to one of the Census Bureau-defined MSAs.

Finally, our observations are not distributed evenly across MSAs. In fact, the ten largest MSAs, in terms of number of observations where the inventor is located in that MSA, account for almost half the sample. This is illustrated in Table 1. As described in the methods section, we deal with the heavily skewed nature of these data by constructing a set of control patents that is intended to account for the uneven distribution of innovative activity across geographic space.

V. Results

In this section we report results from testing our two hypotheses. In addition, since our methodology builds heavily on that employed by JTH, we compare the results generated from our method with those generated from their method in the context of their hypothesis concerning the localization of knowledge spillovers. The results from this comparison are reported in the Appendix. Our data and methods confirm their earlier findings.

H1: The Enduring Social Capital Hypothesis

Here we focus our attention on the premium captured by the inventor's previous location. Once an inventor has moved, they are gone — but are they forgotten? We hypothesize that inventors who move are likely to maintain social ties with some individuals from their former location. Since new knowledge is characterized as having an important tacit component, and tacit knowledge is often communicated

[22] City and country information is used for assigning Canadian inventors to a CMA.

[23] While MSAs and CMAs are similar in spirit, they are defined slightly differently. The Canadian criterion requires that the urban core have a population of at least 100,000 for a metropolitan area to exist. In contrast, for the period 1990 to 2000, the United States had two criteria to determine whether or not a metropolitan area existed: 1) where there is either a city of 50,000 or more inhabitants or 2) where there is a Census Bureau-defined urban area, i.e., a population of at least 50,000 and a total metropolitan population of at least 100,000 (75,000 in New England). Thus, the Canadian approach is the more restrictive of the two.

informally through social ties, individuals from the inventor's former location are likely to be better able to build on that inventor's spillovers.

We identify movers as individuals who have patented prior to their original patent in a North American MSA that is different from the MSA they were in at the time of their current patent. We follow the same procedure as the one described earlier for the post-1990 movers, except we search for previous rather than subsequent patents (pre-1990 movers). Again, if there is more than one, we select the patent with the application date that is closest in time to that of the original patent.

Our results, presented in Table 2, support the hypothesis. The frequency of citing patents matching the inventor's previous location is significantly greater than the frequency of control patents matching the inventor's previous location. The premium associated with the prior location is 1.6% and 1.7% for 1989 and 1990, respectively. Stated another way, for the 1990 data citing patents are 50% more likely than control patents to be located in the inventor's previous location.[24] Using equation (5), this finding can be interpreted as indicating that prior co-location increases the probability of a citation by 50 percent. Note that though the prior location premium is only half the magnitude of the current co-location premium, it is still highly statistically significant.

Also, note that the proportion of control location matches is substantially larger in the current location than in the prior location (5.1% compared to 3.6% in 1989 and 5.4% compared to 3.4% in 1990). This suggests that, on average, movers relocate to regions where there is more activity in their technology area. While this may not seem surprising, the magnitude of the difference in levels of activity between prior and current locations is quite large. For example, on average, a control 1990 patent is 59% more likely to be found in the inventor's current location than in their prior location.[25]

Impact of ex-ante differences in mobility

As argued above (equation (4)), if local communications are more valuable than long-distance communications, an increase in the probability of separation will lower the equilibrium level of investment in co-located social relationships: Thus we expect that the current location premium will be smaller for inventors with a higher *ex ante* probability of moving. This probability is difficult to measure directly, but we can proxy for it by looking at the average *ex ante* differences in location premia for groups of inventors who we know *ex post* will move at different points in the future. Specifically we

[24] 1.7/3.4 = 0.50.

[25] (5.4-3.4) / 3.4 = 0.588

assume that inventors who will *ex post* turn out have moved in the near future have higher prospective mobility *ex ante* than do inventors who will *ex post* turn out to have moved in later years, who in turn have higher prospective mobility *ex ante* than inventors who we know *ex-post* did not move.

Table 3 presents results of computing the current location premium for these three groups: immediate movers (who will move to a new location within one year), future movers (who will move to a new location more than one year into the future), and non-movers. For both 1989 and 1990 samples, the differences in co-location premia are consistent with the prediction of the model. Non-movers have the highest current location premium, followed by future movers, with immediate movers having the lowest current location premium.

