WORKING PAPER

Six Degrees of GM Bacon: Network Analysis of Biotechnology Inventors

By
Daniel K.N. Johnson

Colorado College Working Paper 2006-02
August, 2006

Department of Economics and Business
Colorado College
Colorado Springs, Colorado 80903-3298
www.coloradocollege.edu/dept/EC

Six Degrees of GM Bacon:
Network Analysis of Biotechnology Inventors

By

Daniel K.N. Johnson

Abstract

Network analysis shows 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

Daniel K.N. Johnson is an Assistant Professor of Economics at Colorado College, 14 East Cache La Poudre Street, Colorado Springs, CO 80903, tel: (719) 389-6654, fax: (719) 389-6927, email:djohnson@coloradocollege.edu.

Special thanks to Milena Mareva and David Popp for assistance and suggestions on data and programming, as well as to anonymous referees for their helpful comments. Financial support was provided by a National Science Foundation Award for the Integration of Research and Education, a Wellesley College Faculty Grant, a Colorado College Faculty Grant and a Mrachek Research Fellowship.

Six Degrees of GM Bacon: Network Analysis of Biotechnology Inventors

Abstract

Network analysis shows 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.

According to a popular game, actors and actresses in Hollywood are measured by their degrees of separation away from Kevin Bacon. Co-stars with Kevin Bacon are one degree removed, and actors starring in films with anyone one degree removed are themselves two degrees removed. Popular wisdom suggests that no actor or actress is more than six degrees removed. Similar counts have been made for the biology, physics and mathematics disciplines (Newman, 2004).

This paper shows that the same wisdom holds for biotechnology. While we expand our analysis beyond genetically modified (GM) food research, the analogy to the popular game should be clear: networks between individual inventors, between firms, and between cities hosting those firms and inventors, have become more tightly knit than ever before. There is less periphery (fewer inventors with high degrees of removal) as most inventors display a close connection to the core (low degrees of removal).

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 and methodology, and Section IV presents results of the analysis. Section V concludes with implications for policy and further research.

II. Literature review

Empirical studies emphasize the 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). Those same studies also emphasize the importance of frequent personal contact and research collaboration. In particular, Zucker et al. (1998) show that firms using biotechnology are strongly influenced by the location of superstars in academic research institutions.

Lundvall (1992) points out that the importance of geography, and networks more generally, should differ predictably by technology type. 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 might expect a large geographic impact on knowledge flows and more highly concentrated or clustered activity over time.

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, and therefore actors, is an important driving factor in knowledge transfer.

To our knowledge there has been no social network analysis study of innovators within biotechnology.

III. Data and methodology

This paper relies exclusively on patent citations (Hall et al., 2001) 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 ¼ of all patent citations correspond to a clear spillover of knowledge, another ¼ have some possibility of a spillover, and the remaining ½ do not reflect knowledge transfers. However, their 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.[1]

Regardless of physical proximity, interpersonal linkages are undoubtedly important avenues of knowledge transfer. Based on sociology work studying the "Small

---

[1] 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 (plant patents) due to data limitations on these documents.

World Problem" (Milgram, 1967), and popularized by the play "Six Degrees of Separation" (Guare, 1990), we evaluate 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? A network that is growing tighter across more inventors or locations provides further strong evidence that distance is shrinking in importance, via the integration of less active centers into the research community.

To study this trend, we define the "core" inventors as the most active one percent of all biotechnology inventors (with most patents granted) in a time period. Zucker et al. (1998) performed the same exercise when showing that biotechnology-using firms were located close to superstar academicians in the field. 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.

One reason for geographic clustering of citations in biotechnology would simply be the geographic clustering of biotechnology firms themselves. Naturally, it takes more than the presence of other firms to create a citation, since citations are to particular patents, not firms.

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. 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.

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.

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.

IV. Results

Measures of biotechnology patenting at the state level show remarkable stability in relative rankings over time, while an explosion has occurred in the number of inventors, assignees, and cities represented in biotechnology patents. Furthermore, there is a strong trend toward growing equality between inventors (and between assignees) over time. The Gini coefficient of patent activity among inventors has fallen from 0.29 to 0.21, with a similar fall from 0.64 to 0.56 among assignees. Other concentration measures show the same pattern, suggesting that although states retain their relative rankings, the inventive process in biotechnology is one of increasing participation and lower market shares for each participant.

Tables 1 through 3 display descriptive statistics of the biotechnology innovation network at the level of the inventor, the assignee and the city respectively. Along with a portrait of the network from each degree of connectedness to the core, we also present four important dimensions of the network that are standard in the social network analysis literature (Hanneman and Riddle, 2005). Density is an indicator of the completeness of the network, calculated as the number of actual linkages between actors as a share of all possible connections. Naturally for a network of this type, low values are expected. Fragmentation is another measure of the completeness of the network, measuring the share of all agents who are unable to reach every other agent in the network via any number of consecutive linkages. Finally, the weighted clustering coefficient reflects the relative strength of subgroups within the network as the average local density (measured between agent pairs) of the network, with each agent weighted by their importance to the discipline.

Table 1 shows the network of inventors as it has grown to over three and half times as many members within twenty years. Over that period, the core has become less important as a patenting source, with over nine percent of all patents in 1975-79 and 6.6 percent in 1990-94. At the same time, inventors are connecting more closely to the core: 9 percent (584) of all inventors in 1975-79 are either in the core or within one degree of a core member, while 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.

