Scientometrics (2011) 88:761–770
DOI 10.1007/s11192-011-0414-x

Patent co-citation networks of Fortune 500 companies

Xianwen Wang · Xi Zhang · Shenmeng Xu

Received: 29 November 2010 / Published online: 25 May 2011

Abstract This paper provides an overview of the progression of technology structure based on patent co-citation networks. Methods of patent bibliometrics, social network analysis and information visualization are employed to analyze patents of *Fortune 500* companies indexed in Derwent Innovations Index, the largest patent database in the world. Based on the co-citation networks, several main technology groups are identified, including Chemicals, Petroleum Refining, Motor Vehicles, Pharmaceuticals, Electronics, etc. Relationships among the leading companies and technology groups are also revealed.

Keywords Fortune 500 · Patent bibliometrics · Patent co-citation · Technology structure

Introduction

Fortune 500 are those leading companies in their industries according to the sales revenue ranked by *Fortune* magazine annually. These leading enterprises are intensive in patents, owing to the great emphasis they put on technological innovation for retaining the competitive advantages in corresponding industries. Because of the representativeness for the whole technology frontier around the world, we have great interest in conducting our analysis on patents of these companies.

Patents are becoming increasingly important to commercial organizations, especially for multinational companies. According to the report of World Intellectual Property Organization, nearly 90–95% of the world's R&D outcomes are covered in patent publications, only the rest 5–10% are included in the scientific literatures (papers and monographs) (Liu and Yang 2008).

Patents, which contain a great amount of knowledge on technical innovations, provide a valuable source of information on technology development and innovative activities (Li et al. 2009). It is crucial to analyze patent information to understand industrial trends and set the future developing directions.

X. Wang (✉) · X. Zhang · S. Xu
WISE Lab, Dalian University of Technology, Dalian, China
e-mail: xianwenwang@dlut.edu.cn

Literature review

Patent bibiliometrics

The idea of using bibliometrics methods for analyzing patent information could be dated back to 1940s, when Arthur H. Seidel proposed a citation index of the patent literatures in the Journal of the Patent Office Society and Harry C. Hart endorsed the idea in a later issue (Garfield and Merton 1979). Although this idea was brought up at that time, neither patent citation nor other bibliometrics methods were broadly applied to patent literature analysis until the last decade in 20th century. With his pioneering work on patent citations, Narin opened up a new field of application for bibliometrics methods to patents. He extended the existing interpretative framework for citations in research papers to the field of patent citations successfully (Narin 1994).

Most of the previous works using the patent bibliometrics methods mainly deal with three issues, the productivity, impact and correlation. Patent counting is the most common method used to measure the productivity. Patented innovations which are particularly novel will be the subject of greater citations. For this reason, the number of citations received by a patent is used in the literature as a measure of the innovative output embodied in technology. By counting the number of patents granted each year, the growth of the research productivity could be drawn. The method was applied to analyzing the productivity on countries, assignees, inventors (Narin 1994, 1995; Narin et al. 1994; Karki 1997) and technology levels (Ramani and De Looze 2002; Lopez-Munoz et al. 2003). As for the studies of research impact, citation count is used as an indicator to present the level of impact (Moed 2000; Albert and Plaza 2004). Except for the cited patent count, counting numbers of shared patent citations, as well as patent coupling, were applied to establishing the technological relationships among countries, assignees, inventor and techniques (Lo 2007, 2008).

Patent citation analysis

Generally speaking, there are three types of citations (Small 1997): direct citation, and the other two indirect citation, namely co-citation (Small 1973), and bibliographic coupling (Kessler 1963).

Citation analysis has its origins within bibliometrics, i.e., the study of citation behaviors on behalf of scientific authors or academic journals. A considerable amount of research has been done to identify the similarities and differences between scientific literature and patent (Meyer 2000). A patent citation indicates that an innovation may be partly based on an earlier patented one. All “prior art” related to the patented invention are required to be disclosed. Patent citations allow the analyst to assess the quality and impact of cited material, as well as the linkages between cited and citing countries, between cited and citing companies, and between cited and citing scientific and technological areas (Narin and Olivastro 1988).

