Information Processing and Management 52 (2016) 374–386

Contents lists available at ScienceDirect

Information Processing and Management

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

MuICE: Mutual Influence and Citation Exclusivity Author Rank

Tehmina Amjad [a,*], Ali Daud [a], Dunren Che [b], Atia Akram [a]

[a] Department of Computer Science and Software Engineering, International Islamic University, Sector H-10, Islamabad 44000, Pakistan
[b] Department of Computer Science, Southern Illinois University, Carbondale, IL 62901, United States

ARTICLE INFO

Article history:
Received 23 December 2014
Revised 27 November 2015
Accepted 1 December 2015
Available online 31 December 2015

Keywords:
Academic social networks
Ranking of authors
Author impact
Mutual influence
Citation analysis

ABSTRACT

With constant growth in size of analyzable data, ranking of academic entities is becoming an attention grabbing task. For ranking of authors, this study considers the author's own contribution, as well as the impact of mutual influence of the co-authors, along with exclusivity in their received citations. The ranking of researchers is influenced by the ranking of their co-authors, more so if co-authors are seniors. Tracking the citations received by an author is also an important factor to measure standing of an author. This study proposes Mutual Influence and Citation Exclusivity Author Rank (MuICE) algorithm. We performed a sequence of experiments to calculate the MuICE Rank. First, we calculated Mutual Influence (MuInf) considering three different factors: the number of papers, the number of citations and the author's appearance as first author. Secondly, we computed MuICE incorporating all three factors of MuInf along with the exclusivity in citations received by an author. Empirically, it is shown that the proposed methods generate substantial results.

1. Introduction

Analysis of academic social networks has considerable applications in academic recommendation tasks. Citation, co-citation and co-authorship networks are formed when researchers cite each other's work and work in collaboration. Some general activities in academic social networks are ranking of authors (Ding, Yan, Frazho, & Caverlee, 2009; Liu, Bollen, Nelson, & Van de Sompel, 2005), expert finding (Daud, Li, Zhou, & Muhammad, 2010; Zhang, Tang, & Li, 2007), author interest finding (Daud, 2012) and author name disambiguation (Shu, Long, & Meng, 2009). This study calculates the rank of authors with respect to their mutual influence on each other and exclusivity in their received citations. A novice researcher who may get an opportunity to collaborate with a leading researcher can have more chances to prosper in the future. Considering author's own contribution in a work, as well as the impact of influence of his or her co-authors, gives a comprehensive representation of the position of an author in an academic networks.

Existing approaches find the rank of authors based on their in-links (number of nodes pointing to a node) information (Ding, 2011a; Ding et al., 2009; Gollapalli, Mitra, & Giles, 2011). The proposed method involves in-links information of a node, as well as it also considers out-links (number of nodes pointed to by) of an author. Co-author relationships from network are used to find the out-links information. This study considers a bibliographic network and presents three ways to find out the mutual influence of authors on each other, these are; with respect to the number of papers, the number of citations and the appearance of an author as first author. MuInf is based on PageRank (Brin & Page, 1998; Page, Brin, Motwani, & Winograd, 1999). We also

* Corresponding author. Tel.: +0018123254382.
E-mail addresses: tehminaamjad@iiu.edu.pk, tamjad@indiana.edu, tehminaamjad@hotmail.com (T. Amjad), ali.daud@iiu.edu.pk (A. Daud), dche@cs.siu.edu (D. Che), atiaakram81@gmail.com (A. Akram).

http://dx.doi.org/10.1016/j.ipm.2015.12.001
0306-4573/

tried to track the citations received by an author; i.e. Citations are received from how many exclusive sources. For this purpose, a bibliometric measure "f-index" (Katsaros, Akritidis, & Bozanis, 2009) is used. The term exclusivity here can be explained with the help of a simple example. Suppose there are two articles, if the first article receives three citations from three different authors and the second article receives three citations from the same author, then the first article must receive more weight as the citations received by it are from more exclusive authors. Due to the aforementioned reasons we are motivated to propose Mutual Influence and Citation Exclusivity (MuICE) Rank algorithm. Main contributions of this research are (1) finding the impact of authors based on their mutual influence on each other, with respect to the number of publications, the number of citations, the number of publications as first author and (2) ranking of authors by considering the exclusivity in their received citations along with their mutual influence.