Sample selection

It is also important to recognize that in focusing on movers, our sample size drops dramatically. Though we still have almost 60,000 observations on movers, this is only slightly more than 6% of the full sample of inventor-patent-citation observations, raising the possibility that selection bias is affecting our results. To investigate whether movers are systematically different than the original sample population of inventors, we compare the two samples in terms of spillovers to the inventor's current location, as well as other patenting characteristics. Table 4a presents the results from comparing these two groups in terms of the co-location premiums. Movers' spillovers to their current location do not appear to be measurably different than those associated with the full sample.

Finally, we also compare movers with the general population along other dimensions in Table 4b. We see the two samples are similar in terms of "importance" as measured by the average number of citations received, in terms of distribution across types of assignees, and in terms of distribution across technical categories, although movers seem to be less concentrated in computers/communications and more concentrated in chemical and drugs/medical than the full sample. Though we find no reason to believe that systematic differences between movers and non-movers are driving our results, we intend to look at differences in these two groups in greater detail in future research.

Causal Interpretation: Social Capital versus Distance

Our hypothesis is that knowledge spillovers by movers go disproportionately to their prior locations, relative to the case where they had never lived in that prior location. Moreover, we hypothesize that the effect is *causal* in the sense that the inventor accumulated social capital that is specific to their prior location, and that the enduring element of this social capital facilitates subsequent communication between the inventor in question and other inventors in their prior location.

The findings reported above are certainly consistent with this hypothesis. Citations occur disproportionately from the inventor's prior location when compared to the geographic distribution of a well-specified set of control citations. But we could observe such disproportionate citing to prior locations without the relationship being causal. Our greatest concern is that there is some omitted variable that affects both labor flows and knowledge flows. The most likely candidate for such an omitted variable is distance.

For example, suppose we look at the geographic distribution of citations to a 1990 patent with a New York inventor address, and that this inventor is observed to have applied for a patent in 1985 from a Boston address. Furthermore, we observe that a disproportionate number of cites to the 1990 patent occur from Boston, the inventor's 1985 location. We are tempted to view this as evidence for our enduring social capital hypothesis, in the sense that Boston inventors are disproportionately citing the 1990 New York inventor because they continue to communicate with that inventor through the relationships and networks developed when they were living and working in Boston.

However, an alternative explanation for the disproportionate cites is that Boston is relatively close to New York, so that New York inventors interact more regularly with Boston inventors than they do with inventors who live further away. If it is also true that, conditional on having moved from somewhere, a 1990 inventor is more likely to have moved from somewhere close (Boston to New York in our example), we will observe disproportionate cites to prior locations in our data even without the causal effect we hypothesize being present. Put differently, distance affects both the probability of citation in a given location and the probability of that location being the inventor's prior location, and our results may therefore be confounding the effect of distance with any effect of enduring social capital.

Therefore, we attempt here to identify the causal effect. The identifying assumption is that distance is the omitted variable, and that the social capital-based causal effect is invariant to how far the inventor has moved. This is clearly a strong assumption. It is conceivable that distance affects the durability of social capital, since it may be costlier for former neighbors to maintain social ties the further apart from each other they live.[26] If distance does affect the durability of social capital, however, this identification

[26] It would also be interesting to check for any influence of locations being in the same time zone, or frequency of non-stop flights or other factors affecting the cost of maintaining relationships.

strategy will bias against finding a causal effect, and we thus see our results as providing a lower bound for this effect. To implement this strategy, we choose a matched location for each mover that is approximately the same distance from the inventor's final location as the distance between the inventor's prior location and their final location. The difference between the premia for the actual and matched prior locations is then identified as the causal effect.

