Complex Knowledge Networks and Invention Collaboration

Thomas F. Brantle
Wesley J. Howe School of Technology Management
Stevens Institute of Technology
Castle Point on Hudson
Hoboken, NJ 07030 USA
tbrantle@stevens.edu

M. Hosein Fallah, Ph.D.[†]
Wesley J. Howe School of Technology Management
Stevens Institute of Technology
Castle Point on Hudson
Hoboken, NJ 07030 USA
hfallah@stevens.edu

[†] Corresponding Author

1. Introduction

In today's world we are constructing ever increasingly integrated and interconnected networks for business, technology, communications, information, and the economy. The nature of these networks raises issues regarding not only their significance and consequence but also the influence and risk they represent. As a result it is vital to understand the fundamental nature of these complex networks. During the past several years advances in complex network analysis have uncovered amazing similarities among such diverse networks as the World Wide Web [Albert et al. (1999)], the Internet [Faloutsos et al.(1999)], movie actors [Amaral et al. (2000)], social [Ebel et al. (2002)], phone call [Aielo et al. (2002)], and neural networks [Watts and Strogatz (1998)]. Additionally, over the last few decades we have experienced what has come to be known as the information age or the knowledge economy. At the center of this phenomenon lies a complex and multifaceted process of continuous, widespread and far-reaching innovation advancement and technological change [Amidon (2002)], Cross et al. (2003), and Jaffe and Trajtenberg (2002)]. Understanding this process and what drives technological evolution has been of considerable interest to managers, researches, planners and policy makers worldwide. Complex network analysis offers a new and integrated approach to analyze the information flows and networks underlying this process.

1.1 Knowledge and Innovation Networks

Today, nations and organizations must look for ways of generating increased value from their assets. Human capital and information are the two critical resources. Knowledge networking is an effective way of combining individuals' knowledge and skills in the pursuit of personal and organizational objectives. Knowledge networking is a rich and dynamic phenomenon in which existing knowledge is shared, evolved and new knowledge is created. In addition, in today's complex and constantly changing business climate successful innovation is much more iterative, interactive and collaborative, involving many people and processes. In brief, success depends on effective knowledge and innovation networks. Knowledge collaboration and shared innovation, where ideas are developed collectively, result in a dynamic network of knowledge and innovation flows, where several entities and individuals work together and interconnect. These networks ebb and flow with knowledge and innovation the source and basis of technological advantage. Successful knowledge and innovation networks carry forth the faster development of new products and services, better optimization of research and development investments, closer alignment with market needs, and improved anticipation of customer needs resulting in more successful product introductions, along with superior competitor differentiation. [Skyrme (1999), Amidon (2002), and Cross et al. (2003)]

This paper discusses knowledge and innovation flows as represented by the network of patents and invention collaboration (i.e., inventors and their collaborators) and attempts to bridge recent developments in complex networks to the investigation of technological change and innovation progression. The recent discovery of small-world [Watts and Strogatz (1998)] and scale-free [Barabasi and Albert (1999)] network properties of many natural and artificial real world networks has stimulated a great deal of interest in studying the underlying organizing principles of various complex networks, which has led in turn to dramatic advances in this field of research. We focus on knowledge and innovation flows as represented by the historical records of patents, inventors and collaborators, with future application to technology management, knowledge transfer and innovation advancement.

1.2 Complex Network Analysis

Recent studies in complex networks have shown that the network's structure may be characterized by three attributes, the average path length, the clustering coefficient, and the node degree distribution. Watts and Strogatz (1998) proposed that many real world networks have large clustering coefficients with short average path lengths, and networks with these two properties are called "small world." Subsequently it was proposed by Albert et al. (1999) and Barabasi and Albert (1999) that many real world networks have power law degree distributions, with such networks denoted as "scale free." Specifically scale free networks are characterized by a power law degree distribution with the probability that a node has k links is proportional to $k^{-\gamma}$ (i.e., $P(k) \sim k^{-\gamma}$), where γ is the degree exponent. Thus, the probability that a node is highly connected is statistically more significant than in a random network, with the network's properties often being determined by a relatively small number of highly connected nodes, known as hubs. Because the power law is free of any characteristic scale, networks with a power law node degree distribution are called scale free. For our purposes "power law" and "scale free" are synonymous. [Albert and Barabasi (2002), Newman (2003), and Dorogovtsev and Mendes (2003)] In contrast, a random network [Erdos and Renyi (1959)] is one where the probability that two nodes are linked is no greater than the probability that two randomly chosen nodes happen to be associated, where the connectivity follows a Poisson or Normal distribution.

