

Johnson, Daniel K.N.; Mareva, Milena

**Working Paper**

## It's a Small(er) World: The Role of Geography and Networks in Biotechnology Innovation

Wellesley College Working Paper, No. 2002-01

**Provided in Cooperation with:**

Department of Economics, Wellesley College

*Suggested Citation:* Johnson, Daniel K.N.; Mareva, Milena (2002) : It's a Small(er) World: The Role of Geography and Networks in Biotechnology Innovation, Wellesley College Working Paper, No. 2002-01, Wellesley College, Department of Economics, Wellesley, MA

This Version is available at:

<https://hdl.handle.net/10419/23232>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

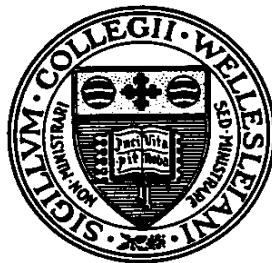
**WORKING PAPER**

**It's a Small(er) World:  
The Role of Geography and Networks  
in Biotechnology Innovation**

**By**

**Daniel K.N. Johnson and Milena Mareva**

**Wellesley College Working Paper 2002-01  
January, 2002**



**Department of Economics  
Wellesley College  
Wellesley, Massachusetts 02481-8203**

## **Abstract**

Using patent citation data for the U.S., we test whether knowledge spillovers in biotechnology are sensitive to distance. Controlling for self-citation by inventor, assignee and examiner, cohort-based regression analysis shows that spillovers are local but that distance is becoming less important with time. Network analysis paints a picture of a stable network between states, but a changing environment between individual actors, with a growing importance of connectedness. The popular maxim that everyone is connected by six degrees of separation is tested with surprising results.

**JEL codes:** L6 --- Industry Studies: Manufacturing  
N9 --- Regional and Urban History  
O3 --- Technological Change  
R1 --- Urban, Rural and Regional Economics

**It's a Small(er) World:  
The Role of Geography and Networks in Biotechnology Innovation<sup>1</sup>**

**I. Introduction**

It has long been noted that firms within an industry often cluster geographically, for a variety of reasons. Localization economies, or Marshall-Arrow-Romer externalities which reduce the cost of inputs to firms in the local industry, have been studied in a variety of contexts (see for example Henderson (1986) or Smith and Florida (1994)). For some industries, it is the nature of the knowledge itself, its speed of progress and the degree to which it is tacit, that encourages firms to locate near other firms in the sector. Both of these aspects of knowledge, diffusion speed and tacitness, have been identified in the literature (see Caballero and Jaffe (1993) for diffusion speed and Von Hippel (1994) for tacitness) but to our knowledge no analysis has measured their strength in biotechnology, nor controlled for all factors which we include.

This paper examines knowledge flows within biotechnology, showing that inter-firm knowledge transfers decrease with distance, but that the impact of physical distance has been diminishing over time. Thus, historically there has been a reason for biotechnology firms to cluster geographically, although that reason is less powerful now than ever before. In addition, we show that the networks between individual inventors,

---

<sup>1</sup> **Daniel K.N. Johnson** is a Visiting Fellow at the Science, Technology and Innovation Policy Program, Center for International Development, Harvard University and Assistant Professor of Economics at Wellesley College, Pendleton East, Room 415, Wellesley College, 106 Central Street, Wellesley, MA 02481-8260, tel: (781) 283-2236, fax: (781) 283-2177, email:djohnson@wellesley.edu. **Milena Mareva** is a Research Analyst, Analysis Group/*Economics*, One Brattle Square, Fifth Floor, Cambridge, MA 02138, tel: (617) 234-8862, fax: (617) 864-3742, e-mail: mmareva@analysisgroup.com. We thank David Popp for suggestions about data and programming. Financial support was provided by a Wellesley College Faculty Grant.

between firms, and between cities hosting those firms and inventors, have become more tightly knit than ever before. The conclusion is that while physical distance is becoming less important, connections with leaders in the field are becoming more important.

Using all biotechnology patents granted in the U.S. between 1975 and 1994, we show that there is a marked tendency to cite patents from nearby areas. Thus, this paper follows the "paper trail" literature (Feldman, 1999) on inter-firm knowledge transfers or "spillovers" in documenting the creation of knowledge via patent citations. We test whether this pattern could naturally arise from a tendency to cite other patents listing the same inventor, the same firm assignee, the same patent examiner, the same technology class or states with many biotechnology patents to cite. We conclude that the geographic clustering of citations holds over and above the effects of these other factors, suggesting that there is a local nature to knowledge spillovers. Since we control for the locations of other active firms in our analysis, this result provides solid evidence that localized knowledge spillovers are a strong factor in the productivity of firms, and therefore in their location decisions. Subsequent network analysis confirms that inter-personal and inter-city connections are becoming more important with time.

In section II of the paper, we briefly review the relevant literature on biotechnology clustering and the geographic nature of knowledge spillovers. Section III describes our data set, designed to permit analyses parallel to both U.S. and European literature on geography-based spillovers, controlling for variables omitted from their analyses while testing the same hypotheses. Section IV presents regression analysis that controls for non-geographic effects in presenting the importance of state-level clustering. Section V uses network analysis to identify the important actors and interactions in the

data, demonstrating that while geographic impacts on knowledge flows may be diminishing, the role of connectedness with the center has been rising in importance. Section VI concludes with implications for policy and further research.

## **II. Literature review**

The literature suggests that knowledge spillovers cluster geographically, with higher spillovers (shown by more patent citations) within a short distance. The underlying supposition is that inventors are more aware of (or find more use for) inventions located close to them, and therefore build more heavily on local inventions, the result being a geographic clustering of citations.

Empirical evidence stresses the important role of geography in the spillover of knowledge from one member of an innovation network to another (see for example a review by Gelsing, 1992). It also emphasizes the importance of frequent personal contact and research collaboration.

The importance of geography differs predictably by technology type (Lundvall, 1992). While geography has little impact on stationary technologies (facing constant needs and opportunities), that importance grows quickly for technologies undergoing incremental innovation and radical innovation. During technological revolutions, there is a dramatic effect on the geographic pattern of subsequent innovations. Since biotechnology has enjoyed aggressive growth, we expect a large geographic impact on knowledge flows.

Geographic proximity has already been used to explain the location of R&D-intensive activities (for Boston's high-technology district, see Dorfman (1988); for France see Carrincazeaux et al., (2001)), due to evidence of localized spillovers within an

industry. However, the location of *firms* is not always a good predictor of the location of *innovation*, even within the same industry. Feldman (1994) tests whether U.S. state patent levels can be predicted simply by the presence of particular industries in that state. Test statistics are decisive, and in fact there is only a 0.42 correlation between innovation measures and value-added in each state by industry. This result is confirmed for the 1975-1994 period in the U.S. (Johnson, 1999) in an exploration of why the northeastern states lost a dramatic share of the national patenting total. It occurred not only due to the location of industries, but also due to industries that did not maintain the patenting rates of the same sector in other regions.

Localization of patent citations has been firmly established by the leading paper on the topic (Jaffe et al., 1993), with a random sample of patents clearly more likely to cite local patents than others at every geographically aggregated level. The effects are small but statistically significant, and are more intense where knowledge becomes obsolescent rapidly, like electronics, optics and nuclear technology (Jaffe and Trajtenberg, 1996). The result has been confirmed for semiconductors (Almeida and Kogut, 1997).

Since biotechnology knowledge becomes obsolescent very rapidly (see Johnson and Santaniello, 2000), one might expect that it will follow the same pattern. However, two factors augur against this quick conclusion. First of all, most biotechnological information is not tacit, so will be relatively easy to communicate across long distances. Second, biotechnology patenting has occurred largely during a period when international and inter-regional communication has been increasingly effective and affordable, so once again we might expect less localization of knowledge spillovers (Feldman, 1999).

Other researchers have demonstrated a geographic pattern to European patent citations. In a limited sample of Swedish patent applications, international trade flows rather than physical distance, was the only variable that robustly explained international references (Sjoholm, 1996). In a larger study of over 100,000 patent citations between European regions, there is strong evidence of geographic clustering (Maurseth and Verspagen, 1999). Regressions show that distance between regions is an important driving factor, along with technological similarity between regions.

Our data, as described in the following section, test both the U.S. (Jaffe and Trajtenberg, 1996) and European (Maurseth and Verspagen, 1999) versions of the geographic-clustering hypothesis, while controlling for factors those studies omitted.

