J. Eng. Technol. Manage. 32 (2014) 129–146

Contents lists available at ScienceDirect

Journal of Engineering and Technology Management

journal homepage: www.elsevier.com/locate/jengtecman

Detecting research fronts using different types of weighted citation networks

Katsuhide Fujita [a,*], Yuya Kajikawa [b], Junichiro Mori [c], Ichiro Sakata [d]

[a] Faculty of Engineering, Tokyo University of Agriculture and Technology, 2-24-16 Naka-cho, Koganei-shi, Tokyo 184-8588, Japan
[b] Graduate School of Innovation Management, Tokyo Institute of Technology, 3-3-6 Shibaura, Minato-ku, Tokyo 108-0023, Japan
[c] Presidential Endowed Chair for Platinum Society, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[d] School of Engineering, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-8656, Japan

ARTICLE INFO

Article history:
Received 1 February 2013
Received in revised form 12 July 2013
Accepted 16 July 2013

JEL classification:
O3

Keywords:
Research front
Citation network analysis
Bibliometrics
Decision support

ABSTRACT

In this paper, we investigate the performance of different types of weighted citation networks for detecting emerging research fronts by a comparative study. Three citation patterns including direct citation, co-citation and bibliographic coupling, have been tested in three research domains including gallium nitride, complex networks, and nano-carbon. These three patterns of citation networks are constructed for each research domain, and the papers in those domains are divided into clusters to detect the research front. Additionally, we apply some measures to weighted citations like difference in publication years between citing and cited papers and similarities of keywords between them, which are expected to be able to effectively to detect emerging research fronts. To investigate the performance of different types of weighted citation networks for detecting emerging research fields, we evaluate the performance of each approach by using the following measures of extracted research fronts: visibility, speed, and topological and textual relevance.

* Corresponding author. Tel.: +81 42 388 7141; fax: +81 42 388 7141.
E-mail address: katfuji@cc.tuat.ac.jp (K. Fujita).

0923-4748/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jengtecman.2013.07.002

Introduction

Over the past several decades, the number of academic papers has increased exponentially (Price, 1965), and each academic area has become specialized and segmented. Davidson et al. (1998) describe the situation as follows: "For most of history, mankind has suffered from a shortage of information. Now, in just the infancy of the electronic age, we have begun to suffer from information excess". Therefore, it is hard for researchers to perceive their specialized fields as a whole, and segmentation occurs simultaneously with specialization, which brings a severe problem and also opportunity to find crucial knowledge by integrating different domains. Naturally, it is hard for researchers and managers to detect a research front in the early stages by human effort only. Therefore, there is a strong need for computational tools of science mapping and emerging topic detection. Previous studies have established effective algorithms for creating academic landscapes and for detecting emerging topics for certain research fronts.

To support the detection of research fronts and visualization of academic landscapes, methods of science mapping by citation analysis have been proposed and developed (Boyack et al., 2005; Klavans and Boyack, 2009). Researchers have also focused on clustering and visualization (Chen, 1999; Chen et al., 2003; Small, 1999). For example, Leydesdorff (2004) and Leydesdorff and Rafols (2009) made a large-scale investigation of a set of academic papers. Not only creating static academic landscapes, topological and semantic analysis of a citation network also helps us to focus on significant movements in research fronts and emerging research fields in a broad context (Shibata et al., 2008).

The other approach is to detect emerging clusters of densely connected papers. Price (1965) employed the concept of a research front, that is, a research domain under development where papers cite each other densely. Scientists tend to cite the most recently published articles in their papers; therefore, the network belonging in a research front becomes very tight. In a given field, a research front refers to the body of articles that scientists actively cite. Researchers have been studying quantitative methods that can be used to identify and track a research front as it evolves over time. Small and Griffith (1974) showed that activated scientific specialists generate clusters of co-cited papers. Braam et al. (1991a,b) also investigated the topics discussed in co-cited clusters by analyzing the frequency of indexing terms and classification codes occurring in these publications.

On the other hand, different citation patterns between papers offer some ways to detect emerging research domains. Shibata et al. (2009) performed a comparative study to investigate the performance of methods for detecting emerging research fronts between three citation patterns, co-citation, bibliographic coupling, and direct citation. When a paper directly cites another as a reference, it is called a direct citation. In other words, the direct citation is the citing of an earlier paper by a new paper. Co-citation is defined as the edge between two documents cited by the same paper(s) (Small, 1973). Bibliographic coupling is defined as the edge between two documents citing the same paper(s) (Kessler, 1963). Three patterns of citation networks were constructed for each research domain, and the papers in those domains were divided into clusters to detect the research front. Direct citation, which could detect large and young emerging clusters earlier, shows the best performance in detecting a research front, and co-citation shows the worst. Small (2006) proposed a method of tracking and predicting growth areas by co-citation analysis that analyzed co-citation networks generated from the top 1% of highly cited papers. Klavans and Boyack (2006) compared the performance of clustering in journal citation networks created by direct citation and co-citation. Their results suggested that a network of co-citation has higher content similarity. Boyack and Klavans (2010) evaluated textual similarity of papers in clusters extracted by four different citation patterns: bibliographic coupling, direct citation, and a hybrid approach of direct citation and co-citation. Bibliographic coupling slightly outperforms co-citation and direct-citation using textual accuracy measures. In a certain cases, especially for large data set, bibliographic coupling might work better. However, it cannot be applied for research front detections in a specific research domain, because hub papers in a specific domain gather larger citations even when there are less common research topics between citing and cited papers. In fact, this bias would have the less effect on results in large corpuses because citation networks become globally sparse and locally dense.

Despite weighted citation networks can capture important information attributes of papers, most of the existing works focus on the non-weighted citation networks. The purpose of this paper is to