Journal of Informetrics 10 (2016) 392–407

Contents lists available at ScienceDirect

Journal of Informetrics

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

Evaluating paper and author ranking algorithms using impact and contribution awards

Marcel Dunaiski[a,*], Willem Visser[b], Jaco Geldenhuys[b]

[a] Media Lab, Stellenbosch University, 7602 Matieland, South Africa
[b] Department of Computer Science, Stellenbosch University, 7602 Matieland, South Africa

ARTICLE INFO

Article history:
Received 21 September 2015
Accepted 22 January 2016
Available online 16 March 2016

Keywords:
Bibliometrics
Ranking
Citation analysis

ABSTRACT

In the work presented in this paper, we analyse ranking algorithms that can be applied to bibliographic citation networks and rank academic entities such as papers and authors. We evaluate how well these algorithms identify important and high-impact entities.

The ranking algorithms are computed on the Microsoft Academic Search (MAS) and the ACM digital library citation databases. The MAS database contains 40 million papers and over 260 million citations that span across multiple academic disciplines, while the ACM database contains 1.8 million papers from the computing literature and over 7 million citations.

We evaluate the ranking algorithms by using a test data set of papers and authors that won renowned prizes at numerous computer science conferences. The results show that using citation counts is, in general, the best ranking metric to measure high-impact. However, for certain tasks, such as ranking important papers or identifying high-impact authors, algorithms based on PageRank perform better.

1. Introduction

Citation analysis is an important tool in the academic community. It can aid universities, funding bodies, and individual researchers to evaluate scientific work and direct resources appropriately. With the rapid growth of the scientific enterprise and the increase of online libraries that include citation analysis tools, the need for a systematic evaluation of these tools becomes more important.

In bibliometrics, citation counts or metrics that are based directly on citation counts are still the de facto measurements used to evaluate an entity's quality, impact, influence and importance. However, algorithms that only use citation counts or are based only on the structure of citation networks can only measure quality and importance to a small degree. What they are in fact measuring is their impact or popularity which are not necessarily related to their intrinsic quality and the importance of their contribution to the scientific enterprise. The difficulty is to obtain objective test data that can be used with appropriate evaluation metrics to evaluate ranking algorithms in terms of how well they measure a scientific entity's impact, quality or importance.

* Corresponding author.
E-mail address: marcel@ml.sun.ac.za (M. Dunaiski).

http://dx.doi.org/10.1016/j.joi.2016.01.010
1751-1577/

In Section 2 background information about the used ranking algorithms is given and related work, in which appropriate test data sets are used, is outlined. It shows that in previous research only small test data sets have been used to validate proposed ranking methods that only apply to one or two fields within computer science.

In this paper we use four different test data sets that are based on expert opinions each of which is substantially larger than those in previous research and apply them in different scenarios:

- 207 papers that won high-impact awards (usually 10–15 years after publication) from 14 difference computer science conferences are used to evaluate the algorithms on how well they identify high-impact papers.
- 464 papers from 32 venues that won best-paper awards at the time of publication are used to see how well venues predict future high-impact papers.
- From a list of 19 different awards, 268 authors that won one or more prizes for their innovative, significant and enduring contributions to science were collected. This data set is used to evaluate author-ranking algorithms.
- A list of 129 important papers, sourced from Wikipedia, is used to evaluate how well the algorithms identify important scientific work.

Therefore, this paper focuses on algorithms that are designed to measure a paper's or an author's impact and are described in Section 3. In Section 4 the MAS (Microsoft, 2013) and ACM (Association for Computing Machinery, 2014) citation data sets are described which are used for the experiments in this article. Section 5 shows the results of evaluating the various ranking algorithms with the above mentioned test data sets followed by a discussion of the results in Section 6.

2. Background information

The idea of using algorithms based on the PageRank algorithm has been applied to academic citation networks frequently. For example, Chen, Xie, Maslov, and Redner (2007) apply the algorithm to all American Physical Society publications between 1893 and 2003. They show that there exists a close correlation between a paper's number of citations and its PageRank score but that important papers, based purely on the authors' opinions, are found by the PageRank algorithm that would not have easily been identified by looking at citation counts only.

Hwang, Chae, Kim, and Woo (2010) modify the PageRank algorithm by incorporating two additional factors when calculating a paper's score. Firstly, the age of a paper is taken into consideration and secondly, the impact factor of the publication venue associated with a paper is also included in the computation. The algorithm was proposed in an article called "Yet Another Paper Ranking Algorithm Advocating Recent Publications". For brevity this algorithm is referred to as YetRank and is described in Section 3.4.

Dunaiski and Visser (2012) propose an algorithm, NewRank, that also incorporates the publication dates of papers similar to YetRank. They compare the NewRank algorithm to PageRank and YetRank and find that it focuses more on recently published papers. In addition, they evaluate the algorithms using papers that won the "Most Influential Paper" award at ICSE conferences and find that PageRank identifies the most influential papers the best.

Sidiropoulos and Manolopoulos (2005) propose an algorithm that is loosely based on PageRank. The authors call their algorithm SceasRank (Scientific Collection Evaluator with Advanced Scoring). SceasRank places greater emphasis on citations than the underlying network structure compared to PageRank. Sidiropoulos and Manolopoulos use a data set of computer science papers from the DBLP library (The DBLP Team, 2014) and compare different versions of the SceasRank algorithm with PageRank and rankings according to citation counts. They evaluate the algorithms using papers that won impact awards at one of the two venues. Firstly, papers that won the 10 Year Award (Very Large Data Base Endowment Inc., 2014) at VLDB conferences, and secondly, the papers that won SIGMOD's Test of Time Award (ACM Special Interest Group on Management of Data, 2014) are used as evaluation data to judge the ranking methods in ranking important papers. Their results show that SceasRank and PageRank perform the best in identifying these high-impact papers but that using citation counts directly performs very close to those methods. They also rank authors by using the best 25 papers of each author and use the "SIGMOD Edgar F. Codd Innovations Award" (ACM Special Interest Group on Management of Data, 2014) as evaluation data. Their results show that SceasRank performs equally well compared to PageRank and improves over the method of simply counting citations to find important authors.

The above mentioned algorithms are designed to rank individual papers and authors or venues. The ranking scores produced by these algorithms can be aggregated to author or venue entities but this entails considerable biases towards certain entities. For example, taking the average score of authors' publications favours authors unfairly that have only published a few highly cited papers which does not reflect their overall contribution or significance.

Therefore, metrics specifically designed for ranking authors are discussed in Sections 3.5 and 3.6. The metrics that are considered and evaluated are the *h*-index (Hirsch, 2005), the *g*-index (Egghe, 2006), the $i10-$index (Connor, 2011) and the Author-Level Eigenfactor metric (West, Jensen, Dandrea, Gordon, & Bergstrom, 2013).

A lot of research has been conducted on variations of PageRank to rank author entities. Fiala, Rousselot, and Ježek (2008), for example, also use the Edgar F. Codd award to evaluate their version of PageRank that includes co-authorship graph information. They find that simply using citation counts performs best at ranking authors.