### **III. Data**

#### ***III.A. Measurement issues***

This paper relies exclusively on patent citations from biotechnology patents as a geographic measure of knowledge spillovers in the sector. When a patent application is submitted for approval, it is accompanied by a list of citations to other patents and literature which have been instrumental in the creation of this technology, or which delineate the legal limits of this application. The intention is twofold: to build a convincing case that this application is novel and unobvious to someone trained in the field, and to provide a legal record of materials consulted during the invention process in order to protect patent rights in the future. To this list of citations, a patent examiner may add his or her own list of citations. The result is a paper trail of knowledge creation.

Of course, patents records do not perfectly reflect the creation of technology, as some innovations are never patented and patents vary greatly in size and importance.



However, within the U.S. on a state-by-state level, patents have a high correlation with other measures of innovative activity. For example, there is a 0.88 correlation between patents and R&D expenditures, 0.99 between patents and research employment records, and 0.93 between patents and a census of innovation citations in scientific and trade journals conducted by the Small Business Administration (Feldman, 1994).

Citations themselves do not perfectly reflect the transfer of knowledge, as they may be inserted for a variety of other reasons including legal protection or examiner privilege. Jaffe et al. (2000) relates survey evidence showing that only  $\frac{1}{4}$  of all patent citations correspond to a clear spillover of knowledge, another  $\frac{1}{4}$  have some possibility of a spillover, and the remaining  $\frac{1}{2}$  do not reflect knowledge transfers. However, statistical tests indicate that overall citations can be interpreted as a signal of spillovers, albeit a noisy signal.

As a final definitional challenge, "biotechnology" definitions differ between nations and over time (see Johnson and Santaniello, 2000). Therefore, we follow the most recent published biotechnology definitions of the U.S. Patent Office (USPTO, 1998), which include portions of eleven separate classes from the U.S. patent classification system.<sup>2</sup>

### ***III.B. Clustering of knowledge citations***

One reason for geographic clustering of citations in biotechnology would simply be the geographic clustering of biotechnology firms themselves. A number of control

---

<sup>2</sup> Specifically, the definition includes U.S. Patent Classes 47/1.1-47/1.4, 47/57.6-47758, 424/9.1-424/9.2, 424/9.34-424/9.81, 424/85.1-424/94.67, 424/130.1-424/283.1, 424/520-424/583, 424/800-424/832, 435/1.1-435/7.95, 435/40.5-435/261, 435/317.1-435/975, 436/500-436/829, 514/2-514/22, 514/44, 514/783, 530/300-530/427, 530/800-530/868, 536/1.11-536/23.74, 536/25.1-536/25.2, 800, 930, 935. We exclude class PLT due to data limitations on these documents.

variables are possible (e.g. number of firms, market value of firms) but in the biotechnology field in particular, those variables are questionable indicators. It takes more than the presence of other firms to create a citation, since citations are to particular patents, not firms. Therefore, in our analysis we control for the number of patents in each state, both biotechnological and other.

However, patent citations may also cluster for non-geographic reasons, coincidentally causing a pattern which appears geographic merely through correlation with other phenomena. For example, inventors may be more familiar with their own patents, citing them more frequently than others, which would give a biased impression of the importance of geography. The same may be true of assignees, if employees of a firm are familiar with other patents held by the firm. While inventor and assignee self-citation may drive a pattern of geographic clustering, they confuse the issue of "local knowledge spillovers," which is the primary focus of our analysis, so we describe and separate it. On the other hand, we do not wish to simply ignore self-citations as being obviously local. If an assignee firm is located in several different locations, high familiarity with other inventions by the same assignee may actually work against a geographic clustering of citations. The same may be true of an inventor who moves during his or her career. Therefore we include self-citations in the analysis but control for them separately.

All patent citations are reviewed, revised and potentially appended by examiners at the U.S. Patent Office. Due to the nature of patent records, it is impossible to verify whether a given citation was originally submitted by the applicant or added by an examiner, so we must treat examiners as another potential source of geographic clustering. While examiners may have less geographic concentration in their knowledge,

they may feel more familiarity with patents that they have examined than with patents that others have examined. This potentially introduces a bias through differences in the geographic zones of examiner caseloads. Since applicants do not know which examiner will be assigned to their case, it is unlikely that applicants will include a large number of citations to any particular examiner. Thus, we can infer an "examiner self-citation effect" to distinguish it from any geographic pattern we may observe.

Using U.S. patent data from a combination of sources (NBER website as described in Hall et al., 2001 in addition to raw data collected by the independent firm MicroPatent), we collected citations from all biotechnology patents granted between 1975 and 1994. We then traced all self-citations by inventors, allowing for some flexibility in name spellings (since the USPTO does not standardize name format). These include not only first inventors, but all inventors listed for each patent. We found that self-citation accounted for almost precisely one percent of all citations from biotechnology patents, suggesting that while some self-citation is present, there are strong inter-inventor knowledge spillovers. Unlike academic citations, there is very little reason here to self-cite as a means of advertising, so we can be fairly sure that self-citations are indicators of useful capital or legal protection.

To add to this measure, we investigated citations between assignees. Using the same biotechnology patents, we found that nine percent of all citations were to the same assignee firm or person, again checking rigorously for variations of firm names. This share varied from nearly 14 percent in 1975-79 to a low of less than 8 percent in 1985-89, with no obvious trend. This intra-firm pattern is obviously an important component of the flow of knowledge, but they still imply that 9 of 10 citations are to other firms.

Conceivably, those citations to other firms and inventors were simply added by patent examiners, with no relationship to knowledge flows between inventors. We found that five percent of all biotechnology citations were made to other patents sharing the same examiner with considerable variation between examiners. In fact, for one examiner over seventy percent of the citations made to patents he examined hail from other patents he reviewed. Clearly then, examiners are inserting or confirming references to material with which they are familiar. While some of their citations are based on the technology itself (and since they are biotechnology experts, many of the citations should be to biotechnology examiners), there is a definite emphasis on personal contact with other patents cited.

It is also possible that geographic patterns of citations are not due to spatial variables but are instead due to technologies produced in different regions, which subsequently create citation patterns that appear spatially oriented. To investigate the correlation of geography and technology, we counted patents in each state by 7-digit International Patent Class (IPC), a measure of technological groups. We then calculated pair-wise correlations between states, based on the number of patents in each of the top 635 IPCs, to provide an overall measure of the similarity between patented technologies by region. Similarly, we perform the calculation for biotechnology patents alone, using the 249 patent classes showing data for biotechnology. A summary of these variables is presented in Table 1. The table distinguishes between all patents and biotechnology patents, but also distinguishes between neighbor states (those sharing a boundary), and distant states (those not sharing a boundary, and excluding the state itself which has a correlation of 1 by definition).

Notice that while the average pair-wise correlation is fairly high (0.6 to 0.75), it is higher on average for biotechnology patents, with more variation between states. From the spatial point of view, averages are always higher for neighbor states, and while standard deviations are fairly similar for neighboring and distant states, the lowest correlations are always with distant states. Thus, we will include information on technology correlations as a control variable in our analysis of geographic spillovers below.

#### **IV. Regressions**

We approach regression analysis of citation flows in two ways. First, we use an aggregate analysis to determine whether geography plays a role in citations. Next, we use a more sophisticated approach, controlling for the age of knowledge, as well as for self-citation by inventors, assignees and examiners. Both approaches illuminate important aspects of the data, so we present them in turn. However, for ease of comparison they are presented together in Table 2.

##### ***IV.A. Aggregate Analysis***

Following the work of Maurseth and Verspagen (1999), we first present a simple test of whether geographic variables have an effect on citation patterns, using the citations between each state-to-state pairing as observations. Knowledge flows ( $KF_{ij}$ ) between states  $i$  and  $j$  are expressed as actual citations as a share of all possible citations between patents of states  $i$  and  $j$ . That is, every patent from state  $i$  could potentially cite every patent from state  $j$ , creating  $P_i P_j$  possible citations. Since this value depends critically upon the patenting activity of the states involved, knowledge flows have been normalized as a share of this total  $P_i P_j$ .

To test whether scale effects are important, we include variables  $P_i / (P_i + P_j)$  and  $P_j / (P_i + P_j)$  to measure the relative size of the citing and cited states. Since the dependent variable  $KF_{ij}$  is already expressed as a share of possible citations, unequal coefficients on these two variables indicate disproportionate clustering of citations among actively (or inactively) patenting states.

