Journal of Informetrics 10 (2016) 183–199

Contents lists available at ScienceDirect

Journal of Informetrics

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

Ranking authors using fractional counting of citations: An axiomatic approach☆

Denis Bouyssou[a,*,1], Thierry Marchant[b,1]

[a] CNRS (LAMSADE, UMR 7243) & Université Paris Dauphine, Place du Maréchal de Lattre de Tassigny, F-75 775 Paris Cedex 16, France
[b] Ghent University, Department of Data Analysis, H. Dunantlaan, 1, B-9000 Gent, Belgium

ARTICLE INFO

Article history:
Received 26 October 2015
Received in revised form
16 December 2015
Accepted 16 December 2015

Keywords:
Bibliometric ranking of authors
Field normalization
Fractional counting of citations
Axiomatization

ABSTRACT

This paper analyzes from an axiomatic point of view a recent proposal for counting citations: the value of a citation given by a paper is inversely proportional to the total number of papers it cites. This way of fractionally counting citations was suggested as a possible way to normalize citation counts between fields of research having different citation cultures. It belongs to the "citing-side" approach to normalization. We focus on the properties characterizing this way of counting citations when it comes to ranking authors. Our analysis is conducted within a formal framework that is more complex but also more realistic than the one usually adopted in most axiomatic analyses of this kind.

1. Introduction

With the increasing use of bibliometric indices for evaluation and monitoring purposes, there are many situations in which one has to compare papers, authors, research units, journals belonging to different fields of research (see, e.g., Ruiz-Castillo & Costas, 2014). But each field of research has its own "citation culture". Without trying to be exhaustive, elements of the culture of a field include: (i) its more or less complete coverage by bibliometric database, (ii) its growth rate, (iii) its tendency to import citations from or to export citations to other fields, (iv) its size, (v) its publication intensity, i.e., the typical number of papers published per year, (vi) its citation intensity, i.e., the typical length of the list of references in a publication, (vii) the tendency to publish research in other outlets than peer-reviewed journals, (viii) the existence of journals specialized in publishing review papers, (ix) its co-authorship practices, (x) the length of the period during which a paper published is usually cited.

This makes the problem of "normalization" between fields a difficult but important challenge: it is well-known that comparing citations counts in Mathematics with citations counts in Biology is inherently difficult (Amin & Mabe, 2000). Although this is an old theme (Schubert & Braun, 1986, 1996) in the bibliometric literature, there has been a recent increase in the literature devoted to this question. Two main paths have been followed to tackle the question.[2]

☆ We thank Michel Zitt for his comments on an earlier draft of this text. We are most grateful to Ludo Waltman for his detailed comments and his very useful pointers to the literature. Two anonymous referees of this journal made useful comments. The usual caveat applies.

* Corresponding author. Tel.: +33 1 44 05 48 98; fax: +33 1 44 05 40 91.
E-mail addresses: bouyssou@lamsade.dauphine.fr (D. Bouyssou), thierry.marchant@UGent.be (T. Marchant).

[1] These authors have equally contributed to the paper.

[2] As pointed out to us by a referee, the work of Kaur, Ferrara, Menczer, Flammini, and Radicchi (2015), extending Crespo, Ortuño-Ortí n, and Ruiz-Castillo (2012), can be seen as a third path to tackle the question. It is based on the comparison between a given set of papers and carefully chosen sets of papers of the same size, obtained via simulation. This path concentrates on the "quality" of a set of papers of a given size. It should be adapted before it can be used for the comparison of sets of papers of different sizes, as we do here.

http://dx.doi.org/10.1016/j.joi.2015.12.006
1751-1577/

The first path (sometimes called "a posteriori" normalization or "cited-side" normalization) consists in comparing citation counts with what is observed (usually, on average) in the field. This requires to have at hand a delineation of fields. This is not as obvious as it may seem at first sight (see, e.g., Small & Sweeney, 1985; van Raan, 2004, 2005; Waltman & van Eck, 2012a; Waltman, van Eck, & Noyons, 2010). For instance, some papers (authors, journals) will be at the border between two fields. Moreover, the adequate granularity of this partition into fields is not easy to determine. Even worse, important differences may exist within what is usually considered to be a field (Franceschini & Maisano, 2014; van Eck, Waltman, van Raan, Klautz, & Peul, 2013). This line of research has recently been revived by the debate about "universality of citations distributions" (Radicchi, Fortunato, & Castellano, 2008; Waltman, van Eck, & van Raan, 2012) and has generated numerous empirical studies (Abramo, Cicero, & D'Angelo, 2012; Albarrán, Crespo, Ortuño, & Ruiz-Castillo, 2011; Bornmann & Daniel, 2009; Crespo, Herranz, Li, & Ruiz-Castillo, 2014; Crespo, Li, & Ruiz-Castillo, 2013; Li, Radicchi, Castellano, & Ruiz-Castillo, 2013; Li & Ruiz-Castillo, 2013; Radicchi & Castellano, 2012b; Ruiz-Castillo, 2012; Ruiz-Castillo & Waltman, 2015) (see also the debate around the "crown indicator" of the Leiden group (CWTS) and the references cited in Section 6 on this subject). Nonparametric approaches based on percentiles have also been investigated (Bornmann, 2013; Bornmann, Leydesdorff, & Mutz, 2013; Leydesdorff & Bornmann, 2012; Leydesdorff, Bornmann, Mutz, & Opthof, 2011). All approaches relating to this first path require to be able to allocate papers to fields.