Patent co-citation analysis

Co-citation analysis, firstly introduced by Herry Small in 1973, is the most influential technique in bibliometrics, which is used widely for macro S&T policy planning and evaluation.

Limited research has been done about patent co-citation analysis. Mogee and Kolar used patent co-citation clustering to identify the major technology fronts of Eli Lilly and Co. (Mogee 1991). Patent co-citation was also employed to identify the core patents and evaluate the similarities between the core patents, and to further investigate core technology capabilities through the co-citation structures in the patents (Wu and Chen 2008). Verspagen (2007) proposed the frameworks for analyzing technology transition based on an enhanced clustering method to indentify the citation route combing coupling and co-citation analysis. Kuei-Kuei Lai developed a patent classification system based on patent similarities assessed by the frequency of the patent pairs (Lai and Wu 2005). Though application of patent co-citation has proliferated in the filed of recognizing the core technology and technology similarities in recent years, effort in analyzing enterprises' structures of technology competition, especially based on patent co-citation analysis, are limited.

Compared to co-citation in scientific literatures, only a few studies on patent co-citation is found. Most patent-related studies focus on statistical analysis on the counts, issued time, IPC Classification, time of cited patent et al. Patent citation and co-citation approaches were rarely applied in the establishment of technological relationships among countries, assignees, inventor and techniques.

Among all the three types of citation network analysis, co-citation is "*particularly well suited to obtaining a rapid and broad overview of the research areas represented in a large set of highly cited documents*" (Small and Upham 2009). In this research, we are going to get a broad overview of the technology landscape in a very large set of highly cited patents, so, we adopt co-citation analysis to analyze the network.

Research gaps and research questions

As an important R&D evaluation issue, patent analysis has been conducted by numerous researchers. During the past decade, the theory of patent citation has developed a lot, of which few studies on patent co-citation application is carried out. And, most previous studies focus on a certain company or technical field. Not any previous studies about analyzing the patent co-citation relationship of cross-industry are found.

We are interested in methods that will capture and map the structure of patent co-citation networks for technology development. In this research, we mainly focus on the following research questions:

What is the macro-level structure of technology? What is the evolution process for the technology landscape during the last 3 decades? Which company is the pivotal point in the network? What is the emergent technology field in the recent decade?

Data and methodology

The way a patent being cited was considered as an indicator of one patent's impact over others. The patents with a higher cited frequency are generally regarded to have higher R&D influence. The same conception was applied to the analysis on patent assignees (most patent assignees are companies). The highly cited companies were assumed to have greater R&D influence on the development of essential technology than others. Besides, co-citation is used to reveal the correlation among the core assignees. We assume that the assignees citing the same patents are more technological related than the assignees without co-citation relation. In other words, higher co-cited frequency presents a higher correlation of the assignees.

In this study, we conduct a patent search on the database of the Derwent Innovations Index (DII), the largest patent database in the world. This database contains almost 20 million patent records that were issued from 1963 to now, with over 30,000 patent records added in weekly. In DII, information can be found not only for patent according to patent code, assignee, Assignee Code and inventor, but also for patent citation according to Cited Patent Number, Cited Assignee Name, Cited Assignee Code.

We use Assignee Code, a unique four-digit code for each top patenting company given by DII, to select the basic patents. For example, *SIEI* is an assignee code of *Siemens*. Considering the strategy adjustment during the company's development, mergers (e.g. *Exxon Mobil*, *Royal Dutch Shell*, *BHP Billiton*, *Bristol Myers Squibb*) or joint ventures (e.g. *Bosch Siemens Home Appliances Corporation*, *Fujitsu Power Equipment Company*) may occur among companies.

In order to overcome the traditional limitations due to the incompleteness and unreliability of a dataset, we introduce a proper search strategy to process these source data, taking merger and joint venture into account. Figure 1 gives an example for *Siemens*. If we search *Siemens* by the assignee name, only those patents granted to those whose companies' names begin with "Siemens" are listed. In this case, some joint ventures also belonging to *Siemens* without a "Siemens" beginning name are left out, e.g., names of *Siemens* and *Bosch*, *Voith* begin with *Bosch* and *Voith* respectively.