To test the importance of geography, three variables have been included. A measure of distance  $d_{ij}$  between the citing and cited states (kilometers between state capitals) will determine whether physical distance between states is an issue. While distances between capital cities is not the most accurate distance measure (e.g. distances between economic centers, or biotechnology innovation centers would be preferable) it is the approximate size that matters, to distinguish neighboring states in the southwest from neighboring states in the northeast. Finally, there are dummy variables for intra-state and state-to-neighbor citations.

The literature on absorptive capacity suggests that lower-income regions may not be able to adopt from (or contribute to) higher-income regions effectively. Regions require some level of competence in the technology before knowledge flows can be realized. While the variation in income levels between US states is not large, it is worth considering here, so the gap in GDP per capita income between the citing and cited states is included, along with the square of that variable.

Finally, to test whether the specific technological profile of each state is important to the creation of citation flows, two variables are included to measure the technological correlation between citing and cited states. These variables measure the correlation

between all patented technologies, and between patented biotechnologies, of states  $i$  and  $j$  (as described above). Each of these measures ranges between 0 and 1.

The resulting estimation equation is therefore:

$$\begin{aligned}
 KF_{ij} = \ln\left(\frac{C_{ij}}{B_i P_j}\right) &= \alpha_C + \alpha_{BI} \ln\left(\frac{B_i}{B_i + P_j}\right) + \alpha_{BJ} \ln\left(\frac{B_j}{B_i + P_j}\right) + \alpha_{PJ} \ln\left(\frac{P_j}{B_i + P_j}\right) \\
 &+ \alpha_D \ln d_{ij} + \alpha_S \text{SELF} + \alpha_N \text{NEIGH} + \alpha_{G1} \text{GAP} + \alpha_{G2} (\text{GAP})^2 \\
 &+ \alpha_T \text{TECH} + \alpha_B \text{BIO} + \varepsilon
 \end{aligned} \tag{1}$$

where  $C_{ij}$  is the number of citations from patents of state  $i$  to patents of state  $j$

$B_i$  is the number of biotechnology patents granted to applicants from state  $i$   
(citing state)

$B_j$  is the number of biotechnology patents granted to applicants from state  $j$   
(cited state)

$P_j$  is the number of all patents granted to applicants from state  $j$   
(cited state)

$d_{ij}$  is the distance between capitals of states  $i$  and  $j$

*SELF* is a dummy for intra-state citations

*NEIGH* is a dummy variable for citations to neighboring states

*GAP* is the log of the ratio between the citing and cited state GDP per capita

*TECH* is the correlation between patented technologies in states  $i$  and  $j$

*BIO* is the correlation between patented biotechnology in states  $i$  and  $j$

Results (corrected for heteroskedasticity) are presented in the first column of Table 2. While most coefficients are statistically significant, the overall validity of the model is unsatisfactory due to at least two inherent limitations of aggregate analysis. First, the analysis does not allow us to see whether the citation pattern has changed over time. In fact, if state A patented very infrequently before 1990 but has become a powerhouse in biotechnology in the 1990's, it will show up as cited infrequently simply because its patents have not had enough time to be cited by others. Thus, the preceding

analysis penalizes late-bloomers and may bias the results insofar as there have been geographic shifts in patenting over time.

Second, there is no consideration in this analysis for the bias toward clustering due to self-citation (by inventors, assignee firms or patent examiners). The geographic clustering evidenced here may not be geographic, but rather of a more personal or firm-level nature.

To overcome these limitations, we present a second, more comprehensive regression analysis. As a result, we forego examination of the coefficients here, preferring to compare them to the next section's coefficients.

#### ***IV.B. Citation-Cohort Analysis***

Our complementary regression analysis follows the seminal work of Jaffe and his co-authors (Caballero and Jaffe, 1993; Jaffe and Trajtenberg, 1996). The model estimates the likelihood that a patent  $k$  granted in year  $t$  will be cited by a subsequent patent,  $K$ , granted in year  $T$ . Using exponential rates of decay and diffusion to model the flow of knowledge over time, the probability can be written as:

$$p(k, K) = \alpha(k, K)\delta(k, K) \exp[-\beta_1(T-t)][1 - \exp(-\beta_2(T-t))] \quad (2)$$

where  $\beta_1$  represents the decay rate (in our estimation based on knowledge stocks measured in units of millions of patents available for use),  $\beta_2$  represents the rate of diffusion (in our estimation based on the lag between citing and cited patents measured in years), and both exponential terms depend (directly or indirectly) upon the time elapsed between granting of the cited and citing patents. The term  $\alpha(k, K)$  represents other non-geographic attributes of patents  $k$  and  $K$  that affect the probability of citation, while the term  $\delta(k, K)$  refers to the geographic attributes, the key variables in which we are



interested. For example, patents with the same examiner may cite each other more frequently than other attributes would predict, so we include that variable to eliminate potential bias in our analysis of geographic clustering. In particular, we include five control variables ( $\alpha$  parameters) and three geographic ( $\delta$ ) variables :

*Control variables*

- whether or not patents  $k$  and  $K$  have the same inventor ( $\alpha_{SI}$ ),
- whether or not patents  $k$  and  $K$  have the same assignee ( $\alpha_{SA}$ ),
- whether or not patents  $k$  and  $K$  have the same examiner ( $\alpha_{SE}$ ),
- degree of similarity in patented technology from states represented by patents  $k$  and  $K$  ( $\alpha_{TECH}$ ),
- degree of similarity in patented biotechnology from states represented by patents  $k$  and  $K$  ( $\alpha_{BIO}$ ),

*Geographic variables*

- whether or not patents  $k$  and  $K$  have the same state origin ( $\delta_S$ ),
- whether or not patents  $k$  and  $K$  have neighboring state origins ( $\delta_N$ ),
- distance between state origins of patents  $k$  and  $K$  ( $\delta_D$ ),

As in the analysis above, we also include control variables to measure the technological similarity between the states involved. While the correlation between technology patterns of two states may take any value between 0 and 1, they have been grouped into three categories: high (above 0.67), medium (between 0.33 and 0.67) and low (below 0.33) correlation to permit estimation of the model. We normalize the parameter for low correlation to zero, give cohorts with medium correlation a rank value of unity, and cohorts with high correlation a rank value of two. Thus, we estimate the values of the medium and high parameter levels, but enforce a linear relationship between them based on one estimated coefficient.

We include two types of geographic variables: dummy variables for same-state and neighboring-state citation, and physical distance. To permit estimation,  $(k, K)$  patent

pairs are grouped into ten cohorts ranging from high distance (over 2,250 kilometers between state capitals of  $k$  and  $K$ ) to low distance (less than 250 kilometers), in increments of 250 kilometers. This last variable's parameter has been normalized so that low distance has a parameter value of zero and the ten cohorts are given a distance value in tens of thousands of kilometers. Again, this permits estimation of parameters, but enforces a linear relationship between them. Unfortunately, we cannot include an income gap variable between citing and cited states, simply due to estimation limitations.

Using these parameters, the probability of a patent  $k$  granted in year  $t$  being cited by a patent,  $K$ , granted in year  $T$  can be estimated as:

$$p_{k,K} = \alpha_{SI} \alpha_{SA} \alpha_{SE} \alpha_{ST} \alpha_{TECH} \alpha_{BIO} \delta_S \delta_N \delta_D \exp[-\beta_1(T-t)][1 - \exp(-\beta_2(T-t))] + \varepsilon_{k,K} \quad (3)$$

A true geographic effect of clustering would be evidenced by values of the  $\delta$  parameters greater than unity, since other relevant attributes of patents  $k$  and  $K$  have been controlled.

Since most patents are never cited, if we were to estimate equation (9) for individual pairings of citing/cited patents, the dependent variable would be zero for the vast majority of all observations. Thus, following the earlier work of Jaffe and his coauthors, we group the patents into cohorts of potential citations. For example, one cohort may be all patents granted in 1976 and cited in 1978 by a patent that shares the same inventor, the same assignee, a different examiner, a different technology group, and hails from a neighboring state within 250 kilometers which happens to have a low similarity of technology but a high similarity of biotechnology. This list of variables means that there are 5,760 cohorts for every combination of cited-citing years, for a total of over one million cohort citation observations if we group cohorts around individual

years. For estimation, we group the data into five-year cohorts by year of the citing patent (1975-79, 1980-84, 1985-89, 1990-94).

