Evaluation of Technological Influence Power of Enterprises through the Enterprise Citation Network

Shu Wang, Jiangning Wu and Donghua Pan
Institute of Systems Engineering, Dalian University of Technology, Dalian, 116024, PR China
shuwang@mail.dlut.eud.cn, jnwu@dlut.edu.cn, gyise@dlut.edu.cn

Abstract: In this paper, we establish a weighted directed enterprise network based on the enterprise citation relations, and accordingly provide a novel algorithm to evaluate the technological influence power of enterprises by means of the network structure. The proposed evaluation method not only adopts the traditional centrality assessment parameter like the in-degree but also takes more information from the enterprise citation network into account. Experiments with such method have been done by using the patent dataset in the field of chemical fertilizer during the 25-year period from 1975 to1999 from United States Patent and Trademark Office. The evaluation results on the technological influence power of the selected corporations have been found to be more reasonable than the traditional evaluation methods have. This study is helpful for the enterprise to identify its potential technological competitors and fix on its own technological niche in the whole industry.
Key words: Technological Influence Power; Patent Analysis; Enterprise Citation Network; Evaluation

1 Introduction

In the era of globalization and knowledge-based economy, the competition between enterprises has shifted from product manufacturing to innovative creation. Technological innovation strategy has been thought to be the key to the development of enterprises. Patents, as the core of the product and process of technological innovation, contain more than 90% of the latest technical information in the world (Robert, 2006). With the fairly large scale, extensive new contents, complete descriptions and the precise classification, patents become the unique source for discovering new technological information and competitive intelligence (Chakrabarti, 1991). When making the patent strategic, the primary task is to discover the technologically influential enterprises in the industry, which is an important support for enterprises to identify their actual and potential competitors and make a thorough and effective innovation strategic deployment.

In the study, we utilize the enterprise network based on the patent citation information to evaluate the technological influence power of enterprises. The idea comes from the fact that the dominant status in interaction with other enterprises is a key factor in evaluation of the enterprises' technological influence power. Nowadays, the enterprises have attached more and more importance to make connections with other corporations in order to enlarge their market share (Margherita, 2004， Gay, 2005). The enterprise that dominates and controls many other enterprises in cooperation and negotiation occupies an important strategic status in the market and has a large technological influence power. This inspires us to establish an enterprise network based on the influence and dependence relation between enterprises to discover the technological influential enterprises

Another motivation for this study comes from the widely used citation information of patents where more technical information and influence relations between patents, inventors, and enterprises are involved. The behavior of citation can be viewed as the recommendation or endorsement of the later inventors (Organization for Economic Co-operation and Development, 1994). An important patent will be cited repetitively by a large number of improved patents over a long period of time. The number of citations can reflect the innovation level of the patent and its contribution to the progress of the industry. If an enterprise has many important patents which are cited by many other corporations, it will have a great influence power in negotiation with these corporations and have a great competitive advantage in the market competition (Karki, 1997). Therefore, an enterprise network by means of the patent citation information is developed in the paper to assess the technological influence power of enterprises in the whole industry.

In the enterprise citation network, both weight and direction information of links are introduced, which correspond to the strength of citation relations and the citation direction from one enterprise to another one, respectively. Hence a weighted and directed enterprise citation network can be obtained. Based on this network, a new algorithm is put forward to evaluate the technological influence power of enterprises by giving a comprehensive consideration of the number of citing enterprises, the importance of the citing enterprises and the citation strength between enterprises. The experimental results have verified the effectiveness of the proposed evaluation method.

2 Related works

With the severe competition in the global market and the complex process in product innovation, enterprises seek to the opportunity for cooperation and engage in co-investment or joint R&D frequently. The power of the enterprise to influence other enterprises in cooperation and negotiation has received much attention by plenty of scholars. Kinds of related models and indicators for evaluating the technological influence power of enterprises have been proposed by analyzing the external characteristics in patent literature (Bart, 2003, Narin, 1987, Schmoch, 1999). However, these indicators are provided only on the basis of quantitative information of patents, which is not objective to some extent due to the different patent strategies of different enterprises.

