

## Tracing knowledge diffusion

CHAOMEI CHEN,<sup>a</sup> DIANA HICKS<sup>b,c</sup>

<sup>a</sup>*College of Information Science and Technology, Drexel University, Philadelphia PA (USA)*

<sup>b</sup>*School of Public Policy, Georgia Institute of Technology, Atlanta GA (USA)*

<sup>c</sup>*CHI Research, Inc., Haddon Heights, NJ (USA)*

Knowledge diffusion is the adaptation of knowledge in a broad range of scientific and engineering research and development. Tracing knowledge diffusion between science and technology is a challenging issue due to the complexity of identifying emerging patterns in a diverse range of possible processes. In this article, we describe an approach that combines complex network theory, network visualization, and patent citation analysis in order to improve the means for the study of knowledge diffusion. In particular, we analyze patent citations in the field of tissue engineering. We emphasize that this is the beginning of a longer-term endeavor that aims to develop and deploy effective, progressive, and explanatory visualization techniques for us to capture the dynamics of the evolution of patent citation networks. The work has practical implications on resource allocation, strategic planning, and science policy.

### Introduction

Knowledge diffusion can be defined as the adaptations and applications of knowledge documented in scientific publications and patents. Tracing the transfer of knowledge from science to technology, from technology to technology, or from defense to civil sectors is currently the most popular area of research in relation to citation analysis.<sup>1</sup> Citations in both scientific publications and patents have been regarded as the most fundamental indicators of impact. However, tracing knowledge diffusion between science and technology remains a challenging issue. Conventional approaches are often qualitative in nature, including interviews, questionnaires, and in-depth case studies. Such methods are often time-consuming, expensive to use, and requiring a substantial level of prior domain knowledge.

Crane<sup>2</sup> identified the crucial role of scientific communities in understanding the growth of knowledge. The growth of scientific knowledge is largely due to a diffusion process in which new ideas are transmitted from person to person. The exponential increase in the number of publications is a good indicator of the existence of such diffusion processes. In contrast, the absence of a diffusion process is more likely to demonstrate a linear growth pattern.

Received June 24, 2003

*Address for correspondence:*

CHAOMEI CHEN

College of Information Science and Technology, Drexel University, Philadelphia PA, 19104-2875, USA

E-mail: chaomei.chen@cis.drexel.edu

0138–9130/2004/US \$ 20.00

Copyright © 2004 Akadémiai Kiadó, Budapest

All rights reserved

The ‘references cited’ section in a U.S. patent is required by patent law in order to establish that an issued patent meets three general criteria: its usefulness, its novelty, and that it is not obvious from prior art. Non-patent references (NPRs) on the front page of U.S. patents contain a mixed list of references to scientific journal papers, books, and other forms of documentations. These references are listed as “other references cited.” Because they appear on the front pages of U.S. patents, they are also known as front-page patent-to-paper citations.

Knowledge Domain Visualization (KDViz) is a rapidly growing field, primarily concerned with the analysis and modeling the structure and dynamics of a scientific and technological domain.<sup>3-10</sup> Related issues have been also addressed by researchers over the last decade.<sup>11-15</sup> In this article, we will describe an integrative approach to tracing knowledge diffusion based on progressive and explanatory visualization of the evolution of patent citation networks. We will demonstrate the potential of integrating complex network theory, network visualization, and KDViz through an illustrative study of patent citations associated with tissue engineering.

### Science and technology linkage

Carpenter et al.<sup>16</sup> studied patent-to-paper citations in 319 gas laser patents and 399 prostaglandin patents from the U.S. Patent Office (USPTO) and found that nearly 90% of all journal references made by patent applicants and examiners refer to basic or applied scientific journals, as opposed to engineering and technological literature. They also found that the time between publication of a journal article and the patent application citing that article was relatively short – generally, three to five years. In addition, patent applicants and examiners tend to cite scientific articles in the central core of the scientific literature covered by the *Science Citation Index* (SCI).

Narin and his colleagues at CHI Research provided large-scale evidence detailing a massive, contemporary linkage between industrial technology and public science, with a tripling of the knowledge links from U.S. technology to U.S. science in just six years, based on an analysis of 430,226 NPRs from 397,660 U.S. patents issued in 1987–1988, and 1993–1994.<sup>17,18</sup> They concluded that public science plays an essential role in supporting U.S. industry, across all the science-linked areas of industry, amongst companies large and small, and is a fundamental pillar of the advance of U.S. technology.

