International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:2, 2013

Identifying Potential Partnership for Open Innovation by using Bibliographic Coupling and Keyword Vector Mapping

Inchae Park, and Byungun Yoon

***Abstract*—**As open innovation has received increasingly attention in the management of innovation, the importance of identifying potential partnership is increasing. This paper suggests a methodology to identify the interested parties as one of Innovation intermediaries to enable open innovation with patent network. To implement the methodology, multi-stage patent citation analysis such as bibliographic coupling and information visualization method such as keyword vector mapping are utilized. This paper has contribution in that it can present meaningful collaboration keywords to identified potential partners in network since not only citation information but also patent textual information is used.

***Keywords*—**Open innovation, partner selection, bibliographic coupling,Keyword vector mapping, patent network.

I. Introduction

OPEN innovation has been proposed as a new paradigm for the management of innovation [1], [2]. It is defined as 'the use of purposive inflows and outflows of knowledge to accelerate internal innovation and to expand the markets for external use of innovation, respectively' [3]. It thus comprises both outside-in and inside-out movements of technologies and ideas, also referred to as 'technology acquisition' and 'technology exploitation'[4]. While Studies about the concept and application of open innovation mostly have been conducted in previous research [5], the research on matching firms for open innovation does not exist.

To enable open innovation, innovation intermediaries such as online technology transfer exchanges, which list technologiesavailable for licensing to interested parties, or technology brokers, who solicit new innovations by posting problems requiring asolution, have emerged to address specific technology needs. However, these intermediaries typically concentrate only on afraction of the technology intelligence relevant to an organization's strategy [6].Technology intelligence has been defined as 'the capture and delivery of technological information as part of the processwhereby an organization develops an awareness of technological threats and opportunities' [7].

Patent is one of the widely used sources which have technology information. Patent citations have been used extensively in measuring the impact of a patent. The more times that a patent is cited, the more impact it has on other patents [8]. The impact of a patent is also referred as its quality. Citation paths are the useful in understanding knowledge flow, industrial trends, and technology developments [9]. The patent citations are analyzed by means of multi-stage measurements of the inventive progress. Multi-stage measurement considers not only direct citations, but also indirect citations and bibliographic coupling [10].

It is studied the causes of unused relevant information among scholarly papers in previous research Previous [11], and then three reasons for unused relevant information are proposed, namely: (1) failure to find, (2) information overload, (3) non-use policy. The citing motivation of patent inventors should be similar to that of paper authors. However, unlikely papers, a patent for an invention is the grant of a property right to the technique or design innovation of an inventor or assignee. It brings out the seriously rivalrous relationship among patents. An assignee may deliberately choose not to cite the relevant patents of competitors. Moreover, citable materials such as relevant prior publications or patents may be unused due to the failure to find them. Missing relevant patent citation links are identified by using bibliographic coupling with consideration for above mentioned circumstances [9]. Although bibliographic coupling is proper method to identify missing patent link, collaboration contents cannot be suggested after linking firms because contents is not considered.

In this paper, we suggest a methodology to identify the interested parties for open innovation as one of technology intelligence tools. Patent network at assignee level is generated with bibliographic coupling which is suggested methodology in previous research. Taking it a step further, this paper presents collaboration keywords between potential partners for open innovation by developing patent network which includes patent text information by using keyword vector mapping as a patent contents analysis.

II. Background

A. Bibliographic Coupling

There may be some missing relevantpatents, and the analyses of patent citations will be inaccurate due to incomplete information on the relationship amongpatents. Adding the missing relevant patent links (MRPLs) in the citation network would provide a more comprehensiveview of the relationship among patents.Bibliographic coupling(BC) [12] and co-citation

Department of Industrial & Systems Engineering,Dongguk University-Seoul,3-26 Pil-dong, Chung-gu, Seoul, 100-715, Republic of Korea(e-mail: inchae.j.park@gmail.com).
Department of Industrial & Systems Engineering,Dongguk University-Seoul,3-26 Pil-dong, Chung-gu, Seoul, 100-715, Republic of Korea (e-mail: postman3@dongguk.edu).