Finally, 5 assignee codes are obtained by searching *Siemens*, which are *BOSC*, *BSHB*, *SIEI*, *VOIJ* and *SITS*. However, not all codes belong to *Siemens*, e.g. *BOSC* covers 44 assignee names, of which only two, *Bosch Siemens AG* and *Bosch Siemens GMBH*, are joint ventures of *Siemens* and *Bosch*. Thus, based on all assignee codes related to *Siemens*, we make "*CN* = (*siemens or bosch siemens or bsh bosch & siemens or voith siemens or sits soc tal telecom siemen*s) *or CC* = *siei*" as the final search query. In this study, we make unique search strategies for each company according to their peculiar process of development.

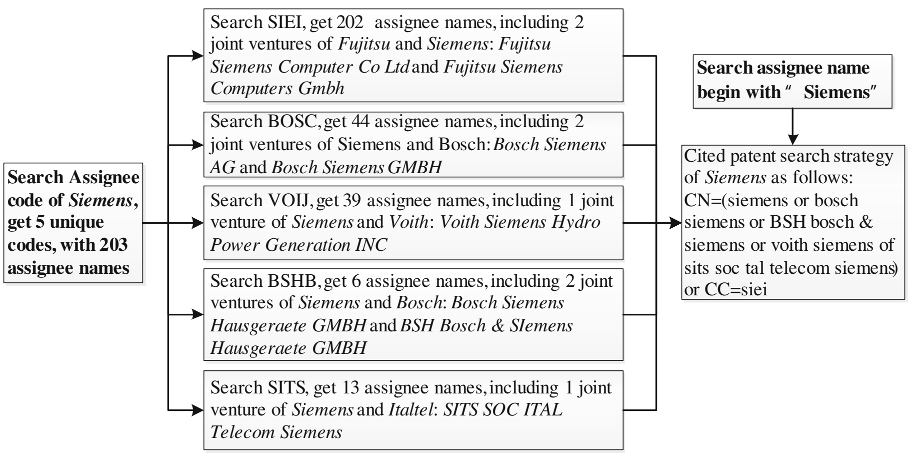

Fig. 1 Patent searching strategy of Siemens in Derwent innovations index

Given the search queries made from the previous step, we conduct patent citation search in DII database for the 500 companies. And we find that for the companies in some industries, for example, insurance companies, food companies and energy companies, they have very few patents published and receive few citations each year. Thus, in this research, we only investigate those companies with high patenting and high R&D influences. From the citation perspective, we select the companies with over 100 patent citations as our research objects. Finally, 217 leading companies are picked out.

Then, we conduct patent co-citation queries one by one for these 217 companies in DII, and a 217 $\times$ 217 co-citation matrix, including 47,089 cells, is constructed by calculating the co-cited frequency of each assignee pair.

Similarly to document co-citation strength in bibliolimetrics (Small 1973), we calculate the Jaccard coefficient matrix of patent assignees.

$$S_J(i,j) = \frac{\mathrm{coc}(i,j)}{\mathrm{cit}(i) + \mathrm{cit}(j) - \mathrm{coc}(i,j)} \tag{1}$$

when S_J (i,j) denotes the jaccard coefficient of patents of company i and company j, coc(i,j) denotes the co-citation times of company i and company j, and cit(i) and cit(j) is the cited times of patents of i and j.

The graphs are laid out with the Spring-embedding algorithm (Kamada and Kawai 1989) in NetDraw (Borgatti 2002). This layout can get maximum spaces for the nodes and avoid edges overlapping (Leydesdorff 2007).

Results

In a network, different threshold will get different visualization results. If we set the threshold too low, too many nodes and lines would make the network very dense, and the network structure, including nodes, clusters, and their relations, would be very difficult for us to identify. As the threshold becomes larger, fewer nodes and lines get into the network, which would make it more distinct. However, too few nodes and lines may also make the network too sparse, and make it meaningless in the opposite way. In this study, besides giving considerations to the validity of the network, we select a fixed threshold of 0.03 in order to compare three evolution progresses in the same level.

