INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

IJARIIT

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 2)
Available online at: www.ijariit.com

Finding rising stars in co-author networks

Akil Reddy Gopavarapu
gakilreddy.15.cse@anits.edu.in
Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam, Andhra Pradesh

K. S. D. Sai Sowmya
ksridurgasaisowmya.15.cse@anits.edu.in
Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam, Andhra Pradesh

B. Shanmuk Abhishek
bshanmukabhishek.15.cse@anits.edu.in
Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam, Andhra Pradesh

P. Vinod Babu
vinodbabu.cse@anits.edu.in
Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam, Andhra Pradesh

ABSTRACT

Finding rising stars is a challenging and interesting task which is being investigated recently in many organizations and Academic networks. This paper addresses the problem of finding rising stars in Co-author networks. Predicting future rising stars is to find talented researchers in Co-author Networks. An effort for finding rising stars named PubRank is proposed, which considers the mutual influence and static ranking of conferences or journals. In this work, modification of PubRank is proposed by considering attributes like mutual influence based on authors' contribution, co-authors' citations, the order of appearance and publication venues. This method can be effectively used for mining the Co-author networks to search for rising stars in the research communities and organizations.

Keywords— *Rising stars, Co-author networks, Citations, Author order, Publication venue score*

1. INTRODUCTION

The rapid evolution in scientific research works has created large volumes of publications every year and will increase in future [6]. Online databases such as (DBLP & AMiner) provide a large collection of publications containing information like publication's title, abstract, author, published year and venues [7]. Extra features such as co-authorship information in research communities can be considered to provide further services for online databases. The rising stars are scholars with expertise and capabilities to achieve a high reputation in their respective fields in the near future. With this method, the research communities know about their potential researchers before time. Many organizations and research communities are concerned with searching "rising stars". The rising star is one who has relatively low profile currently but who may subsequently emerge as prominent contributors in their workplaces in future.

Most of the related social network mining research has focused on discovering groups or communities from social networks [1-2] and on the study of how these communities grow, overlap and change over time [3]. In this paper, we explore the likelihood of identifying rising stars from the co-author networks. These networks are constructed using interactions like research associations.

Finding future rising stars is a new research area and there is little work done on this. Initially, this idea was first formulated by PubRank [4], PubRank only incorporates authors and papers mutual influence and static ranking of publication venues. Later, PubRank is enhanced by StarRank [5], which incorporates two new features: the author's contribution based mutual influence and dynamic ranking lists of publications. Later, a prediction method is proposed, that applies machine learning techniques to find rising stars using publications, co-author and venue based features combination. Yet, the author's contribution based mutual influence is not deeply explored in terms of co-author's papers, paper citations, the order of appearance and publication venues.

The research networks are typically based on co-authorship and co-citation relationship among authors and publications. There are online websites that process the stored information of publications, such as ArnetMiner and DBLP. However, there is a little or no work done for differentiating authors and maintaining the evolution of researcher's profile over time. In this paper, we propose a new way of finding rising stars. The method incorporates the mutual influence of co-author's papers, co-author citations, the order of appearance and publication venue score. This method of using mutual influence enables us to effectively find rising stars.

Our technique is potentially useful for academics and research institutions in their recruitment and grooming of junior researchers in their organizations and institutions. It is also useful to fresh PhDs and postdocs for selecting promising and good supervisors. Finally, it can be useful for tracking one's relative performance in the research community, and for deciding whom to associate more with.

2. RELATED WORK

PubRank is the first method derived for finding rising stars in bibliographic networks [4]. It is based on two features, the mutual influence among researchers and track record of author's publications at different levels of venues. Consider an example of two authors' A_k and A_l with 4 and 3 publications respectively. Both authors are co-authors of two papers. The mutual influence weights between authors are calculated as follows:

$$W(A_l, A_k) = (A_l, A_k)/PA_k = 2/4 = 0.5$$
$$W(A_k, A_i) = (A_k, A_i)/PA_i = 2/3 = 0.66$$

Where, weight $W(A_l, A_k)$ describes the influence of author A_l on author A_k and PA_l and PA_k are the total number of publications by authors A_l and A_k. The weight $W(A_l,A_k)$ value is smaller than the weight $W(A_k,A_l)$ value, because author A_k has more publications than author A_l. So, author A_k has more influence on author A_l. Then worth of scholar's publications is computed on the basis of reputation/prestige of publications' venues. For the computation of publication quality score, venue of publication is considered i.e. publishing in high -level venue in the start of career shows scholar has bright chances to become future rising star. Long et al. suggests static ranking listings with following rank information. That is, Rank 1 (premium), Rank 2 (leading), Rank 3 (reputable), Rank 4 (unranked) [10]. Finally, the publication quality score for an author with a set P of publications is computed as follows.

