Journal of Informetrics 11 (2017) 748–765

Contents lists available at ScienceDirect

Journal of Informetrics

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

Disaggregated research evaluation through median-based characteristic scores and scales: a comparison with the mean-based approach

Gabriel-Alexandru Vîiu

Research Institute of the University of Bucharest – Social Sciences Division, University of Bucharest, Panduri 90, Bucharest 050663, Romania

ARTICLE INFO

Article history:
Received 11 March 2017
Received in revised form 17 April 2017
Accepted 17 April 2017

Keywords:
Characteristic scores and scales (CSS)
Scientometrics
Research evaluation
Disaggregation

ABSTRACT

Characteristic scores and scales (CSS) were proposed in the late 1980s as a powerful tool in evaluative scientometrics but have only recently begun to be used for systematic, multi-level appraisal. By relying on successive sample means found in citation distributions the CSS method yields performance classes that can be used to benchmark individual units of assessment. This article investigates the theoretical and empirical consequences of a median-based approach to the construction of CSS. Mean and median-based CSS algorithms developed in the R language and environment for statistical computing are applied to citation data of papers from journals indexed in four Web of Science categories: Information Science and Library Science, Social work, Microscopy and Thermodynamics. Subject category-level and journal-level comparisons highlight the specificities of the median-based approach relative to the mean-based CSS. When moving from the latter to the former substantially fewer papers are ascribed to the poorly cited CSS class and more papers become fairly, remarkably or outstandingly cited. This transition is also marked by the well-known "Matthew effect" in science. Both CSS versions promote a disaggregated perspective on research evaluation but differ with regard to emphasis: mean-based CSS promote a more exclusive view of excellence; the median-based approach promotes a more inclusive outlook.

1. Introduction

Research evaluation and the specific instruments used in its service constitute one of the main topics of debate in contemporary academia and in higher education policy. Although motivated by many political, social and economic reasons, the increased attention towards assessing research can be explained by governments' need to monitor and manage the performance of higher education institutions, by the need to elicit accountability of these institutions to stakeholders and also by the quest to base funding decisions on objective evidence (Penfield et al., 2014). While there is consensus among academics and policy makers on the importance of competitive, high quality research for economic and social prosperity, there is no universally accepted instrument for assessing research performance, quantifying scientific impact or measuring scholarly influence. The lack of a unique answer to such scientometric problems and the idea that divergent, even contradic-

E-mail addresses: gabriel.viiu@yahoo.com, gabriel.viiu@icub.unibuc.ro

http://dx.doi.org/10.1016/j.joi.2017.04.003

tory evaluations are always possible is a recurrent theme in the recent literature – see for instance Leydesdorff et al. (2016), Waltman et al. (2016) and Abramo and D'Angelo (2015).

On a fundamental level all types of evaluation, including scientometric appraisal and the many indicators in its toolkit, hinge on the idea of aggregation and on the specific form that the aggregative process takes. The way in which specific informational inputs are combined to yield evaluative outcomes is critical. This fact has sparked an already long debate in scientometrics, particularly in the wake of the *h*-index (Hirsch, 2005) and its many variants like the *g*-index (Egghe, 2006) which, in essence, only offer an alternative aggregation of the underlying citation data. The moral of the continuing debates surrounding the *h*-index, of the separate debates around the journal impact factor as well as the moral of the more recent discussions hosted in the pages of this very journal regarding size-independent indicators versus efficiency indicators (Abramo and D'Angelo, 2016) is that in order to be confident in the outcomes of an evaluation it is crucial to be confident in the instrument used to conduct it. As aggregated scientometric indicators have become more important within national assessment processes and international university rankings, their properties, advantages and limitations have attracted increased attention, and the meaningful use of citation data has become a critical issue in research evaluation and in policy decisions (van Raan, 2005).

There seems to be an indisputable consensus in the scientometric community regarding the fact that aggregated indicators are inadequate for the purpose of research evaluation since each indicator, taken separately, can only provide a partial and potentially distorted view of the performance attained by a specific unit of assessment (Hicks et al., 2015; Moed and Halevi, 2015; Van Leeuwen et al., 2003; van Raan, 2006; Vinkler, 2007). This wisdom has been affirmed a fortiori following the introduction of the Hirsch index in 2005 and the wave of Hirsch-type indicators (Bornmann et al., 2011; Schreiber, 2010) that were subsequently proposed as improvements. The overt consensus regarding the rejection of single-number indicators such as the *h*-index has as its corollary an implicit consensus around a more general principle: when faced with the option between an aggregated approach and a *disaggregated* approach to research evaluation the latter is to be preferred to the former. In other words, one should use research evaluation instruments that discard as little information as possible and offer a wide and comparatively rich picture of performance.[1] One of the contemporary research evaluation instruments that adhere to these desiderata is given by characteristic scores and scales (CSS) for scientific impact assessment (Glänzel and Schubert, 1988; Schubert et al., 1987) which represent an effort towards achieving a multi-dimensional, disaggregated perspective regarding research performance.

The CSS method was proposed in the late 1980s to assess the eminence of scientific journals on the basis of the citations received by the articles they publish and its cornerstone idea is that of allowing a parameter-free characterization of citation distributions in such a way that impact classes are defined recursively by appealing to successive (arithmetic) means found within a given empirical distribution. The approach is highly relevant to scientometric evaluation because it addresses one of the fundamental problems associated with the adequate statistical treatment of citation data – the skewness of science (Albarrán et al., 2011; Seglen, 1992) which makes analysis through standard statistical practice difficult and potentially biased.

The aim of this article is to explore a theoretically grounded proposal to modify CSS by changing the reference thresholds used in this evaluative instrument from arithmetic means to medians. To the knowledge of the author this possibility has only been noted in a single previous study (Egghe, 2010b) where it received only a formal, theoretical exploration in a continuous Lotkaian framework. All empirical studies devoted to the application of CSS (see Section 2.1) have so far relied on the original mean-based approach. As a result, to date there are neither empirical analyses that leverage median-based CSS, nor factual comparisons of any results produced by this instrument with results relying on the mean-based approach. This article addressed these knowledge gaps by examining both mean and median-based CSS in an application to citation data of journals indexed in the Web of Science categories Information Science and Library Science, Social Work, Microscopy and Thermodynamics. The article also offers a practical implementation of the CSS algorithms in the freely available R language and environment for statistical computing. More generally, the article argues in favor of a disaggregated, inclusive approach to research evaluation and performance assessment.

The article is structured as follows: Section 2 presents the CSS mechanism in more detail, reviews the state of the art with regard to the use of this instrument and puts forward the arguments that justify the need for the alternative, median-based approach; this section also examines the theoretical implications of this shift and provides information on the data used in the empirical investigation together with adjacent methodological notes. Section 3 presents the comparative results of the empirical analyses and highlights the distinctiveness inherent in the application of median-based CSS to citation data. Section 4 summarizes the results and provides a few concluding remarks.

[1] Note for instance that widely used indicators like the *h*-index and impact factor discard very much information specifically due to their underlying aggregation.