Contents lists available at ScienceDirect

Journal of Informetrics

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

Individual and field citation distributions in 29 broad scientific fields[☆]

Javier Ruiz-Castillo [a,*], Rodrigo Costas [b]

[a] *Departamento de Economía, Universidad Carlos III of Madrid, Spain*
[b] *Centre for Science and Technology Studies, Leiden University, The Netherlands*

ARTICLE INFO

ABSTRACT

Using an initial dataset consisting of 18.5 million distinct authors and 15 million distinct articles published in the period 2000–2016, which are classified into 29 broad scientific fields, we search for regularities at the individual level for very productive authors with citation distributions of a certain size, and for the existence of a macro-micro relationship between the skewness of a scientific field citation distribution and the characteristics of the individual citation distributions of the authors belonging to the field. Our main results are the following three. Firstly, although the skewness of individual citation distributions varies greatly within each field, their average skewness is of a similar order of magnitude in all fields. Secondly, as in the previous literature, field citation distributions are highly skewed and the degree of skewness is very similar across fields. Thirdly, the skewness of field citation distributions is essentially explained in terms of the average skewness of individual authors, as well as individuals' differences in mean citation rates and the number of publications per author. These results have important conceptual and practical consequences: to understand the skewness of field citation distributions at any aggregate level we must simply explain the skewness of the individual citation distributions of their very productive authors.

1. Introduction

At any aggregation level, bibliometric studies using citation counts may reveal statistically significant macro-patterns in the communication process that cannot be seen from the limited perspective of the individual researcher in peer review exercises. Given a classification system of publications in the periodical literature into a set of scientific fields, field citation distributions consist of the citation counts for all publications in each field. Similarly, an individual citation distribution consists of the citation counts for an author's publications. In this paper, we search for regularities at the level of individual authors, and the nature of the macro-micro relationship between a field citation distribution and the individual citation distributions of the authors in the field.

Information available for large datasets for journal-based or publication-level classification systems indicates that field citation distributions are typically highly skewed in the sense that a large proportion of articles receives no or few citations while a small percentage of them account for a disproportionate amount of all citations. Furthermore, in spite of

---

wide differences in production and citation practices, the degree of skewness is very similar across fields at very different aggregation or granularity levels (Albarrán & Ruiz-Castillo, 2011; Albarrán, Crespo, Ortuño, & Ruiz-Castillo, 2011; Glänzel, 2007; Li, Castellano, Radicchi, & Ruiz-Castillo, 2013; Radicchi, Fortunato, & Castellano, 2008; Radicchi & Castellano, 2012; Ruiz-Castillo & Waltman, 2015; Schubert, Glänzel, & Braun, 1987). In so far as field citation distributions consist of articles published by individual authors, the three following questions naturally arise.

Firstly, which are the basic characteristics of individual citation distributions? In particular, are they typically as skewed as field citation distributions, or are they normally, uniformly, or otherwise symmetrically distributed? Alternatively, are authors so different that it is impossible to assign them any systematic pattern at all? Clearly, independently of the conceptual interest of answering this first question, the existence of a typical pattern for individual citation distributions in a given field would facilitate the choice of reference standards for the citation-impact assessment of individual authors as recently suggested, for example, in Thijs, Debackere, and Glänzel (2017).

Secondly, is there any relationship between the characteristics of individual citation distributions in a given field and the characteristics of the field citation distribution of the articles they publish? In the context of science as a system of highly interconnected entities at different levels (individual researchers, research groups, university departments, research institutes, universities), Costas, Bordons, van Leeuwen, and van Raan (2009) have emphasized the importance in large networked systems of the relations between large-scale attributes and local patterns (i.e. between field and individual citation distributions in our case). More generally, Katz (2016) views the global research system as a complex innovation system exhibiting a variety of scale-invariant properties that are statistically similar at many levels of observation. Costas et al. (2009) study the scaling relationship between the number of citations and the number of scientific publications. Specifically, they investigate whether the scaling behavior identified at the research group level (van Raan, 2006a, 2006b, 2008) is also observed at the individual level. As for Katz (2016), he studies scale-invariant correlations between the growth of impact and size over time, and between impact and size across fields and sub-fields at a point in time. In this paper, we investigate the possibility of explaining the skewness of field citation distributions in terms of the characteristics of individual citation distributions. For assessing the skewness of citation distributions, we use the Characteristic Scores and Scales (CSS hereafter) technique for grouping ranked observations into ranked-specific categories (Glänzel & Schubert, 1988; Schubert et al., 1987).

Thirdly, is the macro-micro relationship between the characteristics of field and individual citation distributions common to all sciences, or is the authors' research experience quite different in more basic or more applied fields, in fields with a high or a low citation-density, or in the natural, the engineering and the social sciences?

These are key questions for understanding the communication process in any science. However, the systematic study of the characteristics of individual citation distributions has traditionally been hampered by the lack of appropriate information. In this paper, we have largely overcome this difficulty by constructing a large dataset along the lines initiated in Ruiz-Castillo and Costas (2014a) –RCC hereafter. Our initial dataset consists of 15 million distinct articles indexed by Clarivate Analytics, formerly the IP & Science business of Thomson Reuters, and published by 18.5 million distinct authors in the period 2000-2016. Citations of articles published in a given year are recorded up to the year 2016 in a variable citation window.

It should be noted that, in order to study the skewness of entire citation distributions at the individual level, we must ignore authors with few publications. Thus, as in Thijs et al. (2017), we must restrict our attention to researchers with a citation distribution of a certain size. Specifically, we focus on *very productive* authors with a number of publications above a certain relative benchmark that takes into account that the average number of articles per author varies widely across fields. We also consider *merely productive* authors, defined as those who publish at least five articles during our 16-year period. On average over all fields, these two types of productive authors only represent 5.2% and 9.4% of the population but are responsible for 38.0% and 47.9% of all publications.

The remainder of the paper is organized into five Sections and an Appendix A. Section 2 presents the data, the notation, some descriptive statistics, and a brief description of the CSS method. Section 3 contains the within- and between-field results concerning individual citation distributions among very productive authors. Section 4 presents the within- and between-field results concerning the macro-micro relationship between field and individual citation distributions with the help of some illustrative examples presented in the Appendix A. Section 5 discusses the main findings of the paper, while Section 6 offers some concluding comments.

In order to facilitate the reading of the paper, three issues have been relegated to a Supplementary Material section. In Part 1 of the Supplementary Material, we assess the reliability of our dataset by comparing some of its key characteristics with those of the RCC dataset. Part 2 of the Supplementary Material is devoted to the following problem. We solve the assignment of individual responsibility in cases of co-authorship in a multiplicative manner. However, previous research on field citation distributions need not contend with this problem. This means that the size of field citation distributions in the two cases are very different. Fortunately, in Part 2 of the Supplementary Material we establish that the characteristics of field citation distributions are independent of the co-authorship problem. Finally, in Part 3 of the Supplementary Material we study the robustness of our results in Section 3 for individual citation distributions when we consider merely productive authors.