Technological Forecasting & Social Change 103 (2016) 71–82

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

A novel approach to identify the major research themes and development trajectory: The case of patenting research

Louis Y.Y. Lu [a,b,*], John S. Liu [c]

[a] College of Management, Yuan Ze University, 135 Yuan-Tung Road, Chung-Li 32003, Taiwan
[b] Innovation Center for Big Data and Digital Convergence, Yuan Ze University, 135 Yuan-Tung Road, Chung-Li 32003, Taiwan
[c] Graduate Institute of Technology Management, National Taiwan University of Science and Technology, 43, Section 4, Keelung Road, Taipei 10607, Taiwan

ARTICLE INFO

Article history:
Received 16 March 2015
Received in revised form 24 October 2015
Accepted 26 October 2015
Available online xxxx

Keywords:
Main path analysis
Clustering
Research theme
Development trajectory
Citation network
Patenting

ABSTRACT

This paper presents a novel approach to identify the major research themes and development trajectory of a targeted field and takes the case of patenting research as an example. Edge-between clustering and key-route main path analysis are employed to complementarily accomplish the task. This study retrieves patenting related articles covering 1970 to 2013 from Web of Science (WOS) and constructs the citation network among them. The edge-betweenness clustering technique and key-route main path analysis are then applied to identify the major research themes and development trajectories of patenting research. Eight major research themes are identified: 'citation network analysis', 'patent law', 'patent valuation', 'academic patenting', 'gene patenting', 'patent policy', 'patent protection', and 'technology analysis'. The linkage among these eight research themes is exhibited, along with a presentation of the statistics of top influential journals and authors. This study demonstrates that the approach used herein is a powerful way to determine the major research themes and development trajectories of a target academic field. The approach is also applicable to any other data with citation relationships.

1. Introduction

The bibliometric approach is one of the methodologies to quantitatively analyze academic literature and can handle a large amount of studies in the literature. Kajikawa et al. (2007) and Kajikawa et al. (2008) adopted Girvan and Newman's (2002) edge-betweenness clustering concept to identify the research fronts of a specific field. A research front is a theme on which many researchers have targeted their efforts at around the same time. More specifically, a research front is formed when a group of researchers publishes many articles that address the same or similar issues during a specific time span. Liu and Lu (2012) proposed an integrated method of main path analysis to track the development trajectory of a scientific or technological field. Although Kajikawa et al. (2007) and Liu and Lu (2012) improved the clustering and main path analyses respectively, there is still room for improvement. This study conducts a literature survey through a novel approach that adopts both edge-betweenness clustering and main path analysis, demonstrating that this approach can deliver more valuable information than previous methods.

To demonstrate the usefulness of this novel approach, this study applies it to patenting research. A patent is the exclusive right granted by a government to an inventor or assignee for a limited period of time in exchange for disclosing the details of an invention. A patent right is a form of intellectual property rights (IPRs), and the protection of them is very critical for multinational corporations (MNCs) to keep their technological superiority, competitiveness, and return on innovation investments. Patenting is one of the most important means among the many varieties of IPRs that firms adopt to protect their valuable intangible assets and maintain competitive advantages. Many researchers have investigated the issues of patenting from different perspectives over the past decades. It is now time to conduct a complete survey of the published literature to understand the knowledge development trajectories in patenting.

There have been quite a few review papers on patenting over the past decade. Somaya (2003) reviewed the research on patent strategy and highlighted two major themes: generic patent strategies and strategic management of patents. He also mapped two major themes onto three domains. He carefully read the literature on patent strategy to identify the research themes, but excluded many research of them due to resource limitation. Hanel (2006) surveyed the policy changes regarding intellectual property protection in the U.S. and the use and strategies of IPRs in the U.S., Canada, EU, Japan, and Australia. From 125 papers, Baldini (2006) reviewed university patenting and licensing activities and achieved some findings via a thorough reading of these papers. The above review papers contributed significantly to the patenting field, but they adopted the traditional qualitative method which is hard to survey literature with a large amount of articles.

* Corresponding author at: College of Management, Yuan Ze University, 135 Yuan-Tung Road, Chung-Li 32003, Taiwan.
E-mail addresses: louislu@saturn.yzu.edu.tw (L.Y.Y. Lu), johnliu@mail.ntust.edu.tw (J.S. Liu).

http://dx.doi.org/10.1016/j.techfore.2015.10.018
0040-1625/

This study applies the novel method to 8025 research articles in patenting and successfully identifies the major research themes and the overall development trajectory. First, patenting related articles are retrieved from ISI Web of Science (WOS) databases, and then a citation network through their citation data is constructed. Next, this study applies the edge-betweenness based clustering technique (Newman, 2006; Newman and Girvan, 2004) to split the citation network into several sub-networks (groups). After clustering, the articles in a group are tightly connected (citing or cited) within the group, but loosely connected to articles outside the group. Articles with different research themes are thus distinguished through edge-betweenness based clustering. The word clouding technique is conducted on the title of all the articles in each group to extract the key research topics and to name each group. The key-route main path analysis (Liu and Lu, 2012) helps identify the overall development trajectory of patenting research.

The aims of this paper are to describe the novel approach and to answer the following questions in patenting research. Which authors and journals are involved in the development of patenting research? What are the major research themes in patenting research? What is the overall development trajectory of patenting research?