The second path was suggested by Zitt and Small (2008) and Zitt (2010, 2011) for ranking journals (see also Pinski & Narin, 1976; Small & Sweeney, 1985; Zitt, Ramanana-Rahary, & Bassecoulard, 2005). It is referred to as "citing-side", "a priori", or "source" normalization. Compared with the above approaches, its does not require the a priori division of research into fields. The normalization is performed based on the "referencing behavior of citing publications or citing journals" (Waltman & van Eck, 2013a, p. 834). Following Waltman and van Eck (2013a, 2013b), who reformulate proposals often made for journals in terms of publications, it is possible to distinguish three main techniques belonging to this second path.

This first one is similar to the audience factor proposed by Zitt and Small (2008). The citations emitted by a paper have a weight that is inversely proportional to the average number of citations emitted by papers published in the same journal and in the same year.

The second technique in this category consist in the "fractional counting of citations" proposed in Leydesdorff and Opthof (2010a, 2010b) and Glänzel, Schubert, Thijs, and Debackere (2011). For each paper p citing a paper q, the "value" of this citation is inversely proportional to the total number of citations made by the paper p. Leydesdorff and Opthof (2010a) suggest using all references for the computation of this index. Waltman and van Eck (2013b) and Waltman and van Eck (2013a) use only "active" references, i.e., references pointing to a paper in the database within a given time-window. They argue that, since the coverage of fields in bibliometric database is uneven, considering all references may penalize fields with a low coverage.

The third technique is similar to the Source Normalized Impact per Paper (SNIP) proposed by Moed (2010) and revised in Waltman, van Eck, van Leeuwen, and Visser (2013). The value of a citation emitted by a paper is inversely proportional to the product of the number of active references made by this paper with the proportion of papers published in the same journal and in the same year as this paper having at least one active reference.

This paper is only concerned with fractional counting of citations. It is described in Leydesdorff and Opthof (2010a, p. 2367, col. 2) as "simple and elegant" and in Leydesdorff, Radicchi, Bornmann, Castellano, and de Nooy (2013, p. 2300, col. 1) as a "radicalized" version of the other techniques. There has been a number of empirical studies on fractional counting of citations. Although some of them give evidence in its favor (Leydesdorff & Shin, 2011; Leydesdorff, Zhou, & Bornmann, 2013; Zhou & Leydesdorff, 2011), others lead to a less positive picture (Bornmann & Marx, 2015; Leydesdorff, Radicchi, et al., 2013; Radicchi & Castellano, 2012a; Ruiz-Castillo, 2014; Waltman & van Eck, 2013a, 2013b).

Our interest in "fractional citation counts" lies in the fact that it is a new way of counting citations that is simple and seems intuitive. A much related idea can also be found in the literature on "domination" in social networks. The "β measure" proposed in van den Brink and Gilles (2000, 2002) is indeed quite reminiscent of fractional counting with links having a dual interpretation: if a node p dominates a node q, this raises the domination score of p by a factor that is inversely proportional to the number of nodes dominating q. Moreover, this idea is also at work in PageRank (see Page & Brin, 1998: in this algorithm, the probability that a "random surfer" goes from page p to page q is inversely proportional to the number of outgoing links in page p. It should be noted that this idea and, more generally, the underlying algorithm was anticipated by Pinski & Narin, 1976, in the context of ranking journals).

Our aim in this paper is to contribute to a formal study of the properties of fractional counting of citations. This will allow us to analyze the pros and cons of this particular way to count citations from a formal point of view. As in Bouyssou and Marchant (2010), the framework that we use is rich, i.e., much richer[3] than a framework in which each author is viewed as a collection of papers each having received a certain number of citations as done, e.g., in Bouyssou and Marchant (2011b), Marchant (2009a), Marchant (2009b), Quesada (2010), or Woeginger (2008a). We will concentrate on bibliometric *rankings* instead of bibliometric *indices*. We study bibliometric rankings of *authors* instead of bibliometric rankings of *journals*, as done, e.g., in Bouyssou and Marchant (2011a), Zitt and Small (2008) and Zitt (2010, 2011). We acknowledge the fact that librarians

[3] As pointed out to us by a referee, our framework, although rich, does not include a temporal dimension that would allow us to say that a paper p is written before paper q and that, therefore, q cannot cite p.