Whereas the citation relation between patents contains more meaningful information on the mutual influence or dependence between enterprises which is more objective and accurate than the quantitative information of patents. As the citing enterprise makes an attempt to get authorization from the cited enterprise to use and transfer the required technology in its patents, the cited enterprise can take the initiative in the cross-licensing negotiations with the citing enterprise (Anthony, 2002, Gavin, 2005). Therefore, more and more research has been concerned with the use of the patent citation information to evaluate the technological influence of enterprises (Yuya, 2009, Christian, 2008, Xin, 2007, Bernard, 2009). According to the structure of enterprise networks established by means of the citation information, in which nodes represent enterprises and directed links represent citation relations between enterprises, kinds of centrality assessment indicators in social network analysis are applied to evaluate the technological influence power of enterprises. And the most well-known indicator in this field is in-degree, which is defined as the number of links that have the node as an end, namely the numbers of citing enterprises (Xin, 2007, Bernard, 2009).

Although, the in-degree counting is a natural measure of an enterprise's technological influence power, it could only identify the technologically competent enterprises in a limited area. For instance, the citations from authoritative enterprise or from ordinary enterprise are treated equally, and the citing strength of different enterprises is ignored.

In this case, based on a different rank mechanism concerning deeper relations among enterprises, we are motivated to study an improved index of in-degree by distinguishing the different contribution of citation from different enterprises and taking the citation strength into consideration, in order to shed light on evaluating the enterprises' technology influence power objectively and identifying the potential important competitors from another perspective.

3 Evaluation Model and Method

Our evaluation method includes two main parts: enterprise citation network creation and an evaluation algorithm.

3.1 Enterprise citation network creation

First of all, we establish a weighted directed enterprise network based on the citation relations of patents. Given a network $G = (V, E, W)$, where $V = \{ v_1, v_2, \ldots\ldots, v_n \}$ represents the set of enterprises, n represents the number of enterprises in set V, directed link $E = \{ (v_i, v_j) \}$, $1 \leq i,j \leq n$, $i \neq j$, represents a citation link from a citing enterprise to a cited enterprise, *weight of link* $W = \{ w_{ij} \mid (v_i, v_j) \in V \times V \}$ represents the citation frequency that measures the numbers of citation times between enterprises. As we focus on the citation relation between enterprises, we ignore the self-citation situation when the patents of one enterprise cite with each other.

3.2 Technological influence power evaluation

Our algorithm to evaluate the enterprise's technological influence power is based on the assumptions below.

(1) The enterprise's technological influence power is proportional to the number of the citing enterprises. This assumption is consistent with the idea of in-degree indicator.

(2) The technological influence power still has to do with the importance of the citing enterprises. Citation from a much more authoritative enterprise will contribute more to the cited enterprise's technological influence power than from less authoritative ones. Due to the strong capability of an influential enterprise to research and develop high-tech products, the enterprise cited by these influential ones must have some unique and advanced technologies to be reckoned with.

(3) The enterprise's technological influence power is also proportional to the citation frequency received from the citing enterprises in their total citation percentage, which indicates the technological dependence between them

Based on an overall consideration of various factors, we can define the technological influence power of an enterprise by the following formula:

$$TI_A = p * \sum_{B \in IN_A} (D_{BA} * TI_B) + (1 - p) / n \qquad (1)$$

$$D_{BA} = \frac{W_{BA}}{\sum_{C \in ON_B} W_{BC}} \qquad (2)$$

Where, the technological influence power is denoted by TI, IN_A denotes the set of all the enterprises that have citing links pointing to A, D_{BA} denotes the citation strength, which is the ratio of the citation frequency between B and A to the sum of the citation frequencies between B and all its neighbors, ON_B denotes the set of all the enterprises that have cited links pointed from B. p is a damping factor which can be set between 0 and 1 and controls the performance of the algorithm. The second term $(1\text{-}p) / n$ is used to make the sum of all the TI values equal to 1 and complement the losing fraction in the first term multiplied with p. It also means that an enterprise cited by no one can also receive a contribution $(1\text{-}p) / n$ at each step.

