Journal of Informetrics 8 (2014) 840–853

Contents lists available at ScienceDirect

Journal of Informetrics

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

A comparison of journal similarity across six disciplines using citing discipline analysis

Dietmar Wolfram*, Yuehua Zhao

School of Information Studies, University of Wisconsin–Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA

ARTICLE INFO

Article history:
Received 5 February 2014
Received in revised form 14 August 2014
Accepted 18 August 2014
Available online 14 September 2014

Keywords:
Journal similarity
Citer analysis
Citation analysis
Citing discipline analysis

ABSTRACT

A similarity comparison is made between 120 journals from five allied Web of Science disciplines (Communication, Computer Science-Information Systems, Education & Educational Research, Information Science & Library Science, Management) and a more distant discipline (Geology) across three time periods using a novel method called citing discipline analysis that relies on the frequency distribution of Web of Science Research Areas for citing articles. Similarities among journals are evaluated using multidimensional scaling with hierarchical cluster analysis and Principal Component Analysis. The resulting visualizations and groupings reveal clusters that align with the discipline assignments for the journals for four of the six disciplines, but also greater overlaps among some journals for two of the disciplines or categorizations that do not necessarily align with their assigned disciplines. Some journals categorized into a single given discipline were found to be more closely aligned with other disciplines and some journals assigned to multiple disciplines more closely aligned with only one of the assigned disciplines. The proposed method offers a complementary way to more traditional methods such as journal co-citation analysis to compare journal similarity using data that are readily available through Web of Science.

1. Introduction and previous research

Aspects of scholarly communication may be investigated from different levels of granularity to reveal and better understand relationships between researchers, research groups, institutions, regions/nations, specializations/disciplines, publications or publication outlets. Connections that exist between sources of interest may take the form of direct citations, co-citations, co-authorship, co-occurrence of words or subjects, or more recently, latent topics. The presence of these connections and their frequency can provide an indication of the strength of similarity between sources. The network of relationships between sources based on citations, in turn, can help to better understand the structure of the larger environment (e.g., disciplines) in which the sources exist (Small & Crane, 1979).

A publication outlet such as a journal serves as a primary venue through which products of scholarship are disseminated. Journal-level relationships may be studied using different publication features such as citations or subjects associated with journals. Journal citation analysis and co-citation analysis to identify related groups or journal similarity have been employed by a number of authors. Journal similarity comparison has been frequently studied using co-citations. Journal co-citation studies have been carried out on a number of fields including economics (McCain, 1991), information retrieval

* Corresponding author. Tel.: +1 414 229 6836; fax: +1 414 229 6699.
E-mail addresses: dwolfram@uwm.edu (D. Wolfram), yuehua@uwm.edu (Y. Zhao).

http://dx.doi.org/10.1016/j.joi.2014.08.003
1751-1577/

(Ding, Chowdhury & Foo, 2000), information systems (Marion, Wilson & Davis, 2005), medical informatics (Morris & McCain, 1998), neural networks (McCain, 1998), and semiconductor research (Tsay, Xu & Wu, 2003). Much of the literature has focused on individual fields or specializations. One reason co-citation studies tend to focus on individual fields is that the journal–journal co-citation matrix that emerges when multiple disciplines are employed can be quite sparse (Boyack, Klavans, & Börner, 2005). Co-citation data can also be labor-intensive to extract and are not easily available through citation database sources such Thomson Reuters Web of Science (WoS) without downloading all references from a corpus of articles.