The Barabasi and Albert (BA) (1999) model suggests two main ingredients of self-organization within a scale-free network structure, i.e., growth and preferential attachment. They highlight the fact that most real world networks continuously grow by the addition of new nodes, and new nodes are then preferentially attached to existing nodes with large numbers of connections, a.k.a., the rich get richer phenomenon. In addition, Barabasi et al. (2002) and Newman (2004) have studied the evolution of the social networks of scientific collaboration with their results indicating that they may generally be characterized as having small world and scale free network properties.

1.3 Econometrics and Statistical Analysis

Patents have long been explored and recognized as a very useful and productive source of data for the assessment of technological and innovation development. A number of ground-breaking efforts and recent systematic empirical studies have attempted to conceptualize and measure the process of innovation advancement, knowledge spillover and technological change, plus the impact of the patenting process on patent quality, litigation and new technologies to our knowledge and information driven economy [(Griliches (1990), Jaffe and Trajtenberg (2002), Cohen and Merrill (2003)]. However, these historical and ongoing studies have primarily relied upon traditional statistical analysis and econometric modeling. It is believed that the application of complex network analysis will reveal not before seen associations and relationships leading to a further and improved understanding of these processes.

1.4 Invention, Knowledge and Technology

Patents provide a wealth of information and a long time-series of data about inventions, inventors, collaboration, prior knowledge, and the assigned owners. Patents and the inventions they represent have several advantages as a technology indicator. In particular, patents and patent citations have long been recognized as a very rich and fertile source of data for studying knowledge, innovation and technological change. As such, providing a valuable tool for public and corporate technology analysis, planning and policy decisions [Griliches (1990), Jaffe and Trajtenberg (2002), Cohen and Merrill (2003)]. Nevertheless, patents and invention collaboration have undergone limited investigation, thus offering a very rich information resource for knowledge and innovation research that is even less well studied and is yet to be fully exploited [Jaffe and Trajtenberg (2002)]. In a companion paper we analyze and discuss patents and patent citations from a complex network analysis perspective [Brantle and Fallah (2006)].

1.5 Organization of Paper

First, we review the network of knowledge and innovation flows as represented by the invention collaboration process. Next, the probability distribution of inventors and collaborators is examined, and then the power law network property of invention collaboration is analyzed. Finally, a summary and conclusions is presented.

2 Invention Collaboration

Patents and invention collaboration data contains relevant information allowing the possibility of tracing multiple associations among patents, inventors and collaborators. Specifically, invention collaboration linkages allows one to study the respective knowledge and innovation flows, and thus construct indicators of the technological importance and significance of individual patents, inventors and collaborators. For these reasons, an item of particular meaning, is the connections between patents and invention collaborators. Thus, if inventor 1 collaborates with inventor 2, it implies that inventor 1 shares or transfers a piece of previously existing knowledge with inventor 2, and vice versa, along with the creation of new knowledge as represented by the newly patented invention. As a result, not only is a flow of knowledge shared between the respective invention collaborators, but an invention link or relationship between the individual collaborators is established per the patented invention.

The supposition is that invention collaboration is and will be informative of the relationships between inventors and their collaboration in creating new knowledge and innovation. We discuss the construction of the invention collaboration network (i.e., the undirected network created by collaboration links) and discuss its relevance to knowledge and information. Next, summary statistics, probability distributions and the power law node degree distribution are analyzed.