$$\lambda\,(A_i) = \frac{1}{|P|}\sum_{i=1}^{|P|}\frac{1}{\alpha^{(r(pubi)-1)}}$$

Where, pub_i is the i^{th} publication, $r(pubi)$ is the rank of paper and value of α is $(0 < \alpha < 1)$. The value of α is low for a low-rank paper. The larger λ (A_i) is, the higher the average quality of papers published by the researcher. Finally, PubRank is formulated as follows:

$$\text{PubRank}(A_i) = \frac{(1-d)}{n} + \text{d}\,.\,\text{X}$$

$$X = \sum_{j=1}^{|v|}\frac{\text{W (Ai , Aj) . }\lambda\text{ (Ai) . PubRank (Aj)}}{\sum_{k=1}^{|v|}\text{W (Ak , Aj) . }\lambda\text{ (Ak)}}$$

Where, n is the total number of authors, $W(A_j,A_i)$ and $\lambda\ (A_i)$ are influential weight and publication quality score respectively.

The second method is Star Rank, in this the order in which authors appear in papers is also considered with the first author as a maximum contributor [5]. The author with less contribution gets a low score and with more contribution gets a higher score. Based on this intuition, author contribution based mutual influence is calculated.

The third method proposed for finding rising stars is WMIRank [8]. It is derived by combining three attributes of co-authorship, i.e. co-author's citations based mutual influence, co-author's order based mutual influence and co-author venues' citations based mutual influence. The mathematical formulation of these features and composite ranking method (WMIRank) description are also part of this section.

Afterwards, a prediction method is proposed [9], which uses machine learning techniques for rising star prediction in the co-author network. The goal of this research is mainly to apply classification models (discriminative and generative) to predict the rising stars in the domain of Co-author Network in the near future. The final result of the classifier is "Does an author has chances of becoming a future rising star or not?". Three classes of features (authors, venues and co-authorships) are explored and their mathematical formulation is also derived.

3. PROPOSED WORK

In this section, a new method is proposed for finding rising stars. It is derived by combining four attributes of co-authorship. In this method, attributes considered four are:(1) Publication venue score which is calculated by static ranking of different venues based on their popularity, (2) Mutual influence among the researchers in terms of co-authorships, (3) Order of appearance of the authors for each paper and (4) citation count, for the journals and conferences.

3.1 Mutual influence based on co-authorships

The mutual weight among the authors is calculated. When authors (A_k,A_l) co-author in any article, to calculate the weight of author A_k, we put the weight $A_k=(A_k,A_l)$ as a fraction of A_l when co-authored with A_k. Moreover, the weight of $A_l= (A_l,A_k)$ is a fraction of A_k. This weighting strategy is based on the intuition that a junior researcher will influence it's senior-less and senior will influence more as he has more publications which reduce the junior researcher fraction of co-authored work [8]. Let us take an example to understand how mutual influence is calculated.

Consider 2 authors A_k with 4 publications and A_l with 3 publications. If they have co-authored 2 papers with each other, the value with which they influence each other is calculated as:

$$\text{W}(A_l, A_k) = \frac{(\text{Al, Ak})}{PAk} = 2/4\ = 0.5$$

$$\text{W}(A_k, A_l)\frac{(\text{Al, Ak})}{PAk} = 2/3\ =\ 0.66$$

Here, PA_l and PA_k are the total numbers of papers written by authors A_l and A_k. (A_l,A_k) is the number of co-authored papers between authors A_l and A_k. The value $W(A_l,A_k)$ with which author A_l influences author A_k is less as compared to the value $W(A_k,A_l)$ with which author A_k influences author A_l. As the number of papers written by author A_k is 4 which are more in number as compared to A_l, So, A_k influences A_l more.

3.2 Mutual influence based on order of appearance

Next, the order in which the authors appear in papers is also considered. Generally, the first person is the highest contributor to the paper. So, the author with less contribution gets a low score and the author with more contribution gets a higher score. Based on this intuition, author contribution based mutual influence is calculated. Consider two authors A_l and A_k with 4 and 3 publications respectively. Let the paper number and order of author information be,

$$A_k =\ \text{P1(1), P2(3), P3(2), P4(1)}$$

$$A_l = \text{P1(2), P2(2), P3(1)}$$

Here, $P1,P2,..$ are the paper numbers and the number in brackets is the order of appearance in that paper. The author A_l has fewer papers than A_k and A_l appears as the first author in one paper and as the second author in two papers but author A_k appears as the first author in two papers and second and third author in two other papers. So, A_k is senior to A_l in co=authorship. The author contribution to weight value can be calculated as,

