Journal of Informetrics 10 (2016) 82–97

Contents lists available at ScienceDirect

Journal of Informetrics

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

Interrelations among scientific fields and their relative influences revealed by an input–output analysis

Zhesi Shen[a], Liying Yang[b], Jiansuo Pei[c], Menghui Li[d], Chensheng Wu[d], Jianzhang Bao[a], Tian Wei[a], Zengru Di[a], Ronald Rousseau[e], Jinshan Wu[a,*]

[a] School of Systems Science, Beijing Normal University, Beijing 100875, PR China
[b] National Science Library, Chinese Academy of Sciences, Beijing 100190, PR China
[c] School of International Trade and Economics, University of International Business and Economics, Beijing 100029, PR China
[d] Beijing Institute of Science and Technology Intelligence, Beijing 100044, PR China
[e] KU Leuven, Department of Mathematics, 3000 Leuven, Belgium

ARTICLE INFO

Article history:
Received 9 June 2015
Received in revised form 3 November 2015
Accepted 3 November 2015

Keywords:
Informetrics
Economics
Input–output analysis
Physics

ABSTRACT

In this paper, we try to answer two questions about any given scientific discipline: first, how important is each subfield and second, how does a specific subfield influence other subfields? We modify the well-known open-system Leontief Input–Output Analysis in economics into a closed-system analysis focusing on eigenvalues and eigenvectors and the effects of removing one subfield. We apply this method to the subfields of physics. This analysis has yielded some promising results for identifying important subfields (for example the field of statistical physics has large influence while it is not among the largest subfields) and describing their influences on each other (for example the subfield of mechanical control of atoms is not among the largest subfields cited by quantum mechanics, but our analysis suggests that these fields are strongly connected). This method is potentially applicable to more general systems that have input–output relations among their elements.

1. Introduction

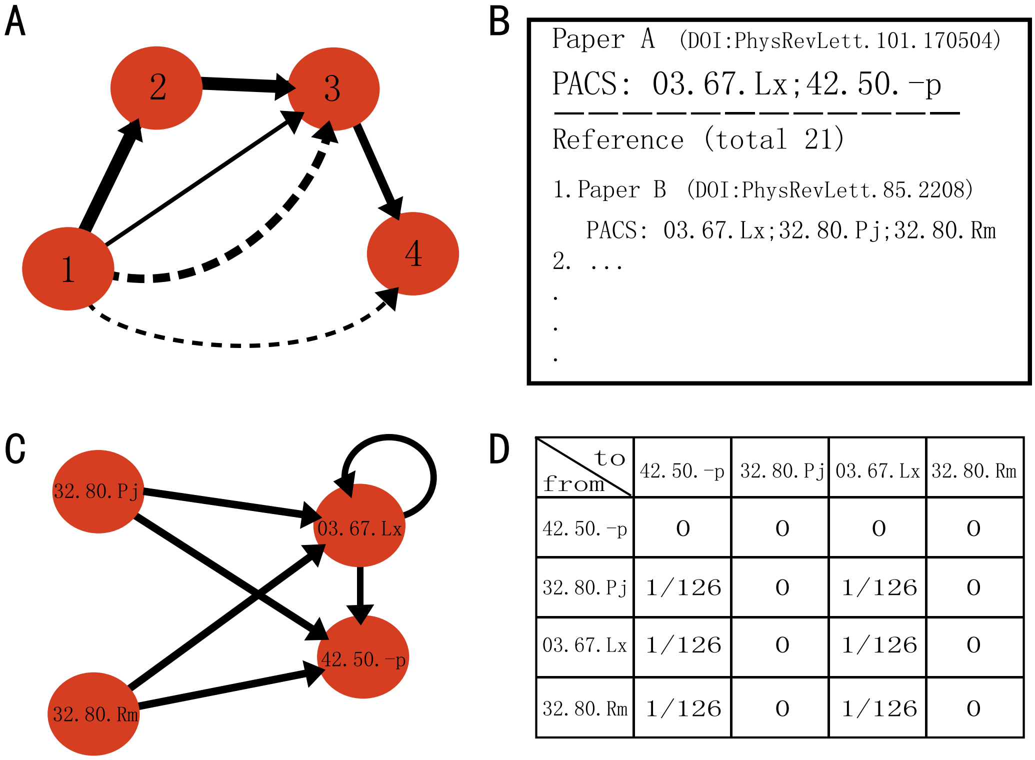
Science funding agencies and science policymakers often have to decide on which science or technology fields to prioritize for a period of time for an efficient management of scientific resources and activities. To answer this question, the funding agencies need to assess the (future) relative importance of all scientific fields. Furthermore, once the target, i.e. the prioritized field, is chosen, it becomes an important consideration to find other fields which support the target field, as these be supported too.