Above all, the TI value of enterprise A is recursively defined by the TI value of B which has cited A. With the algorithm, the TI value of B is always weighted by the citation strength D_{BA}, leading to a smaller TI value which B transfers to A. Thereby, the TI value of A is the sum of all TI values receiving from all its citing enterprises.

When calculating the TI value, it can be recognized as the calculation of eigenvector of the matrix. And the process can be expressed as the following matrix representation.

$$I = pWI + (1 - p)\frac{e^t}{n} \tag{3}$$

Where I donates the TI vector of enterprises, $I = \{I_1, I_2, \ldots, I_n\}^T$, n donates the number of enterprises and e^t represents an n-vector whose elements all equal to 1. A $n \times n$ matrix transposing from the adjacent matrix of the enterprise citation network is given by W with the element w_{AB} defined as following:

$$w_{AB} = \begin{cases} \dfrac{x_{AB}}{\deg(B)} & if \quad B \quad cites \quad A \\ 0 & if \quad B \quad do \quad not \quad cite \quad A \end{cases} \tag{4}$$

Where x_{AB} is the citation frequency from B to A and $deg(B)$ is the sum of citation frequencies from B to other enterprises. If B cites A x times, then w_{AB} is normalized to $deg(B)$, else w_{AB} is zero. However, if B does not make any citation to any enterprises, in order to guarantee the convergence to a unique vector I, the column B in matrix W is substituted for a n-vector whose elements equal to $1/n$ which has nothing to do with the result.

As we know that $\sum_{A=1}^{n} I_A = eI = 1$, so Eq. (3) can be rewritten as,

$$\begin{aligned} I &= pWI + (1 - p)\frac{e^t}{n} \\ &= pWI + \frac{1-p}{n}e^t * 1 \\ &= pWI + \frac{1-p}{n}e^t * eI \\ &= (pW + \frac{1-p}{n}e^t * e) * I \\ &= UI \end{aligned} \tag{5}$$

Where $U = pW + \frac{1-p}{n} e^t e$.

Therefore, we can see that the final value of vector *I* converges to the principle eigenvector of matrix *U*. The solution of *I* equals to the computation of the principle eigenvector corresponding to the maximum eigenvalue of matrix *U*. As *U* is irreducible and stochastic, the computation can be carried out through an iterative updating process.

Then we can assign the same initial value to each enterprise in the network and calculate the final value of vector *I* with the dynamic equation as following:

$$I(t) = pWI(t-1) + (1-p)\frac{e^t}{n} \tag{6}$$

Eventually, the stationary unique solution is reached which is not relevant to the given initial value.

We can regard the iterative computation as a global vote. Each enterprise votes for other enterprises that are worthy of technological concerning and technological cooperation. It is relatively objective and fair to determine the technological influence power of an enterprise by the number of votes from others.

As mentioned above, if enterprise *B* has cited some technology of enterprise *A* in a large number, then *B* has a pressing need for this technology and *B* will vote *A* for cooperation and negotiation. The more enterprises have cited *A*'s technology, the more votes *A* will receive. In addition, as different enterprises will receive different numbers of votes from others, the amount of votes they can cast are different too. If *B* has received many votes from others, it will vote *A* a lot more. It indicates that the credibility of its recommendation to *A* is increasing due to the powerful influence of *B*. Furthermore, it's possible for *B* to demand technology not only from *A* but also from other enterprises. Thus, the more times *B* has cited *A*, the lager number of votes it will give to A.