There are often multiple factors that may influence the predominant route and direction of knowledge transfer between particular scientific disciplines and technological sectors. In fields such as health and semiconductor research there tends to

be a strong positive connection between basic research and technological innovations, whereas in fields such as information technology it is technology that leads science by more than a year according to the publication dates of cited patents and scientific publications.<sup>19</sup> In contrast, Meyer<sup>20</sup> studied the relationship between nanoscience and nanotechnology based on patent-to-paper citations, and concluded that they appear to be two different disciplines. Al-Thubaity and Ahmad<sup>21</sup> studied the emergent domain of nano-structured tunnel diodes in semiconductor physics based on patent descriptions retrieved from U.S. Patent and Trademark Office (USPTO). They examined the use of terms at the lexical level as in order to identify a consensus in the use of these terms and thereby understand how knowledge evolves in an emergent domain. However, much of their work was carried out manually, which tends to be limited in terms of flexibility, cost-effectiveness, and scalability.

### Statistical mechanics of networks

Statistical mechanics of complex networks have recently become the center of the attention in several scientific communities, including statistical physics, computer science, and information science. These studies focused on the topological properties of large networks (the Internet, the Web, scientific networks) and found some surprising similarities. The latest advances are primarily rooted in two types of networks known as small-world networks<sup>22</sup> and scale-free networks.<sup>23</sup> A particularly relevant line of research is the study of mechanisms that can explain, in statistical terms, topological properties demonstrated by a class of networks. Barabási and his colleagues<sup>24</sup> found that preferential attachment mechanisms could produce the topological properties of the co-authorship networks of mathematicians and neuroscientists over an 8-year period (1991–1998).

The degree of a node is the number of links to the node. Scale-free networks are characterized by an extremely skewed distribution with a long tail.<sup>25</sup> Mathematically, such distributions can be described by a power law, which means that the probability of finding a node with  $k$  links to other nodes is proportional to  $k^{-\gamma}$ . The size of the exponent  $\gamma$  has been the focus of a large number of studies. For instance, it was found to be 1.5 for networks of words, 2.2 for metabolic networks, 2.5 for protein-protein interactions, 2.5 for collaboration networks, and between 2.5 and 3.0 for citation networks.<sup>26</sup> Redner<sup>27</sup> found a power-law degree distribution, with an exponent of 3, in two large citation networks. He suggested that the citation distribution is an appealing venue for theoretical modeling. It is a reasonable expectation that patent citations are likely to follow a power law as well.

### Paper- and patent-citation analysis

Price<sup>28</sup> introduced the notion of research fronts – the collection of highly cited papers that represent the frontiers of science at a particular point of time. Based on an examination of citation patterns of scientific papers, he conjectured that it is possible to identify objectively defined subjects in citation networks. He particularly emphasized the significance of understanding the nature of such moving frontiers in the development of a quantitative method for delineating the topography of current scientific literature.

In the 1970s, Small and Griffith examined issues concerned with identifying specialties by mapping the structure of scientific literatures, especially through analyses of co-citation networks.<sup>29</sup> Small subsequently found rapid changes of focus in collagen research.<sup>30</sup> Documents clustered by their co-citation links can represent leading specialties. The abrupt disappearance and emergence of such document clusters indicate rapid shifts in research focus. By tracing key events through a citation network, Hummon and Doreian<sup>31</sup> successfully re-constructed the most significant citation chain in the development of DNA theory.

Tracing the transfer of knowledge requires us to consider how we can move from one realm of citation analysis to the other through a consistent and meaningful framework. Therefore, in addition to the two types of citation research, one must consider citation analysis of interrelationships of heterogeneous structures (See Table 1).

Table 1. Citation analysis of knowledge transfer between science and technology

	Knowledge transfer	Citation analysis	Patent citation analysis	Research focus of our study
1	from science to science	Most often	Rare	Secondary
2	from science to technology	Rare	Rare	Primary
3	from technology to science	Rare	Often	Primary
4	from technology to technology	Rare	Most often	Secondary
5	absence of a strong connection	Rare	Rare	Primary

In Table 1, each type of knowledge transfer is ranked by the frequency of being primarily studied in terms of rare, often, and most often in science citation analysis and patent citation analysis, respectively. For example, the first type of knowledge transfer

is a typical topic of citation analysis of scientific citations, whereas the fourth one has been the focus of the majority of patent citation analysis. The primary and secondary focuses of our research are identified in the table. The current table is based on our understanding of the contemporary practice in citation research.

