Journal of Informetrics 11 (2017) 32–45

Contents lists available at ScienceDirect

Journal of Informetrics

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

Regular article

A comparison of the Web of Science and publication-level classification systems of science

Antonio Perianes-Rodriguez[a], Javier Ruiz-Castillo[b,*]

[a] Departamento de Biblioteconomía y Documentación, Universidad Carlos III, SCImago Research Group, Spain
[b] Departamento de Economía, Universidad Carlos III, Spain

ARTICLE INFO

Article history:
Received 5 September 2016
Accepted 24 October 2016
Available online 18 November 2016

Keywords:
Classification systems of science
Journal-level *versus* publication-level systems
Field-normalization

ABSTRACT

In this paper, we propose a new criterion for choosing between a pair of classification systems of science that assign publications (or journals) to a set of clusters. Consider the standard target (cited-side) normalization procedure in which cluster mean citations are used as normalization factors. We recommend system A over system B whenever the standard normalization procedure based on system A performs better than the standard normalization procedure based on system B. Performance is assessed in terms of two double tests – one graphical, and one numerical – that use both classification systems for evaluation purposes. In addition, a pair of classification systems is compared using a third, independent classification system for evaluation purposes. We illustrate this strategy by comparing a Web of Science journal-level classification system, consisting of 236 journal subject categories, with two publication-level algorithmically constructed classification systems consisting of 1363 and 5119 clusters. There are two main findings. Firstly, the second publication-level system is found to dominate the first. Secondly, the publication-level system at the highest granularity level and the Web of Science journal-level system are found to be non-comparable. Nevertheless, we find reasons to recommend the publication-level option.

1. Introduction

For many theoretical and practical purposes in the evaluation of research activities in current society, we need a *classification system* of science, that is, an assignment of individual publications (or journals) to a set of clusters or sub-fields. As is well known, the choice of a classification system remains an open question in Scientometrics (see *inter alia* Boyack, Klavans, Börner, 2005 ; Leydesdorff, 2004, 2006; Small, 1999; Leydersdorff & Rafols, 2009, as well as the references they contain). Together with the classification systems included in Thomson Reuters' Web of Science (WoS) and Elsevier's Scopus databases, there are a number of interesting proposals suggested by individual researchers (see *inter alia* Börner et al. (2012), as well as the references in Waltman & Van Eck, 2012).[1]

In this paper, we contribute to the search for an appropriate classification system begun in Ruiz-Castillo and Waltman (2015). The main idea is the following. Given a classification system, it is well known that differences in production and

* Corresponding author.
E-mail address: jrc@eco.uc3m.es (J. Ruiz-Castillo).

[1] The historical background section of Klavans and Boyack (2015) contains an illuminating guide to the literature on the construction of "research fronts" and publication-level or journal-level "taxonomies" (or classification systems).

http://dx.doi.org/10.1016/j.joi.2016.10.007

citation practices preclude the direct comparison of the raw citations received by any pair of publications belonging to different clusters. In this situation, one way to evaluate the performance of research units working in different clusters begins with the normalization of the original citation counts. Consider the standard target (or cited-side) normalization procedure in which normalized citation scores in every cluster are equal to the original raw citations divided by the cluster mean citation. If one could establish that the standard normalization procedure based on system A performs better than the standard normalization procedure based on system B, then we would recommend the use of system A over system B. In this paper, we use the graphical and numerical methods introduced in Li and Ruiz-Castillo (2013) for that purpose. Following up Sirtes (2012) and Waltman & Van Eck (2013a), these methods include the possibility of using a third, independent classification system C for the evaluation of any pair of systems A and B.

We illustrate this strategy by comparing a Web of Science (WoS) journal-level classification system, consisting of 236 journal subject categories (or simply categories hereafter), with two alternatives arising from the publication-level algorithmic methodology introduced in Waltman & Van Eck (2012) that classifies individual publications into clusters solely based on direct citations between them.