2.1 Bipartite Graphs and Affiliation Networks

An invention collaboration network similar to that produced by the movie actor network [Watts and Strogatz (1998)] may be constructed for invention collaboration where the nodes are the collaborators, and two nodes are connected if two collaborators have coauthored a patent and therefore co-invented the invention. This invention affiliation or collaboration relationship can be easily extended to three or more collaborators. The relationship can be completely described by a bipartite graph or affiliation network where there are two types of nodes, with the edges connecting only the nodes of different types. A simple undirected graph is called bipartite if there is a partition of the set of nodes so that both subsets are independent sets. Collaboration necessarily implies the presence of two constituents, the actors or collaborators and the acts of collaboration denoted as the events. So the set of collaborators can be represented by a bipartite graph, where collaborators are connected through the acts of collaboration. In bipartite graphs, there are no direct connections between nodes of the same type, and the edges or links are undirected.

Figure 1 provides a bipartite graph or affiliation network representation with two sets of nodes, the first set labeled "patents" which connect or relate the second set labeled "invention collaborators" who are linked by the shared patent or invention. On the left we have the two mode network with three patents, labeled P_A, P_B and P_C, and seven collaborators, C_1 to C_7, with the edges joining each patent to the respective collaborators. On the right we show the one mode network or projection of the graph for the seven collaborators. It is noted that singularly authored patents would not be included in the bipartite graph and resulting invention collaboration network.

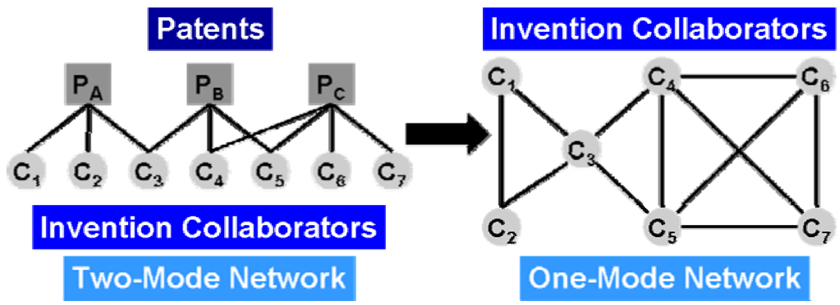

Figure 1 – Invention Collaboration Bipartite Graph or Affiliation Network

2.2 Knowledge and Innovation Flows

Patents and invention collaboration constitute a documented record of knowledge transfer and innovation flow, signifying the fact that two collaborators who coauthor a given patent, or equivalently co-invent said invention, may well indicate knowledge and innovation flowing between the respective collaborators along with the creation of new knowledge and innovation as represented by the new invention. The patent invention link and collaboration knowledge and innovation flow is illustrated Figure 2 and can be easily extended to three or more collaborators as appropriate. Therefore, knowledge and innovation information made publicly available by the patent has not only flowed to the invention, but has significantly influenced the invention's collaborators. Several network measures may be applied to the collaboration network in order to both describe the network plus examine the relationship between and the importance or significance of individual inventors and collaborators [Newman (2004)].

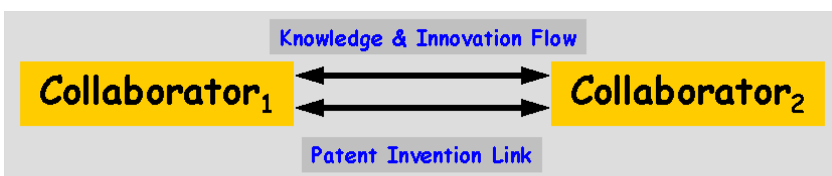

Figure 2 Invention Collaborator Knowledge & Innovation Flows and Invention Links

2.3 Patents, Inventors and Data

For this analysis we construct the invention collaboration network using the inventor data provided by the NBER (National Bureau of Economic Research) patent inventor file [Jaffe and Trajtenberg (2002)]. This file contains the full names and addresses of the inventors for patents issued from the beginning of 1975 through the end of 1999, comprising a twenty-five year period of patent production and invention collaboration. This includes approximately 4.3 million patent-inventor pairs, consisting of about 2.1 million patents and roughly 1.4 million inventors.

