PNCOIRank: Evaluating the Impact of Scholarly Articles with Positive and Negative Citations

Xiaomei Bai, Jun Zhang, Hai Cui, Zhaolong Ning and Feng Xia
School of Software, Dalian University of Technology, Dalian 116620, China
zhaolongning@dlut.edu.cn

ABSTRACT

Evaluating the impact of an article is a significant topic and has attracted extensive attention. Citation-based assessment methods currently face a limitation, i.e. the anomalous citations patterns still remain poorly understand. To remedy this drawback, we propose a Positive and Negative Conflict of Interest (COI)-based Rank algorithm, named PNCOIRank, to acquire positive COI, negative COI, positive suspected COI and negative suspected COI relationships. We investigate the citation relationships by the following scholarly factors: citing times, the interval of citing time, collaboration times, the interval of collaboration time, and team of citing authors with the purpose of weakening the COI relationships in citation network. A weighted PageRank is finally constructed and employed, with HITS algorithm to assess the impact of articles. Through experiments on American Physical Society (APS) dataset, we show that PNCOIRank significantly outperforms the existing methods in terms of recommendation intensity.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

Keywords

Citation analysis; Conflict of Interest; Team relationships

1. INTRODUCTION

Citing previous work is a cumulative nature as far as knowledge generation is concerned. However, citations are inevitably manipulated intentionally, which results in the difficulty of evaluating the impact of scholarly articles objectively. In order to seek creditworthy citations of one paper with true representative, we leverage Conflict of Interest (COI) relationships and team relationships to distinguish different citation weights. In previous studies, less attention has been paid to the various reasons behind one citation. In particular, as oppose to making citations according to the relevance of research, none-essential citations could be made due to peer-influence; for instance, authors of citing and cited paper are research collaborators with joint publications in the past, or they belong to the same affiliation. These none-essential citations can distort the impact assessment of papers, scholars, and institutions. However, limited investigations in COI and team relationships have been found in the literature.

In this poster, we propose a Positive and Negative COI-based Rank algorithm, called PNCOIRank, to identify anomalous citations to fairly quantify the impact of a paper. The main novelty of our method is to mine different categories of COI relationships and citing authors' team relationships for distinguishing the citing strengths. Furthermore, we construct a weighted PageRank algorithm, with HITS algorithm to assess scholarly publications objectively.

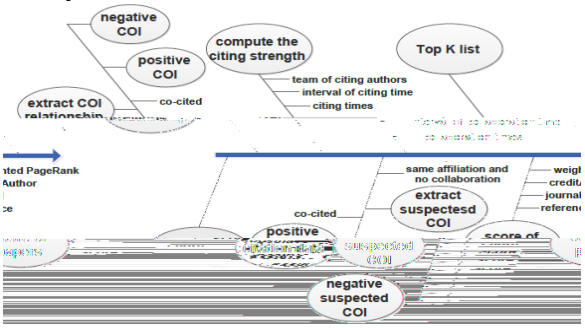

Figure 1. The structure of PNCOIRank.

2. PNCOIRANK DESIGN

As shown in Figure 1, our PNCOIRank mainly consists of four parts. *Firstly*, we identify the four categories of COI relationships. As shown in Figure 2, let p and a represent the list of papers and authors, respectively. Different types of COI are defined as below:

- *Positive/negative COI*: for scenarios where the authors a_i and a_j are co-authors, and subsequently a_i's paper p_i cites a_j's paper p_j, considered as COI. To further differentiate the citation behavior, if there's one or more papers with author(s) other than a_i and a_j cite both p_i and p_j, that is, other authors recognize the relevance between p_i and p_j, then the p_i-to-p_j citation is considered as a positive COI. Otherwise if p_j only attracts citations from collaborators, the citation is viewed as a negative COI.
- *Positive/negative suspected COI*: similar to the concept of positive/negative COI which exploits citation behavior among collaborators, suspected COI exploits the citation behaviors for authors belonging to the same affiliation. Likewise, if there are independent authors recognizing the relevance of papers by authors from the same affiliation, the citations is considered as a positive suspected COI, otherwise the citation is viewed as a negative suspected COI.

Secondly, we identify the team relationships, which are composed of citing authors of one paper. If citing authors have co-authored one or more papers, or if they belong to the same affiliation, they are considered being in a common virtual team, otherwise, they do not belong to a team. *Thirdly*, according to different COI relationships and team relationships of citing authors of one paper, citing strength is defined. If two papers have positive COI or positive suspected COI relationship, the citing strength is set as 1. If two papers have a negative COI relationship, the citing strength is defined as follows:

WWW 2016 Companion, April 11-15, 2016, Montréal, Quebéc, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2889383

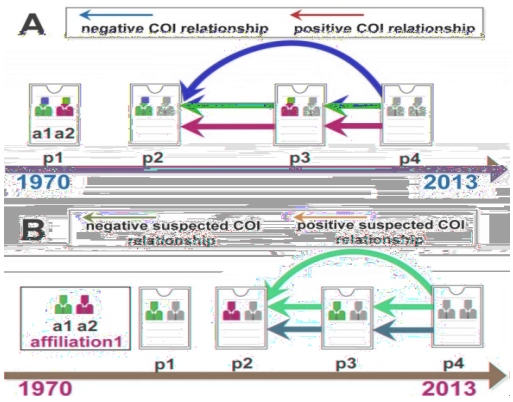

Figure 2. Identifying different COI relationships.

