Finding Rising Stars in Co-Author Networks via Weighted Mutual Influence

Ali Daud [a,b]
[a] Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
adfmohamad@kau.edu.sa

Naif Radi Aljohani [a]
[a] Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
nraljohani@kau.edu.sa

Rabeeh Ayaz Abbasi [a,c]
[c] Department of Computer Science, Quaid-i-Azam University, Islamabad, Pakistan
rabbasi@kau.edu.sa

Zahid Rafique [b]
[b] Department of Computer Science and Software Engineering, International Islamic University, Islamabad, Pakistan
zahid@ogdcl.com

Tehmina Amjad [b]
[b] Department of Computer Science and Software Engineering, International Islamic University, Islamabad, Pakistan
tehminaamjad@iiu.edu.pk

Hussain Dawood [d]
[d] Faculty of Computing and Information Technology, University of Jeddah, Jeddah, Saudi Arabia
hdaoud@uj.edu.sa

Khaled H. Alyoubi [a]
[a] Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
kalyoubi@kau.edu.sa

ABSTRACT

Finding rising stars is a challenging and interesting task which is being investigated recently in co-author networks. Rising stars are authors who have a low research profile in the start of their career but may become experts in the future. This paper introduces a new method Weighted Mutual Influence Rank (WMIRank) for finding rising stars. WMIRank exploits influence of co-authors' citations, order of appearance and publication venues. Comprehensive experiments are performed to analyze the performance of WMIRank in comparison to baseline methods, which have ignored weighted mutual influence. AMiner[1] data for years 1995-2000 is used for experiments. List of top 30 authors as per proposed and baseline methods are compared for their average number of papers, average number of citations and achievements. Experimental results provide convincing evidence of the effectiveness of the investigated weighted mutual influence.

CCS CONCEPTS

- **Information systems~Link and co-citation analysis**
- *Information systems~Retrieval models and ranking*
- Information systems~Social networks

[1] https://aminer.org/

WWW 2017 Companion, April 3-7, 2017, Perth, Australia.
ACM 978-1-4503-4914-717/04.
DOI: http://dx.doi.org/10.1145/3041021.3054137

KEYWORDS

Bibliographic databases; Citations; Author Order; Rising Stars; Mutual Influence; Co-Author Networks

1. INTRODUCTION

The rapid evolution of scientific research has created large volumes of publications every year, and is expected to continue in near future [21]. Online databases such as (DBLP or AMiner) provide large number of publications containing information about publication's title, abstract, author, published year and venues [17].

Informally, co-author relationships and citation links can be called Academic Social Networks (ASNs). Formally, an academic social network consists of people in research community where they share ideas, research publications, comments and ask questions. Several areas are well investigated for web databases, such as, expert finding / author ranking [2,5,6,19], author interest finding [4], research collaboration [9], citation content analysis [23], author name disambiguation [10,15,16,20], identifying influential entities [1,3,11,21], etc. However, finding rising-stars problem [7,8, 12,18] still needs to be explored.

The rising stars are the scholars with expertise and capabilities to achieve high reputation in their respective fields in near future. Rising star finding in a specific domain is an emerging research direction that may enable research communities to highlight potential researchers before time.

Rising star finding emerges as a new research area and there is little work done in this regard. Initially, this idea was first formulated by PubRank [12], PubRank only incorporates authors and papers mutual influence and static ranking of publication venues. Dynamics of authors' research profiles are modeled in [18]. In this method, researchers are classified into four groups using unsupervised learning techniques. Later, PubRank is enhanced by StarRank [7] which incorporates two new features: author's contribution based mutual influence and dynamic ranking lists of publications. Afterwards, a prediction method is proposed [8], that applies machine learning techniques for finding rising stars using publication, co-author and venue based features combination. However, author's contribution based mutual influence is not deeply explored in terms of co-author's citations, order of appearance and publication venue's citations.

In this paper, we propose WMIRank for finding outstanding researchers. It incorporates weighted mutual influence of co-author's citations, order of appearance and publication venue's citations. This weighted mutual influence enables us to effectively find the outstanding researchers.