2.4 Invention Collaboration Distributions

In this section we discuss some basic statistics and distributions to provide a view of the basic structure and shape of invention collaboration. First the numbers of patents, inventors and collaborators, then the distribution of the number of inventors per patent, and finally the numbers of inventors that have collaborated or have not done so is presented.

2.4.1 Inventors per Patent Distribution

The average number of inventors per patent over the stated 25 year period is approximately 2.0, which is at the lower end of the range as compared to that discussed by Newman (2004) for authors per paper distributions for several scientific publications, where a range of two to almost four was observed. We conjecture that this is as intuitively anticipated, concluding that the number of inventors per patent is expected to be less than the number of scientists per publication. Additionally almost 49% of patents have only one inventor and are thus singularly invented without the aid of collaboration. The remaining 51% held by two or more inventors. The average number of inventors per patent for this set yields roughly 3.0, fifty percent more the previous average.

Our intuition and observation leads us to recognize that the reward in publishing journal articles by scientists is driven by a dual objective to produce both quality and perhaps quantity to as much or even greater extent, with the sharing of authorship (co-authorship) not necessarily having a diminishing or minimizing affect on these objectives. However, with regard to patents and the ownership of an invention, quality and the realized market value and economic potential are of a higher order of magnitude than scientific publication [Hall et al. (2005)]. So, invention collaboration followed by the sharing of patent rights and perhaps resulting economic reward and financial gains might have a minimizing or optimizing effect on the number of inventors per patent so as to not distribute the recognition and any resulting financial reward further than absolutely necessary. While as regards scientific publication, this goal is much less a concern, as having another author collaborate on a paper is not viewed as a negative factor, but as a potentially positive influence if it can both increase the quality and quantity of publication.

In addition to the average number of inventors per patent, it is even more interesting to observe the distribution p(k) of number of k inventors per patent. Figure 3 provides the distribution for the number of inventors per patent on logarithmic scales. It may be seen that the best fit line for this distribution follows approximately a power law distribution, with an exponent of about 4.5 [Dorogovtsev and Mendes (2003) and Newman (2005)]. We conclude that a power law distribution provides a reasonable fit to the data. However, as proposed by Newman (2004) and recently by Borner et al. (2004), a truncated power law distribution with an exponential cutoff distribution may also provide a suitable representation. In fact, this model when fit to the data provides excellent improvement in total variance explanation ($R^2 \approx 1.0$). This orderly deviation from a power law distribution is that the highest partnering (co-inventing) inventors are partnering less often than predicted by the power law, similarly the lowest partnering (co-inventing) inventors are partnering more often than predicted. As discussed by Borner et al. (2004) a realistic basis for this deviation is that in many networks where aging occurs, these networks show a connectivity distribution that possess a power law organization followed by exponential or Gaussian decay [Amaral et al. (2000)].

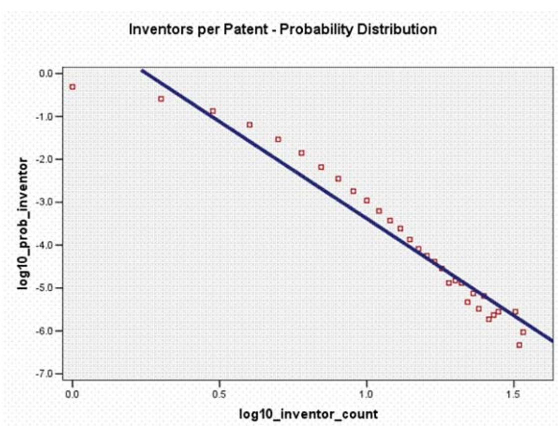

Figure 3 – Inventors Per Patent: Probability Degree Distribution

As a further explanation for the improved fit of the truncated power law with exponential cutoff model over simply a power law model might be attributed to the reduced incentive to partner (co-invent), and thus limit the number of partnering inventors per patented invention to that which is both necessary and sufficient. Again, so as to not further allocate the potential market value, possible resulting economic reward and financial gains across a larger set of inventors, i.e., not wanting to split the reward beyond an individual's contribution factor. Thus, since patents tend to have a higher variance in terms of their market value and financial reward, as compared to scientific publication, the number of partnering inventors is impacted significantly by the value of an additional inventor's contribution factor.