For the patents co-cited annually in the 31 industries, we divide 30 years into three equal time spans: 1980–1989, 1990–1999, 2000–2009, in order to compare the decade-by-decade progression of the industry fields. In Figs. 2, 3 and 4, the co-citation relations among the 217 companies in each period are illustrated, and the three co-citation networks together draw us a dynamic changes over 30 years.

Figure 2 is the visualization result of the patent co-citation network during the first decade, 1980–1989. Each node represents a company, while each line for a co-citation relation. The size of a node corresponds to its betweenness centrality in the network.

We set a threshold of 0.03, as a result, 198 nodes and 412 lines emerge in the network in Fig. 2.

As Fig. 2 shows, the whole network consists of several main groups:

(1) Motor vehicles group. This group consists of *Ford Motor*, *Nissan Motor*, *Honda Motor*, *Bosch*, *GM* and so on. With the highest betweenness centrality, *Bosch* connects electronics group and motor vehicles group, since *Bosch* is involved not only in motor vehicle equipment, but also in consumer goods and building technology.

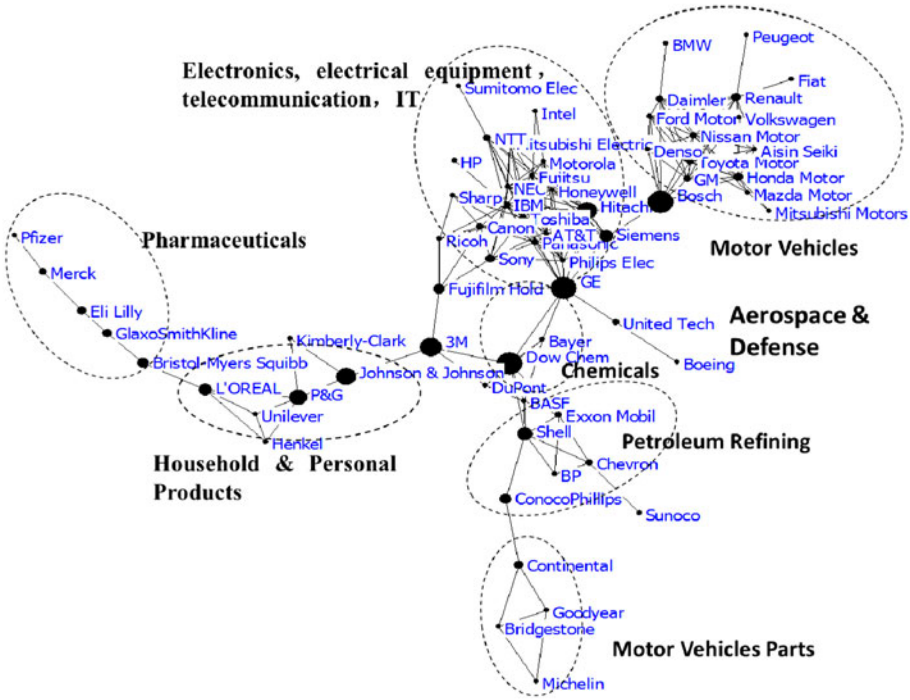

Fig. 2 Patent co-citation network of Fortune 500 (1980–1989, cutoff is 0.03)

(2) Electronics, electrical equipment, telecommunication, IT group. There is no exact boundary between electronics, electrical equipment industries (such as *Mitsubishi Electrics, Siemens*) and IT (*IBM, Hp, Intel, Microsoft*). *GE* is a pivotal node to link this group with chemical group and aerospace and defense group.

(3) Chemicals group. Companies in this group are *Dow Chem, Bayer, BASF*, etc.

(4) Household and personal products group. It is very close to chemicals group. This group includes *P&G, Johnson& Johnson, and Unilever*, etc.

(5) Pharmaceuticals group. Companies in this group are Pfizer, Merck, Eli Lilly, GlaxoSmithKline, Bristol-Myers Squibb, etc.