At the beginning, we distribute votes to every enterprise equally. After some rounds of voting, there is a change in the distribution of the votes among enterprises. When the distribution becomes stable, the number of votes in hand reflects the enterprise's technological influence power ultimately.

4 Evaluation Results and Discussions

4.1 Data collection

In this section, we set up an enterprise citation network based on the patent data in the field of chemical fertilizer during a 25-year period from 1975 to1999, and then apply our evaluation algorithm with the goal of measuring the technological influence power of the corresponding enterprises.

First of all, the query formulation like ((ISD/19750101->19991231)AND (CCL/71)) is applied in the United States Patent and Trademark Office(USPTO). Then 1763 patents collected are matched up to the patent information in NBER database to extract the bibliographic data and citation data (Hall, 2001). The citation relations which are outside the dataset will be filtered out, and the patents whose patentees are not enterprises will also be excluded. Once the cited–citing relations have been built up, a directed weighted enterprise citation network comes into existence. There are 402 nodes representing enterprises and 767 directed links representing the citations made by all these enterprises. In the established network, the average weight of the directed links is 2.5 and the maximal weight is 64, which represents the citation frequency between enterprises.

4.2 Visualization of the enterprise citation network

To implement the computation, we compile a computer program under the environment of Visual Studio.NET, using Microsoft SQL Server 2000 as the supporting database. During the calculation, we start with a uniform value which equals to 1 and choose the parameter p to be 0.5, and then act 30 times in the database. Eventually, a steady set of *TI* values for all enterprises are reached.

Then, the results are input into the open source graph visualization soft PAJEK (Batagelj, 2007). Hence a global visualization of the enterprise citation network can be obtained as shown in Fig. 1, where the node represents the enterprise, the node size represents its *TI* value, the directed link represents the citation relation from one enterprise to another one, and the width of the directed links represents how many times one enterprise cites another one. It can be found that there are a large number of enterprises with small *TI* values and a small number of enterprises with high technological influence power in the centre of the network. It is necessary to find the fewer but more influential enterprises out in the industry, which may be the competitors with large technical advantage and strength.

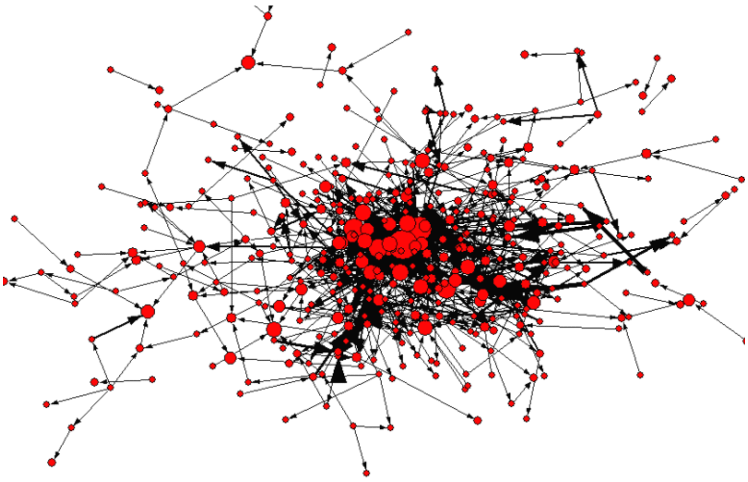

Fig.1.The distribution of enterprise citation network

4.3 Comparisons with the traditional evaluation method

As our method makes an improvement of the in-degree based method, the comparisons with the in-degree counting (the numbers of the citing enterprises) are then analyzed in detail to demonstrate the correctness and validity of our proposed algorithm.

To see how our algorithm performs, we may first take a look at the correlation relationship between the two kinds of algorithms. Firstly, the average *TI* value for each group of enterprises with k in-degree, namely $TI(k)$, is calculated as a function of k. It has been found in Fig. 2 that the plot of $TI(k)$ versus k is smooth and goes up linearly, that is to say $TI(k)$ value and the in-degree value are positively correlated. This is because the computation of *TI* value gives in consideration of the number of the citing enterprises which is the same as in-degree counting. The consistency between *TI* and in-degree demonstrates that our algorithm can be considered as a reliable method which cannot lead to the perverse results.