(CC) [13] are methods currentlyused for retrieving relevant documents. Both occur when two works reference a common third work in their bibliographies.BC is constructed by the citing relationship, while CC is constructed by the citedrelationship.Many studies [14]-[17] employ BC or CC to discover the relevant literatures that were not found during ordinary studies. Also, BCor CC clustering are used to explore the research fronts [18]-[21]. Comparisons between these two methods have been performed in severalresearch works [20]. BC is immediately available upon publication ofthe later-issued patent from a BC pair; however, it takes time to retrieve the CC between a pair of patents. Compared withCC, BC provides more current and immediate information about patents. Therefore, BC is chosen for identifying MRPLs inthis research.

B. Keyword Vector Mapping

There is a keyword vector mapping(KVM) in the one way that visualizes unstructured documents. This method includes the following procedure, keyword information is extracted from documents through text-mining that is one of data mining method, documents are presented as keyword vector, and keyword vector is visualized through some methodologies such as social network analysis(SNA), self-organizing map(SOM), and multi-dimensional scaling (MDS). In this process, a keyword vector mapping is applied with social network analysis. There are some attempts to visualize documents using this method, for patent analysis in particular. A new approach is proposed to discovering new technology opportunities using patent map which is developed from patent documents using keyword vector of patents and PCA method[22]. Visualization of patent analysisthrough keyword vector mapping of patent documents is applied in other previous research [23]. They also established a semantic network of keywords from patent documents to visualize a clear overview of patent information in a more comprehensible way. There was also attempt to develop a self-organizing feature map (SOFM)-based patent map [24]. In their research, word vectors and document vectors which are generated and positioned in an identical vector space and relevant degree between any two words or documents can be computed as a cosine coefficient of two vectors.

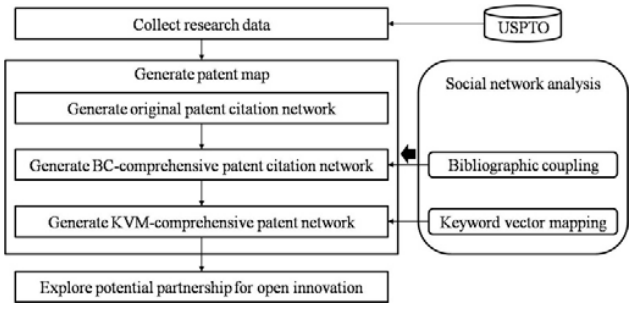

Fig. 1 Overall process of the research

III. Research Methodology

A. Research Ideas

This research aims at proposing method for generating patent network that utilizes bibliographic coupling and keyword vector mapping to find the potential partnership for open innovation. Patent bibliographic and textual information is utilized to generate three types of patent networks- original patent citation network, BC-comprehensive patent citation network, and KVM-comprehensive patent network. The potential partnership is identified after generating patent network by interpreting the patent network.

B. Overall Process

The overall research process consists of several steps like Fig. 1, the first of which is data collection from the U.S. Patent and Trademark Office (USPTO). Second, several types of patent networks are generated by using social network analysis. The relationships between nodes in patent network are established by bibliographic coupling (BC) and keyword vector mapping (KVM). Finally, the potential partnership for innovation is identified by analyzing the comprehensive patent network.

C. Data Sources

Patent bibliographic information including such as patent number, assignee and citation etc. is collected for generating original patent citation network and BC-comprehensive patent citation network from USPTO using patent information analysis system (PIAS) which is useful patent analysis software. Additionally, 230 Patent documents are downloaded for KVM from USPTO.

TABLE I
Example of Keyword Vector

Keyword	K_1	K_2 …		K_{N-1}	K_N
Patent1	(0	11	…	1	0)
Patent2	(1	1	…	30	0)
…			…		…
Patent m-1	(27	2	…	0	0)
Patent m	(36	0	…	0	0)

D. Original Patent Citation Network

Original patent citation network (OPCN) is constructed by the citing and cited relationship among patents. The patents are represented by vertices and the citations by arcs. An arc from vertex i to vertex j denotes that the patent j is in the reference list of patent i. the patent i is defined as the citing patent, and the patent j is defined as the cited patent. A patent can be both a citing and cited patent if it has references and also appears as a reference in other patents' reference list. The vertex-adjacency matrix O for an OPCN at the patent level can be defined as:

$$O_{ij} = \begin{cases} 1, \text{if the patent j is in the reference list of the patent i} \\ 0, \text{otherwise} \end{cases} \quad (1)$$

Where O is an asymmetric m × m matrix, i.e., $o_{ij} \neq o_{ji}, m =$

|P|, and P is the set of patents.