$$\mathrm{OA}(A_l, A_k) = \frac{\sum AOl + \sum AOk}{\sum PAOk}$$

$$\mathrm{OA}(A_l, A_k) = \frac{(0.5+0.5)+(1+0.33)}{1+0.33+0.5+1} = 0.823$$

$$\mathrm{OA}(A_k, A_l) = \frac{\sum AOk + \sum AOl}{\sum PAOl}$$

$$\mathrm{OA}(A_k, A_l) = \frac{(1+0.33)+(0.5+0.5)}{0.5+0.5+1} = 1.165$$

Where *OA* is the author order weight based mutual influence, $\sum PAOl$ and $\sum PAOk$ are total cothe ntribution of authors A_l and A_k and AO_k and AO_l are co-authored contribution of authors A_k and A_l. The author A_k has higher ia nfluence on author A_l as compared to influencthe e of author A_l on author A_k. The reason is that author A_k has more number of papers than A_l and author A_k also has more papers as a first author than A_l.

3.3 Mutual Influence based on co-Authored citations

The reputation of a researchers work can be judged by the number of citations his papers have. But it is biased towards earlier publications as publication needs time to be cited by other papers. Rising stars cannot have highly cited papers. So, the citation count is given the least priority while finding the rising stars in our approach. Citation count is considered in our approach to removing any ambiguity among the researchers.

Let us consider two authors A_l and A_k with each where two of their publications are co-authored,

$$A_k = \mathrm{P1(7), P2(8), P3(1)}$$

$$A_l = \mathrm{P1(7), P2(8), P4(0), P5(6)}$$

Here, *P1,P2,...* are the paper numbers and the value in the brackets is the citation count of the papers. The mutual influence based on the co-authored paper's citations can be calculated as,

$$\mathrm{CA}(A_l, A_k) = \frac{(\mathrm{Acl, Ack})}{PAck} = 15/16 = 0.937$$

$$\mathrm{CA}(A_k, A_l) = \frac{(\mathrm{Ack, Acl})}{PAck} = 15/21 = 0.714$$

Here, *(Ac_l,Ac_k)* is the total number of co-authored citations between A_k and A_l. PAC_k is the number of citations of all papers written by author *K* and PAC_l is the number of citations of all papers written by author *L*. The weight $C(A_l,A_k)$ with which author Al influences author A_k is more as compared to the weight $C(A_k,A_l)$ with which author A_k influences author A_l. Since A_l has more citations, A_l has more influence on A_k compared to A_k on A_l.

3.4 Publication venue score

Rising stars, being junior researchers, are unlikely to have many highly cited papers. We, therefore, opted for an alternative process based on the popularity of its publication venue. The prestige of publication venue is considered for finding out the rising star. Many publications will be done in several venues and conferences. We use the concept of static ranking listings [10]. All the possible venues are taken and divided into four categories 1, 2, 3 and 4.We ranked these categories according to their popularity. Here, the first category consists of popular venues like Scopus, IEEE etc. The second category has next level venues. All the unranked venues are put into the last category. The main idea is that if a researcher is publishing in high-level venues at the beginning of his career, he has bright chances to be an expert in the future.

$$\mathrm{PVS}(A_i) = \sum_{j=1}^{|n|} \frac{1}{rank(n)}$$

Where *PVS(A_i)* is the publication venue score for author *Ai*, *n* is the number of papers published by A_i.

3.4 Hybridization

Finally, the score for each author is calculated to rank them. The final hybrid equation to find the score for each author is defined as follows:

$$\mathrm{Score}(A_i) = \frac{(1-d)}{n} + \mathrm{d.T}$$

Where T is,

$$T = \sum_{j=1}^{|v|} \frac{\mathrm{W(Ai\,,Aj).OA(Ai\,,Aj).C(Ai\,,Aj).PVS(Ai)}}{\sum_{\mathrm{k}=1}^{|\mathrm{v}|} \mathrm{W(Ak\,,Aj).OA(Ak\,,Aj).C(Ak\,,Aj).PVS(Ak)}}$$

Where A_i is the author *i*,*d* is damping factor and its value is between the range *(0<d< 1)[pub]*, usually d is configured to 0.85 value and n is the number of authors. *v* is the set of co-authors for author A_j. *v* is the set of co-authors for author A_i.

4. EXPERIMENTS

4.1 Dataset

To evaluate the performance of our method, our college faculty data is taken. For experiments, we selected the author's data for six years (2013-2018) which contains 159 authors and 179 Publications data. The authors who have publication before 2013 are excluded. Finally, our dataset contains titles of publications, author names, and venues including conferences and journals, citations. Data pre-processing is carried to make sure that every author has the same name throughout the data.

