Information Sciences 185 (2012) 18–31

Contents lists available at SciVerse ScienceDirect

Information Sciences

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

An Internet measure of the value of citations

Boleslaw K. Szymanski [a,b], Josep Lluis de la Rosa [b,c,*], Mukkai Krishnamoorthy [b]

[a] Społeczna Wyższa Szkoła Przedsiębiorczości i Zarządzania, ul. Sienkiewicza 9, 90-113 Łódź, Poland
[b] Network Science and Technology (NeST) Center, Rensselaer Polytechnic Institute, 118, 8th Street, Troy, NY, USA
[c] EASY Innovation Center, University of Girona, Campus de Montilivi, E17071 Girona, Catalonia (EU), Spain

ARTICLE INFO

Article history:
Received 7 June 2010
Received in revised form 4 February 2011
Accepted 10 August 2011
Available online 26 August 2011

Keywords:
Bibliometrics
Scientometrics
Citation analysis
Author ranking
Impact factor
PageRank

ABSTRACT

A new method for computing the value of citations is introduced and compared with the PageRank algorithm for author ranking. In our proposed approach, the value of each publication is expressed in CENTs (sCientific currENcy Tokens). The publication's value is then divided by the number of citations made by that publication to yield a value for each citation. As citations are the acknowledgements of the work by authors other than oneself (indicating that it has been useful), self-citations count as zero in acknowledged citation value. Circular citations, a generalized type of self-citation, are considered to have a reduced acknowledged citation value. Finally, we propose a modification of the h-index to define it as the largest integer such that the *i-th* publication (on the list of publications sorted by their value in CENTs) is worth more than *i* CENTs. This new index, termed the *i-index* or i^2 in short, appears to be a more precise measure of the impact of publications and their authors' productivity than the h-index.

1. Introduction

Currently, the impact of a scientific publication is often measured by the number of citations it receives. Perhaps we are suffering from an over-analysis of citations for the purposes of assessing scientists and universities productivity, impact, or prestige—the examination of citations of scientific publications has become a cottage industry in higher education. This approach has been taken to extremes both for the assessment of individuals and as a measure of the productivity and influence of entire universities or even academic systems. Pioneered in the 1950s in the United States, bibliometrics was invented as a tool for tracing research ideas, the progress of science and the impact of scientific work. First developed for the "hard" sciences, it was later expanded to include the social sciences and humanities.

The citation system was invented mainly as a way to understand how scientific discoveries and innovations are communicated and how research functions [1]. It was not initially seen as a tool for evaluating individual scientists, entire universities or academic systems. Hence, the citation system is useful for tracking how scientific ideas are propagated among researchers and how individual scientists use and communicate research findings. The use of citation analysis for the assessment of research productivity or impact questionably extends the original reasons for creating the bibliometric system. Evaluators and rankers need to go back to the drawing board in considering a reliable system for the accurate measurement of the scientific and scholarly work of individuals and institutions. The unwieldy and inappropriate use of citation analysis and bibliometrics for the evaluation and ranking of research and researchers does not serve higher education well and it entrenches existing inequalities.

* Corresponding author at: EASY Innovation Center, University of Girona, Campus de Montilivi, E17071 Girona, Catalonia (EU), Spain.
E-mail address: peplluis@eia.udg.edu (J.L. de la Rosa).

0020-0255/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2011.08.005

More recently, a new index based on citations, the h-index, has been proposed as an indicator of overall productivity and impact of the published work of a researcher [15]. The h-index of a researcher is the largest integer h such that at least h publications by this researcher have no less than h citations each. For example, an author with an h-index of 20 must have at most 20 publications with 21 or more citations and at least 20 publications with 20 citations each.[1] This index can easily be determined from the "times cited" in the Thomson ISI Web of Science or Google Scholar and it provides a metric for the author's productivity in terms of citations.