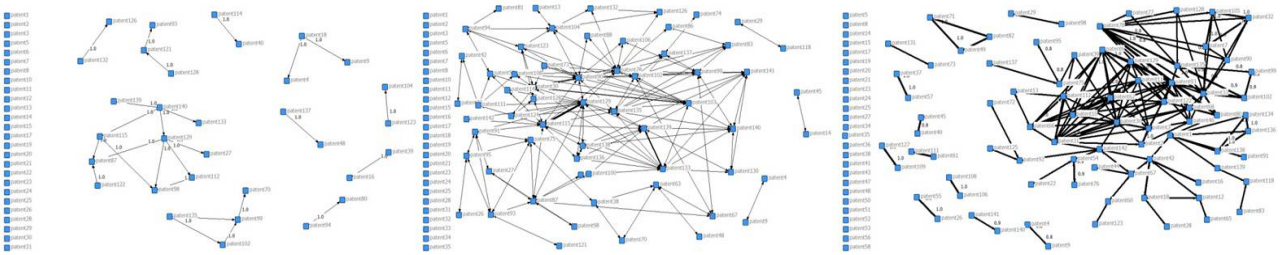

Fig. 2 (a) OPCN at patent level, (b) BC network at patent level, (c) KVM network at patent level in portable fuel cell technology

E. Comprehensive Patent Network

In this research, two types of comprehensive patent network which are BC-comprehensive patent citation network (CPCN) and KVM-comprehensive patent network (CPN) are generated. BC-CPCN for finding missing relevant patent links (MRPL) is generated in previous research[9]. Two or more patents are said to be bibliographically coupled if they have cited the same references. The strength of the BC is defined as the number of common references. In general, the more references they both cite, the more common technical background they both based on for development [12]. That is to say, the higher the BC strength of between the two patents, the higher the relevance of them [25]. The vertex-adjacency matrix B for the BC-CPCN at the patent level is defined as:

$$b_{ij} = \begin{cases} 1, \text{if there are r BC pairs between i and j for } r \geq \alpha \\ 0, \text{otherwise} \end{cases} \quad (2)$$

Where matrix B is a symmetric m× m matrix, i.e., $b_{ij} = b_{ji}$. The value of b_{ij} is one if the BC strength of patents i and j is larger than a specific threshold α, otherwise, it is zero. In this paper, sensitive analysis is used to determine a reasonableα.

For generating KVM-comprehensive patent network (CPN), collected patent documents are transformed into keyword vectors. Keyword vector mapping is conducted to visualize patent contents into two-dimensional network by using a set of keywords which are selected by technology domain experts. Each patent document is represented the combination of keyword frequency like TABLE I. The vertex-adjacency matrix K for the KVM-CPN at the patent level is defined as:

$$k_{ij} = \frac{\vec{V}(p_i) \cdot \vec{V}(p_j)}{|\vec{V}(p_i)||\vec{V}(p_j)|} \quad (3)$$

Where matrix B is a symmetric m× m matrix, i.e., $k_{ij} = k_{ji}$, and $\vec{V}(p_i)$ is a keyword vector of patent i. The value of k_{ij} is a similarity between patent i and patent j.

If the patent pairs without existing citations have the greater BC strength and the greater patent similarity, MRPL can be identified by using mathematical vertex-adjacency matrix C for the BC-CPCN and KVM-CPN at the patent level can be defined as:

$$\begin{cases} BC - c_{ij} = o_{ij} + b_{ij} & (4) \\ KVM - c_{ij} = o_{ij} + k_{ij} & (5) \end{cases}$$

Where matrix C is an asymmetric m × m matrix, i.e., $c_{ij} \neq c_{ji}$. The same equations can be applied for analyzing at assignee level using aggregated coupling input values. While bibliographic information is only considered by using BC-CPCN, meaningful textual information is considered by using KVM-CPN.