Consider an originating patent filed in 1990 by an inventor living in Austin, TX. Suppose that the likelihood of citation to the 1990 patent by any given location is, all else equal, negatively related to the distance of that location from Austin. Thus, for example, the likelihood of a citation from Denver, CO is greater than the likelihood of a citation from Portland, OR since Denver is closer than Portland to Austin. Suppose further that given an inventor has moved to Austin, the likelihood that the inventor came from any particular location is, all else equal, also negatively related to the distance of that location from Austin. It follows that citations will disproportionately occur from the locations from which inventors moved, even when there is no social capital-based causal effect.

An obvious way to isolate the causal effect is to find control locations that match the inventor's previous location in terms of distance from their 1990 location. We use a two-step procedure for identifying "matching MSAs." First, we measure the distance between the inventor's prior location and their 1990 location.[27] Then, we identify all other MSAs that are the same distance from the inventor's 1990 location, plus or minus 100 miles.[28] From this set of MSAs, we select the MSA that is closest to the inventor's previous location in terms of number of patents.[29,30] Thus, we select a control MSA that is similar to the inventor's previous MSA in terms of both its distance from their 1990 location and its level of technological activity. Figure 2 illustrates an example. In this case, the inventor moved from Portland, OR to Austin, TX. A band is created to identify all other MSAs that are approximately the same distance as Portland is from Austin. Over 20 such MSAs are identified. Portland is a mid-sized MSA in terms of

[27] For this step, we measure the distance from city to city.

[28] For this step, we measure the distance from MSA to MSA. The distance between MSAs is measured between the largest city within each of the MSAs.

[29] For the purposes of comparing patenting activity across MSAs, we use patents with 1990 (or 1989) application dates and assign patents to MSAs by inventor location, using the "majority rules" location determination method, as described in the text.

[30] Not all candidates within this set are likely to be equally good matches for the actual prior location. Given the tendency for inventive activity to cluster, one concern is that the actual prior locations are more likely to be major metropolitan areas than a randomly chosen location from the candidate set. Although this concern is partly allayed if we compare the actual citation pattern with the control citation pattern for both the actual and the matching prior location—that is, use a difference-in-difference estimation approach—we are still concerned that matching prior locations are systematically different from the actual prior locations. For this reason, we apply the following rule for choosing the matching location from the candidate set: Choose the location that comes closest to the actual prior location in terms of overall 1990 patent applications.

inventive activity. From the set of MSAs that satisfy the distance from Austin criteria, Seattle, WA has the closest number of 1990 patents and is therefore selected as the control MSA for that observation.[31]

We then compute the difference-in-differences between citing and control patents for prior and matching MSAs. The results are shown in Table 5 for both 1990 and 1989 data. Again, we focus only on the 1990 results, as the two sets of results are almost identical. The prior location premium in the actual prior locations (1.6%) is substantially higher than it is in the matched prior locations (0.6%). The difference in the prior location premium for actual and matched samples—i.e., the difference-in-difference—is clearly statistically significant with a z-stat of almost 14.

H2: Knowledge Transfer across Technology Fields

As argued above, spillovers of knowledge across technology fields (or between communities of practice) may rely more strongly on co-location facilitated social relationships than do spillovers within a technology field. Because inventors have fewer alternative means for accessing new knowledge between different technology fields than they do for accessing new knowledge within their own field, the marginal benefit of geographically based social relationships may be higher for spillovers across fields than within fields.

We test this hypothesis by splitting the sample into two groups, those observations in which the citing patent is from the same field as the original patent and those in which it is not. We employ two schemes for classifying patents as being from the same field. First, we classify those citing patents with the same 2-digit NBER patent subcategory classification as the original patent as being from the same field. Next, we classify those citing patents with the same 3-digit US patent classification as the original patent as being from the same field. The two methods produce similar results[32]; only the NBER subcategory results are reported here.

The results are presented in Table 6 for both the current and prior co-location premia. Table 6a examines the difference in the current co-location premium for both the 1989 and 1990 data for the full sample of inventors; Table 6b examines the difference in the both the current and the prior co-location premia for

[31] Just as we only consider movers who have moved from one North American MSA to another, we restrict our search for matching MSAs to North America.