The expected number of citations to a cohort with attributes  $(SI, SA, SE, ST, TECH, BIO, S, N, D; t, T)$ , hereafter abbreviated  $(X; t, T)$ , is the likelihood of a single citation times the number of potentially citing (or cited) patents:

$$E[C_{X;t,T}] = (N_{X;t})(A_{X;T})(p_{k,K}) \quad (4)$$

where  $C$  is the number of citations to the cohort of patents described by the list of attached parameters,  $N$  is the number of (all) patents in that cohort available to be cited, and  $A$  is the number of potentially citing (biotechnology) patents granted in year  $T$ . This equation can now be rewritten to use what we know about patents in the cohort to which the pairing  $(k, K)$  belongs:

$$p_{k,K} = p_X = \left( \frac{C_{X;t,T}}{(N_{X;t})(A_{X;T})} \right) \quad (5)$$

and combining (3) and (5) gives us an equation:

$$\left( \frac{C_{X;t,T}}{(N_{X;t})(A_{X;T})} \right) = \alpha_{SI} \alpha_{SA} \alpha_{SE} \alpha_{ST} \alpha_{TECH} \alpha_{BIO} \delta_S \delta_N \delta_D \exp[-\beta_1(T-t)][1 - \exp(-\beta_2(T-t))] + \varepsilon_X \quad (6)$$

which can be estimated by non-linear least squares as long as the error term,  $\varepsilon_X$ , is well behaved. Because the data are grouped, we weight each observation by  $\sqrt{(N_{X;t})(A_{X;T})}$  to avoid heteroskedasticity issues (Greene, 1993).

To calculate the dependent variable fractions, we evaluated close to ten billion possible patent-to-patent citations, summarized into 19,330 observations of cohort pairings. The second column of Table 2 presents those results.

The cohort analysis results are broadly consistent with those of Johnson and Popp (2000). For example, our regression yields constants which diminish over time, implying that although patents per year increase, there is a decreasing probability of citation to any particular one of them.

As expected, according to the aggregate regression results (in the first column, omitted in second column due to functional form of estimation), the number of citing and citable biotechnology patents both add to the number of citations between states. Interestingly, it is not the size of the citable patent pool that matters, but also its relevance (note that while the relationship with cited *patents* overall is positive and significant, there is also a significant, positive relationship with citable *biotechnology patents*). It is also interesting that the aggregate regression shows significance to income gaps between states (a large gap is bad for citation flows), which were omitted from the cohort analysis.

Turning to the cohort analysis (second column) there is a positive but statistically insignificant effect of same-inventor citations and a high, positive and significant effect of same-assignee citations. This implies that inventors have a stronger propensity to cite other laboratory co-workers and not necessarily themselves, which is probably a better signal of useful knowledge acquisition than the self-promotion effect we would suspect in the case of same-inventor citations.

Furthermore, the same-examiner effect, which we introduce for the first time in analysis of this type, is strong, positive and significant, meaning that there is a

demonstrable correlation between subject material and the particular examiner. Since applicants presumably would not be able to predict the examiner in advance, and would have little reason to cite patents by the examiner even if they knew, we can conclude that examiners apparently add those citations, adding disproportionately more citations to patents that they have read themselves.

A puzzling result is the strong, negative coefficient for states with similar technologies (and weak but still negative for states with similar biotechnology), a result which stands in sharp contrast with the strong, positive coefficient for the same variables in the aggregate regression. One possible explanation of these counterintuitive results is that the aggregate regression lacks three self-citation variables (same inventor, same assignee, and same examiner) all likely to capture the effect of citations to similar states. Since inventors, assignees and examiners all have subjects of expertise, it is likely that they will offer citations to patents in the same subject, thereby citing states with similar (bio)technological fields. In the aggregate analysis, this effect is only seen through the variables capturing technological similarity between states. The cohort analysis shows that once we control for self-citation by experts within a technology, and have controlled for citations within the state and to neighboring states, long-distance citations are more likely to cite different technologies. That is, once local knowledge has been fully explored, citations jump to another state precisely because it is technologically different. Of course, other explanations for this puzzling pair of coefficients are welcome.

There is home state bias in citations, as shown in the cohort regression. It is troubling that this effect was not demonstrated by the aggregate regression, and is presumably due to other confounding factors in the aggregate (e.g. self-citation by

assignee across state lines, as when a researcher for IBM cites another patent by IBM although the patents were developed in different states).

Most importantly, and as expected, the effects of distance are negative, significant and extensive. This confirms our hypothesis that geography plays a role in the creation of patents, that local knowledge spillovers play an important role in knowledge transfer, even after controlling for other factors.

It does not appear important whether distance is measured to a neighboring state or another more distant border, as evidenced by the weak negative and insignificant coefficient on neighbor distance. This variable is also critical in dismissing the potential argument that distance matters only in the Northeastern states where neighbors are close, or only in the Southwestern states where neighbors are far apart. It appears to matter everywhere.

Finally, both decay and diffusion are significant and strikingly faster than evidenced elsewhere in the literature (Johnson and Popp, 2000), indicating that the biotechnology field moves at a fast pace in its creation and decay of knowledge. Thus, we conclude that biotechnological knowledge decays and diffuses quickly across time, but diffuses in a more limited manner across distance.

#### ***IV.C. Citation-Cohort Analysis: Tests for Change In Coefficients Over Time***

The last column of Table 2 presents a cohort-based analysis permitting the coefficients to change over time. The F-statistic on the restrictions is 5.0645 with  $df=(24,19293)$ , so we can reject the hypothesis of "no change in coefficients over time" with 99% confidence. Results are similar to the previous section, with several exceptions.

Only in the last period does the effect of same-inventor citations become significant (and positive), suggesting that prior to the 1990's, there was no predilection for inventors to cite themselves. On the other hand, the positive coefficients on same-assignee and same-examiner effects have been decreasing constantly, precipitously in the case of examiners, suggesting that those agents are becoming much less likely to cite themselves (and therefore more likely to cite the work of others). This may be due to better electronic access to patent records, or due to greater knowledge of assignee firms about what other firms are doing in the field.

States with dissimilar technologies (either all patents, or biotechnology alone) seem to have cited each other often in the 1970's, abandoned that pattern in the 1980's, and returned to it in the 1990's.

Most importantly, a combination of results points to the decreasing importance of distance as a factor determining citation probability. First of all, the coefficient on same-state citation dropped from an enormous level to insignificance between the first and last periods. Second, neighbor coefficients remain insignificant in each period by themselves. Lastly, the distance coefficient is only half of its 1970's value by the 1990's (although it is still very significant and negative, even more so than in the late 1980's).

While this change may not seem enormous, an illustration may help. In the late 1970's, the coefficients indicate that the half-life distance of a biotechnology patent (i.e. the distance at which it would be half as likely to be cited as at the origin) was 3100 kilometers. By the late 1980's, that distance had stretched to nearly 8200 kilometers.

This time-varying result points unquestionably to the fact that distance is becoming less important with time. Perhaps the trend is due to the nature of the

knowledge being created, but we suspect that it is more due to the nature of communication, which allows easier transmission of information across great distances in the era of computerization, internet, teleconferencing and cellular communication. In short, the principles underlying the inter-firm transfer of knowledge are changing in a striking fashion, making spillovers easier than ever before.

## **V. Network Analysis**

One possible explanation for this evidence on distance is that connections of any type, geographic or otherwise, have become less important over time. After all, the probability of any given patent receiving a citation is lower than in previous periods, self-citation by assignees and examiners has fallen, and distance has decreased in importance. Is it possible that citations are simply becoming random noise between patent applicants?

This section presents a picture that contradicts that possibility, showing the increased importance of networks over time. Four frames use network analysis to identify the key actors (or regions), evaluate inequality between them, describe the knowledge flows between them, and examine the degrees of separation between them.

### ***V.A. Key Regions***

By any measure, California tops the list. It has almost three times the aggregate citation activity of any other state, nearly surpassing the next three states combined, and accounts for almost one-quarter of all citations. The other top states are equally unsurprising (New Jersey, New York, Massachusetts, Illinois, Pennsylvania).

A commonly accepted measure of the centrality of a network member  $i$  (Knoke and Kuklinski, 1982) is



$$centrality_i = \frac{\sum_{j=1}^N (C_{ij} + C_{ji})}{\sum_{i=1}^N \sum_{j=1}^N C_{ij}} * 100, \forall i \neq j \quad (7)$$

where  $C_{ij}$  and  $C_{ji}$  each take the value of 1 if there is a citation from state  $i$  to state  $j$ . In short, centrality measures the percent of all active citation routes which involve region  $i$ . A star in the network, to and from which all citations flow, would warrant a score of 100, while a score of 20 would be an enormous hub of the network. The calculated values for biotechnology are presented in the first column of Table 3, and show that while there are certainly centers and periphery zones, the network reaches a wide number of regions with good coverage. A variation on the centrality measure, called flow centrality, uses actual citation numbers instead of ones and zeroes for  $C_{ij}$ , and unsurprisingly shows stronger hubs but the same overall pattern.