As two different measures of importance assessment, the distinctions between them are then studied as follow. When we calculate both the *TI* value and the in-degree value for each enterprise, we can find out that as the increasing of *TI* value, some extreme outliers with high *TI* values and relatively low in-degree values come into existence (see Fig. 3). The disparity between the *TI* and in-degree methods lies in that, as mentioned in the previous section, the former involves not only the in-degree but also the *TI* values of the citing enterprises and the citation strength between them. Therefore, our algorithm does well in revealing some exceptional enterprise with a modest value of in-degree but being an influential and potential competitor in the reality.

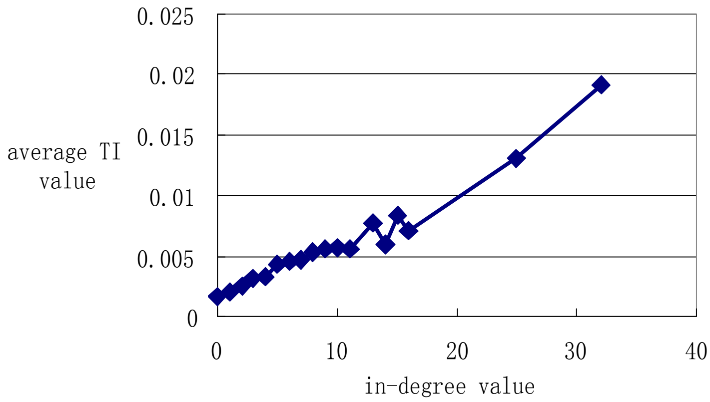

Fig.2. The average *TI* value as a function of the in-degree value

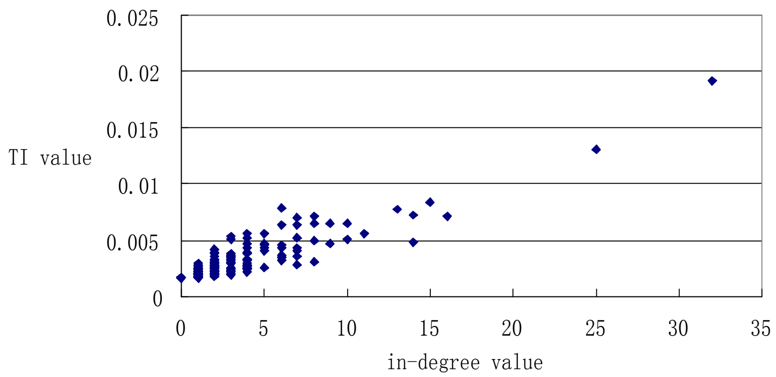

Fig.3. The scatter plot of the *TI* value vs. the in-degree value.

For further analysis, we list the top-20 highest ranking enterprises according to their *TI* values in Table 1, together with their ranks of in-degree values and their brief bibliographic information. We can find out that all the top-20 enterprises have remained in the top-100 in-degree list. Besides, there are 14 enterprises in the top-20 in-degree list and 17 enterprises in the top-50 in-degree list. Obviously, there is a positive relevance between two groups of results obtained from the above two different kinds of algorithms. On the other hand, while many highly-cited enterprises appear in the list, there are also several modestly-cited enterprises that are highly ranked according to our algorithm. In addition to compare the advantage of our algorithm superior to the in-degree counting, let us take the results of two enterprises TEXACO INC and BAYER AKTIENGESELLSCHAFT as an example. The former has a higher rank of *TI* value while the later has a higher rank of in-degree value.