[32] We compare classification schemes in the context of the “Drugs & Medical” category to offer some sense of the relationship between these two. This NBER defined *category* encompasses four *subcategories* and fourteen 3-digit *US classifications*. For example, the four subcategories include drugs, surgery & medical instruments, biotechnology, and miscellaneous. One of these subcategories, drugs, encompasses two US classifications, 424 and 514, which are both described as “drug, bio-affecting and body treating compositions.”

the sample of movers using only the 1990 data.[33] Clearly, the percentage of (current and prior) location matches is very similar for the two types of spillovers. Given the aforementioned fact of geographical concentration of activity by technological field, however, it is not surprising that the percentage of location matches for the controls is greater for the within-field spillovers in all cases.[34] It follows that the co-location premium (% citing matches - % control matches) is greater for the across-field spillovers. For example, the current co-location premium is 2.1 percentage points higher for the 1990 data.

It is perhaps more revealing to look at the ratio co-location premium (current and prior) to the percentage of matches in the control sample. As discussed in Section III, using Bayes Rule this ratio can be interpreted as the proportionate increase in the probability of a citation that is associated with co-location. Looking at the 1990 data, we see that co-location results in a 76 percent increase in the probability of a cross-field citation, but only a 34 percent increase in the probability of a within-field citation. The results are similar for prior co-location. Prior co-location results in a 63 percent increase in the probability of a cross-field citation, but only a 36 percent increase in the probability of within-field citation. Thus co-location (current and prior) does indeed appear to be most important in supporting knowledge spillovers when the inventors are working in different technological fields.[35]

VI. Conclusion

Much has been written about the "death of distance." Modern information and communications technologies are thought to have diminished the obstacles to economic interaction created by geographic separation. Yet the tendency for high technology industries to be geographically clustered[36]—industries whose knowledge-intensive outputs are essentially weightless—suggests that proximity to sources of knowledge spillovers as inputs to R&D is critically important.

[33] Almost identical results were obtained for the 1989 data.

[34] Recall that the control patents are chosen to maximize the likelihood that they have the same technology classification as the actual citing patent. The reason that we expect to see large number of location matches for the control patents is that technological activity is geographically concentrated by field. If, however, we limit attention to citations that are in a different field from the originating patent, our control citations will also be (by construction) in a different field. Thus we are less likely to see location matches for this subset of the data.

[35] We hasten to add that this does not imply that co-location leads to a larger absolute increase in the probability of a citation for cross-field inventor-patent pairs. The reason is that the unconditional probability of a citation is likely to be smaller for cross-field pairs than for within-field pairs (see equation (5) above).

[36] See, for example, the evidence in Audretch and Feldman (1996).

In this paper, we develop and test a model of knowledge spillovers that depends on social ties between inventors. In our model, proximity is essential to developing the social ties that facilitate communication, but we allow for the possibility that social ties endure even after individuals have become separated. In effect, geographical proximity works to overcome social distance and, once relationships are established, individuals can remain socially close even when they become geographically separated. We explore empirically how the prospect of separation affects the extent and form of social relationships that people develop and find evidence to support the hypothesis of an enduring social capital effect.

We think these results are interesting in the context of increasingly knowledge-based and geographically mobile societies. But we have only touched on the interesting questions they raise. What does an economic region lose when a portion of its highly skilled workforce leaves—but is not (completely) forgotten? What does an economic region gain when it attracts highly skilled workers who remain networked to their former peers? How does the shift to a more mobile society affect individual incentives to develop economically useful social relationships? What are the implications of increased intranational and international mobility for the diffusion of technological knowledge and thus for regional and national government incentives to fund research and development? Answers to these questions may have great significance for policy makers interested in regional differences in growth and prosperity, as well as for individuals or firms making privately optimizing location decisions.

Appendix

Here we confirm that although we have modified the JTH methodology by using a different unit of analysis (inventor-patent-citation rather than patent-citation) our method produces similar results with respect to estimating the localization of spillovers. As described in the methodology section, we compare the probability that citing patents are from the same location as the originating patent with the probability that control patents selected to match the citing patents in terms of timing and technology classification are from the same location as the originating patent. The results are presented in Table 7a.