The prestige of each region in the network can be measured as the degree to which other regions cite it as opposed to the reverse. That is, it is a mark of prestige when others refer (or defer) to a region's accomplishments, instead of the reverse. For a simple prestige formula

$$simple\ prestige = I_i / (centrality_i)(N - 1) \quad (8)$$

will suffice, where  $I_i$  is the number of regions interacting either directly or indirectly with region  $i$ . A direct interaction involves a state which cites or is cited by state  $i$ , while an indirect interaction involves one or more degrees of separation from state  $i$ . For example, state  $k$  is one degree of separation from state  $i$  if it does not interact directly with state  $i$ ,

but does with state  $j$ , and state  $j$  interacts with  $i$ )<sup>3</sup>. For a more complete (weighted) prestige formula, we consider the prestige of each region interacting with region  $i$ , so that

$$prestige_i = \sum_{j=1}^{N-1} prestige_j C_{ij} \quad (9)$$

is the weighted average of prestige of interacting regions. This is of course a system of  $N$  simultaneous equations, which is solved for solutions to the set of prestige values for all regions. After rescaling those solutions so that the highest prestige value is unity, these values are presented in the last columns of Table 3.

The most important conclusion (other than the fact that the three measures of centrality and prestige rank states in very similar fashion) is that the measures have changed remarkably little over time. Despite the lower importance of distance and self-citation by state, states have retained their relative rankings, suggesting that an underlying pattern of citations remains constant, supporting the structure.

### ***V.B. Flows Between Regions***

Next, we consider the flows of the network as a whole. The centralization of the entire citation network is calculated as the scaled difference between the centrality score of the most central actor (\*) and those of the other actors:

$$centralization = \frac{\sum_{i=1}^N (centrality^* - centrality_i)}{(N-2)(N-1)} \quad (10)$$

which has a value of between 0.10 and 0.11 for three of the four periods (and only swerves upward slightly to 0.14 in the early 1980's).

---

<sup>3</sup> These simple prestige scores use five degrees of separation, at which point all states in our data are connected indirectly to every other state. However, scores using only two degrees of separation (direct plus

The density of the inter-state citation network is measured as

$$density = \frac{\sum_{i=1}^N \sum_{j=1}^N C_{ij}}{N^2} \quad (11)$$

where  $C_{ij}$  takes the value of unity if there are citations between regions  $i$  and  $j$  (0 otherwise), and  $N$  is the number of regions. Its value has stayed in a narrow band ranging between 9.1 and 12.3 except for the early 1980's. Network cohesion, a measure of density incorporating information about symmetry, is calculated as

$$cohesion = \frac{\sum_{i=1}^N \sum_{j=1}^N (C_{ij} + C_{ji})}{N^2 / 2}, \forall i \neq j \quad (12)$$

where  $(C_{ij} + C_{ji})$  takes the value of unity if there are citations in both directions between regions  $i$  and  $j$  (0 otherwise). The cohesion measure stays between 0.63 and 0.66 over the period, indicating again that there has been no network-based systemic change in the way that states cite each other.

### ***V.C. Inequality Between Agents***

Now that we have ascertained that the network between states (including state prestige and network cohesion) has remained stable over time, we ask what other changes may have occurred in parallel with the demise of distance. In fact, there has been a trend toward growing equality between inventors (and between assignees, and between examiners) over time. Figure 1 shows the Lorenz curves for patenting, and the associated Gini coefficients, which measure inequality. The top curve illustrates perfect equality, where each percentile of the population is responsible for one percent of all patents. The

---

one indirect connection degree) show almost precisely the same values, with only four of 51 scores changing by more than 0.005.

Gini coefficient for this first curve (measured as the area between the diagonal and the curve) is therefore zero.

The second curve shows cumulative patents by each percentile of the spectrum of inventors, ranked from least active to most active, with a Gini coefficient of 0.21. This shows less inequality than the previous period, when the Gini coefficient was 0.29.

Inequality is greater overall among assignees, since several large assignee firms garner a large share of all patents, but these too have become slightly more equal (from a Gini of 0.64 to 0.56). Finally, some examiners bear the burden of most biotechnology patents, making the distribution of patents among examiners highly inequitable. However, this distribution has also become slightly more equal over time, at least among very active examiners.

The figure therefore shows that there have been (potentially important) distributional changes in terms of who patents in biotechnology, which firms retain the rights to those patents, and who examines those patents. This increase in equality has occurred simultaneously with the fall in the importance of distance.

#### ***V.D. Degrees of separation***

Although the cohesion and density of the state-to-state network has not changed since 1975, and there has been a movement towards equality among active agents in the patenting process, we have not yet shown evidence that the person-to-person or city-to-city network has remained stable. It would be impractical to calculate cohesion measures between such a large number of agents, but thankfully, another alternative is open.

Based on sociology work studying the “Small World Problem” (Milgram, 1967), and popularized by the play “Six Degrees of Separation” (Guare, 1990), we investigate

the question of how far removed each biotechnology inventor is from the next. That is, how tight is the network between associated inventors (as measured by citations to each other) and has that measure changed with time?

To study this trend, we define the “core” inventors as the most active one percent biotechnology inventors (with most patents granted) in a time period. We then define the “first degree of separation” as the core plus all inventors who were co-inventors with a member of the core, or who cited (or were cited by) a patent invented by someone in the core. The second through sixth degrees of separation are defined analogously, each including the previous degrees. Separately, we tabulate all biotechnology patents involving a member of each degree.

Table 4 shows the number of inventors within a certain “distance” of the core. For example, 9 percent (584) of all inventors in 1975-79 are either in the core or within one degree of a core member. In contrast, 16 percent (3,750) are in the same position by 1990-94. Over time, there has been a uniform increase in the number and percent of inventors at each degree, a trend that has been especially marked from the second degree and upwards. In other words, inventors are becoming more tightly networked, with fewer degrees of separation required to link an ever growing number.

A similar presentation can be made for assignees in Table 5. There is tighter clustering overall, with most assignees included by the sixth degree, and very little increase between the third and sixth degrees. Although the increase in assignee connectedness is substantially more gradual over time, the trend is still apparent. Notice too that virtually all patents are connected by the fourth degree of separation.

Table 6 considers examiners as the base for analysis, and shows that there has been little change here since the late 1970's. Virtually all patents are included by the third degree due to the limited number of examiners (compared to inventors or assignees). This table is presented merely to document the fact that any growing connectedness between biotechnology patents has not been reflected in a growing connectedness of examiners.

Finally, we consider cities as the base for analysis. The results in Table 7 suggest *dramatic* increases in geographic connectedness between knowledge creation centers. The degrees of geographic separation results are consistent with our cohort analysis conclusions concerning the changing significance of distance, and once again point out that distance is becoming less important as knowledge creation centers become more tightly interwoven over time.

According to a popular maxim, every person on earth is connected to every other by no more than six degrees (Guare, 1990). While that is not the case here, the trend is approaching that conclusion. In fact, it is virtually true of all patents by the fourth degree of separation, if we consider cities as the base of analysis. Thus, while the patenting and citation patterns between states has remained relatively stable, distance has become less important and networks between actors or cities have become more important.

## **VI. Conclusion and Policy Implications**

We are left with a striking picture of the inter-firm transfer of biotechnological knowledge. Controlling for other factors, knowledge flows diminish with physical distance, but that the importance of distance has been receding with time. That is, knowledge is more likely to transfer over long distances now than it was twenty years

ago. However, this has occurred in tandem with the increasing importance of inter-firm, inter-personal and inter-city network connections. It has truly become a smaller world for biotechnology, as distance is less relevant and active researchers and firms are more connected to each other than ever before.

In a sense, the policy recommendations of this paper have therefore been heard before. In an age of more intense and distance-free communication, the conduits of knowledge transmission take on a new importance. Researchers and firms have obviously benefited tremendously from the movement to electronic patent searches and filings. In fact, that trend may have partially driven our results.

The policy implications of this paper presents are important not only for regions of the U.S. but for less developed nations as well. First, strong geographic clustering suggests that innovation will be more difficult and more costly if pursued at a great distance from a "biotechnology center". For less developed regions (or nations), that are by their nature far from a center, it may be wiser to license existing technology, or to rent research space and hire researchers close to the core, than to perform independent research at remote locations. However, the importance of physical distance is diminishing dramatically over time, facilitating innovation in less traditional areas, including less developed nations. In addition, licensing existing technology brings with it the attendant difficulties of applicability to local environments and conditions (see Johnson and Evenson, 2000).