The h-index focuses more on measuring productivity than on measuring the impact and influence of the dissemination of a publication. However, some h-index variations attempt to capture the latter [4,5]. Measuring *impact* by the number of *new* authors who cite a publication appears to be a more accurate measure than measuring it by the h-index because it reflects the utility of an author's work to various individuals rather than only *the same* people. Thus, any type of direct or *indirect* self-citations should be discounted to a certain degree. Moreover, if *impact* signifies the importance of knowledge dissemination in publications citing the given publication, then citing a publication with a greater impact should in turn endow a higher impact to the cited publication.

In this work, we propose a new approach for measuring the impact of publications and compare it with an author ranking computed using the PageRank algorithm [17]. To the best of our knowledge, PageRank was originally inspired by the scientific bibliometric system (citations), but only recently has it been applied to measure the impact of journals, publications and scientists. The success of Google's method of ranking web pages has inspired numerous measures of journal impact that apply social network analysis to citation networks. Pinski and Narin [20] (predating PageRank) proposed ranking journals according to their eigenvector centrality in a citation network. Extending this idea, we propose a more accurate measure of impact than those based on the h-index. Our measurement is based not on a row citation count but on the impact of the citing publications and their distance from self-citations. Section 2 provides a precise explanation of our approach, introducing scientific currency tokens as a measure of the impact of citations. Section 3 presents an algorithm for estimating this value from a network of publications and authors connected by citations. Section 4 presents an example of how many tokens would be assigned to each citation in a network of nine citations among 6 publications by four authors. Section 5 describes an application of the PageRank algorithm to the same example followed by a comparison of the values of the citations calculated by both algorithms. Section 6 describes a method to compute the citation earnings of each author when there are multiple authors for a publication and shows an example of how to apply the h-index to CENTs instead of to citations. The conclusions and prospective future work are provided in Section 7.

2. CENTs – scientific currency tokens

We first describe the heuristics behind our model. We advocate measuring the value of each citation in sCientific currENcy Tokens (CENTs). The introduction of this currency was inspired by complementary currencies for the scientific communities proposed in [8,11,12] and also conceptualized as tokens or measure of reputation by [10,18]. Scientists are assumed to hold a new scientific currency and to have rational expectations with it [7]. The initial value of a publication is one CENT and then each raw non-self-citation received by the publication increases its value by one CENT. The initial value of each citation in a publication, called the raw value of citation and denoted $r_{i,j}$ when publication i cites publication j, is equal to the inverse of the total number of citations in publication i, denoted R_i, so $r_{i,j} = 1/R_i$ for $j \in [1, R_i]$. The raw value of citation is constant and therefore not affected by future publications.

The value of each citation of a cited publication is proportional to the value of the citing publication. Hence, every citation of a publication has a value in CENTs that is computed by multiplying the value of the publication by the raw value of this citation. Both the value of the publication and the value of citation increase with each publication that cites the publication, either directly or indirectly via a chain of citations from a new publication to the original publication.

Let P_i be the value (in CENTs) of Publication i, and, as previously noted, let R_i be the number of its citations; the value $w_{i,j}$ of citation of publication j by publication i is then:

$$W_{i,j} = P_i/R_i = P_i r_{i,j} \quad \text{CENTs} \tag{1}$$

Eq. (1) captures the notion that a high-impact publication endows its citations with high values. For example, if a publication receives 99 CENTs of citations after its publication, then its value, which includes the initial one CENT, becomes $P_i = 100$ CENTs. If this publication cites $R_i = 10$ other publications, then each of its citations is worth 10 CENTs.

There is a problem with Eq. (1) when all citations in the publication are self-citations. In such a case, the real value of the publication citations should be zero CENTs because there is no real external acknowledgement of the cited work. Another case of overvaluing arises when all 10 raw citations are of the work of another author who always cites back to the author of the citing publication. In other words, these two authors cite each other every time. Thus, they are in fact half self-citations, and intuitively the value of each citation should be halved. The following section shows how to eliminate any type of self-citation prior to the conversion of the raw value of a citation into its value in CENTs.

[1] Interestingly, a researcher with 20 publications with 100 citations each and 20 publications with 19 citations each has the same h-index as a scientist with only 20 publications with 20 citations each or a researcher with 100 publications with 20 citations each.