Journal of Informetrics 12 (2018) 315–329

Contents lists available at ScienceDirect

Journal of Informetrics

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

Research assessment by percentile-based double rank analysis

Ricardo Brito[a,*], Alonso Rodríguez-Navarro[a,b]

[a] *Departamento de Estructura de la Materia, Física Térmica y Electrónica and GISC, Universidad Complutense de Madrid, Plaza de las Ciencias 3, 28040, Madrid, Spain*
[b] *Departamento de Biotecnología-Biología Vegetal, Universidad Politécnica de Madrid, Avenida Puerta de Hierro 2, 28040, Madrid, Spain*

ARTICLE INFO

Article history:
Received 4 December 2017
Received in revised form 26 January 2018
Accepted 26 January 2018
Available online 5 February 2018

Keywords:
Research assessment
Bibliometric analysis
Citation distribution
Double-rank analysis
Percentile distribution

ABSTRACT

In the double rank analysis of research publications, the local rank position of a country or institution publication is expressed as a function of the world rank position. Excluding some highly or lowly cited publications, the double rank plot fits well with a power law, which can be explained because citations for local and world publications follow lognormal distributions. We report here that the distribution of the number of country or institution publications in world percentiles is a double rank distribution that can be fitted to a power law. Only the data points in high percentiles deviate from it when the local and world μ parameters of the lognormal distributions are very different. The likelihood of publishing very highly cited papers can be calculated from the power law that can be fitted either to the upper tail of the citation distribution or to the percentile-based double rank distribution. The great advantage of the latter method is that it has universal application, because it is based on all publications and not just on highly cited publications. Furthermore, this method extends the application of the well-established percentile approach to very low percentiles where breakthroughs are reported but paper counts cannot be performed.

1. Introduction

Research assessment is an absolute requirement to perform a competent research policy. States and private institutions invest large amounts of funds in research, and society and private investors must know the efficiency of these investments by evaluating research outputs (Garfield & Welljams-Dorof, 1992; Martin & Irvine, 1983; Martin, 1996). In the case of applied research directly focused on the improvement of products or services these outputs have many possibilities of assessment attending to their economic benefits. In contrast, this assessment is much more difficult for basic research. In this case, the assessment can be analyzed in two contexts: the achievement of discoveries and scientific advancements, and the economic benefits as in applied research. However, the latter neither can be easily established nor is the only target of basic research (Bornmann, 2012; Salter & Martin, 2001); even the method that should be applied to this economic analysis is under debate (Abramo and D'Angelo, 2014, 2016; Bornmann & Haunschild, 2016b; Glanzel, Thijs, & Debackere, 2016). Therefore, it seems that the best evaluation of basic research must be done by attending to its scientific achievements. However, even by focusing the assessment of basic research exclusively on these achievements, the assessment is intrinsically difficult because of the intangible nature of the product to be measured (Martin & Irvine, 1983).

---

* Corresponding author.
E-mail addresses: brito@ucm.es (R. Brito), alonso.rodriguez@upm.es (A. Rodríguez-Navarro).

https://doi.org/10.1016/j.joi.2018.01.011

Scientific publications are tangible and easily measured. However, although scientific achievements are communicated in publications not all publications communicate real scientific advances. In fact, a large proportion of the published research is “normal science” (Kuhn, 1970) that supports real achievements, but a very low proportion of all publications reports important achievements.

As a consequence of the described needs and difficulties, in the last twenty years, there has been a Cambrian explosion of metrics (van-Noorden, 2010) or metric tide (Wilsdon et al., 2015). In this scenario, it has been suggested that no more metrics should be added unless their added value is demonstrated (Waltman, 2016). Many of these metrics are based on the number of publications, but, using a sports simile, counting publications in research is somewhat like counting the kicks in European football rather than counting the goals (Rodriguez-Navarro & Narin, 2017). The weakness of this simile is that football goals are easily recognizable but this easiness does not apply to scientific achievements. Therefore, many metrics and indicators “are based on count what can be easily counted rather than what really counts” (Abramo and D’Angelo, 2014, p. 1130). In fact, 45 years ago, Francis Narin stated that “the relationship between bibliometric measures and other measures may only be validated using a “rule of reason approach” (Narin, 1976, p. 82), which explains the causes for a more recent feeling of Harnad (2009, p. 149): “so we have thus far been rather passive about the validation of our scientific and scholarly performance metrics, taking pot-luck rather than systematically trying to increase their validity, as in psychometrics.”