These results strongly support the hypothesis that spillovers are geographically localized; the proportion of citing patents that match the location of their originating patents is significantly greater than that of control patent location matches. The z-statistics, which test the equality of the proportion of citing-original versus control-original location matches, are large, with p-values less than 0.001. Thus we confirm the JTH result in this much larger and somewhat more finely observed sample.

It is interesting (and intuitively appealing) to note that our results also support the notion that the localization of spillovers decreases over time as knowledge diffuses across geographic space. The co-location premium is 3.1% for 1990 data which measures spillovers over a 12-year period, but only 2.3% for 1975 data which measures spillovers over a 27-year period.

It is also interesting to note that the 1990 and 1989 data generate almost identical results even though they are based on distinctly different data sets. While they include spillovers over approximately the same duration, the data sets are based on two completely different sets of originating patents. Nonetheless, the 1990 and 1989 samples generate very similar co-location premiums of 3.1% and 3.2%, respectively. The magnitude of this premium represents a 56% (1990) to 62% (1989) increase in the number of patents that cite the original patent in the location of the original patent, relative to the case where the patent had not come from that location.[37]

We described in the methodology section how our unit of analysis differs from the unit of analysis in JTH. In Table 7b we present the co-location premium from our data using the JTH patent-citation unit of analysis. Note that the numbers of observations in Table 7a are larger than those in 7b, reflecting the average 1.5 to 2.0 inventors per patent in the years under investigation. Most importantly, note that the

[37] *3.1/5.5 = 56.4%, 3.2/5.2=61.5%*

premiums listed in 7a are very similar to those listed in 7b, suggesting a similarity in the magnitude of spillovers as measured using these different units of analysis.

Finally, we directly compare our data to that of JTH in Table 7c using their unit of analysis. The primary difference is that we use all patent-citation observations available for that year (approximately 160,000), whereas they use small samples (approximately 2,000) categorized by assignee type. Note that the sample used to generate Table 7c is smaller than that used to generate 7b since we only include citations up to 1989 in order to be consistent with JTH.

There are two important observations from Table 7c. First, our proportion of citing patent matches falls within the range measured by JTH. Our measured values likely reflect the weighting of particular types of assignees (e.g., university versus corporate) in the general distribution of patents from 1975 and 1980. Second, our proportion of control patent matches is measurably higher (4% compared to 1-3%) than the proportion measured by JTH. This may suggest that general patenting activity is more geographically concentrated than that reflected in the categories focused on by JTH, namely university, top corporate, and other corporate. The implication of Table 7c is that the nature of our sample (higher proportion of control patent matches and thus lower co-location premiums) decreases the likelihood of finding support for our hypotheses.

References

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

Borjas, G. (1995) "The Economic Benefits of Immigration," *Journal of Economic Perspectives*, Vol 0, Issue 2, Spring, pp. 3-22.

Brown, J.S. and P. Duguid (1991) "Organizational Learning and Communities of Practice: Towards a Unified View of Working, Learning, and Innovation," *Organization Science*, Vol. 2, pp.40-57.

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

Cockburn, I., S. Kortum, and S. Stern (2002) "Are All Patent Examiners Equal? The Impact of Examiner Characteristics on Patent Statistics and Litigation Outcomes," NBER Working Paper No. 8980.

Crane, D. (1965) "Scientists in major and minor universities: a study of productivity and recognition," *American Sociological Review*, Vol. 30, pp. 699-714.

Crane, D. (1969) "Social Structure in a Group of Scientists: A Test of the 'Invisible College' Hypothesis," *American Sociological Review*, Vol. 34, pp. 335-352.

Desai, M., D. Kapur, and J. McHale (2002) The Global War for Talent: Implications and Policy Options for Developing Countries, forthcoming as a monograph from the Center for Global Development/Institute of International Economics, Washington D.C.

Fukuyama, F. (1995) Trust, Basic Books: New York, NY.

Glaeser, E.L, D. Laibson, and B. Sacerdote (2002) "The Economic Approach to Social Capital," *Economic Journal*, August.

Granovetter, M.S. (1973) "The Strength of Weak Ties," *American Journal of Sociology*, Vol 78, pp. 1360-1380.