Table 1. The top 20 highest ranking enterprises according to the *TI* value

Rank based on *TI* value	Company name	Rank based on in-degree value	In-degree value	Amount of patents
1	UNION OIL COMPANY OF CALIFORNIA	1	32	44
2	ALLIED-SIGNAL INC.	2	25	36
3	PHILLIPS PETROLEUM COMPANY	4	15	12
4	TEXACO INC.	30	6	5
5	EXXON RESEARCH ENGINEERING CO.	7	13	17
6	BAYER AKTIENGESELLSCHAFT	5	14	13
7	OLIN CORPORATION	14	8	6
8	O. M. SCOTT AND SONS COMPANY	3	16	11
9	ITT CORPORATION	18	7	7
10	KAO SOAP CO., LTD.	11	9	1
11	CONOCO, INC.	10	10	8
12	W. R. GRACE & CO.-CONN.	13	8	14
13	TEXACO TRINIDAD, INC.	19	7	5
14	IMPERIAL CHEMICAL INDUSTRIES PLC	31	6	6
15	JUDD RINGER CORPORATION	60	4	3
16	FMC CORPORATION	38	5	2
17	CHISSOASAHI FERTILIZER CO., LTD.	8	11	3
18	NOVEX RT	87	3	1
19	COMPAGNIE NEERLANDAISE DE L'AZOTE	21	7	5
20	INTERNATIONAL MINERALS AND CHEMICAL CORPORATION	23	7	8

Tables 2 and 3 show the *TI* ranks and the citation strength of the citing enterprises which have cited the above two enterprises. It is showed in Table 2 that patents of TEXACO INC are cited by 6 enterprises, 3 of which rank in the top-20 *TI* list with citation strength more than 50%. On the other side, in Table 3, we can see that although BAYER AKTIENGESELLSCHAFT has 14 citing enterprises, more than TEXACO. There are only 2 enterprises in the top-50 *TI* list and only one in the top-20 *TI* list with citation strength lower than 50%. The empirical observations demonstrate that the citing enterprises of BAYER AKTIENGESELLSCHAFT are less influential than TEXACO and do not rely on the technology of BAYER AKTIENGESELLSCHAFT very much. Thus, for an enterprise with a large *TI* value, its citing enterprises should have a great technological influence power and highly depend on its technology.

Table 2. The *TI* ranks and the citation strength of the citing enterprises of TEXACO INC

Company name	Rank based on	Citation

	TI values	strength
W. R. GRACE & CO.-CONN.	12	0.33
UNION OIL COMPANY OF CALIFORNIA	1	0.13
TEXACO TRINIDAD, INC.	13	0.5
KALO LABORATORIES, INC.	101	1
RECKITT + COLMAN PRODUCTS LIMITED	240	1
NORSK HYDRO A/S	112	0.14

Table 3. The *TI* ranks and the citation strength of the citing enterprises of BAYER AKTIENGESELLSCHAFT

Company name	Rank based on *TI* values	Citation strength
MELAMINE CHEMICALS, INC.	23	0.33
SOCIETE CARBOCHIMIQUE SOCIETE ANONYME	310	0.33
SOILIZER CORPORATION	322	0.5
BIOTHERM INTERNATIONAL, INC.	326	1
LANG & CO., CHEMISCH-TECHNISCHE PRODUKTE KOMMANDITGESELLSCHA	330	0.2
MITSUBISHI KASEI CORP.	234	1
IGENE BIOTECHNOLOGY, INC.	333	0.5
COLLOIDS, INC.	334	0.25
PHOSYN PLC.	338	0.5
EXXON RESEARCH + ENGINEERING CO.	5	0.11
TIOXIDE EUROPE, S.A.	356	0.2
HAIFA CHEMICAL SOUTH LTD.	238	0.05
PLANTAGENET HOLDINGS PTY LTD.	364	0.4
RSA MICROTECH, INCORPORATED	377	0.2

Moreover, NOVEX RT, whose rank is 18 in the top-20 *TI* list, is particularly striking. This enterprise has few citing enterprises and a low in-degree rank but its *TI* rank is relatively high. What accounts for this high *TI* rank? The high *TI* rank of NOVEX RT stems from the average high *TI* contribution from its influential citing enterprises, especially from BAYER AKTIENGESELLSCHAFT, which ranks 6 in the top-20 *TI* list and has the citation strength of 37.5% with NOVEX RT. Furthermore, we find out that it is the unique core technology of reactive tenside soil conditioners that lets patents of NOVEX RT to receive so much concern from other authorities.