So, the number of inventors per patent roughly follows a power law distribution, where the numbers of inventors per patent falls off as $k^{-\gamma}$ for some constant $\gamma \approx 4.5$, indicating that some inventions necessitate increased partnership and a larger number of inventors, while most inventions are achieved by a smaller or more modest number of inventors.

2.4.2 Inventors With and Without Collaboration

Although our focus is primarily on the network of invention collaboration, we also examined the distribution for the number of collaborators per inventor which includes inventors that did not collaborate and thus would simply be disconnected isolated nodes in the network, i.e., would have degree or number of collaborators equal to zero (0). Figure 4 presents the distribution for the number of collaborators per inventor which considers all inventors independent of collaboration or not. We first note that around 22% (≈ 0.3 million) of the total number of inventors (≈ 1.4 million) did not collaborate, with a corresponding 78% (≈ 1.1 million) having one or more collaborators. The average number of collaborators per inventor when considering all inventors with and without collaboration is about 3.6, however if we restrict the inventor set to only those that did collaborate, this average increases to just about 4.7.

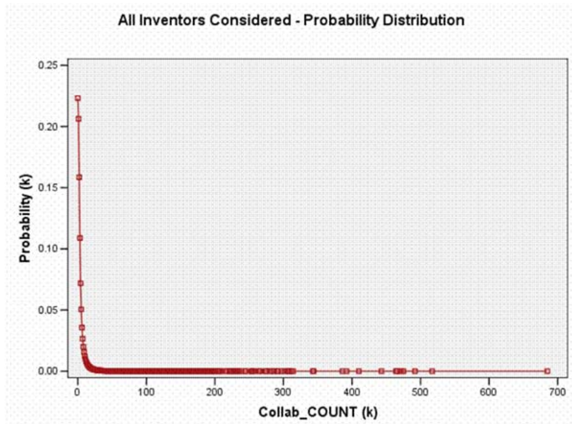

Figure 4 – Invention (w & w/o) Collaboration, Collaborators Per Inventor Probability Degree Distribution

2.5 Invention Collaboration Network

This section discusses the scale free or power law degree distribution of the invention collaboration network.

2.5.1 Power Law Degree Distribution

Figure 5 provides the probability distribution for the invention collaboration network on logarithmic scales. It may be seen that the best fit line for this distribution follows roughly a power law distribution with an exponent of 2.9 [Dorogovtsev and Mendes (2003) and Newman (2005)]. Hence we conclude that a power law distribution provides a reasonable fit to the data. We again note that a truncated power law distribution with an exponential cutoff may provide a suitable representation, with an associated improvement in the explanation of total variance ($R^2 \approx 1.0$). As discussed this methodical deviation from a power law distribution is that the highest collaborating (co-inventing) inventors are collaborating less often than predicted and correspondingly the lowest collaborating (co-inventing) inventors are collaborating more often than predicted. Once more, a reasonable rationale for this deviation is that in many networks where aging occurs, these networks show a connectivity distribution that possess a power law organization followed by an exponential or Gaussian decay distribution [Amaral et al. (2000)].

Following our previous discussion, the improved fit of the truncated power law with exponential cutoff model, may be attributed to a distinction in the objectives of invention patenting versus scientific publishing. Accordingly the collaboration on patenting an invention followed by the sharing of patent rights and the further dividing of potential economic rewards and financial gains might have a rather minimizing or at least optimizing effect on any incentive to increase the number of collaborators. It would be expected that inventors would evaluate and weigh the potential technical contribution against the economic and financial impact of the prospective collaboration on the invention and its shared ownership. Again, with respect to scientific publication this objective is much less of a consideration.