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

Jaffe, A., M. Trajtenberg, and M. Fogarty (2002) "The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees" in Patents, Citations, and Innovations, A. Jaffe and M. Trajtenberg, eds., MIT Press, Cambridge, MA.

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

Krugman, P. (1991) Geography and Trade, The MIT Press: Cambridge, MA.

Lave, J. and E. Wenger (1993) Situated Learning: Legitimate Peripheral Participation, New York, NY, Cambridge University Press.

Marshall, A. (1920) Principles of Economics, Macmillan: London, UK.

Portes, A. (2000) “Social Capital: Its Origins and Applications in Modern Sociology” in Knowledge and Social Capital, E. Lesser ed., Butterworth-Heinemann, Woburn, MA.

Putnam, R. D. (2000) Bowling Alone, Simon & Schuster: New York, NY.

Rauch, J.E. (2001) “Business and Social Networks in International Trade,” *Journal of Economic Literature*, Vol. XXXIX, No. 4, pp. 1177-1203.

Romer, P. M. (1986) "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, Vol. 94, pp. 1001-37.

Romer, P. M. (1990) "Endogenous Technological Change," *Journal of Political Economy*, Vol. 98, supplement to No. 5, pp. 71-102.

Saxenian, A. (2002) *Local and Global Networks of Immigrant Professionals in Silicon Valley*, Public Policy Institute of California.

Trajtenberg, M., A. Jaffe, and R. Henderson (1997) “University versus Corporate Patents: A Window on the Basicness of Invention,” *Economics of Innovation and New Technology*, 5(1), pp. 19-50.

Zucker, L., M. Darby, and M. Brewer (1998) “Intellectual Capital and the Birth of U.S. Biotechnology Enterprises,” *American Economic Review*, Vol. 88, pp. 290-306.

Table 1
10 Largest MSAs in terms of Number of Observations in Dataset

MSA	**Number of observations with inventor of originating patent from specified MSA**	**Percentage of total observations**
SAN FRANCISCO OAKLAND SAN JOSE, CA	99,414	10.0%
NEW YORK NORTHERN NEW JERSEY LONG ISLAND, NY NJ CT	98,833	10.0%
BOSTON WORCESTER LAWRENCE, MA NH ME CT	57,503	5.8%
LOS ANGELES RIVERSIDE ORANGE COUNTY, CA	53,186	5.4%
CHICAGO GARY KENOSHA, IL IN WI	43,238	4.4%
MINNEAPOLIS ST. PAUL, MN WI	29,059	2.9%
PHILADELPHIA WILMINGTON ATLANTIC CITY, PA NJ DE MD	28,000	2.8%
DETROIT ANN ARBOR FLINT, MI	24,240	2.4%
DALLAS FORT WORTH, TX	22,681	2.3%
ROCHESTER, NY	20,714	2.1%
Total for 10 largest MSAs	**476,868**	**48.1%**

Table 2
Spillover Premiums associated with Movers: Current versus Prior Locations
Percent of citing/control patents in current/previous MSA

	1989		1990	
	Matching with current (1989) location	Matching with previous location	Matching with current (1990) location	Matching with previous location
% Citing matching	8.1	5.2	8.4	5.1
% Controls matching	5.1	3.6	5.4	3.4
Co-location premium	3.0	1.6	3.0	1.7
z-statistic	20.75	13.78	20.64	14.45
Co-location Premium / % Controls Matching	0.58	0.44	0.56	0.50
n	57,878	57,878	59,734	59,734

Table 3
Expected Mobility and Investments in Social Capital
Percent of citing and control patents in the same MSA as the originating patent

	1990			1989		
	Non-Movers	Immediate Movers	All movers	Non-Movers	Immediate Movers	All movers
Years citation data	12	12	12	13	13	13
% Citing matching (excl. self-cites)	8.6	6.6	8.1	8.4	7.2	8.2
% Controls match	5.5	4.8	5.7	5.2	5.1	5.8
Co-location prem.	3.2	1.7	2.5	3.2	2.1	2.4
z-statistic	82.66	5.7	21.14	84.07	6.47	19.92
Co-location Premium / % Controls Matching	0.58	0.35	0.44	0.66	0.41	0.41
n	896,522	11,663	95,840	848,588	11,071	89,831