The study conducts a detailed experimentation which shows that proposed MuICE method generates satisfying results when compared to existing methods. The rest of the paper is organized as follows: Section 2 provides a review of link analysis based ranking methods in academic social networks. Section 3 gives an overview of the existing methods used as baselines and the details of our proposed method. Section 4 describes the dataset, performance evaluation, parameter settings with results and discussions and at the end Section 5 concludes the study.

The following sets of terms are used interchangeably throughout the text, (academic social networks, bibliography network, co-author networks), (paper, article, publication) and (author, researcher) etc.

2. Related works

Author ranking methods based on analysis of link structure of a network can be classified into two groups: (1) Iterative methods and (2) Non iterative methods. This study includes the review of the former methods only.

Iterative link analysis methods execute a set of instructions iteratively for a decided number of times or until convergence of the algorithm. PageRank (Brin & Page, 1998; Page et al., 1999) is a state-of-the-art iterative link analysis algorithm and is the foundation of a large number of approaches that have been proposed for author ranking. These methods consider the authors or the publications of authors as vertices instead of pages while forming a graph. These graphs can represent co-authorship, citation or co-citation relationships. PageRank asserts a page to be significant if there are many other significant pages referencing it. The rank of a page is evenly dispersed among all the pages it is linking.

Another prevalent iterative technique based on link structure analysis is Hyperlink Induced Topic Search (HITS) which discriminates the pages as Hubs and Authorities (Kleinberg, 1999). The pages that function as directories by providing links to informative pages are known as Hubs. The pages that contain actual information and are pointed to by the hubs are called Authority pages. Fiala, Rousselot, and Ježek (2008) provided a modification of PageRank for ranking of authors in a bibliographic networks that considers the co-authorship graphs and citations. A variation of PageRank was proposed for finding experts from digital libraries to include various available facts from different objects and relations (Gollapalli et al., 2011). An alternative of centrality measure for analyzing the properties of academic networks was presented as Author-Rank algorithm (Liu et al., 2005). Author-Rank finds the impact of an author in an undirected co-authorship network contemplating collaboration frequency. Two variations of weighted PageRank were proposed for ranking of authors by Ding et al. (2009) and Yan and Ding (2011) where they studied a co-citation network and a co-authorship network respectively. Li and Tang (2008) explored the temporal dimension for the problem of expert finding. A generalization of PageRank for bibliographic networks was presented that included the time based statistics by exercising the forward and backward propagation process and combined the social networks with the random walk model. Fiala (2012) presented a method that weighs the citation between two authors on the basis that whether and when these authors have collaborated with each other. The time of publication and citation is also considered in this time-aware algorithm. Radicchi, Fortunato, Markines, and Vespignani (2009) proposed a new weighted version of PageRank, in which ranking was conducted by considering the diffusion of credits traded by the authors. Ding (2011a) measured the popularity and prestige of an author in a co-citation network. The primary focus of study was to assign high weight to the citations from prestigious authors as compared to the citations from less known authors. Wei, Barnaghi, and Bargiela (2011) semantically ranked documents and demonstrated the web surfing activities of a scholar instead of a random surfer. For this purpose, they introduced a knowledge base which contains a terminological topic ontology and academic research entities like authors, journals/conferences and papers. Ding (2011b) presented a topic sensitive extension of PageRank algorithm. The novelty was to enhance the semantics of authors ranking by introducing topic dependent weights in PageRank algorithm. Recently, a topic based model was presented for simultaneous modeling of academic entities including authors, papers and journals in a heterogeneous network (Amjad, Ding, Daud, Xu, & Malic, 2015).

3. Mutual Influence and Citation Exclusivity Author Rank (MuICE)

The concept of mutual influence of authors was presented by Li, Foo, Tew, and Ng (2009). They introduced PubRank algorithm for finding the rising stars. Using the same concept of mutual influence Daud, Abbasi, and Muhammad (2013) proposed StarRank for finding the rising star in co-authors network. This study adopts the term, mutual influence, in an intuitive manner for ranking of authors. To identify the exclusivity in citations received by an author, we used f-index (Katsaros et al., 2009), which is a bibliometric measure to evaluate an author by considering the citations received by an author. The f-index introduced the concept of co-terminal citations. Co-terminal citations are a generalization of co-citations and are introduced as an attempt to find the