National Bureau of Economic Research (NBER) – <http://www.nber.org/patents/> – has made available patent citation datasets, comprising nearly 3 million U.S. patents granted between January 1963 and December 1999, and over 16 million citations made to these patents between 1975 and 1999. Figure 1 depicts the citation degree distribution based on the 16 million citations to patents. The distribution follows a power law, with an exponent of 2.89, suggesting that patent citation networks are scale-free, not a real surprise.

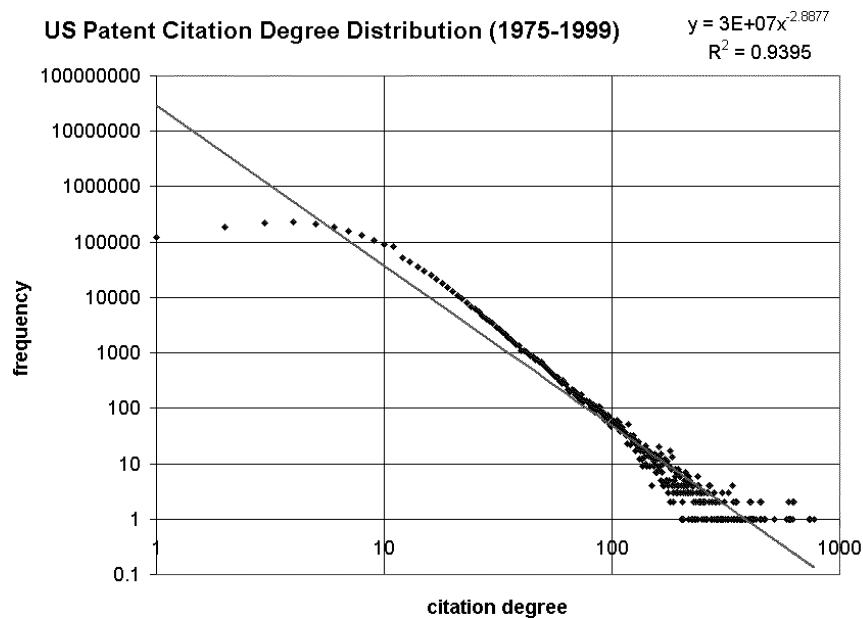


Figure 1. Patent citation networks are scale-free.  
 A patent-to-patent citation degree distribution based on the NBER dataset follows a power law with an exponent of 2.89

### Tissue engineering

Tissue engineering is defined as the application of principles and methods of engineering and life sciences toward fundamental understanding of structure-function relationships in normal and pathological mammalian tissues, and the development of biological substitutes to restore, maintain, or improve tissue function.<sup>32,33</sup> The term “tissue engineering” was coined in 1987. Starting from a few modest NSF grants in the mid-1982s, followed by major funding from NIH and NIST, the field has over \$3 billion in funding over the past decade, much of it from private sources.<sup>34</sup> Some products are already in use clinically, and their number is rapidly increasing. Given that the national health care cost associated with tissue loss and organ failure is estimated to be over \$400 billion per year, tissue engineering has the potential to revolutionize methods of health care treatment and dramatically improve the quality of life for millions of people throughout the world. Tissue engineering combines both basic research and technological inventions. In addition, it is a multidisciplinary field of research firmly rooted in a number of fields of basic research as well as the strong pharmaceutical industry “pull.” There are a large number of patents in this field.

Paper-to-paper citation data on tissue engineering was extracted from the Web of Science, using a simple query “tissue engineering.” It should be noted that although such simple queries may not produce a comprehensive collection of articles published in the field of tissue engineering, as the first step essential patterns and dynamics of the field cannot escape the search unnoticed. Patent citation data on tissue engineering, especially front-page patent-to-science citations, were extracted from the patent databases maintained by CHI, containing 267 U.S. patents which in turn made 5,387 patent-to-science citations to 562 unique scientific articles. Figure 2 shows that the paper-to-paper citation distribution follows a power law, with an exponent of 2.31, which is in line with the available findings in the literature.