Second, while the importance of physical distance has been decreasing, the importance of communication links to the center have been increasing. To foster innovation, policy must emphasize the importance of networks with leading inventors and

their firms. Fewer and fewer successful inventors exist in isolation from research centers, as connections drive more and more of the research activity. For less developed regions or nations, this points to the importance of educational/training exchange programs, joint venture initiatives and international conferences.

The outlook for less developed regions is open to interpretation. They are far from the center of research, so will tend to receive knowledge in more limited fashion. However, that situation is changing and access to the networks of biotechnological information is less and less location-dependent. Income gaps may be an impediment to knowledge creation, as seen in the aggregate regression results. On the other hand, there is a propensity for knowledge to flow to areas with different technological patterns, if it overcomes the challenge of distance, and this will undoubtedly favor less developed regions. It is a small world for biotechnology, and with time it is growing ever smaller, ever more interconnected. It is therefore imperative to become an active part of the system quickly, before its networks are even more tightly woven.



## VII. References

- Caballero, R.J. and A.B. Jaffe (1993), "How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth," in Olivier J. Blanchard and Stanley Fischer, eds., NBER Macroeconomics Annual 1993, MIT Press, Cambridge, MA.
- Dorfman, N.S. (1988) "Route 128: The Development of a Regional High Technology Economy". The Massachusetts Miracle: High Technology and Economic Revitalization. D. Lampe, editor. MIT Press, Cambridge.
- Feldman, M.P. (1994) The Geography of Innovation. Kluwer Academic Publishers, Dordrecht, Netherlands.
- Feldman, M.P. (1999) "The New Economics of Innovation, Spillovers and Agglomeration: A Review of Empirical Studies". Economics of Innovation and New Technology. Vol 8, p.5-25.
- Gelsing, L. (1992) "Innovation and Development of Industrial Networks". National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning. B. Lundvall, editor. Pinter Publishers, London.
- Greene, W.H. (1993) Econometric Analysis, Macmillan Publishing Company, New York.
- Guare, J. (1990) Six Degrees of Separation. Vintage Books, New York.
- Hall, B.H., A.B. Jaffe and M.Trajtenberg (2001) "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools". NBER Working Paper #8498.
- Henderson, J.V. (1986) "Efficiency of Resource Usage and City Size". Journal of Urban Economics. Vol 19, p.47-70.
- Jaffe, A.B., M.Trajtenberg and R.Henderson (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations". Quarterly Journal of Economics. Vol 108, p.577-598.
- Jaffe, A.B. and M.Trajtenberg (1996) "Flows of Knowledge From Universities and Federal Labs: Modeling the Flow of Patent Citations Over Time and Across Institutional and Geographic Boundaries", NBER Working Paper #5712.
- Jaffe, A.B., M.Trajtenberg and M.S.Fogarty (2000) "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors". American Economic Review. Vol 90, p.215-218.

- Johnson, D.K.N. and D.Popp (2001) "Forced Out of the Closet: The Impact of the American Inventors Protection Act on the Timing of Patent Disclosure". NBER Working Paper #8374.
- Johnson, D.K.N. and R.E.Evenson (2000) "How Far Away Is Africa? Technological Spillovers to Agriculture and Productivity". American Journal of Agricultural Economics. Vol 82, p.743-749.
- Johnson, D.K.N. and V.Santaniello (2000) "Biotechnology Inventions: What Can We Learn from Patents?". Agriculture and Intellectual Property Rights: Economic, Institutional, and Implementation Issues in Biotechnology. V. Santaniello et al., editors. CABI Publishing, New York.
- Johnson, D.K.N. (1999) "150 Years of American Invention: Methodology and a First Geographical Application". Wellesley College Department of Economics Working Paper 99-01. January, 1999.
- Knoke, D. and J.H.Kuklinski (1982) Network Analysis. Sage Publications, London.
- Lundvall, B.A. (1992) "User-Producer Relationships, National Systems of Innovation and Internationalisation". National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning. B. Lundvall, editor. Pinter Publishers, London.
- Maurseth, P.B. and B.Verspagen (1999) "Europe: one or several systems of innovation? An analysis based on patent citations". Economic Challenge to Europe: Adapting to Innovation-Based Growth. Edward Elgar Publishing, London.
- MetLife.com (2001) Distance Calculator. Accessed August 2001.
- Milgram, S. (1967) "The Small World Problem." Psychology Today. p. 60 - 67.
- Sjoholm, F. (1996) "International Transfer of Knowledge: The Role of International Trade and Geographic Proximity". Weltwirtschaftliches Archiv. Vol 132, p. 97-115.
- Smith, D. and R.Florida (1994) "Agglomeration and Industry Location: An Econometric Analysis of Japanese-Affiliated Manufacturing Establishments in Automotive Related Industries". Journal of Urban Economics. Vol 36, p. 23-41.
- USPTO Technology Assessment and Forecast Program (1998) "Technology Profile Report: Patent Examining Technology Center Groups 1630-1650, Biotechnology". United States Patent and Trademark Office report.
- Valente, T.W. (1995) Network Models of the Diffusion of Innovations. Hampton Press, Inc. Cresskill, New Jersey.

Von Hippel, E. (1994) "Sticky Information and the Locus of Problem-Solving".  
Management Science. Vol 40, p. 429-439.

**Table 1: Technological Correlations between States**

	<i>Mean</i>	<i>Minimum (Maximum)</i>	<i>Standard Deviation</i>
All patents, neighbor states	0.712	0.336 (0.916)	0.149
All patents, distant states	0.624	0.096 (0.916)	0.172
Biotech patents, neighbor states	0.732	0.135 (0.969)	0.196
Biotech patents, distant states	0.699	0.011 (0.982)	0.198

**Table 2: Regression Results for Aggregate and Cohort Analysis**

<i>Variable</i>	<i>Aggregate Analysis</i>		<i>Cohort Analysis</i>		<i>Time-Varying Cohort Analysis</i>	
	<i>Coefficient</i>	<i>(t-statistic)</i>	<i>Coefficient</i>	<i>(t-statistic)</i>	<i>Coefficient</i>	<i>(t-statistic)</i>
Constant	4.76	(5.93) <sup>***</sup>	-----	-----	-----	-----
1980-1984	-----	-----	-1.86	(1.91) <sup>*</sup>	-9.45	(2.92) <sup>***</sup>
1985-1989	-----	-----	-4.75	(3.93) <sup>***</sup>	-11.07	(2.86) <sup>***</sup>
1990-1994	-----	-----	-6.90	(4.42) <sup>***</sup>	-12.63	(2.64) <sup>***</sup>
Citing patents	-0.34	(7.18) <sup>***</sup>	-----	-----	-----	-----
Cited biotechnology	0.94	(10.14) <sup>***</sup>	-----	-----	-----	-----
Cited total patents	0.09	(0.37)	-----	-----	-----	-----
Same inventor	-----	-----	0.62	(1.39)	-----	-----
1975-1979	-----	-----	-----	-----	-1.60	(0.90)
1980-1984	-----	-----	-----	-----	1.02	(0.92)
1985-1989	-----	-----	-----	-----	-0.25	(0.30)
1990-1994	-----	-----	-----	-----	1.13	(1.75) <sup>*</sup>
Same assignee	-----	-----	2.67	(6.13) <sup>***</sup>	-----	-----
1975-1979	-----	-----	-----	-----	6.92	(3.80) <sup>***</sup>
1980-1984	-----	-----	-----	-----	5.48	(4.97) <sup>***</sup>
1985-1989	-----	-----	-----	-----	1.81	(2.28) <sup>**</sup>
1990-1994	-----	-----	-----	-----	1.69	(2.73) <sup>***</sup>
Same examiner	-----	-----	1.81	(4.11) <sup>***</sup>	-----	-----
1975-1979	-----	-----	-----	-----	7.68	(4.82) <sup>***</sup>
1980-1984	-----	-----	-----	-----	3.54	(3.39) <sup>***</sup>
1985-1989	-----	-----	-----	-----	1.99	(2.49) <sup>**</sup>
1990-1994	-----	-----	-----	-----	0.12	(0.18)
All patent correlation	2.45	(7.12) <sup>***</sup>	-1.12	(3.45) <sup>***</sup>	-----	-----
1975-1979	-----	-----	-----	-----	-4.14	(3.26) <sup>***</sup>
1980-1984	-----	-----	-----	-----	-0.24	(0.30)
1985-1989	-----	-----	-----	-----	-0.51	(0.87)
1990-1994	-----	-----	-----	-----	-1.31	(2.78) <sup>***</sup>
Biotechnology correlation	5.68	(19.56) <sup>***</sup>	-0.45	(1.51)	-----	-----
1975-1979	-----	-----	-----	-----	-2.15	(1.85) <sup>*</sup>
1980-1984	-----	-----	-----	-----	0.99	(1.37)
1985-1989	-----	-----	-----	-----	-0.23	(0.43)
1990-1994	-----	-----	-----	-----	-0.78	(1.78) <sup>*</sup>