Consequently, with the patent invention collaboration network viewed as an undirected network, the degree exponent of the number of patent invention collaborators is approximately 2.9. Therefore, we have demonstrated that the number of invention collaborators roughly follows a power law distribution. That is, the numbers of collaborators per inventors falls off as $k^{-\gamma}$ for some constant $\gamma \approx 2.9$, implying that some inventors account for a very large number of collaborators, while most inventors collaborate with just a few and smaller number of additional collaborators. These results are consistent with the theoretical and empirical work of Albert and Barabasi (2002), Newman (2003), Dorogovtsev and Mendes (2003), and others where a degree exponent of approximately 3.0 is expected and predicted for very large networks, under the assumptions of growth and preferential attachment.

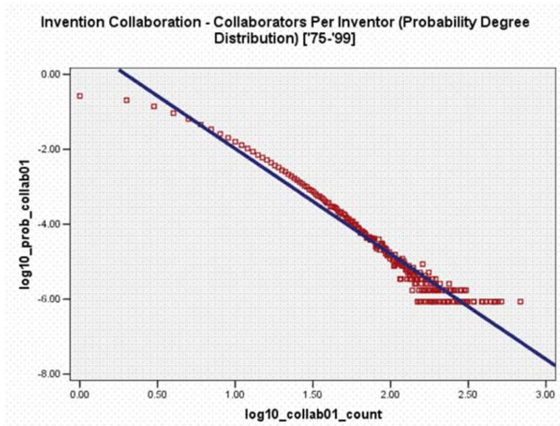

Figure 5 – Invention Collaboration, Collaborators Per Inventor

3 Summary, Discussion and Conclusions

This section provides a brief summary, a few suggested areas of application, along with conclusions.

3.1 Summary

We have studied knowledge and innovation as typified by the network of patents and invention collaboration, as well as discussed the importance of this network along with its underlying processes to the advancement of knowledge exchange and technological innovation. In particular we have suggested that this area of research while traditionally investigated via statistical analysis and econometric modeling may well be further investigated and advanced via complex network analysis. We also have demonstrated the scale free power law property of the invention collaboration network, showing that the network may be characterized by a power law degree distribution, where the probability that an inventor or collaborator being highly connected is statistically more significant than would be expected via random connections or associations. As a result the network's properties now being determined by a relatively small number of highly connected inventors and collaborators known as hubs.

3.2 Areas of Application

The benefit of further network analysis of the invention collaboration process however has even further implications for technological expansion and global economics. Immediate areas of continued investigation and potential application include: technology clusters and industrial districts [Saxenian (1994), Porter (1998), Jaffe and Trajtenberg (2002)], knowledge spillover and technology transfer [Cockburn and Henderson (1998), Jaffe et al. (2000), Jaffe and Trajtenberg (2002)], and patent quality, litigation and new technology patenting [Cohen and Merrill (2003), Lanjouw and Schankerman (2004)]. Analyses of invention collaboration and application to these areas from a complex network analysis perspective should provide a deeper understanding as to their underlying structure and evolution which may influence private and public policy decision making and planning initiatives.

3.3 Conclusions

The patent invention collaboration process results in the sharing of knowledge along with the creation of new technological innovations. Similarly the patenting process and resulting invention collaboration further advances and accelerates the diffusion of the knowledge exchanged and innovation created. Thus, there has been significant effort and research placed both recently and in the past into investigating the organization, development and progression of knowledge and innovation, and their impact on technology advancement and the global economy. However, only recently has this been possible due to the increased access and availability of very large high quality datasets plus the computer algorithms, tools and techniques necessary to analyze this vast array of data and information.

This paper is a work in-progress. It is our hope that our continued research in this area helps provide additional foundation and motivation for the wide-scale use of complex network analysis as an important approach to both empirical as well as theoretical studies of the structure, evolution and dynamics of knowledge and innovation. It is believed that invention collaboration as considered by complex network analysis measures offers tremendous potential for providing a theoretical framework and practical application to the role of knowledge and innovation in today's technological and information driven global economy.

R References

Aielo, W., Chung, F., and Lu, L. (2002). Random Evolution of Massive Graphs. In Abello, J. Pardalos, P.M., and Resende, M.G.C. eds., *Handbook of Massive Data Sets*. (pp. 97-122) Dordrecht, The Netherlands: Kluwer Academic Publishers.