(6) Petroleum refining group, which is also extended from the chemicals group. *Shell* and *Exxon* mobile are pivotal nodes to connect this group to Chemicals group.

During the period of 1980–1989, the network appears to be rather dispersed. Technology communication among industries is infrequent. However, we can still see that the connections among Pharmaceuticals, Chemicals and Petroleum Refining are very closely, and there are also close connections between Motor vehicles group and Electronics and Electrical equipment group.

As Fig. 3 shows, during the period of 1990–1999, as more companies emerge in the co-citation network, several changes enrich the whole network.

(1) The group of electronics, electrical equipment, telecommunication, IT becomes much larger.

(2) Pharmaceuticals group and chemicals group establish direct connections.

(3) The amount of chemicals companies also increases a lot.

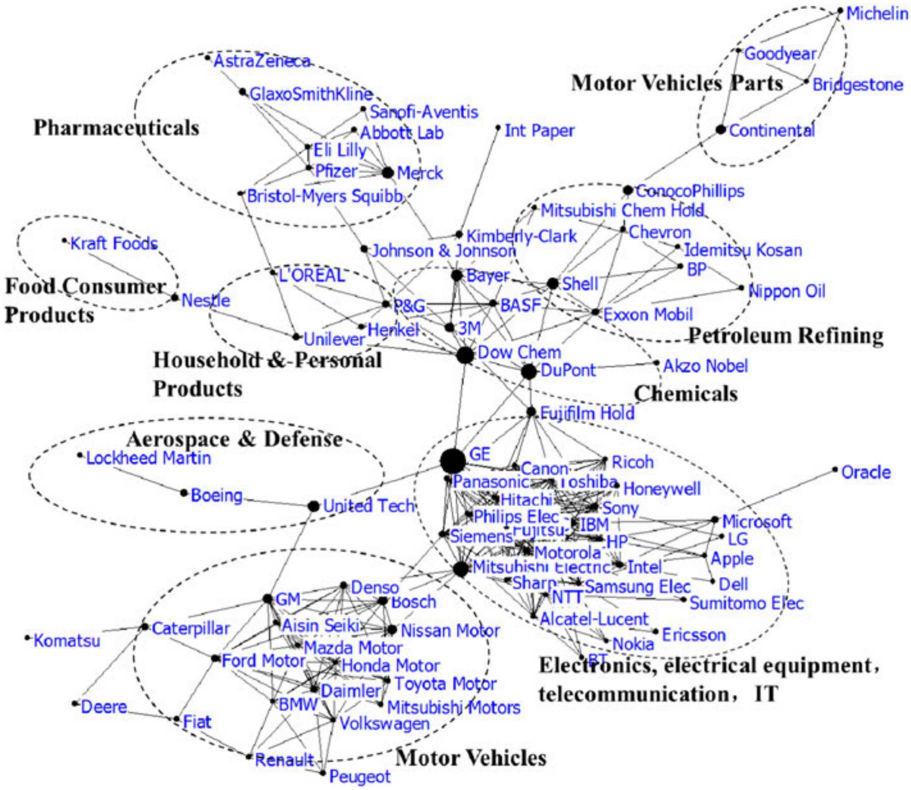

Fig. 3 Patent co-citation network of Fortune 500 (1990–1999, cutoff is 0.03)

It is obviously that chemicals industry plays a more important and fundamental role in pharmaceuticals industry, petroleum refining industry, and household and personal products industry. As a result of the technical requirements and consumer demands, products in these industries becomes more reliable on chemical techniques. With the improvement of biotechnology and chemical techniques, pharmaceutical research becomes more effective, and the relation between biotechnology and chemical pharmaceutical industry is also strengthened.

For the network of 2000–2009 with the same threshold of 0.03, more nodes and links emerge in the network.

The left part is a group which comes into our sight during the latest decade for the first time, including financial enterprises such as *J.P. Morgan Chase & Co., Morgan Stanley* and insurance enterprises such as *Nationwide*, being linked by Oracle and Amazon.com with others. The emergence of these companies may be attributed to increasing technology content of some financial and insurance products, e.g., some patents involving the key ring and account security.