4.2 Rising Stars

After calculating a score for every author in our dataset by using the above-mentioned formula, we then list out the top rising stars of our dataset by sorting them in decreasing order of their score as the higher score author has high chances of becoming a rising star.

Table 1: Top 5 Rising Stars

Authors	Papers	Citations	Score
Dr. S. C. Satapathy	28	338	23.91578
Pritee Parwekar	25	21	15.24026
Dr. K. Selvani Deepthi	10	6	7.97627
B. Ravi Kiran	8	8	7.70455
Dr. M. Ramakrishna Murty	13	33	7.59375

From the above result, the 2nd author has fewer citations when compared to the author at the 5th position because the 2nd author has a higher score than author at 5th position, this is because the 2nd author has published his/her papers at highly rated publication venues.

4.3 Effect of Damping Factor (d)

Damping factor value of 0.85 is usually used for ranking pages on Web. Google itself uses this value because it is easy to get refined results [11]. The score of every author is calculated for various values of damping factor to see its effect on the final output. The citations of authors gradually increase from lower to the high value of the damping factor. In this paper, we used the damping factor value of 0.85 through one can use 0.7 and 0.8 too. We study the effect of damping factor on average citations only as they are the most important factor to judge the quality of research [5].

5. FUTURE WORK AND CONCLUSION

5.1 Conclusion

In this research, we used four types of features which are hybridized for the formulation to find the Score for each author. The attributes are co-author papers, co-author citations, co-author order of appearance and publication quality score of each author. It can be concluded from the results that the proposed features are highly effective in finding rising stars. Mutual influence of co-related entities and venues helps in finding rising stars in other social networks. Since a rising star is very useful in the fields of research for effectively collaborating with people who can be expert in a particular field, finding a rising star is useful. Rising stars also help in educational sectors and universities while doing projects to implement this approach in Python.

5.2 Future work

Since research is a never-ending area, a rising star would be equally useful in every research work as the work to be done can become easier when done with more capable people. So, this would be a better idea to find rising stars in a co-author network in the areas of research or education where a lot of people can benefit from it. In future, we plan to empirically investigate why some potential rising stars do not become actual rising stars and accordingly make our method more accurate. Further, this concept can be applied in other domains e.g. finding rising products in online shopping, finding rising reviewers in review community, finding rising bloggers [8]

6. REFERENCES

[1] M.E.J. Newman: "Detecting community structure in networks", European Physical Journal B, vol. 38, pp. 321-330, 2004.

[2] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek: "Uncovering the overlapping community structure of complex networks in nature and society", Nature, vol. 435, pp. 814-818, 2005.

[3] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan: "Group Formation in Large Social Networks: Membership, Growth, and Evolution," in Proceedings of Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006.

[4] L.L. Xiao, F.S. Chuan, T.L. Kar, N. See-Kiong: “Searching for Rising Stars in Bibliography Networks”, Proc. of the 14th International Conference on Database Systems for Advanced Applications (DASFAA), pp.288-292 (2009).

[5] A. Daud, R. Abbasi, F. Muhammad:” Finding rising stars in social networks”, Database Systems for Advanced Applications (LNCS), 7825, 13–24, 2013.

[6] R. Yan, J. Tang, X. Liu, D. Shan, X. Li:” Citation Count Prediction: Learning to Estimate Future Citations for Literature.”, CIKM, 1247-1252, 2011.

[7] J. Tang, J. Zhang, L. Yao, j. Li, L. Zhang, Z. Su: “Arnetminer: extraction and mining of academic social networks.”, KDD, 990-998, 2008.

[8] Ali Daud, Naif Radi Aljohani, Rabeeh Ayaz Abbasi, Zahid Rafique, Tehmina Amjad, Hussain Dawood, Khaled H. Alyoubi: “Finding Rising Stars in Co-Author Networks via Weighted Mutual Influence”, Published in WWW, April 3-7, 2017, Perth, Australia. ACM 978-1-4503-4914-717/04 2017.

[9] A. Daud, M. Ahmad, M. Malik, D. Che:” Using machine learning techniques for rising star prediction in coauthor network”, Scientometrics, 102,2 (2015), 1687-1711.

[10] P.M. Long, T.K. Lee, J. Jaffar: "Benchmarking Research Performance in Department of Computer Science”, School of Computing, National University of Singapore, 1999.

[11] B.P. Boldi, M. Santini, V. Sebastiano: “PageRank as a Function of the Damping Factor.”, In: Proc. of the 14th International Conference on World Wide Web, pp.557-566, 2005.