$$W_{i,j}^{NCite} = e^{-(\rho/\tau)\left(T^{Now}-T^{Cite}+1\right)W_{i,j}^{NCOI}} \quad (1)$$

where $W_{i,j}^{NCite}$ ranges (0,1], ρ is a constant value to indicate predefined decay parameter. τ represents the number of citing papers produced from *authors* either with prior collaborative experiences or belonging to the same affiliation. T^{Now} denotes present time, T^{Cite} denotes citing time of a paper, $T^{Now} - T^{Cite} + 1$ is the citing time interval, $W_{i,j}^{NCOI}$ is the negative COI strength between citing paper and cited paper. Its value is determined by citing times, the interval of citing time, collaboration times, the interval of collaboration time and the team of citing authors. If two papers have negative suspected COI relationship, the citing strength is defined as follows:

$$W_{i,j}^{NSCite} = e^{-(\rho/\tau)\left(T^{Now}-T^{Cite}+1\right)W_{i,j}^{NSCOI}} \quad (2)$$

where $W_{i,j}^{NSCite}$ ranges (0,1], indicating the negative suspected COI strength between citing paper and cited paper. $W_{i,j}^{NSCOI}$ is determined by above mentioned five factors. *Lastly*, based on CAJTRank algorithm [1], we construct a weighted PageRank algorithm, with HITS algorithm to rank publications and list top K publications. Meanwhile, in order to reasonably allocate the impact of one publication to different signed authors, a credit allocation algorithm is adopted [2]. The score of a publication is demonstrated as follows:

$$S(P_i) = \alpha \cdot WP(P_i) + \beta \cdot CA(P_i) + \gamma \cdot J(P_i) + \delta \cdot R(P_i) + (1-\alpha-\beta-\gamma-\delta)/n \quad (3)$$

where $S(P_i)$ is the prestige score of a manuscript. $WP(P_i)$ is the score of weighted PageRank. $CA(P_i)$ represents the credit score of the different signed authors of one paper. $J(P_i)$ and $R(P_i)$ are the prestige scores of journal and reference of one paper, respectively. Experimentally, the probability of random jump is set as 0.15,α, β, γ and δ are constants, ranging from 0 and 1, the sum of $\alpha + \beta + \gamma + \delta$ is set as 0.85.

3. EXPERIMENTS & RESULTS

We evaluate the performance of the proposed PNCOIRank on two subsets PRC and PRE of American Physical Society dataset (http://publish.aps.org), including 71287 papers. Furthermore, we compare PNCOIRank with two representative schemes: CAJTRank and FutureRank algorithms, as shown in Figure 3. A multivariate linear regression is used to estimate the parameters of the PNCOIRank, CAJTRank and FutureRank algorithms [3], and three groups of optimal parameters are estimated for the comparison of the accuracy of Recommendation Intensity (RI) [4]. According to Figure 3, we observe that the accurate rates of RI in CAJTRank and FutureRank algorithms are in between 0.557 and 0.7, and between 0.28 and 0.5 respectively. In comparison, accurate rates of RI of PNCOIRank are between 0.615 and 0.8, indicating that PNCOIRank outperforms CAJTRank and FutureRank over various K. The above observations demonstrate that PNCOIRank is better than the other two algorithms in terms of RI. By comparing CAJTRank and FutureRank, we find that the journal dimension is beneficial to improve the evaluation performance. Meanwhile, by comparing PNCOIRank and CAJTRank, the preceding results indicate that COI relationships, weighted PageRank and credit allocation algorithm of the authors can enhance the evaluation performance as well.

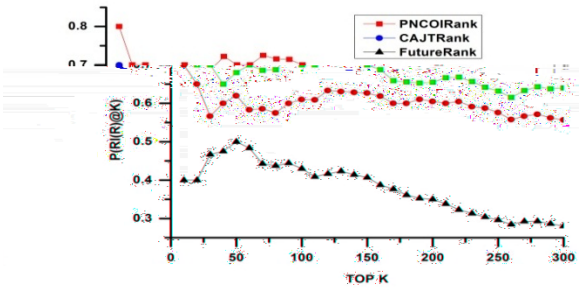

Figure 3. The probabilities of Recommendation Intensity of PNCOIRank, CAJTRank, and FutureRank algorithms.

4. CONCLUSION

This poster explores the effect of COI relationships and team relationships in evaluating the impact of scholarly publications. By joining the factors of citing times, the interval of citing time, collaboration times, the interval of collaboration time, and team of citing authors into the positive and negative COI-based evaluation method, we illustrate that COI relationships are universal phenomenon, and understanding the COI and team relationships is of great importance to effectively evaluate scientific entities, such as scholarly publications, scholars, journals and institutions. Meanwhile, our PNCOIRank algorithm can evaluate the impact of scholarly articles more objectively.

5. REFERENCES

[1] Wang, Y., Tong, Y., and Zeng, M. 2013. Ranking scientific articles by exploiting citations, authors, journals, and time information. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence* (Washington, USA, July 14-18, 2013), AAAI, 933-939.

[2] Shen, H. and Albert-László Barabási. 2014. Collective credit allocation in science. *Proc. Natl. Acad. Sci. U.S.A* 111, 5 (Aug. 2014), 12325-12330.

[3] Sayyadi, H. and Getoor, L. 2009. Future Rank: Ranking scientific articles by predicting their future PageRank. *2009 SIAM International Conference on Data Mining* (John Ascuaga's Nugget, Sparks, Nevada), SIAM, 533-544.

[4] Wang, S., Xie, S., Zhang, X., Philip S.Y., and Shu, A. 2014. Future influence ranking of scientific literature. In *Proceedings of the 2014 SIAM International Conference on Data Mining* (Philadelphia, Pennsylvania, USA, April 24-26, 2014), SIAM, 749–757.