This paper is organized as follows. After the Introduction, Section 2 briefs the methodologies used in this study — in particular, edge-betweenness clustering and key-route main path analysis. Section 3 describes the data and presents the basic statistics. Section 4 elaborates upon the major research themes and overall development of patenting research and the linkage among all research groups. Section 5 concludes the findings and contributions and discusses the future perspectives.

2. Methodology

There are two ways to reveal the patterns of a citation network: bibliographic coupling and co-citation analysis. Bibliographic coupling analyzes the relationship from the citing point of view. When different authors cite one or more papers in common, these papers are bibliographically coupled and are highly likely to be related. Co-citation analysis counts the frequency of co-citation to identify pairs of highly cited papers. Bibliographic coupling is backward-looking whereas co-citation analysis is forward-looking (Garfield, 2001). Bibliographic coupling and co-citation analysis involve either cited or citing papers, and both analyses examine the relationship among papers from a one-sided lens. This study sees the patterns of a citation network from a holistic view and takes into consideration both citing and cited relationships at the same time. Such an approach allows us to identify the development trajectory and the major research themes of a scientific field.

When a paper cites a previous work, knowledge presumably flows from the previous work to the citing paper. By establishing the citation links among a citation network and then tracing the main paths of knowledge flow, the development trajectory or the knowledge diffusion path can be identified (Liu and Lu, 2012; Hummon and Doreain, 1989). In a citation network, the importance of each citation is regarded as the same along with the knowledge flows from a cited article to the article that cites it, and hence a citation network is a non-weighted and directed network.

This study uses two methods to analyze the citation network built from the collected dataset and explores the embedded attributes. The first one is the clustering method based on edge-betweenness (Girvan and Newman, 2002) and optimal modularity concepts (Newman, 2006) to group similar articles. The second one is the key-route main path analysis proposed by Liu and Lu (2012) to identify the overall evolutionary trajectories of patenting development and exhibits the linkage among all research groups. The concepts of these two methods are briefly explained in the following sections.

2.1. Edge-betweenness clustering

In a citation network, if two articles connect to (cite or are cited by) the same other articles, then it is highly possible that they address a similar issue. Those articles that address the same issue form a tightly knitted 'community' in the citation network. If one can separate the communities that address different issues from a citation network, then the research groups of a target field are identified. Edge-betweenness clustering achieves this purpose by removing the edges that are 'between' different groups. The edge-betweenness of an edge is defined as 'the number of shortest paths between pairs of vertices that run along it' (Girvan and Newman, 2002). The edge-betweenness clustering method divides network nodes into groups within which the network connections are dense, but between which the network connections are sparse.

Girvan and Newman (2002) introduced the concept of edge-betweenness and applied it to cluster a social network and a biological network. They further enhanced their clustering method by introducing a fast algorithm in their later studies (Newman and Girvan, 2004; Newman, 2004). Newman (2006) further proposed the modularity index to effectively identify the 'optimal' community structure of a network. Modularity is defined as 'the number of edges (links) falling within groups minus the expected number in an equivalent network with edges placed at random'. Applying the modularity concept, one can search for the division with the maximum modularity in order to obtain the optimal division of a network. This study applies the optimal modularity concept in edge-betweenness clustering to group patenting articles that discuss similar issues.

The algorithm of edge-betweenness clustering is as follows: First, calculate the betweenness for all edges in the network; second, remove the edge with the highest betweenness; third, recalculate the betweenness for all edges affected by the removal; and fourth, repeat from step 2 until no edge remains. At this point, the method traces back the abovementioned process and selects the network division that has the largest modularity. This study directly uses the function provided in

Table 1
List of the top 30 influential authors.

g-Index	h-Index	Total papers	1st author	Active years[a]	Name
26	17	26	15	1997–2013	Lemley, MA
23	13	23	13	1994–2011	Lerner, J
20	12	20	13	2000–2013	Meyer, M
17	11	24	1	2006–2013	Park, Y
15	10	15	10	1993–2013	Jaffe, AB
14	8	17	4	2004–2013	Matthijs, G
13	8	13	8	1999–2012	Harhoff, D
12	9	12	11	2004–2012	Czarnitzki, D
12	9	12	11	1992–2011	Eisenberg, RS
12	9	12	9	1995–2012	Ernst, H
12	9	12	6	1999–2010	Verspagen, B
12	8	12	1	2003–2013	Chen, HC
12	8	12	10	1996–2005	Lanjouw, JO
12	6	13	13	1996–2012	Denicolo, V
12	6	12	5	1999–2013	Malerba, F
12	6	12	3	2006–2013	Vanoverwalle, G
12	5	18	8	2008–2013	Lee, S
11	8	14	12	1996–2002	Steele, P
11	8	11	8	1984–2000	Narin, F
11	6	15	10	2001–2009	Marinova, D
11	6	11	11	2001–2013	Hall, BH
10	8	12	3	2002–2013	Cook-Deegan, R
10	7	10	1	1999–2013	Debackere, K
10	7	10	6	1998–2011	Mowery, DC
10	7	10	2	1986–2013	Schankerman, M
10	7	10	6	2002–2013	Wagner, RP
10	6	14	4	2006–2013	Delapotterie, BV
10	6	12	2	2002–2007	Mcaleer, M
10	6	11	9	2001–2013	Bhattacharya, S
10	6	10	6	2001–2013	Sampat, BN

[a] Only the volumes are indexed in the ISI database.