F. Exploring Potential Partnership

The potential partnership for open innovation is identified by analyzing two types of comprehensive patent network. Both networks suggest an enhanced link, new link, and unchanged link in comparison to OPCN. Firms with enhanced link has citation relationship currently, these have possibility to collaborate each other since there patents cite common patents and similarity of patent is very high. Although firms with new link have no citation relationship or low citation relationship currently, these have possibility to start to collaborate with existing firms since the firms appear in network. The firms with unchanged link are not the target of collaboration even if these have the citation relationship currently.

New links appear between existing assignees to enhance the relationship, with new assignee and between new assignees to start to relationship. For existing assignees, direct linked assignees are the first targets of collaboration partner to be considered than indirect linked assignees.

Threshold should be manipulated less than one since the similarity based on Keyword vector mapping is more than zero and less than one. Then, new links affected by similarity value appear. The notable difference of KVM-CPN is that it is able to suggest common keyword list between linked two assignees so that user can obtain information regarding collaboration contents before they investigate target of collaboration firm.

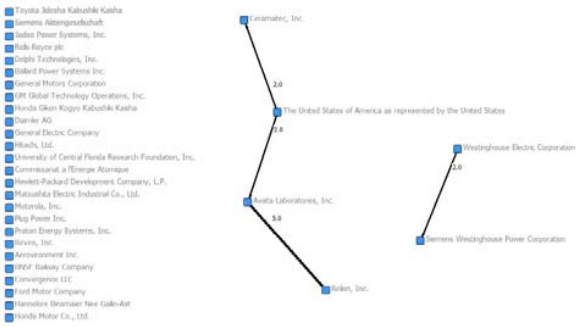

Fig. 3 OPCN at assignee level in portable fuel cell technology

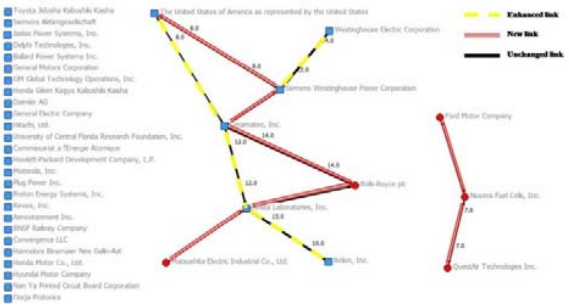

Fig. 4 BC-CPCN at assignee level in portable fuel cell technology

IV. Illustration

In order to demonstrate the feasibility of the research methodology, the portable fuel cell technology which is one of the promising technology domains as a green technology is chosen for the case study in this paper. 230 patents documents which are owned by 116 assignees and bibliographic information are collected. 143 patents among 230 patents which 37 assignees whose patents are more than two have are utilized for demonstration.

A. Original Patent Network

Fig. 2 is three types of original patent network when threshold is 1, 1, and 0.8 respectively. Each original patent network has different patterns of linking since the arcs means different relationship in each network. Furthermore, it is too complex to identify the potential partnership because of too many nodes and self citation. Thus, this paper focuses on analyzing the network at the assignee level.

Fig. 3 shows OPCN at assignee level when threshold is 2. There are 6 assignees in the OPCN which have citation relationship each other.

B. Comprehensive Patent Network

Fig. 4 and Fig. 5 are BC-CPCN and KVM-CPN when the threshold is 4 and 0.95 respectively. Both show enhanced links and new links around 6 existing assignees from OPCN. There are no unchanged links since threshold is manipulated to focus on the purpose of the research which is to explore potential partnership. All of the 6 existing assignees from OPCN have enhanced links. Different assignees are linked with existing assignees in each network. Circular nodes appear in BC-CPCN while diamond shaped nodes appear in KVM-CPN.

TABLE II
Properties of Comprehensive Patent Network

properties	OPCN	BC-CPCN		KVM-CPN	
		total	Add-in	total	Add-in
# of Nodes	6	11	5	15	9
# of Arcs	4	11	7	14	10
Enhanced link	-	4		4	
New link	-	7		10	
Unchanged link	-	0		0	
Direct linked assignees	-	· Rolls-Royce Plc · Matsushita Electric Industrial Co., Ltd.		· Delphi Technologies, Inc. · Hewlett-Packard Development Company, L.P. · Convergence LLC · General Motors Corporation	
Indirect linked assignees	-	· Ford Motor Company · Nuvera Fuel Cells, Inc. · QuestAir Technologies Inc.		· Plug Power Inc. · Matsushita Electric Industrial Co., Ltd. · Motorola, Inc. · Toyota Jidosha Kabushiki Kaisha · Daimler AG	