In practice, the choice of the WoS classification system is often made because, together with the Scopus system, it is readily available. However, a number of studies question the appropriateness of this system for normalization purposes.[2] Among the publication-level alternatives, Klavans and Boyack (2015) conclude that classification systems based on direct citation using the Waltman & Van Eck (2012) methodology are more accurate than classification systems based on bibliographic coupling or co-citation. Ruiz-Castillo and Waltman (2015) apply the publication-level algorithmic methodology introduced by Waltman & Van Eck (2012) to a WoS dataset consisting of 9.4 million publications from the 2003–2012 period. They construct a sequence of twelve independent classification systems, in each of which the same set of publications is assigned to an increasing number of clusters. In this paper, we use the versions obtained at granularity levels 6 and 8 (the G6 and G8 classification systems hereafter) consisting of 1363 and 5119 clusters, respectively. Therefore, we have three standard normalization procedures based on three classification systems, and two interesting comparisons to make: the G6 versus the G8 system, and the winner in this contest versus the WoS system.

We focus on the 3.6 million articles published in the 2005–2008 period, and the citations they receive during a five-year citation window for each year in that period. However, two complications should be noted. Firstly, approximately 45% of the articles in the WoS system are assigned to two or more categories up to a maximum of six. To deal with this problem, we adopt a multiplicative strategy in which articles classified into several categories are wholly counted in all of them. In this way, the space of articles is expanded as much as necessary beyond the initial size. Secondly, since the methods for the evaluation of normalization procedures in Li and Ruiz-Castillo (2013) require the partition of cluster (and category) citation distributions into, say, 100 quantiles, we eliminate clusters (and categories) with less than 250 articles.

The remainder of this paper consists of four Sections. Section II presents the data. Section III serves two purposes: the description of the graphical and numerical methods for the comparison of the performance of two normalization procedures based on two different classification systems, and the application of these methods to the comparison between the G6- and G8-normalization procedures. Since the G8 system performs better than the G6 system, Section IV compares the performance of the WoS- and the G8-normalization procedures. Finally, Section V discusses the results and offers some concluding comments. The Supplementary material (SM hereafter) includes some descriptive statistics, a numerical example illustrating the various citation distributions used in the paper, and a method to evaluate the differences between a pair of classification systems in different circumstances.

2. Data

Our dataset results from the application of a publication-level algorithmic methodology to 9,446,622 distinct articles published in 2003–2012. Publications in local journals, as well as popular magazines and trade journals, have been excluded (for details, see Ruiz-Castillo and Waltman, 2015). We work with journals in the sciences, the social sciences, and the arts and humanities, although many arts and humanities journals are excluded because they are of a local nature.

In this paper, we focus on the set of 3,614,447 distinct articles published in the period 2005–2008, and the 31,290,249 citations received by these articles during a five-year citation window for each year in that period. The percentage of distinct articles assigned to two or more categories is very high: 45.2% of the total (for details, see Section A in the SM).[3] To deal with the problem of multiple assignment of articles to WoS categories, we adopt a multiplicative strategy in which articles classified into several sub-fields are wholly counted in all of them. In this way, the space of articles is expanded as much as necessary beyond the initial size. As a matter of fact, the total number of articles in what we call the *extended count* for the 236 WoS categories is 5,944,533, or 64.5% larger than the original dataset. The number of citations in the extended count is 50,307,834, or 60.8% more than in the original dataset.

[2] See *inter alia* Neuhaus and Daniel (2009) for Chemistry and related fields, Van Leeuwen and Calero-Medina (2012) for Economics & Business, Van Eck, Waltman, Van Raan, Klautz, and Peul (2013) for Clinical and Basic Medical Research, and Leydesdorff and Bornmann (2015) and Wang and Waltman, 2016 for Library and Information Science, and Science & Technology Studies.

[3] This amount is of the same order as that found in other comparable datasets. For example, this percentage is 42% in the WoS dataset of 3.7 million articles published in the 1998–2002 period that was used in Albarrán, Crespo, Ortuño, and Ruiz-Castillo, 2011a.