Therefore, our *TI* algorithm is helpful to identify the hidden potential enterprises which would probably be ignored due to the information from limited number of citing enterprises.

Conclusions

In the paper, we establish an enterprise citation network based on the patent citation information to show the mutual technological influence between enterprises and provide an improved evaluation method based on the in-degree, which is a common centrality assessment indicator in social network analysis, to evaluate the enterprises' technological influence power. Our algorithm takes the number of the citing enterprises into consideration as well as the influence of the citing enterprises and the citation strength between them. In order to investigate the correctness and validity of our own algorithm, we compare our method with in-degree counting by using the patent data in the field of chemical fertilizer during a 25-year period from 1975 to 1999. It indicates that our algorithm is a better measure for the enterprise to discover the potential competitive enterprises which take the lead in the industry.

Acknowledgements

This work is sponsored by the National Science Foundation of China (No. 70771019), and supported by the National High Technology Research and Development Program of China (No.2008AA04Z107).

References

Anthony F. Breitzman, Mary Ellen Mogee (2002). The many applications of patent analysis, Journal of Information Science 28(3), 187-205.

Bart Van Looy, Julie Callaert (2003). Patent Related Indicators for Assessing Knowledge-Generating Institutions:Towards a Contextualised Approach, Journal of Technology Transfer, 28, 53-61.

Batagelj, V, Mrvar, A(2007). Pajek - program for large network analysis, http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

Bernard Gress (2009). Properties of the USPTO patent citation network: 1963–2002, World Patent Information, 32(1), 3-21.

Chakrabarti,A.K (1991). Competition in High Technology： Analysis of Problems of US, Japan, UK, France, West Germany and Canada, Transactions on Engineering Management, 38, 78-84.

Christian Sternitzke, Adam Bartkowski (2008). Visualizing patent statistics by means of social network analysis tools,World Patent Information 30, 115-131.

Gavin Clarkson(2005). Patent Informatics for Patent Thicket Detection: A Network Analytic Approach for Measuring the Density of Patent Space, Academy of Management, August 2005.

Gay B, Dousset B (2005). Innovation and network structural dynamics study of the alliance network of a major sector of the biotechnology industry, Research Policy, 34(2), 1457-1475.

Hall,B.,Jaffe,A.,and Trajtenberg,M (2001). The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools," Manuscript, September.

Margherita Balconi, Stefano Breschi (2004). Networks of inventors and the role of academia: an exploration of Italian patent data, Research Policy 33, 127-145.

M. M. S. Karki(1997). Patent citation analysis: A policy analysis tool, World Patent Information19(4), 269-272.

Narin, F, Noma E, Perry R(1987). Patents as Indictors of Corporate Technological Strength, Research Policy, 16, 143-155.

Organization for Economic Co-operation and Development (1994). The measurement of scientific and technological activities using patent data as science and technology indicators patent manual, Paris.

Robert M, Sherwood (2006). Intellectual Property and Economic Development, London.

Schmoch U (1999). Impact of international patent applications on patent indicators, Research Evaluation, 8(13), 119-131.

Xin Li, Hsinchun Chen (2007). Patent citation network in nanotechnology (1976–2004), Journal of Nanoparticle Research, 9, 337-352.

Yuya Kajikawa, Yoshiyuki Takeda (2009). Citation network analysis of organic LEDs, Technological Forecasting & Social Change 76, 1115-1123.