Figure 3 shows a front-page patent-to-paper citation matrix. The patents in our dataset cited papers published ranging from 1950s up to date. For each patent-to-paper citation, we plotted both the application date and the issue date of the citing patent. There are several interesting results. First, the lag between the publication year of a paper and the first front-page citation in a patent has a mean of 9.6 years, and a mode of 2 years. The oldest paper cited by a patent was published 82 years ago, whereas the youngest one was published within the same year. Second, the waiting from the application date to the issue date is about 3 years in average, with a mode of 2 years.

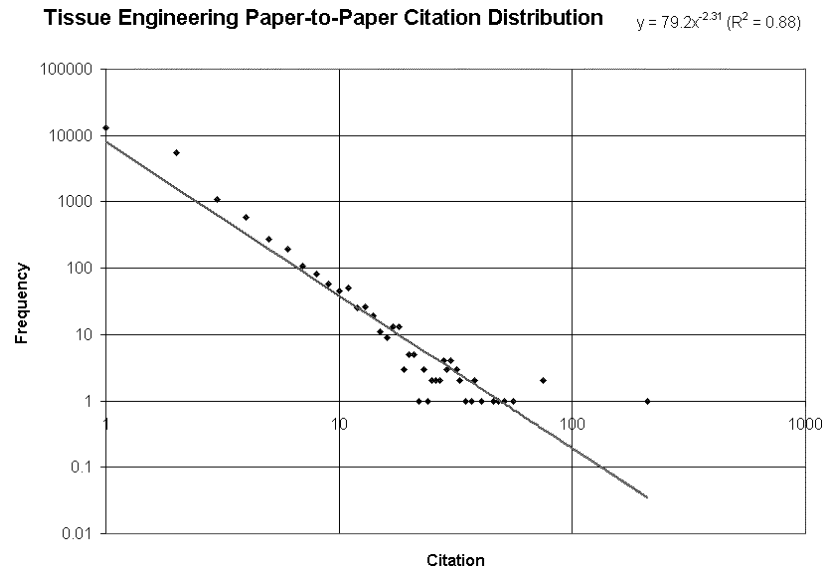


Figure 2. The paper-to-paper citation degree distribution for tissue engineering

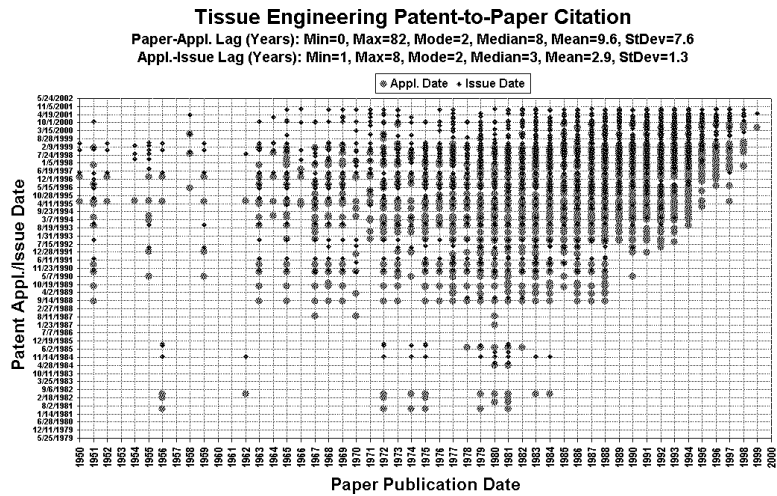


Figure 3. Patent-to-paper citations by patents' application dates and issue dates

The longest one took 8 years, and the quickest one just 1 year. The figure appears to show a surge of patent applications between the beginning of 1993 and the end of 1995. Patent applications made in this period cited widely scientific papers that were mainly published across the period of 1960–1995. This pattern suggests that this must be a significant stage of the tissue engineering field.

Figure 4 represents three distributions associated with patent-to-paper citations: two for in-degree citations and one for out-degree citations. Since patent-to-paper citations involve two heterogeneous types of entities, patents are the source and papers are the destination, in-degree refers to the number of coming-in citations received by a paper and out-degree refers to the number of going-out citations from a patent, or the number of papers the patent cites. We considered two types of in-degree citation distributions because some papers have acknowledgements to funding sources and we conjectured that papers resulted from funded research are more likely to have an impact on inventions described in a patent. One of the in-degree distributions is based on all papers cited; the other is based on papers that have explicitly acknowledged the support of funding. Both in-degree distributions follow power laws: the one with funding has an exponent of 1.97 (the line fitted on the void squares in the figure), whereas the one without has an exponent of 2.16 (the line fitted on the solid ‘diamonds’ in the figure), suggesting that funded research tends to be cited relatively more frequently by patents than research without explicit funding information.