Table 4a
Comparing Movers with the General Population: Patenting Characteristics (1990)

	General population	Movers
	n=990,524[38]	n= 62,817
Average number of citations received	21.8	20.5
Assignee Code		
Unassigned	13.5%	9.4%
Assigned to a US non-government org.	81.8%	87.0%
Assigned to a non-US, non-government org.	2.1%	1.7%
Assigned to a US individual	1.0%	0.7%
Assigned to a non-US individual	0.0%	0.0%
Assigned to the US (Federal) Government	1.5%	1.0%
Assigned to a non-US government	0.1%	0.1%
Technological Category		
Chemical	14.3%	19.8%
Computers & Communications	25.3%	14.4%
Drugs & Medical	15.4%	20.9%
Electrical & Electronic	16.4%	18.4%
Mechanical	12.4%	12.0%
Other	16.2%	14.7%

[38] This dataset was generated by merging our database with the patent data available on the NBER website.

Table 4b
Comparing Movers to General Population: Spillovers to their Current Location
Percent of citing and control patents in the same MSA as the originating patent

	1989		1990	
	Full Sample	Movers	Full Sample	Movers
% Citing matching	8.4	8.1	8.6	8.4
% Controls matching	5.2	5.1	5.5	5.4
Co-location premium	3.2	3.0	3.1	3.0
z-statistic	86.08	20.75	85.15	20.64
Co-location Premium / % Controls Matching	0.62	0.58	0.56	0.56
n	938,419	57,878	992,362	59,734

Table 5
Spillover Premiums:
Comparison of Actual Prior Locations with Prior Locations Matched on Distance

	1989	1990
Actual Prior Locations		
% Citing matching	4.2	4.1
% Controls matching	2.7	2.6
Prior co-location premium	1.5	1.6
z-statistic (p > z)	11.98 (0.00)	12.97 (0.00)
Prior co-location premium / % Controls matching	0.55	0.61
n	43100	44592
Matched Prior Locations		
% Citing matching	1.3	1.5
% Controls matching	0.8	0.9
Prior co-location premium	0.5	0.6
z-statistic (p > z)	6.69 (0.00)	8.30 (0.00)
Prior co-location premium / % Controls matching	0.62	0.67
n	43100	44592
Difference in Differences		
Difference in prior co-location premium between actual and matched prior locations	1.0	0.9
z-statistic (p > z)	15.24 (0.00)	13.84 (0.00)

Table 6a
Spillover Premiums (Full Sample):
Co-location Premiums for Cross-Field versus Within-Field Spillovers

	1989	**1990**
Cross-Field Spillovers		
% Citing matching	8.6	8.8
% Controls matching	4.7	5.0
Co-location premium	3.8	3.8
Co-location premium / % Controls matching	0.79	0.76
z-statistic	79.33	80.25
N	528,047	577,851
Within-Field Spillovers		
% Citing matching	8.2	8.3
% Controls matching	5.9	6.2
Co-location premium	2.3	2.1
z-statistic	40.62	37.44
Co-location premium / % Controls matching	0.39	0.34
N	410,372	414,511

Table 6b
Spillover Premiums (Movers):
Prior Co-location Premiums for Cross-Field versus Within-Field Spillovers (1990)

	Current Location	**Prior Location**
Cross-Field Spillovers		
% Citing matching	8.3	5.2
% Controls matching	4.6	3.2
Prior co-location premium	3.8	2.0
z-statistic	19.85	12.77
Prior co-location premium / % Controls matching	0.82	0.63
N	33,618	33,618
Within-Field Spillovers		
% Citing matching	8.4	5.0
% Controls matching	6.4	3.6
Prior co-location premium	2.1	1.3
z-statistic	9.00	7.40
Prior co-location premium / % Controls matching	0.33	0.36
n	26,116	26,116