The main contribution of our work is summarized as follows.
1. It is the first attempt to consider co-author's citations, co-author's order of appearance and citations of co-author's publication venue.
2. Mathematical formulations for the computation of weighted mutual influence.
3. Performance evaluation of proposed and baseline methods in terms of average number of papers and average number of citations.
4. Qualitative analysis of top ranked 30 authors in terms of their achievements.
5. The effect of parameters for the computation of ranks of authors is examined and comparison of methods for ranking position relocation is provided.

This paper is organized as follows. The problem definition is presented in section 2 followed by the brief review of two existing methods and detailed description of proposed method in section 3. In section 4, dataset, performance evaluation, experiments to examine the accuracy and efficiency of the proposed method and results comparison with two baseline methods is provided. Finally, section 5 concludes this work.

2. PROBLEM DEFINITION

Formal problem definition of calculating rising star scores is provided here. The ranking problem structure is quite similar to the famous page rank problem.

Given a set $\{A_1, ..., A_n\}$ of n authors, we have to calculate the ranking score (rising star score) for each author by calculating different mutual influence values. Here, $W = \{W_1, ..., W_n\}$ be an $n \times n$ matrix, where W_{ij} represents the mutual influence of author A_i on author A_j. Let Y be a vector representing ranking scores of n authors. Then ranking function is defined as follows.

$$y(A_i) = \frac{1-d}{n} + d \,.\, \sum_{j=1}^{|v|} \frac{(A_i, A_j)}{\sum_{k=1}^{|v|} (A_k, A_j)} \,.\, y(A_j) \tag{1}$$

Where, d is damping factor, v is the set of co-authors for author A_i and $W(A_i, A_j)$ is the influential weight. The $y(A_i)$ score is calculated for each author to rank the authors.

We propose three features i.e. co-author's citations based mutual influence, co-author order based mutual influence and co-author venues' citations based mutual influence for the computation of rising star score of an author A_i.

3. METHODS

Before describing WMIRank, a brief introduction of existing methods related to rising stars finding is presented. These methods are PubRank [12] and StarRank [7]. These methods are derived from the famous Page Rank algorithm. The details of these methods are presented in Section 3.1 and 3.2.

3.1 PubRank

PubRank is the first method derived for finding rising stars in bibliographic networks [12]. It is based on two features, the mutual influence among researchers and track record of author's publications in form of publishing in 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:

$$\begin{aligned} W(A_l, A_k) &= (A_l, A_k) \,/PA_k = 2/4 = 0.5 \\ W(A_k, A_l) &= (A_k, A_l) \,/PA_l = 2/3 = 0.66 \end{aligned} \tag{2}$$

Where, weight $W(A_l, A_k)$ describes 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. i.e. Rank 1 (premium), Rank 2 (leading), Rank 3 (reputable), Rank 4 (unranked) [13]. Finally, the publication quality score for an author with a set P of publications is computed as follows.

$$\lambda(A_i) = \frac{1}{|P|} \sum_{i=1}^{|P|} \frac{1}{\alpha^{(r(pub_i)-1)}} \tag{3}$$

Where, pub_i is the i^{th} publication, $r(pub_i)$ 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 $\lambda(A_i)$ is, the higher the average quality of papers published by the researcher. Finally, *PubRank* is formulated as follows.

$$PubRank(A_i) = \frac{1-d}{n} + d \,.\, X \tag{4}$$

$$X = \sum_{j=1}^{|v|} \frac{W(A_i, A_j) \,.\, \lambda(A_i) \,.\, PubRank(A_j)}{\sum_{k=1}^{|v|} W(A_k, A_j) \,.\, \lambda(A_k)}$$

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.

3.2 StarRank

In StarRank, the order in which authors appear in papers is also considered with first author as maximum contributor [7]. Author with less contribution gets low score and with more contribution gets higher score. Based on this intuition, author contribution based mutual influence is calculated. It also recommends a dynamic way of computing the publication venue score instead of static ranking. Consider two authors A_k and A_l with *4* and *3* publications respectively. The paper number and order of author information are presented in Table 1.