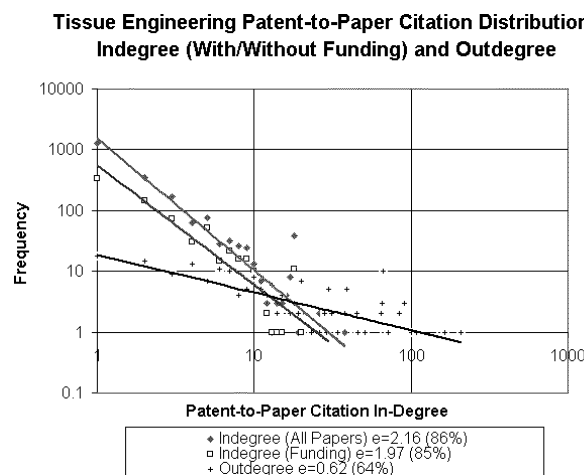


Figure 4. Three patent-to-paper citation distributions, including in-degree with explicit funding acknowledgement, in-degree without, and out-degree of citations



### Visualization of diffusion paths

Figure 5 is a visualization of the underlying co-citation network based on scientific literature. Each sphere in the scene denotes a published paper. A link between a pair of spheres represents a link preserved by Pathfinder network scaling.<sup>4,7</sup> The height of the semi-transparent vertical bar on a sphere depicts the number of citations the underlying article has received, in our case, since 1995. Articles with over 50 citations are labeled.

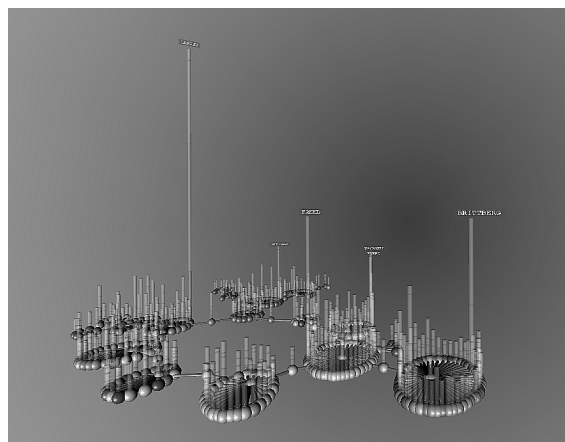


Figure 5. A landscape of a co-citation network of scientific papers in tissue engineering. Articles with over 50 citations are labeled in the scene. From left to right: Langer (1993), Niklason (1999), Freed (1994), Vacanti (1991), and Brittberg (1994)

Figure 6 shows four frames from an animated visualization sequence. The overall topology of the tissue engineering topic consists of two clusters of papers. The color of a paper indicates its membership of a primary specialty. Unlike what we have seen in our studies of scientific citation networks, there are only two major specialties instead of three or more. One specialty consists of papers colored in light red; the other in green. The light-red one tends to be more predominant in recent years, whereas the green one is likely to be secondary in terms of its impact.\*

---

\* For a colored version of Figures 5 and 6, please, contact the authors.

It is particularly interesting to note that there is a special link that provides not only a crucial section of the path between the two clusters, but also the only link that connects two spheres of different colors. The arrow in each frame in Figure 6 points out this link. The western endpoint of the link, in light red, is an article published by Wozney et al. in 1988; the eastern endpoint of the link, in green, is by Sirica, Hwang, Sattler, and Pitot in 1980. Their bibliography is as follows:

SIRICA, A., HWANG, C., SATTLER, G., PITOT, H. (1980) Use of primary cultures of adult rat hepatocytes on collagen gel-nylon mesh to evaluate carcinogen-induced unscheduled DNA synthesis. *Cancer Research*, 40, 3259–3267.

WOZNEY, J. M., ET AL. (1988) Novel regulators of bone formation: Molecular clones and activities. *Science*, 242 : 1528–1534.

In order to identify the implications of this link, it is necessary to examine the context in which the two articles were cited in a patent. To illustrate how this can be done, we located the six U.S. patents that have co-cited the two articles (See Figure 6). From the point of view of knowledge diffusion, such links are potentially important because of their strategic position in the network. We can trace the nature of the link and uncover underlying connections. Just as in citation analysis of scientific literature, it is impossible to fully understand the implications of each instance without in-depth analysis of the context of such citations. We will continue to investigate knowledge diffusion in association with tissue engineering in the near future.