Fig. 5 KVM-CPN at assignee level in portable fuel cell technology

C. Exploring Potential Partnership

Some changes from OPCN to CPN can be confirmed in TABLE II. There are quite different new assignee appearance patterns since BC-CPCN is based on citation information and KVM-CPN is based on text information. When company selects partner for open innovation using this network, collaboration contents can be predicted in KVM-CPN since KVM-CPN has advantage in that it suggests common keyword list between linked assignee pairs in the network like TABLE III. The common keywords are from 47 keywords which are extracted to construct the keyword vectors when keyword vectors are constructed.The common keywords can be interpreted as common interest between connected assignees in specific technology area. Some common keywords are general keywords such as material, air, type etc. However, some features can be extracted by analyzing the keyword list. For example, Both Siemens Westinghouse Power Corporation andWestinghouse Electric Corporation are interested in technology on temperature control since they have temperature,

thermal, insulation etc. in the list as common keywords. Also, user can guess that General Electric Company andAvista Laboratories, Inc. are interested in fuel cell technology for safety or protection from the dangerous circumstance since there are keywords such as resistance, pressure, condition, vapour, gas, protection, humidity, safety device, hazard etc. Thus, these common keywords will be the meaningful information for two connected assignees which are identified as potential partners to suggest the collaboration guide.

TABLE III
EXAMPLE OF COMMON KEYWORD LIST BETWEEN LINKED ASSIGNEE PAIRS

Linked Assignee Pairs	Common Keyword List
Avista Laboratories, Inc. - The United states of America as represented by the united states	air, material, storage, type, surface, resistance, mechanical
Avista Laboratories, Inc. -Relion, Inc.	Air, gas, material, metal, pressure, condition, control, energy, thermal, cell system, storage, oxygen, temperature, type, stack, surface, resistance, concentration, cooling, emission, mechanical
Relion, Inc. - The United states of America as represented by the united states	Air, material, storage, type, surface, resistance, enclosure, mechanical
Siemens Westinghouse Power Corporation -Westinghouse Electric Corporation	Air, gas, oxygen, temperature, type, material, metal, stack, surface, pressure, discharge, energy, thermal, insulation, cell module, concentration, condition, control, voltage, enclosure, fuel supply, fuse, power system
General Electric Company -Avista Laboratories, Inc.	Material, emission, battery, outdoor use, indirect contact, stack, resistance, normal operation, pressure, condition, vapour, power system, monitoring, gas, protection, enamel, humidity, mechanical, air, enclosure, safety device, surface, hazard

V. CONCLUSIONS

This research presented a new approach to identify missing relevant patent links. Although there are many endeavors to identify MRPLs by using bibliographic coupling and co-citation in previous research, there is a limitation in that available information is only bibliographic information. Contents analysis is conducted by using keyword vector mapping methodology to overcome the limitation of utilizing information in this research. The property of method presented by comparing between citation analysis approach and contents analysis approach. Thus, the potential partnership for open innovation is identified by finding missing relevant patent links from two types of comprehensive patent network at assignee level. Furthermore, collaboration keyword is provided as a result of analyzing the keyword vector mapping comprehensive patent network. These keywords will be meaningful supporting information to set up strategic cooperation direction once the partnersare selected then start to cooperate each other.

However, type of open innovation or cooperating strategy is not suggested for selected partners although the relationship between assignees is identified in network. Additionally, the guideline for utilizing the extracted collaboration keyword needs to be suggested for the future research.

In future research, type of cooperation strategy for open innovation will be suggested through literature surveys on open innovation. Then the type will be used to interpret the network and draw meaningful implications with collaboration keywords.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(No. 2012R1A1A1011934).