Table 1: Author with papers and order of appearance.

Author	Paper No. (Order of Appearance)
A_k	**P1(1), P2(3)**, P3(2), P4(1)
A_l	**P1(2), P2(2)**, P5(1)

The author A_l has fewer paper than A_k and A_l appears as first author in one paper and as second author in two papers but author A_k appears as first author in two papers and second and third author in two other papers. So, we can summarized that A_k is senior than A_l in co-authorship. Then author contribution weight value is calculated as follows.

$$AOWI\ (A_l\ ,\ A_k) = \frac{(\Sigma AO_l + \Sigma AO_k)}{\Sigma PAO_k} = \frac{(0.5{+}0.5) + (1{+}.33)}{1{+}.33{+}.5{+}1} = 0.823 \tag{5}$$

$$AOWI\ (A_k\ ,\ A_l) = \frac{(\Sigma AO_k + \Sigma AO_l)}{\Sigma PAO_l} = \frac{(1{+}.33) + (0.5{+}0.5)}{0.5 + 0.5 + 1} = 1.165 \tag{6}$$

Where, *AOWI* is author order weight based mutual influence, $\sum PAO_l$ and $\sum PAO_k$ are total contribution 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 influence on author A_l as compared to influence 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. Given a paper, StarRank computes a measure of paper quality based on the entropy value. Entropy of venue is computed using the following equation.

$$Entropy\ (v) = -\sum_{i=1}^{m} w_i \log_2 (w_i) \tag{7}$$

Where, w_i is the probability of word i in a venue v and entropy value is lower for high level venues and higher for low level venues, while publication quality score for each author who has a publication set P, can be computed using the following equation.

$$\lambda\ (dpq_i) = \frac{1}{|p|} \cdot \sum_{i=1}^{|p|} \frac{1}{\alpha^{Entropy\ of\ Venue}} \tag{8}$$

Where, $\alpha\ (0 < \alpha < 1)$ is a damping factor so that lower ranked publications have lower scores. The larger $\lambda\ (dpq_i)$ is, the higher the average quality of papers published by author.

Finally, StarRank for each author node is computed by merging both author contribution based mutual influence and dynamic publication score as follows.

$$StarRank\ (A_i) = \frac{1-d}{n} + d\ .\ S \tag{9}$$

$$S = \sum_{j=1}^{|v|} \frac{AOWI\ (A_i\ ,A_j)\ .\ \lambda\ (dpq)\ .\ StarRank\ (A_j)}{\sum_{k=1}^{|v|} AOWI\ (A_k\ ,\ A_j)\ .\ \lambda\ (dpq)}$$

3.3 Weighted Mutual Influence Rank (WMIRank)

In this section, a new method WMIRank is proposed for finding rising stars. 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.

3.3.1 Co-Author Citations based Mutual Influence

Here, mutual influence between authors is computed and weights are derived in terms of how much an author influences another author based on the number of citations. The mutual influence weight of an author describes the impact of his/her contribution on another author in co-authorship. The intuition is based on the fact the more the citations a co-author has the more he / she will influence his / her collaborators.

Suppose there are three authors A_k, A_l and A_m with 3, 4 and 4 publications respectively. The total number of publications and their citations' information are presented in Table 2 (order of publications in parentheses). The authors A_k and A_l co-authored two papers whereas author A_k and A_m also coauthored two papers. The co-authored papers are highlighted in bold letters in Table 2. The mutual influences between authors A_k and A_l and mutual influence between author A_k and A_m based on the number of citations can be calculated as follows.

Table 2: Author's Papers & Citations.