Citation-based data may also be used to identify disciplinary or specialization affiliations for journals. This is particularly important for informetrics studies, where the misclassification of journals may affect the ranking of journals within a given field. Leydesdorff and Cozzens (1993) explored the feasibility of delineating and attributing journals to specialties based on journal–journal citations and their changes over time. They demonstrated how the citation data could be used to construct macro-journals, consisting of aggregations of journals around a central journal. Pudovkin and Garfield (2002) developed a journal relatedness factor based on citing and cited journals. The goal of their proposed method was to help identify thematically related journals. Similarly, Glänzel and Schubert (2003) developed a three-step process for the categorization of journals that involved pre-defined categories, journal classification and article classification for articles in journals with ambiguous subject assignments based on references. More recently, Rafols and Leydesdorff (2009) compared the outcomes of two algorithms for the decomposition of large matrices against Web of Science Subject Categories and Glänzel and Schubert's categorization. The four methods they used resulted in similar map outcomes on a large scale. Leydesdorff and Rafols (2009) also investigated the relationships among 170+ Web of Science Subject Categories using a citation matrix consisting of the subject category citation frequencies. They concluded that a classification scheme could be developed using analytical arguments. Similarly, Leydesdorff and Schank (2008) visualized and animated the disciplinary ties of three seed journals over time to demonstrate relationships among journals and their interdisciplinarity. In a comparison study of different methods for assessing research fronts in biomedical literature, Boyack and Klavans (2010) compared results from co-citation analysis, bibliographic coupling, direct citation and a hybrid approach for accuracy of outcomes. They concluded that bibliographic coupling performed the best in representing the research fronts.

Co-citation analysis relies on citing articles to identify the strength of relationships between the units of interest, whether authors, papers or journals; however, it does not consider any attributes of the source of the citations – only that the citations or co-citations exist. Authors such as White (2001) and Ajiferuke, Lu, and Wolfram (2010) have called for a shift in the focus of citation-based research away from citation counts received by an author of interest to the origin of the citation and its characteristics to assess author impact from a different perspective. Earlier, White (2000) had proposed using the source of the citation to identify characteristics of the cited author's research. He introduced the idea of citation image-makers, who represent all the authors who reference a given author. Citation image-makers may also be identified for journals, where citing authors or journals can serve as the unit measure (Bonnevie-Nebelong, 2006; Bonnevie-Nebelong & Frandsen, 2006).

This research investigates the use of data derived from citing journals to assess the similarity of cited journals. The journal citation image of a target journal, which is determined by the list of journals that cite the target journal, provides an indicator of the reach of a journal. When combined with the frequencies of citation by the citing journals, the frequency distribution of citations provided by the citing journals creates a "signature" for each cited journal. These signatures may be compared using various analytical methods. One possible challenge associated with using the citing journals themselves to create a signature for a cited journal is the potentially high number of citing journals that an influential and prolific journal might attract. If the data could be combined to reduce the computational overhead associated with the comparison of journal signatures while still preserving the signature of the cited journal created by the citing journals, the similarity comparison could be simplified. Wang and Wolfram (forthcoming) proposed a method to reduce the computational overhead associated with the citing journal data. Their method of citing discipline analysis uses the subjects/disciplines assigned to the citing journal and the resulting citation frequencies of the citing disciplines to constitute the cited journal's signature.

Thomson Reuters Web of Science assigns each of the journals it indexes to one or more subject designations. There are approximately 250 "Web of Science Category Terms" (http://images.webofknowledge.com/WOKRS57B4/help/WOS/hp_subject_category_terms_tasca.html), representing disciplines and specializations. Each journal is also assigned one or more "Research Areas" (http://images.webofknowledge.com/WOKRS57B4/help/WOS/hp_research_areas_easca.html), which are at the discipline or sub-discipline level. A majority of the 151 Research Area headings (86.8%) also appear in the WoS Category Terms. The Research Areas lack topical designations within disciplines and represent a higher level of aggregation (Leydesdorff, Carley & Rafols, 2013). As an example, Computer Science has seven listings by sub-area and Materials Science has eight listings by sub-area within the WoS Category Terms, but each only has one listing in the Research Areas list. By relying on the subject designation of a citing journal to define a cited journal's signature, the size of the signature can be reduced from potentially thousands of elements (represented by journals) to no more than 151 designations (represented by subjects/disciplines). These smaller signatures may then be compared using analytical and visualization techniques to determine the journal similarity. Citing Research Areas, which we refer to as citing disciplines for simplicity, were used for the present study; however, a number of the Research Areas may represent sub-disciplines.

Wang and Wolfram (forthcoming) employed citing discipline analysis to explore journal similarity among 40 high impact journals in Information Science and Library Science (ISLS) as classified in Journal Citation Reports (JCR). They found that some of the journals classified into the ISLS category did not map in close proximity to one another based on multidimensional scaling and cluster analysis. A number of the journals included were also classified into allied fields, but did not cluster or appear in close proximity to a number of journals only classified in ISLS. The authors noted that how journals are classified