REFERENCES

[1] G H. Chesbrough, *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business School Press, Boston, MA: 2003.
[2] O.Gassmann, “Opening up the innovation process: towards an agenda”,*R&D Management*, vol. 36, no.3, pp. 223–228, 2006.
[3] H. Chesbrough, W. Vanhaverbeke, J. West, *Open Innovation: Researching a New Paradigm*, Oxford University Press, London. 2006.
[4] U. Lichtenthaler, “Open innovation in practice: Analysis of strategic approaches to technology transactions”, *IEEE Transactionson Engineering Management*, vol.55, no.1, pp.148–157, 2008.
[5] S. Lee, G. Park, B. Yoon,J. Park, "Open innovation in SMEs—An intermediated network model", *Research Policy*, vol.39, pp.290–300, 2010.
[6] M. Veugelers,J. Bury, S. Viaene, "Linking technology intelligence to open innovation", *Technological Forecasting & Social Change*, vol.77, pp.335–343, 2010.
[7] C.I.V. Kerr, L. Mortara, R. Phaal, D.R. Probert, "A conceptual model for technology intelligence", *Int. J. of Technology Intelligence Planning* vol.2, no.1, pp.73–93, 2006.
[8] A. B. Jaffe, M. Trajtenberg, M. S.Fogarty, "Knowledge spillovers and patent citations: Evidence from a survey of inventors", *American EconomicReview*, vol.90, no.2, pp. 215–218, 2000.
[9] D. Chen,M. Huang, H. Hsieh, C. Lind,"Identifying missing relevant patent citation links by using bibliographic coupling in LED illuminating technology", *Journal of Informetrics*, vol.5, pp.400–412, 2011.
[10] I. V. Wartburg, T. Teichert, K. Rost, “Inventive progress measured by multi-stage patent citation analysis”,*Research Policy*, vol. 34, no.10, pp.1591–1607, 2005.
[11] P.Wilson, “Unused relevant information in research and development”,*Journal of the American Society for Information Science*, vol.46, no.1, pp.45–51, 1995.
[12] M. M. Kessler, “Bibliographic coupling between scientific papers”,*American Documentation*, vol.14, no.1, pp.10–25, 1963.
[13] H. G.Small, “Co-citation in the scientific literature: A new measure of the relationship between two documents”,*Journal of the American Society forInformation Science*, vol.24, no.4, pp.265–269, 1973.
[14] C. W. Cleverdon, “The Cranfield tests on index language devices”,*Aslib Proceedings*, vol.19, no.6, pp.173–194, 1967.
[15] S. P.Harter, “The Cranfield II relevance assessments: A critical evaluation”,*The Library Quarterly*, vol.41, no.3, pp. 229–243, 1971.
[16] D. R. Swanson, “Some unexplained aspects of the Cranfield tests of indexing performance factors”,*The Library Quarterly*, vol.41, no.3, pp.223–228,1971.
[17] R. R.Braam, H. F. Moed, A. F. J.vanRaan, “Mapping of science by combined co-citation and word analysis. I. Structural aspects”,*Journal of theAmerican Society for Information Science*, vol. 42, no.4, pp.233–251, 1991.
[18] H. G.Small, B. C.Griffith, “The structure of scientific literatures. (I) Identifying and graphing specialties”,*Science Studies*, vol. 4, no.1, pp.17–40, (1974).
[19] O.Persson, “The intellectual base and research fronts of JASIS1986–1990”, *Journal of the American Society for Information Science*, vol.45, no.1, pp. 31–38, 1994.
[20] S. A.Morris, G. G. Yen, Z.Wu, B.Asnake, “Time line visualization of research fronts”,*Journal of the American Society for Information Science andTechnology*, vol.54, no.5, pp.413–422, 2003.

[21] B.Jarneving, "Bibliographic coupling and its application to research-front and other core documents", *Journal of Informetrics*, vol.1, no.4, pp.287–307, 2007.
[22] S. Lee, B. Yoon, Y. Park,"An approach to discovering new technology opportunities: Keyword-based patent map approach", *Technovation*, Vol. 29, no. 6-7, pp.481-497, 2009.
[23] Y.Kim, J. Suh, S. Park,"Visualization of patent analysis for emerging technology", *Expert Systems with Applications*, vol. 34, no. 3 pp.1804-1812, 2008.
[24] B. Yoon, C. Yoon, Y. Park, "On the development and application of a self-organizing feature map-based patent map", *R&D Management*, vol. 32, no. 4, pp.291-300, 2002.
[25] M. H. Huang, L.Y. Chiang, D. Z. Chen, "Constructing a patent citation map using bibliographic coupling: A study of Taiwan's high-tech companies", *Scientometrics*, vol.58, no.3, pp.489-506, 2003.