Author	Paper (# of Citations)
A_k	**P1(07)**, **P2(08)**, P3(13)
A_l	**P1(07)**, **P2(08)**, P4(15), P5(18)
A_m	**P1(07), P2(08)**, P6(02), P7(03)

$$CACWI\ (A_l\ ,\ A_k\ ) = \frac{(AC_l\ ,\ AC_k\ )}{TAC_k} = \frac{15}{28} = 0.53 \tag{10}$$

$$CACWI\ (A_k\ ,\ A_l\ ) = \frac{(AC_k\ ,\ AC_l\ )}{TAC_l} = \frac{15}{48} = 0.31 \tag{11}$$

$$CACWI\ (A_k\ ,\ A_m\ ) = \frac{(AC_k\ ,\ AC_m\ )}{TAC_m} = \frac{15}{20} = 0.75 \tag{12}$$

$$CACWI\ (A_m\ ,\ A_k\ ) = \frac{(AC_m\ ,\ AC_k\ )}{TAC_k} = \frac{15}{28} = 0.53 \tag{13}$$

Where, TAC_k, TAC_l, TAC_m are the total numbers of citations of authors A_k, A_l, *and* A_m respectively, $(AC_l, AC_k) = (AC_k, AC_l)$ is the number of co-authored citations of papers for Author A_l and A_k and $CACWI\ (A_k,\ A_m)$ is the co-author citation weight based influence of A_k on A_m.

Thus, the influence of author A_l on author A_k is greater than influence of author A_k on author A_l, therefore author A_l is more influential due to having more number of citations. Similarly, influence of author A_k on author A_m is greater than influence of author A_m on A_k. Therefore, author A_k is more influential than author A_m in co-authorship due to having more number of citations.

3.3.2 Co-Author Order based Mutual Influence

Here, author's order in a paper is considered and mutual influence is computed for each author based on co-author order. The idea was derived by considering coauthor contributions in research publications [14]. An author that appears as first author in a paper is referred as a maximum contributor. Therefore, we applied first author order as the main idea for the computation of influence of an author on another author.

Suppose there are two authors A_k and A_l with 4 and 3 papers, respectively. The authors A_k and A_l co-authored each other and co-authored papers are highlighted in bold letters in Table 3.

We considered only those papers for an author that meet two criteria, first those papers should be co-authored with someone so that mutual influence can be computed. Second, he/she should appear as a first author in those papers. Finally, those papers are assigned with weight value of 1 and all other papers are assigned with weight value of 0. Table 4 shows the weights assigned for each paper. The mutual influences between authors A_k and A_l based on co-author order are calculated as follows.

Table 3: Authors' Papers & Order of Appearance.

Author	Paper # (Order of Appearance)
A_k	**P1(01), P2(03)**, **P3(01)**, P4(02)
A_l	**P1(02)**, **P2(01), P3(03)**

Table 4: Papers Weight Substitutions.

Author	Weight of Papers Set to 1 or 0
A_k	1, 0, 1, 0
A_l	0, 1, 0

$$CAOWI\ (A_k\ , A_l\ ) = \frac{1AP_k + 1AP_l}{TAP_l + 1AP_l} = \frac{(1+0+1)+(0+1+0)}{3+1} = 0.75 \quad (14)$$

$$CAOWI\ (A_l\ , A_k\ ) = \frac{1AP_l + 1AP_k}{TAP_k + 1AP_k} = \frac{(0+1+0)+(1+0+1+0)}{4+2} = 0.50 \quad (15)$$

Where, TAP_k = Total papers written by author A_k, AP_k = Total papers of author A_k as first author and $CAOWI\ (A_k, A_l)$ = Co-author order weight based influence of A_k on A_l.

Hence, influence of author order weight of A_k on A_l is greater than influence of author A_l on A_k because author A_k has 2 papers as first author and author A_l has 1 paper as first author.

Table 5: Author papers and their corresponding Venue' average citations.

Author	Paper # (Venue #, Average Citations of Venue)
A_k	**P1(V1, 0767) P2(V3, 8357)** P3(V2, 1035) P4(V6, 3700) P5(V5, 2178)
A_l	**P1(V1, 0767) P2(V3, 8357)** P3(V4, 8163) P4(V2, 1035)
A_m	**P1(V3, 8357) P2(V4, 8163)** P3(V2, 1035) P4(V1, 767)
A_n	**P1(V3, 8357) P2(V4, 8163)** P3(V1, 767) P4(V2, 1035) P5(V6, 3700)

3.3.3 Co-Author Venue's Citations Based Mutual Influence

For this measure, mutual influence between authors is calculated based on their venues' citations. If a co-author has more papers published in high level venues and those venues have higher citations then that author will be more influential than others.

