Journal of Informetrics 10 (2016) 1055–1078

Contents lists available at ScienceDirect

Journal of Informetrics

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

A critical cluster analysis of 44 indicators of author-level performance

Lorna Wildgaard

Royal School of Library and Information Science, Faculty of the Humanities, Copenhagen University, Birketinget 6, 2300 Copenhagen S, Denmark

ARTICLE INFO

Article history:
Received 28 April 2016
Received in revised form
20 September 2016
Accepted 20 September 2016

ABSTRACT

This paper explores a 7-stage cluster methodology as a process to identify appropriate indicators for evaluation of individual researchers at a disciplinary and seniority level. Publication and citation data for 741 researchers from 4 disciplines was collected in Web of Science. Forty-four indicators of individual researcher performance were computed using the data. The clustering solution was supported by continued reference to the researcher's curriculum vitae, an effect analysis and a risk analysis. Disciplinary appropriate indicators were identified and used to divide the researchers into four groups; low, middle, high and extremely high performers. Seniority-specific indicators were not identified. The practical importance of the recommended disciplinary appropriate indicators is concerning. Our study revealed several critical concerns that should be investigated in the application of statistics in research evaluation.

The strength of the 7-stage cluster methodology is that it makes clear that in the evaluation of individual researchers, statistics cannot stand alone. The methodology is reliant on contextual information to verify the bibliometric values and cluster solution. It is important to do studies that investigate the usefulness of statistical evaluation methodologies to help us as a community learn more about the appropriateness of particular bibliometric indicators in the analysis of different researcher profiles.

1. Introduction

"Quality nowadays seems to a large extent to be defined as productivity," wrote Arensbergen in 2014. Researchers are defined and are defining themselves in assessments in terms of their performance on bibliometric indicators of production and citation impact (Martin & Irvine, 1983). Developing indicators that most accurately capture the researcher's performance has led to an explosion in the generation of author-level indicators, (Wildgaard, 2015a; Iliev, 2014). Yet even though there is now a multitude of indicators available to apply in evaluations, it is still the "famous" ones like the h-index, h, or citations per paper, CPP, that are predominantly used. Given the development in indicator construction, lesser-known indicators have been proposed to offer contextually appropriate solutions, confirming further that the famous indicators are not the best ones to use across the board in author-level evaluations, i.a. (Harzing & Alakangas, 2016; Hicks, Wouters, Waltman, de Rijcke, & Rafols, 2015; Wildgaard, 2015a; Wildgaard, 2015b). Therefore, this paper sets out to define an objective method that will help us sort through the wealth of available indicators, to identify a set of appropriate indicators that fit individual researchers with different academic profiles.

E-mail address: lorna.wildgaard@hum.ku.dk

http://dx.doi.org/10.1016/j.joi.2016.09.003
1751-1577/

Author-level indicators produce a single numerical value, which is used to describe a combined set of publication and citation data. This value typically depicts a central position within that set of data. Indicators are then a summary statistic, which calculate different measures of central tendency using a selected set of data within a dataset or all the available data. Parameters can be defined to determine how certain characteristics relevant to the data and researcher profile are calculated, for example weighting the age of citations or adjusting for citation velocity. Logically this means that under different conditions, because data has different characteristics, some indicators are more appropriate than other indicators. We need to learn more about under what conditions certain indicators are most appropriate. This is important work. The resulting indicator values are interpreted, in evaluations, as measures of the individual's prestige, production, or quality, etc., and they are used to inform decisions about an individual's academic future and as predictors of future performance. Several core developments have heightened the need for intensive action from the bibliometric community, to recommend indicators that support progressive evaluation and support the objectives of the researcher being evaluated:

1. Bibliometrics have become central to economic, political, social and academic evaluation systems, as well as to the individual profile of the researcher (Bloch & Schneider, 2016;; Sivertsen, 2016). Performance and assessment culture has been internalized and institutionalized in university and research institutions and consequently author-level indicators are being used as (self) regulatory tools to monitor and adjust scientific activities, in attempts to optimize the effect of the researcher and their publications (Wouters, 2014; Retzer & Jurasinski, 2009). Because of the critical issues of efficiency and trust in the evaluation system, (Abramo, D'Angelo, & Rosati, 2014), experts on bibliometric indicators do not generally see author-level indices as indicators of research quality. However, socially they seem to partly function like it (van Arensbergen, 2014).
2. Given the diversity of publication and citation cultures within scientific disciplines, the usefulness of an indicator is fluid (Lancho-Barrantes, 2010). An indicator that works well for one particular community of researchers is not necessarily appropriate in another community (apples to oranges) unless well-argued scaling factors are applied when measuring and comparing (Díaz-Faes, Costas, Galindo, & Bordons, 2015); (Abramo, Cicero, & D'Angelo, 2013). As of yet, there is no unified agreement on what these scaling or normalization factors should be and the invalidity of normalization continues to be discussed (Glänzel & Moed, 2013). Thus, bibliometric evaluation of the individual remains framed by culturally influenced norms, disciplinary norms, and "ways of knowing" in the individual's specialty, which is also affected by the individual's visibility or coverage in generic citation databases Harzing and Alakangas, 2016.
3. A prerequisite of informed bibliometric evaluation is that assessors understand the mathematical construction of the indicator and understand how well the mathematical model fits the data used to compute the indicator (Glänzel, 2010). This in turn improves understanding of how the indicator on a particular individual's publication/citation dataset serves as an asset or drawback in summarizing numerically the experiences and achievements of the researcher (Abramo & D'Angelo, 2011; Sandström & Sandström, 2009). The bibliometric community has for many years warned about the volatility of bibliometric statistics at the individual level. Of particular concern is the stability of the indicators and the importance of the numbers they produce, as they are based on limited data. The indicators are only informative with great methodological care (Hicks et al., 2015; IEEE, 2013; Bach, 2011).
4. At the individual level, disciplinary and personal culture has implications for the robustness and appropriateness of the bibliometric indicator in evaluations (Glänzel & Moed, 2013). Indicator values are influenced across disciplines and within academic ranks by the age, nationality, specialty of the researcher, length of career, amount of publications, publication language, number of collaborators, and position on the author by-line (Díaz-Faes et al., 2015; Levitt & Thelwall, 2014;; Claro & Costa, 2011;; Costas, van Leeuwen, & Bordons, 2010; Vinkler, 2007; Archambault & Gagné, 2004). Not to mention the availability of publication and citation data (Meho & Yang, 2007) and method of data-collection (Retzer & Jurasinski, 2009). The aforementioned are vital, non-consistent variables that differ from discipline to discipline, researcher to researcher and their influence must not be under-estimated in a useful and insightful bibliometric evaluation.

Notwithstanding the above complexities of individual bibliometric evaluation, author-level indicators are increasing in popularity, both in invention and in use by bibliometricians, researchers and administrators. Yet it falls to the responsibility of the bibliometric community to identify the stable indicators from the volatile and the true indicators from the spurious. The methodology used to recommend indicators has to be transparent and reproducible, ensuring that the principles for recommending appropriate indicators are not cloaked in math or reliant on fuzzy data. Thus, the research questions are these:

Does cluster analysis provide a useful method to identify disciplinary appropriate author-level indicators?

Does cluster analysis provide a useful method to identify seniority appropriate author-level indicators?

This paper investigates if the applied cluster methodology actually provides an informative approach in grouping researchers based on author-level indicator values. Can we draw informed observations or are the results purely arbitrary? Cluster analysis is a process-based methodology that to give meaningful results builds on seven stages, Bacher et al., 2010. Fig. 1 illustrates these stages which include: 1) data-collection, 2) description of data, 3) presentation, calculation and statistical description of bibliometric indicators, 4) a rationalized choice and application of the cluster algorithm and clustering statistics, 5) presentation and statistical description of clusters, 6) tests of the stability and strength of the clusters and finally, 7) informed interpretation of the clusters. These stages are used to organize the paper as follows: Stages 1 and 2 introduce the Methodology section, followed by a brief presentation of the bibliometric indicators, including their calcu-