(see next page)

**Table 2 (cont.): Regression Results for Aggregate and Cohort Analysis**

<i>Variable</i>	<i>Aggregate Analysis</i>		<i>Time-Varying Cohort Analysis</i>		<i>Cohort Analysis</i>	
	<i>Coefficient</i>	<i>(t-statistic)</i>	<i>Coefficient</i>	<i>(t-statistic)</i>	<i>Coefficient</i>	<i>(t-statistic)</i>
Same U.S. state	-2.96	(4.20) <sup>***</sup>	7.46	(5.54) <sup>***</sup>	-----	-----
1975-1979	-----	-----	-----	-----	34.85	(8.01) <sup>***</sup>
1980-1984	-----	-----	-----	-----	5.96	(1.96) <sup>**</sup>
1985-1989	-----	-----	-----	-----	6.09	(2.48) <sup>**</sup>
1990-1994	-----	-----	-----	-----	1.83	(0.87)
Neighboring state	-0.93	(3.78) <sup>***</sup>	-----	-----	-----	-----
Distance (0000 km)	-0.61	(7.02) <sup>***</sup>	-8.26	(4.14) <sup>***</sup>	-----	-----
1975-1979	-----	-----	-----	-----	-16.06	(2.21) <sup>**</sup>
1980-1984	-----	-----	-----	-----	-9.45	(2.01) <sup>**</sup>
1985-1989	-----	-----	-----	-----	-6.11	(1.68) <sup>*</sup>
1990-1994	-----	-----	-----	-----	-8.27	(2.76) <sup>***</sup>
Neighbor distance (0000 km)	-----	-----	-3.88	(0.67)	-----	-----
1975-1979	-----	-----	-----	-----	9.37	(0.47)
1980-1984	-----	-----	-----	-----	-19.08	(1.43)
1985-1989	-----	-----	-----	-----	-1.46	(0.14)
1990-1994	-----	-----	-----	-----	-1.97	(0.22)
Income gap	-1.07	(3.89) <sup>***</sup>	-----	-----	-----	-----
Squared income gap	-3.59	(3.84) <sup>***</sup>	-----	-----	-----	-----
Decay rate	-----	-----	3.29	(5.91) <sup>***</sup>	5.40	(3.81) <sup>***</sup>
Diffusion rate	-----	-----	0.13	(4.76) <sup>***</sup>	0.13	(4.78) <sup>***</sup>
R-squared	0.31		0.70		0.71	
Observations	2,601		19,330		19,330	

Coefficients are multiplied by 1,000,000 in aggregate analysis and by 1,000 in cohort analysis for more readable table. Significance is indicated as \* for ten percent level, \*\* for five percent level, \*\*\* for one percent level.

**Table 3: State Centrality and Prestige Scores  
in Biotechnology Patent Citations**

	Centrality Scores				Simple Prestige Scores				Weighed Prestige Scores			
	1975-79	1980-84	1985-89	1990-94	1975-79	1980-84	1985-89	1990-94	1975-79	1980-84	1985-89	1990-94
AK	0.09	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
AL	2.69	2.17	2.69	2.69	0.75	0.62	0.75	0.76	0.02	0.01	0.01	0.01
AR	1.30	1.15	1.48	1.30	0.68	0.55	0.63	0.65	0.00	0.00	0.00	0.00
AZ	4.08	2.68	2.78	3.55	0.88	0.67	0.72	0.86	0.01	0.01	0.01	0.00
CA	7.03	9.17	7.60	7.19	1.00	0.94	0.99	1.01	1.00	1.00	1.00	1.00
CO	3.99	4.84	3.89	4.42	0.86	0.76	0.84	0.91	0.03	0.01	0.02	0.01
CT	4.51	4.71	5.00	5.03	0.90	0.75	0.90	0.94	0.04	0.04	0.04	0.04
DC	2.60	2.17	2.41	2.17	0.77	0.68	0.74	0.72	0.00	0.00	0.02	0.01
DE	3.73	3.06	3.61	4.07	0.85	0.75	0.83	0.91	0.02	0.03	0.02	0.03
FL	4.34	5.48	4.08	4.42	0.91	0.84	0.89	0.95	0.02	0.03	0.03	0.03
GA	4.25	3.06	3.52	3.55	0.95	0.73	0.90	0.95	0.01	0.00	0.01	0.01
HI	1.30	0.25	0.65	1.39	0.77	0.51	0.56	0.69	0.00	0.00	0.00	0.00
IA	3.39	2.80	2.78	3.29	0.85	0.77	0.74	0.84	0.00	0.01	0.01	0.00
ID	0.69	0.64	1.11	1.39	0.64	0.61	0.70	0.68	0.00	0.00	0.01	0.00
IL	4.77	5.48	5.19	4.33	1.00	0.88	0.97	0.96	0.09	0.25	0.14	0.11
IN	3.91	4.97	4.63	4.16	0.93	0.84	0.96	0.95	0.07	0.16	0.10	0.08
KS	2.00	2.29	2.22	1.56	0.75	0.71	0.74	0.69	0.00	0.02	0.03	0.00
KY	1.91	1.27	1.48	1.73	0.71	0.67	0.70	0.69	0.01	0.00	0.00	0.01
LA	2.43	1.78	2.41	2.17	0.79	0.67	0.85	0.80	0.00	0.04	0.02	0.01
MA	4.25	5.48	4.54	4.25	0.99	0.93	0.98	1.00	0.24	0.11	0.21	0.19
MD	3.82	3.95	3.71	3.90	1.00	0.84	0.95	0.99	0.10	0.06	0.10	0.11
ME	1.48	1.40	1.76	1.13	0.78	0.68	0.77	0.70	0.01	0.02	0.01	0.01
MI	2.60	3.18	3.06	3.21	0.96	0.81	0.88	0.92	0.04	0.07	0.04	0.03
MN	2.60	3.06	2.69	2.69	0.92	0.84	0.90	0.95	0.03	0.02	0.04	0.06
MO	2.60	2.42	2.59	2.34	0.90	0.77	0.88	0.87	0.03	0.04	0.04	0.03
MS	0.52	0.51	0.56	0.52	0.57	0.57	0.51	0.51	0.00	0.00	0.00	0.00
MT	0.52	0.64	0.46	0.95	0.71	0.63	0.66	0.66	0.00	0.00	0.00	0.00
NC	2.26	1.66	2.22	2.08	0.95	0.72	0.94	0.92	0.03	0.01	0.01	0.02
ND	0.17	0.13	0.09	0.35	0.55	0.45	0.00	0.59	0.00	0.00	0.00	0.00
NE	0.61	1.15	1.30	0.78	0.74	0.69	0.81	0.62	0.00	0.00	0.01	0.01
NH	0.95	0.64	1.02	1.04	0.81	0.62	0.70	0.73	0.00	0.00	0.00	0.01
NJ	2.60	2.93	2.59	2.51	0.96	0.92	0.97	0.99	0.14	0.40	0.27	0.20
NM	1.22	0.38	0.74	1.30	0.78	0.57	0.64	0.85	0.00	0.00	0.00	0.01
NV	0.17	0.13	0.09	0.35	0.55	0.54	0.00	0.52	0.00	0.01	0.00	0.00
NY	2.08	2.80	2.41	2.34	0.98	0.93	0.98	1.00	0.16	0.30	0.22	0.23
OH	1.91	1.78	1.67	1.91	0.93	0.83	0.91	0.96	0.05	0.06	0.07	0.04
OK	0.52	0.64	1.20	1.04	0.75	0.76	0.82	0.85	0.02	0.01	0.01	0.01
OR	0.95	0.51	1.02	0.95	0.90	0.67	0.84	0.81	0.00	0.00	0.01	0.00