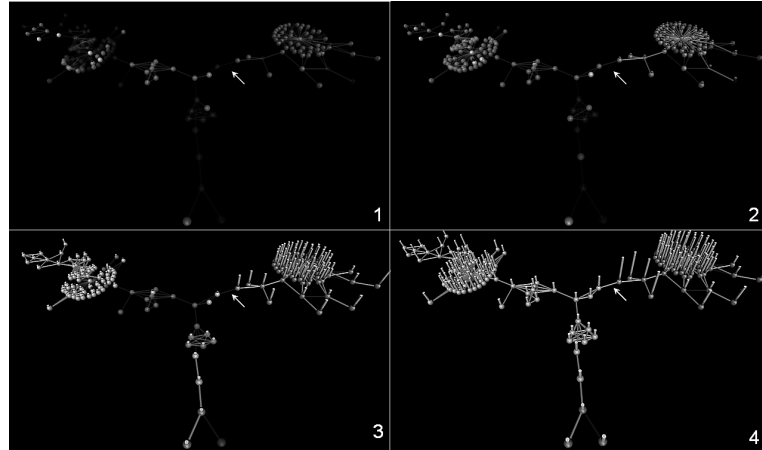


Figure 6. Four frames of animation, showing the evolution of a network of scientific papers derived from front-page patent-to-paper citations. The arrow in each frame points to a potentially interesting “bridge” link that connects two “continents” together

Table 2. US Patents cited both SIRICA et al. (1980) and WOZNEY et al. (1988)

Patent #	Appl. Date	Issue Date	Inventor	Title
5716404	12/16/1994	2/10/1998	Vacanti, et al.	Breast tissue engineering
5741685	6/7/1995	4/21/1998	Vacanti	Parenchymal cells packaged in immunoprotective tissue for implantation
5855610	5/19/1995	1/5/1999	Vacanti, et al.	Engineering of strong, pliable tissues
5863531	6/7/1995	1/26/1999	Naughton, et al.	In vitro preparation of tubular tissue structures by stromal cell culture on a three-dimensional framework
6129761	6/7/1995	10/10/2000	Hubbell	Injectable hydrogel compositions
6140039	1/25/1999	10/31/2000	Naughton, et al.	Three-dimensional filamentous tissue having tendon or ligament function

### Conclusions

We have discussed some of the issues concerning knowledge diffusion and how to trace the process of knowledge diffusion by utilizing patent citation networks. We have illustrated a potentially useful approach, combining statistical mechanics of complex networks, network visualization, and citation analysis. The goal of our research is to improve the understanding of knowledge diffusion and technology transfer, especially with principles and streamlined methodologies for citation analysis, and the expanded scope of citation analysis. This is the beginning of a longer-term research program which aims to improve our understanding of large-scale complex networks by developing and deploying increasingly powerful progressive and explanatory visualization techniques. Comprehensive evaluative studies are necessary to identify strengths and weaknesses of various available approaches to tracing knowledge diffusion.

### References

1. C. OPPENHEIMER, Do patent citations count?, In: *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*, B. CRONIN, H. B. ATKINS (Eds), ASIS, 2000, pp. 405–432.
2. D. CRANE, *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. Chicago, Illinois: University of Chicago Press, 1972.
3. C. CHEN, J. KULJIS, The rising landscape: A visual exploration of superstring revolutions in physics, *Journal of the American Society for Information Science and Technology*, 54 (2003) 435–446.
4. C. CHEN, *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*. London: Springer-Verlag, 2002.