These two questions are relevant not only to policymakers and committees in such agencies, but also to individual scientists, academic committees and university departments. Of course, one can apply peer review, relying on the opinions, feelings and visions of individual experts. However, with the rise of the era of big data, a natural question is whether technical analyses using large collections of published patents and research articles can help answer such questions.

The question of the relative importance of and influences between scientific fields has not yet been answered completely, admitting that investigating connections between scientific fields and technological sectors is one of the areas of investigation

* Corresponding author. Tel.: +86 1062201355.
E-mail address: jinshanw@bnu.edu.cn (J. Wu).

http://dx.doi.org/10.1016/j.joi.2015.11.002



from \ to	42.50.-p	32.80.Pj	03.67.Lx	32.80.Rm
42.50.-p	0	0	0	0
32.80.Pj	1/126	0	1/126	0
03.67.Lx	1/126	0	1/126	0
32.80.Rm	1/126	0	1/126	0

Fig. 1. Citations among all APS papers are converted into an input–output network/matrix of PACS codes. (A) A fictitious network in which there are relatively large differences between the direct and indirect influences between nodes. (B) A piece of the real APS citation network: paper A with PACS codes 03.67.Lx and 42.50.−p cites paper B with PACS codes 03.67.Lx, 32.80.Pj and 32.80.Rm. (C) In the corresponding input–output network of PACS codes, directed links from the PACS codes of Paper B to the PACS codes of paper A are added to the network of PACS codes following the citations from paper A to paper B. (D) A matrix version of B with numbers calculated using Eq. (8).

in the field of informetrics (JST, 2015; Narin, Hamilton, & Olivastro, 1997). In JST (2015), the Japan Science and Technology Agency (JST) was interested in knowing, for a given sector of patents, which scientific fields have been the primary sources of published information. The simple approach used in JST (2015) is to calculate how journal articles cited in a specific sector of patents are distributed across all scientific fields. In Narin et al. (1997), the authors were more focused on how the patterns of citations between patents and scientific publications changed due to national origin and over time. Such analyses based on directly counting the number of articles, patents and citations, are referred to as direct analyses. In this simple, direct statistical approach, an indirect contribution from scientific fields to sectors of patents is missing: if there is one sector of patents T_α, which heavily relies on one scientific field S_i, which in turn makes use of concepts and techniques from another scientific field S_j, then it is clear that even if there are no direct citations from T_α to S_j, S_j is a major contributor to T_α. These connections are referred as indirect connections. They are the main topic of this investigation.

This idea of considering direct as well as indirect relations, though straightforward, cannot be underestimated. Results of such approaches are sometimes described as network effects (West & Vilhena, 2014). In Fig. 1A, we provide an example of a citation relationship between scientific fields in which indirect connections (between node 1 and node 4 or node 1 and node 3) could in principle play a more important role than direct ones, due to the lack of a direct connection between nodes 1 and 4 and a weak connection between nodes 1 and 3. While network science researchers, including those from social network analysis, have often used this perspective (Barabàsi, 2015), the network perspective is not yet a commonplace in informetrics. This remark does not imply that informetricians have not valued the network perspective (Otte & Rousseau, 2002; West & Vilhena, 2014). Indeed, the network effect is the key idea behind Google's PageRank algorithm (Brin & Page, 1998) and its scientific predecessor, the Pinski-Narin influence methodology (Franceschet, 2011; Pinski & Narin, 1976). The PageRank algorithm has been used to measure the relative importance of journals (Bergstrom, West, & Wiseman, 2008) and articles (Chen, Xie, Maslov, & Redner, 2007; Ma, Guan, & Zhao, 2008). However, the PageRank algorithm focuses mainly on ranking the nodes in a network, not on interrelations among the nodes. We consider here both the tasks of ranking as well as describing interrelations.

Now that our work has been placed in its proper context, we first note that we will focus on scientific fields instead of journals and articles. Therefore, we may naively adopt the PageRank algorithm or equivalently the Pinski-Narin influence methodology for our study, by classifying publications into scientific fields.