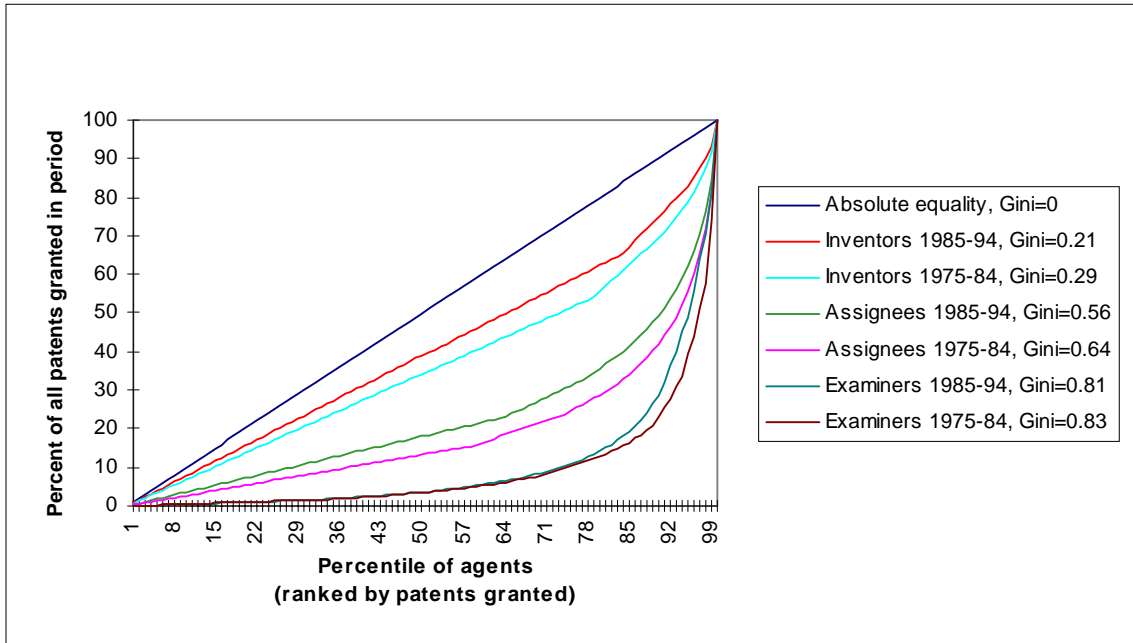
(see next page)

Table 3 (cont.): State Centrality and Prestige Scores  
in Biotechnology Patent Citations

	Centrality Scores				Simple Prestige Scores				Weighted Prestige Scores			
	<u>1975-79</u>	<u>1980-84</u>	<u>1985-89</u>	<u>1990-94</u>	<u>1975-79</u>	<u>1980-84</u>	<u>1985-89</u>	<u>1990-94</u>	<u>1975-79</u>	<u>1980-84</u>	<u>1985-89</u>	<u>1990-94</u>
PA	1.48	1.27	1.39	1.30	0.95	0.83	0.94	0.95	0.10	0.29	0.18	0.12
RI	0.35	0.25	0.09	0.35	0.68	0.52	0.69	0.78	0.00	0.00	0.01	0.00
SC	0.17	0.38	0.74	0.26	0.62	0.57	0.69	0.61	0.01	0.02	0.01	0.01
SD	0.09	0.00	0.00	0.00	0.56	0.00	0.56	0.00	0.00	0.00	0.00	0.00
TN	0.95	0.25	0.65	0.69	0.90	0.68	0.83	0.79	0.01	0.01	0.02	0.01
TX	0.78	1.15	0.83	0.78	0.98	0.88	0.95	1.00	0.07	0.05	0.05	0.04
UT	0.43	0.38	0.28	0.35	0.88	0.75	0.80	0.85	0.01	0.02	0.01	0.01
VA	0.52	0.13	0.37	0.35	0.91	0.73	0.85	0.86	0.01	0.03	0.03	0.01
VT	0.17	0.00	0.00	0.00	0.70	0.47	0.76	0.78	0.00	0.00	0.00	0.04
WA	0.26	0.13	0.19	0.17	0.95	0.73	0.90	0.95	0.06	0.01	0.06	0.07
WI	0.00	0.00	0.00	0.00	0.91	0.76	0.87	0.94	0.02	0.04	0.02	0.02
WV	0.00	0.00	0.00	0.00	0.51	0.00	0.00	0.00	0.00	0.00	0.00	0.00
WY	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00



**Figure 1: Lorenz Curves and Gini Coefficients for  
Key Actors in Biotechnology Patents**



**Table 4: Degrees of Separation for Inventors:  
Percent of Inventors (and Percent of Patents)**

<i>Period</i>	<i>1975-79</i>	<i>1980-84</i>	<i>1985-89</i>	<i>1990-94</i>
Core	1.0 (9.1)	1.0 (8.7)	1.0 (7.0)	1.0 (6.6)
First degree	9.0 (18.6)	11.4 (20.6)	13.4 (19.8)	16.0 (21.3)
Second degree	12.1 (23.7)	20.1 (30.5)	24.9 (32.7)	32.5 (38.9)
Third degree	17.0 (29.4)	32.3 (43.6)	39.9 (48.0)	51.9 (58.9)
Fourth degree	20.3 (33.3)	41.7 (53.8)	51.1 (59.7)	63.8 (71.5)
Fifth degree	22.9 (36.1)	48.0 (59.9)	58.8 (67.5)	70.8 (78.2)
Sixth degree	25.2 (38.7)	52.6 (64.3)	63.1 (71.7)	74.0 (81.1)
Total inventors	6,490	8,284	12,933	23,388

**Table 5: Degrees of Separation for Assignees:  
Percent of Assignees (and Percent of Patents)**

<i>Period</i>	<i>1975-79</i>	<i>1980-84</i>	<i>1985-89</i>	<i>1990-94</i>
Core	1.0 (10.4)	1.0 (7.1)	1.0 (12.2)	1.0 (14.2)
First degree	23.9 (19.4)	20.6 (16.0)	36.2 (30.7)	44.0 (36.8)
Second degree	59.0 (77.0)	66.1 (78.9)	75.3 (87.4)	79.4 (91.1)
Third degree	66.0 (89.5)	75.8 (93.1)	79.3 (94.2)	82.5 (95.8)
Fourth degree	66.5 (90.4)	76.5 (93.7)	79.4 (94.5)	82.8 (95.8)
Fifth degree	66.6 (90.6)	76.5 (93.7)	79.4 (94.5)	82.9 (95.8)
Sixth degree	66.6 (90.6)	76.5 (93.7)	79.4 (94.5)	82.9 (95.8)
Total assignees	943	1,170	1,817	2,666

**Table 6: Degrees of Separation for Examiners:  
Percent of Examiners (and Percent of Patents)**

<i>Period</i>	<i>1975-79</i>	<i>1980-84</i>	<i>1985-89</i>	<i>1990-94</i>
Core	1.2 (24.0)	1.0 (27.3)	1.1 (26.1)	1.1 (18.1)
First degree	21.2 (27.6)	37.0 (38.9)	50.8 (49.9)	24.9 (22.6)
Second degree	47.5 (83.1)	72.8 (95.5)	84.6 (97.1)	74.9 (90.8)
Third degree	54.1 (94.3)	80.0 (98.7)	87.2 (99.4)	84.9 (98.8)
Fourth degree	55.2 (95.4)	80.3 (98.9)	87.2 (99.4)	87.1 (99.3)
Fifth degree	55.2 (95.6)	80.3 (98.9)	87.2 (99.4)	87.4 (99.3)
Sixth degree	55.2 (95.6)	80.3 (98.9)	87.2 (99.4)	87.4 (99.3)
Total examiners	259 (4,707)	305 (6,069)	358 (8,226)	350 (7,605)

**Table 7: Degrees of Separation for Cities:  
Percent of Cities (and Percent of Patents)**

<i>Period</i>	<i>1975-79</i>	<i>1980-84</i>	<i>1985-89</i>	<i>1990-94</i>
Core	1.0 (10.4)	1.0 (13.5)	1.0 (17.2)	1.0 (21.2)
First degree	17.2 (18.6)	25.1 (26.6)	42.3 (41.6)	55.8 (54.2)
Second degree	27.9 (44.7)	58.4 (69.4)	73.5 (86.2)	83.2 (94.2)
Third degree	48.0 (65.4)	79.6 (91.9)	89.5 (96.8)	94.2 (98.8)
Fourth degree	58.0 (78.7)	84.2 (94.6)	92.2 (98.0)	95.0 (99.0)
Fifth degree	65.5 (84.7)	85.3 (95.2)	92.6 (98.1)	95.0 (99.1)
Sixth degree	67.4 (86.5)	85.4 (95.3)	92.7 (98.1)	95.0 (99.1)
Total cities	1,139	1,419	1,879	2,707