Table 7a
Geographic Localization of Knowledge Spillovers
Percent of citing and control patents in the same MSA as the originating patent
Inventor-Patent-Citation Unit of Analysis (ACM)

	1990	1989	1980	1975
No. yrs. citation data	12	13	22	27
% Citing matching (excluding self-cites)	8.6	8.4	6.6	6.5
% Controls matching	5.5	5.2	4.3	4.2
Co-location premium	3.1	3.2	2.3	2.3
Co-location premium / % Controls Matching	0.56	0.61	0.53	0.55
z-statistic	85.15	86.08	54.47	52.22
n	992,362	938,419	555,962	528,148

Table 7b
Geographic Localization of Knowledge Spillovers
Percent of citing and control patents in the same MSA as the originating patent
Patent-Citation Unit of Analysis (JTH)

	1990	1989	1980	1975
No. yrs. citation data	12	13	22	27
% Citing matching (excluding self-cites)	7.4	7.3	5.8	5.8
% Controls matching	4.6	4.4	3.7	3.7
Co-location premium	2.8	2.9	2.1	2.0
Co-location premium / % Controls Matching	0.61	0.66	0.57	0.54
z-statistic ($p > 0$)	59.61 (0.000)	60.64 (0.000)	41.32 (0.000)	39.91 (0.000)
n	499,341	480,744	348,631	345,028

Table 7c
Comparison of ACM to JTH Data
Percent of citing and control patents in the same MSA as the originating patent
Patent-Citation Unit of Analysis (JTH)

	1975 Originating Cohort				1980 Originating Cohort			
	ACM	JTH			ACM	JTH		
	All types	Univ.	Top Corp	Other Corp	All types	Univ.	Top Corp	Other Corp
Years citation data	14	14	14	14	9	9	9	9
% Citing matching (excl. self-cites)	5.9	4.3	4.5	8.7	5.6	6.9	8.8	7.0
% Controls match	4.3	1.0	1.3	1.2	4.1	1.1	3.6	2.3
Co-location prem.	1.6	3.3	3.2	7.5	1.5	5.8	5.2	4.7
z-statistic (ACM) / t-statistic (JTH)	20.95	6.43	4.80	8.24	15.84	9.57	6.28	5.52
Co-location premium / % Controls Matching	0.37	3.30	2.46	6.25	0.37	5.27	1.44	2.04
N	160,726	1,759	1,235	1,050	108,271	2,046	1,614	1,210

Figure 1. Set of profitable social relationship investments (for given C_{ij} and m_{ij})

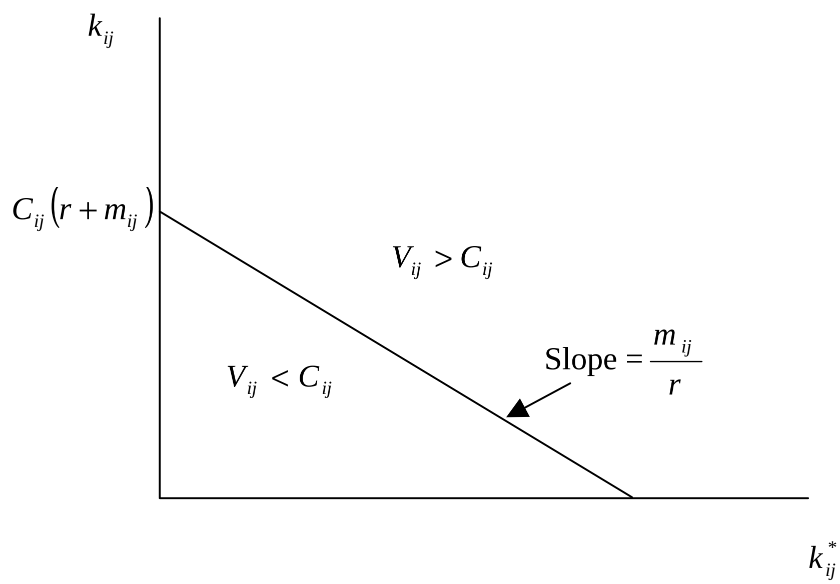

Figure 2 - Selection of Matching MSAs