5. C. CHEN, T. CRIBBIN, R. MACREDIE, S. MORAR, Visualizing and tracking the growth of competing paradigms: Two case studies, *Journal of the American Society for Information Science and Technology*, 53 (2002) 678–689.
6. C. CHEN, J. KULJIS, R. J. PAUL, Visualizing latent domain knowledge, *IEEE Transactions on Systems, Man, and Cybernetics*, 31 (2001) 518–529.
7. C. CHEN, R. J. PAUL, Visualizing a knowledge domain's intellectual structure, *Computer*, 34 (2001) 65–71.
8. H. SMALL, A passage through science: Crossing disciplinary boundaries, *Library Trends*, 48 (1999) 72–108.
9. C. CHEN, Visualising semantic spaces and author co-citation networks in digital libraries, *Information Processing and Management*, 35 (1999) 401–420.
10. K. W. BOYACK, B. N. WYLIE, G. S. DAVIDSON, Domain visualization using VxInsight for science and technology management, *Journal of the American Society for Information Science and Technology*, 53 (2002) 764–774.
11. R. R. BRAAM, H. F. MOED, A. F. J. V. RAAN, Mapping of science by combined co-citation and word analysis. I: Structural aspects, *Journal of the American Society for Information Science*, 42 (1991) 233–251.
12. S. G. STEINBERG, The ontogeny of RISC, *Intertek*, 3(1994) 1–10.
13. R. J. W. TIJSSEN, A. F. J. V. RAAN, Mapping changes in science and technology: Bibliometric co-occurrence analysis of the R&D literature, *Evaluation Review*, 18(1994) 98–115.
14. Y. DING, G. CHOWDHURY, S. FOO, Mapping the intellectual structure of information retrieval studies: an author co-citation analysis, 1987–1997, *Journal of Information Science*, 25 (2000) 67–78.
15. E. NOYONS, Bibliometric mapping of science in a science policy context, *Scientometrics*, 50 (2001) 83–98.
16. M. P. CARPENTER, M. COOPER, F. NARIN, Linkage between basic research literature and patents, *Research Management*, 23 (1980) 30–35.
17. F. NARIN, D. OLIVASTRO, Linkage between technology and science, *Research Policy*, 21 (1992) 237–249.
18. F. NARIN, K. S. HAMILTON, D. OLIVASTRO, The increasing linkage between U.S. technology and public science, *Research Policy*, 26 (1997) 317–330.
19. D. HICKS, T. BREITZMAN, D. OLIVASTRO, K. HAMILTON, The changing composition of innovative activity in the U.S.: A portrait based on patent analysis, *Research Policy*, 30 (2001) 681–703.
20. M. MEYER, Patent citation analysis in a novel field of technology, *Scientometrics*, 51 (2001) 163–183.
21. A. AL-THUBAITY, K. AHMAD, Tracking knowledge of emergent domains, presented at *IEEE International Conference on Information Visualisation*, London, 2002.
22. D. J. WATTS, S. J. STROGATZ, Collective dynamics of 'small-world' networks, *Nature*, 393 (1998) 440–442.
23. A.-L. BARABÁSI, R. ALBERT, H. JEONG, Scale-free characteristics of random networks: The topology of the world-wide web, *Physica A*, 281 (2000) 69–77.
24. A. L. BARABÁSI, H. JEONG, Z. NÉDA, E. RAVASZ, A. SCHUBERT, T. VICSEK, Evolution of the social network of scientific collaborations, *Physica A*, 311 (2002) 590–614.
25. R. ALBERT, A. BARABASI, Statistical mechanics of complex networks, *Reviews of Modern Physics*, 74 (2002) 47–97.
26. S. N. DOROGOVTSSEV, J. F. F. MENDES, Evolution of networks, *Advances in Physics*, 51 (2002) 1079–1187.
27. C. G. CALLAN, L. THORLACIUS, World-sheet dynamics of string junctions, *Nuclear Physics B*, 534 (1998) 121–136.
28. D. D. PRICE, Networks of scientific papers, *Science*, 149 (1965) 510–515.

29. H. G. SMALL, B. C. GRIFFITH, The structure of scientific literatures. I: Identifying and graphing specialties, *Science Studies*, 4 (1974) 17–40.
30. H. G. SMALL, A co-citation model of a scientific specialty: A longitudinal study of collagen research, *Social Studies of Science*, 7 (1977) 139–166.
31. N. P. HUMMON, P. DOREIAN, Connectivity in a citation network: The development of DNA theory, *Social Networks*, 11 (1989) 39–63.
32. R. SKALAK, C. F. FOX, Tissue Engineering, In: *A Workshop on Tissue Engineering*. Granlibakken, Lake Tahoe, CA: New York: Alan Liss, 1988.
33. R. LANGER, J. P. VANCANTI, Tissue engineering, *Science*, 260 (1993) 920–926.
34. L. V. MCINTIRE, H. P. GREISLER, L. GRIFFITH, P. C. JOHNSON, D. J. MOONEY, M. MRKSICH, N. L. PARENTEAU, D. SMITH, *Tissue engineering research: A WTEC Panel Report*, International Technology Research Institute, Baltimore, Maryland, World Technology (WTEC) Panel Report January 2002, 2002.