The middle part of the network is rather dense, which contains industries of motor vehicles and electronics, electrical equip, information technology.

For the right part, there is no distinct boundaries among the pharmaceuticals, chemicals and petroleum refining industries.

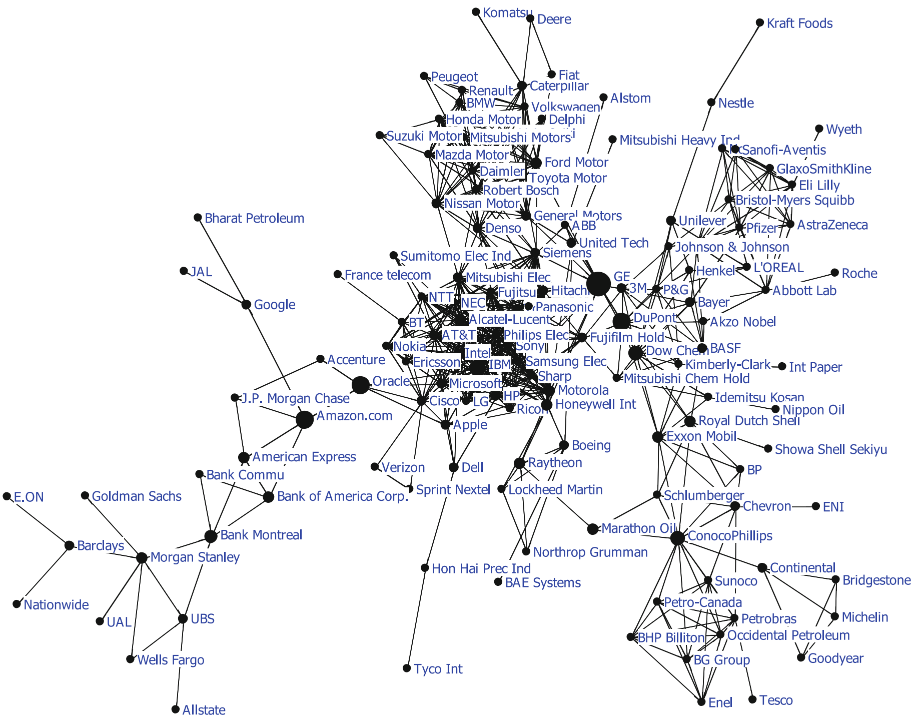

Fig. 4 Patent co-citation network of Fortune 500 (2000–2009, cutoff is 0.03)

The newest phase of the 21st century sees the growing integration of technology in different industries. For example, compared with the past two decades, electronics has become the most fundamental technique in Motor Vehicles. Statistics show the proportion of electronic products used in motor vehicles has been increasing steeply in recent years. Electronic components currently comprise some 20–30% of total costs for all car categories, and this figure is expected to reach 40% or so by 2015 (http://e2af.com/trend/071210.shtml). At present, automobile electronics are used in almost every system of the vehicle production. So, innovations from electronics are playing a more and more important role in Motor Vehicles industry. Applications of automobile electronics are shown in Table 1 below.

Conclusion

This paper has presented a theoretical and experimental study of patent co-citation analysis. In this research, we propose a way to construct large patent co-citation network, and apply it to analyzing the patents of *Fortune 500* companies. Three technology landscapes are visualized decade by decade from 1980 to 2009. Based on the patent co-citation analysis and information visualization, several main technology groups are identified, including chemicals, petroleum refining, motor vehicles, pharmaceuticals, electronics, etc. The relationships among the leading companies and technology groups are also revealed.

Table 1 Applications of automobile electronics

Vehicle functions	Electronic techniques
Dynamic	1. Electronic fuel injection system
	2. Electronic ignition systems
	3. Knock control system
	4. Speed control system
Safety	1. Anti-lock braking system
	2. Acceleration skid control system
	3. Traction control system
	4. Electric brakeforce distribution
Dirigibility	1. Active body control
	2. Crusle control system
	3. Dynamic stability control
	4. Vehicle stability control
Information control	1. Controller regional network communication system
	2. Multi-media directional system transmission network

Based on the comparing of the co-citation networks formed in three periods, the evolution of technology communication structure is studied. The results show that technology connections among industries are becoming closer, and boundaries among different technology groups become blurred. For some industries, e.g., electronics and motor vehicle, pharmaceuticals and chemicals, the merging trend is rather obvious.

