Journal of Informetrics 10 (2016) 254–272

Contents lists available at ScienceDirect

Journal of Informetrics

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

Clustering by publication patterns of senior authors in the social sciences and humanities

Frederik T. Verleysen[a,*], Arie Weeren[b]

[a] Centre for Research & Development Monitoring (ECOOM), Faculty of Social Sciences, University of Antwerp, Middelheimlaan 1, 2020 Antwerp, Belgium
[b] StatUA, University of Antwerp, Prinsstraat 13, 2000 Antwerp, Belgium

ARTICLE INFO

Article history:
Received 21 October 2015
Received in revised form
18 December 2015
Accepted 11 January 2016

Keywords:
Cluster analysis
Bibliometrics
Senior authors
Social sciences
Humanities
Publication practices

ABSTRACT

This study uses cluster analysis as a tool for mapping diversity of publication patterns in the social sciences and humanities (SSH). By algorithmic clustering of 1828 senior authors affiliated with 16 disciplines at five universities in Flanders, Belgium, based on the similarity of their publication patterns during 2000–2011, we distinguish two broad publication styles, both of which are present within each discipline. We conclude that diversity in SSH publication patterns cuts across disciplinary boundaries. Cluster analysis shows promise for application in research evaluation for the SSH.

1. Introduction

Cluster analysis (CA) is a multivariate technique for classifying similar objects into groups. A major strength of algorithmic clustering lies in its ability to unambiguously classify cases by a multitude of attributes within otherwise opaque datasets, and to plot the results in a geometric representation (Kaufman & Rousseeuw, 1990; Johnson & Wichern, 1992). As the development of classification schemes has always played an essential role in science, there are many applications of CA in various scientific fields. Unsurprisingly, in information science as well, the use of CA has been advocated (Egghe & Rousseau, 1990). Well-known applications in this field involve the analysis of bibliographic networks and the cognitive-epistemological structure of the scientific system (Small, Sweeney, & Greenlee, 1985a,b; Waltman, van Eck, & Noyons, 2010; van Eck & Waltman, 2014), or of a single field of research (Lin & Kaid, 2000; Persson, 2015). However, many more applications are conceivable. In 2005, for example, Liu, Li, Xu, and Shi (2005) have used clustering from a research evaluation perspective to identify groups of Chinese scientific research institutions. In the present article, we apply algorithmic clustering to publication patterns at the level of individual senior authors in the social sciences and humanities.

As bibliometric research of the social sciences and humanities (SSH) is growing to maturity, more attention is paid to their internal heterogeneity and dynamics, often within the context of national evaluation and/or performance-based research funding systems (Hicks, 2013; Engels, Ossenblok, & Spruyt, 2012; Hammarfelt & de Rijcke, 2015). Disciplinary publication cultures in the SSH vary to a considerable extent, as has been measured by the share of publication types (books in addition

* Corresponding author.
E-mail addresses: Frederik.Verleysen@uantwerpen.be (F.T. Verleysen), Arie.Weeren@uantwerpen.be (A. Weeren).

http://dx.doi.org/10.1016/j.joi.2016.01.004
1751-1577/

to journal articles), publication language (national or regional languages in addition to or instead of English), target audience (international or local, academic or non-academic), and the frequency of co-authorship. In summary, in many Western countries, most disciplines classified as social sciences tend to show a publication pattern in which international journals, the use of English and frequent co-authorship are starting to predominate, while, in comparison, most humanities disciplines remain more strongly oriented towards book publications and national journals, make use of national or regional languages, and often continue to prefer single authorship (Kyvik, 2003; Nederhof, 2006; Sivertsen, 2009; Sivertsen & Larsen, 2012; Ossenblok, Verleysen, & Engels, 2013; Verleysen & Engels, 2014; Puuska, 2014; Hammarfelt & de Rijcke, 2015).