The inventor network is very sparse, as expected, with density scores well below 0.003, and relatively unchanged over time. It would be rather horrifying to consider a dense network at the inventor level, where each inventor cited every other inventor at least once. The fragmentation of the network is similarly continuously high, meaning that relatively isolated inventors are still not connected into the network. Instead, separate clusters or subgroups have formed without linkage to the rest of the discipline. Notice that the clustering coefficient is roughly one hundred times as high as the density coefficient, meaning fairly concentrated innovation within the network. Again, not much has changed over the two decades presented here.

Table 1: Inventor Network

Period	*1975-79*	*1980-84*	*1985-89*	*1990-94*
Total members	6,490	8,284	12,933	23,388
Percent of Inventors (and Percent of Patents) in each degree				
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)
Network coefficients				
Density	0.0015	0.0024	0.0022	0.0017
Fragmentation	0.973	0.724	0.811	0.959
Weighted clustering	0.115	0.215	0.163	0.105
Ratio clustering: density	76.67	89.58	74.09	61.76

Table 2 presents the same analysis for assignees. There are fewer members, and a slightly lower growth rate in membership than inventors experienced (182 versus 260 percent over the entire period). Unlike inventors, the core firms are becoming more important over time, rising from ten percent of all patents to over fourteen percent of all patents. Connectedness is higher overall (compared to inventors), 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.

The assignee network is substantially more dense than the inventor network, meaning that firm-to-firm linkages are much more likely than inventor-to-inventor linkages, but of course the overall density coefficients are still low. Notice, however, that density has increased sixfold over the period as the core, first and second degree assignees have grown more important. At the same time, the network has become less and less fragmented, to the point where it is virtually impossible to find a firm unconnected to the network. This contrasts heavily with the inventor network.

Finally, clustering is much higher among assignees than for inventors, and has risen steeply over time. Much of that rise is due to the increasing density of the network (density is higher everywhere, so local densities are high as well), so the creation of very pronounced subgroups should be taken in context with a generally increased level of connectivity. The rise in clustering reflects the growing importance of the core at the expense of the periphery, although the periphery is becoming more connected than it was previously.

Table 2: Assignee Network

Period	*1975-79*	*1980-84*	*1985-89*	*1990-94*
Total members	943	1,170	1,817	2,666
Percent of Inventors (and Percent of Patents) in each degree				
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)
Network coefficients				
Density	0.0108	0.0219	0.0343	0.0642
Fragmentation	0.079	0.066	0.040	0.008
Weighted clustering	0.252	0.426	0.555	0.877
Ratio clustering: density	23.33	19.45	16.18	13.66

Finally, we consider cities (locations) as the base for analysis. The results in Table 3 suggest that more cities are involved in biotechnology than ever before (137 percent rise in locations over the period), and a *dramatic* increase in concentration of activity in the core and first degree locations. The core doubles its share of all patents over the period, while jointly the core and first degree triple their share from less than nineteen percent to over 54 percent. By 1990-94, virtually no patents are more than three degrees of location from the core. It would be difficult to imagine a more striking example of growing network concentration.

Table 3: Location Network

Period	*1975-79*	*1980-84*	*1985-89*	*1990-94*
Total members	1,139	1,419	1,879	2,707
Percent of Locations (and Percent of Patents) in each degree				
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)
Network coefficients				
Density	0.0051	0.0240	0.0245	0.0472
Fragmentation	0.137	0.001	0.024	0.016
Weighted clustering	0.147	0.367	0.297	0.590
Ratio clustering: density	28.82	15.29	12.12	12.50

Density of the network has increased markedly over time, as locations cite each other more often. One might propose changes in communication and transportation technology as an obvious reason for this (and other network density) increases. The share of locations cut off from the network dropped radically in 1980 and has remained low ever since, with low fragmentation scores evidencing this pattern.

Like assignees, clustering has risen quickly over the period, and has a similar potency compared to the movement in density scores. A greater connectedness and a stronger core are bringing strong clusters with them.

V. Conclusions

According to the popular maxim, every person on earth is connected to every other by no more than six degrees (Guare, 1990). While that is not quite 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, physical distance has become less important for knowledge spillovers, when measured as connectedness between formerly marginalized inventors, assignees or cities. While other authors have clearly demonstrated the conjoined nature of prominent actors or locations (e.g. Zucker et al., 1998), to our knowledge none has clearly described how this co-location process, or networked nature of the industry, has changed over time.

An obvious next step in this research line is to explain each actor's position relative to the core using measurable characteristics. Are there attributes common to the core which peripheral cities (or assignees or inventors) lack? Is there a monotonic relationship between key variables, such as scientists and engineers per capita, and degree of separation? Hopefully, we can formulate science or business or policy recommendations for biotechnology based on the relationships discovered.

Comparisons of the same patent citation dataset as applied to other industries or innovative activities would also be an instructive exercise. Other industries may have seen less dramatic change over time, and perhaps characteristics of each industry would be helpful in explaining the rates of change in density, fragmentation or clustering.

For now, it is clear that connectedness has become increasingly critical in biotechnology at all network levels of analysis. The implications for the discipline are clear: communication with the core and first degree are critical to success, and physical location in the same city is an advantage. While we are all more connected than ever before, those who are most connected with the core are stronger contributors to the field.

VI. References

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.

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.

Hanneman, R.A. and M.Riddle (2005). Introduction to social network methods. Riverside, CA: University of California, Riverside

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 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.

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.

Milgram, S. (1967) "The Small World Problem." Psychology Today. p. 60 - 67.

Newman, M.E.J. (2004) "Coauthorship networks and patterns of scientific collaboration". Proceedings of the National Academy of Sciences 101: 5200-5205.

Sjoholm, F. (1996) "International Transfer of Knowledge: The Role of International Trade and Geographic Proximity". Weltwirtschaftliches Archiv. Vol 132, p. 97-115.

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.