Acknowledgments The research was supported by the Social Science Foundation of China (Grant No. 10CZX011, Grant No. 08BTQ025), the project of "Specialized Research Fund for the Doctoral Program of Higher Education of China" (Grant No. 2009041110001), as well as the project of "Fundamental Research Funds for the Central Universities" (Grant No. 2009-852009).

References

Albert, A., & Plaza, L. M. (2004). The transfer of knowledge from the Spanish public R&D system to the productive sectors in the field of biotechnology. *Scientometrics, 59*(1), 3–14.

Borgatti, S. (2002). *NetDraw: Graph visualization software*. Harvard: Analytic Technologies.

Garfield, E., & Merton, R. (1979). *Citation indexing: Its theory and application in science, technology, and humanities*. New York: Wiley.

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information processing letters, 31*(12), 7–15.

Karki, M. (1997). Patent citation analysis: A policy analysis tool. *World Patent Information, 19*(4), 269–272.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation, 14*(1), 10–25.

Lai, K. K., & Wu, S. J. (2005). Using the patent co-citation approach to establish a new patent classification system. *Information Processing & Management, 41*(2), 313–330.

Leydesdorff, L. (2007). Mapping interdisciplinarity at the interfaces between the Science Citation Index and the Social Science Citation Index. *Scientometrics, 71*(3), 391–405.

Li, X., Chen, H., et al. (2009). Managing knowledge in light of its evolution process: An empirical study on citation network-based patent classification. *Journal of Management Information Systems, 26*(1), 129–154.

Liu, C., & Yang, J. (2008). Decoding patent information using patent maps. *Data Science Journal, 7*, 14–22.

Lo, S. (2007). Patent analysis of genetic engineering research in Japan, Korea and Taiwan. *Scientometrics, 70*(1), 183–200.

Lo, S. C. (2008). Patent coupling analysis of primary organizations in genetic engineering research. *Scientometrics, 74*(1), 143–151.

Lopez-Munoz, F., Alamo, C., et al. (2003). Bibliometric analysis of biomedical publications on SSRI during 1980–2000. *Depression and Anxiety, 18*(2), 95–103.

Meyer, M. (2000). What is special about patent citations? Differences between scientific and patent citations. *Scientometrics, 49*(1), 93–123.

Moed, H. (2000). Bibliometric indicators reflect publication and management strategies. *Scientometrics, 47*(2), 323–346.

Mogee, M. E. (1991). Patent co-citation analysis of Eli Lilly & Co. patents. *Research-Technology Management, 34*(4), 43–49.

Narin, F., & D. Olivastro (1988). Technology indicators based on patents and patent citations. Handbook of quantitative studies of science and technology, 465–507.

Narin, F. (1994). Patent bibliometrics. *Scientometrics, 30*(1), 147–155.

Narin, F. (1995). Patents as indicators for the evaluation of industrial research output. *Scientometrics, 34*(3), 489–496.

Narin, F., Olivastro, D., et al. (1994). Bibliometrics: Theory, practice and problems. *Evaluation Review, 18*(1), 65–76.

Ramani, S., & De Looze, M. (2002). Using patent statistics as knowledge base indicators in the biotechnology sectors: An application to France, Germany and the UK. *Scientometrics, 54*(3), 319–346.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265–269.

Small, H. (1997). Update on science mapping: Creating large document spaces. *Scientometrics, 38*(2), 275–293.

Small, H., & Upham, P. (2009). Citation structure of an emerging research area on the verge of application. *Scientometrics, 79*(2), 365–375.

Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems, 10*(1), 93–115.

Wu, S., & Chen, H. (2008). *Recognizing the core technology capabilities for companies through patent co-citations*. Singapore, IEEE: IEEE International Conference on Industrial Engineering and Engineering Management.