While contrasting the publication patterns of disciplinary groups has greatly advanced the understanding of research and publication practices in the SSH, and has also contributed to the development of evaluation systems that are better adapted to a scholarly research environment (Hicks, 2013), there are limitations to this approach. Clearly, the aggregation of individual researchers' publications into statistics at the disciplinary level obfuscates intra-disciplinary diversity. The existence of formal disciplinary boundaries throughout science is the outcome of a historical process of cognitive and social structuring (Whitley, 2000; Bod, Maat, & Weststeijn, 2012), and while this process undoubtedly has had its influence on disciplinary *epistemic cultures* (Knorr Cetina, 1999), including publication practices, this in itself does not imply that boundaries between such cultures necessarily coincide with organizational demarcations. Although intra-disciplinary diversity is certainly deserving of more attention by bibliometric research on the SSH, a handful of studies on individual disciplines are available, demonstrating their cognitive fragmentation (Lin & Kaid, 2000; Ahlgren, Pagin, Persson, & Svedberg, 2015) and/or their (related) heterogeneity in terms of publication and citation patterns (Nederhof & Noyons, 1992; Nederhof, 2011; Chi, 2015). Our present study starts from the premise that individual authors in the SSH publish their work in a more idiosyncratic way than an analysis at the aggregation level of disciplines is able to reveal.

To corroborate this in a systematic way, we apply algorithmic clustering to a bibliographic dataset on the comprehensive peer-reviewed publication output of individual senior authors affiliated with 16 SSH disciplines at five universities in Flanders, the Northern Dutch-language part of Belgium. By letting a computer algorithm autonomously form groups of authors, based on the similarity of their publication patterns and regardless of disciplinary affiliations, we demonstrate that epistemic diversity reflected by publication patterns cuts across disciplinary boundaries. From a science policy and research evaluation perspective, both method and result of our analysis offer additional insights.

2. Data

Data used in our analysis is registered in the VABB-SHW, the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (or *Vlaams Academisch Bibliografisch Bestand voor de Sociale en Humane Wetenschappen*). Introduced in 2010, the VABB-SHW comprehensively registers all peer reviewed publications since the year 2000 by researchers affiliated with 16 SSH disciplines and two general categories at the five universities in Flanders (Belgium). The VABB-SHW is used in the regional research funding model for the five universities (Verleysen, Ghesquière, & Engels, 2014).

Five publication types are registered in the VABB-SHW: journal articles, monographs, edited books, book chapters and proceedings papers. For inclusion in the funding model, a weight is attributed to each type: journal articles, edited books and book chapters all receive a weight of 1, whereas monographs have a weight of 4 and proceedings papers one of 0.5.

Two parts comprise the VABB-SHW database. The first, VABB-WoS consists of references to publications (journal articles and proceedings papers) which are also indexed in a journal and/or proceedings index of the Web of Science (WoS). VABB-WoS consists for ca. 95% of English language publications, and concentrates most of the high-profile international journals in the SSH. The second part, VABB-GP consists of references which have additionally been selected as peer reviewed by the Authoritative Panel (*Gezaghebbend Panel* or *GP*), an independent scientific board of university professors, from the whole of the five universities' non-WoS publications. VABB-GP consists for ca. 70% of publications in other languages than English, especially Dutch (Engels et al., 2012). Thus, in the results section below, with regard to journal articles and proceedings papers, a distinction is made between the subsets of VABB-WoS and VABB-GP. Book publications all stem from the VABB-GP subset.

For the present study the VABB-SHW dataset ($N = 10{,}181$ authors) was delimited to the output of the most senior authors ($n = 1828$). These are defined in terms of publication productivity, i.e. as having published at least ten weighted outputs in at least four years. Junior (i.e. less productive) authors ($n = 8353$) were excluded from the analysis, because their less numerous and more sporadic publications – typically a handful of articles or proceedings – have not yet had the chance to crystallize into a clearly discernible pattern. 6171 or 73.8% of junior authors have less than five weighted outputs in the twelve year time span. In the VABB-SHW, disciplines are defined based on the institutional affiliation of authors.

As input for the clustering algorithm (cfr. infra, Section 3), a table was constructed listing the 1828 author names, their main disciplinary affiliation, as well as 11 variables mapping their publication output in the 2000–2011 time frame. These variables belong to three groups of attributes which are known to differentiate SSH publication patterns at the disciplinary level: publication type, publication language and the share of co-authored publications. For the three VABB-SHW book publication types, combined with two publication language groups (English vs. other languages), this resulted in a subtotal of six variables, for each of which the fractional contribution to individual authors' total 12-year weighted output was calculated. For journal articles and proceedings papers, fractions were calculated based on the distinction between VABB-WoS and VABB-GP, resulting in a subtotal of four variables. The 11th variable is the fraction of weighted co-authored publications.