Albert, R. and Barabasi A.L. (2002). Statistical Mechanics of Complex Networks. *Reviews of Modern Physics, 74*, 47-97.

Albert, R., Jeong, H. and Barabasi A.L. (1999). Diameter of the World-Wide Web. *Nature, 401*, 130-131.

Amaral, L.A.N., Scala, A., Barthelemy M., and Stanley, H.E. (2000). Classes of Small World Networks. *Proceedings National Academy Sciences, USA*, *97*, 21, 11149-11152.

Amidon, D.M. (2002). *The Innovation SuperHighway: Harnessing Intellectual Capital for Collaborative Advantage*. Oxford, UK: Butterworth-Heinemann.

Barabasi, A.L. and Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*, 509-512.

Barabasi, A.L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., and Vicsek, T. (2002). Evolution of the Social Network of Scientific Collaborations. *Physica A, 311*, 590-614.

Borner, K., Maru, J.T. and Goldstone, R.L.. (2004). The Simultaneous Evolution of Author and Paper Networks. *Proceedings National Academy Sciences, USA*, *101*, 1, 5266-5273.

Brantle, T.F. and Fallah, M.H. (2006). Scale Free Innovation Networks: Applications to Technology Clusters and Patenting Processes. *Stevens Working Paper.* In Progress.

Cockburn, I.M. and Henderson, R.M. (1998). Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery. *Journal Industrial Economy*, *46(2)*, 157-182.

Cohen, W.M. and Merrill, S.A., eds. (2003). *Patents in the Knowledge-Based Economy*. Washington, DC: The National Academic Press.

Cross, R., Parker, A. and Sasson, L., eds. (2003). *Networks in the Knowledge Economy*. New York, NY: Oxford University Press.

Dorogovtsev, S.N. and Mendes, J.F.F.F. (2003). *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford, Great Britain: Oxford University Press.

Ebel, H., Mielsch, L.I. and Bornholdt, S. (2002). Scale-Free Topology of E-mail Networks. *Physical Review E, 66,* 035103.

Erdos, P. and Renyi, P. (1959). On Random Graphs. *Publicationes Mathematicae, 6*, 290-297.

Faloutsos, M., Faloutsos, P. and Faloutsos, C. (1999). On Power Law Relationships of the Internet Topology. *Computer Communication Review*, *29(4)*, 251-262.

Griliches, Z. (1990). Patent Statistics as Economic Indicators: A Survey. *Journal of Economic Literature*, *28(4)*, 1661-1707.

Hall, B., Jaffe, A. and Trajtenberg, M. (2005). Market Value and Patent Citations. *Rand Journal of Economics 36*, 16-38

Jaffe, A. and Trajtenberg, M., eds. (2002). *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. Cambridge, MA: MIT Press.

Jaffe, A., Trajtenberg, M. and Fogarty, M. (2000). Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors. *American Economic Review, Papers and Proceedings*, *90,* 215-218.

Lanjouw, J and Schankerman, M. (2004). Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators. *Economic Journal*, *114,* 441-465.

Newman, M.E.J. (2003). The Structure and Function of Complex Networks. *SIAM Review, 45*, 167-256.

Newman, M.E.J. (2004). Who is the Best Connected Scientist? A Study of Scientific Co-authorship Networks. In Complex Networks, E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai (eds.), pp. 337-370, Springer, Berlin.

Newman, M.E.J. (2005). Power Laws, Pareto Distributions and Zipf's Law. *Contemporary Physics, 46,* 323-351.

Porter, M.E. (Nov.-Dec. 1998). Clusters and the New Economics of Competition. *Harvard Business Review*, *78*, 77-90.

Saxenian, A. (1994). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Skyrme, D.M. (1999). *Knowledge Networking: Creating the Collaborative Enterprise*. Oxford, UK: Butterworth-Heinemann.

Watts, D.J. and Strogatz, S.H. (1998). Collective Dynamics of Small-world Networks. *Nature*, *393*, 440-442.