Citation analysis is apparently the solution for grading the importance of results of research, because citation counts seem to correlate with expert assessments (a review of old literature is in Narin, 1976; examples of more recent publications are: Aksnes & Taxt, 2004; Allen, Jones, Dolby, & Walport, 2009; Rinia, van-Leeuwen, van-Vuren, & van-Raan, 1998). However, the debate is still open (Adler, Ewing, & Taylor, 2009; MacRoberts & MacRoberts, 1989, 1996) and the conceptual clarity of citation analysis has been questioned (Martin & Irvine, 1983), because it possibly reflects “impact” or “influence” but the relationship of these concepts with “quality”, “importance,” or “scientific advance” is less clear. In any case, although citation counts correlate with certain dimensions of research assessment they do not measure it, which implies that it cannot be applied to low aggregation levels: individual researchers or small groups (Allen et al., 2009; Ruiz-Castillo, 2012; van-Raan, 2005).

Another difficulty of citation analysis is the skewed distribution of publications attending to the number of citations (Albarrán, Crespo, Ortuño, & Ruiz-Castillo, 2011; Seglen, 1992), which makes it difficult to extract relevant information for research assessment from the analysis of simple citation counting. Several approaches have been proposed to extract this information considering citation distribution (Adams, Gurney, & Marshall, 2017; Albarrán, Perianes-Rodríguez, & Ruiz-Castillo, 2015; Bornmann, Mutz, Neuhaus, & Daniel, 2008; Bornmann et al., 2013b; Bornmann & Mutz, 2014; Bornmann & Haunschild, 2016a; Glanzel & Schubert, 1988; Glanzel, Thijs, & Debackere, 2014; Leydesdorff & Bornmann, 2011; Leydesdorff, Bornmann, Mutz, & Opthof, 2011; Li, Radicchi, Castellano, & Ruiz-Castillo, 2013; Schneider & Costas, 2017), including some that specifically attend to both the number of highly and lowly cited papers (Albarrán et al., 2011a, 2011b, 2011c). All these methods have been developed under strict mathematical and statistical considerations but all have the aforementioned problem of difficult validation.

Citation analysis can be focused on counting the number of highly cited papers, which might give an estimate of the number of important scientific achievements (Aksnes & Sivertsen, 2004; Bonaccorsi, 2007; González-Betancor & Dorta-González, 2017; Martin & Irvine, 1983; Martin, 1996; Plomp, 1994; Rodríguez-Navarro, 2012; Rons, 2013; Tijssen, Visser, & van-Leeuwen, 2002). The simplicity of this idea, however, conceals many difficulties, starting with its own definition: “highly cited,” “top-cited,” “most frequently cited,” etc. (Bornmann, 2014), which implies the arbitrariness of selecting the citation level that should be used (Schreiber, 2013a) and, more importantly, with the question about whether highly cited publications really reflects high scientific influence (Waltman, van-Eck, & Wouters, 2013).

Citation counts must be field normalized (Li et al., 2013; Ruiz-Castillo & Waltman, 2015; Waltman, 2016); among the different normalization procedures that can be used there is a method of citation analysis: the percentile rank approach, which intrinsically implies normalization of the citation count data. This approach, which has advantages over other approaches, has been extensively investigated (Bornmann, 2010; Bornmann, Leydesdorff, & Mutz, 2013; Bornmann et al., 2013a; Waltman & Schreiber, 2013), and allows generating a single measure of citation impact by giving different weights to different percentile rank classes (Bornmann & Mutz, 2011; Leydesdorff & Bornmann, 2011; Leydesdorff et al., 2011; Rousseau, 2012; Bornmann, 2013).

With this same idea of obtaining a single measure of citation impact Rodríguez-Navarro (2011) used a different approach. Firstly, he focused only on the percentiles in the high-citation tail of the citation distribution, assuming that this tail contains the information to estimate the number of important scientific achievements, as described above. Secondly, he did not fix the weights for the percentile rank classes but calculated them through linear regression analysis maximizing the correlation of the single measure with the number of Nobel Prize achievements in several high-level research institutions and advanced countries. The resulting index showed high correlation with the number of Nobel Prize achievements and with the articles published in *Nature* and *Science*. Interestingly, a further study of this approach showed that its success occurred because the upper tails of the citation distributions across countries and institutions do not deviate very much from a power law, independent of whether other functions might explain more accurately tail distribution (Brzezinski, 2015; Katz, 2016; Price, 1976; Ruiz-Castillo, 2012). The power law adjusted to the tail allows estimating the frequency of very highly cited papers or the likelihood of publishing them (Rodríguez-Navarro, 2016).

This finding, however, was more conceptual than useful for research assessment. The difficulty lies in the fact that the proportion of publications that can be treated as a power law in the upper tail can be very low (Brzezinski, 2015; Ruiz-Castillo,