Suppose there are four authors A_k, A_l, A_m and A_n with 5, 4, 4, and 5 publications, respectively. We considered here six venues (V1, V2, V3, V4, V5 and V6), where these authors published their papers. The authors A_k and A_l co-authored two papers and authors A_m and A_n also co-authored two papers as highlighted in bold letters in Table 5. We also computed average citations for each venue for the period 1995 to 2000. The author's publications in a venue and his/her average citations information are also described in Table 5.

The co-author venue's citations based influence between author A_k and A_l can be computed as follows.

$$CAVWI(A_k\ , A_l\ ) = \frac{VC_k + VC_l}{TVC_l} = \frac{767+8357}{767+8357+8163+1035} = 0.49 \quad (16)$$

$$CAVWI(A_l\ , A_k\ ) = \frac{VC_l + VC_k}{TVC_k} = \frac{767+8357}{767+8357+1035+3700+2178} = 0.56 \quad (17)$$

Where, $VC_k + VC_l$ = Co-Authored Venue Citations of A_k and A_l, TVC_k = Total citations in venue of author A_k and $CAVWI\ (A_k, A_l)$ = Co-author venue citation's weight based influence of A_k on A_l.

Thus, influence of author A_k venues' citations on author A_l is smaller than influence of author A_l on A_k because author A_k has total 16037 citations of a venue and author A_l has total 18322 citations of a venue. Therefore, we found author A_l more influential than author A_k in terms of total citations of a venue. Although author A_l have fewer number of papers as compared to author A_k.

3.3.4 Hybridization

The WMIRank score is calculated for each author to rank them. The final hybrid equation of WMIRank is defined as follows:

$$WMIRank\ (A_i) = \frac{1-d}{n} + d\ .\ T\ .\ WMIRank\ (A_j) \quad (18)$$

$$T = \sum_{j=1}^{|v|} \frac{CACWI(A_i, A_j)\ .\ CAOWI(A_i, A_j)\ .\ CAVWI\ (A_i, A_j)}{\sum_{k=1}^{|v|} CACWI(A_k, A_j)\ .\ CAOWI\ (A_k, A_j)\ .\ CAVWI\ (A_k, A_j)}$$

Where, d is damping factor and its value is between the range ($0 < d < 1$), 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_i.

4. EXPERIMENTS

4.1 Dataset

To evaluate the performance of WMIRank, the dataset is taken from AMiner which contains data for the year 1949~2000. For experiments, we selected authors data for six years (1995 ~ 2000) which contains 37146 authors and 15403 publications data. The authors which have publication before 1995 are excluded. Finally, our dataset contains features; titles of publications, author names, venues including conferences and journals.

4.2 Performance Evaluation

As it is already mentioned that there are no true values available for our ranking list of rising stars so we can't evaluate our experimental results by the gold standard. Therefore, we adapted the following procedure. In the first step, the rising star score for each author is calculated using Eq. 18. Then a list of top-30 authors is presented by sorting the rising star scores in descending order. As we used authors' data for the period 1995-2000 in the experiments and calculated the rising star scores using data spanning 6 years. Therefore, the presented list of top-30 rising stars is actually predicted list because these authors were not rising star authors during the period 1995-2000 as they were starting their careers at that time.

Then predicted list of top-30 authors (Table 6) is cross checked by the current status of these authors e.g. award and top cited paper citations. If a researcher is a rising star, then his students are also credible researchers because research is a collaborative activity. The awards and goodwill achieved by his students also credits to his research profile. The current status (award, top cited paper citations, designation, etc.) of predicted rising stars confirms our results, the evaluation criteria may be further extended by designing a form and each author may be evaluated by his current total citations, awards and other distinctions. To predict the rising stars status of a solitary researcher, it is a type of outlier detection problem and a new direction for future research in academic social networks.

We presented the detailed comparison of our proposed WMIRank with the existing methods i.e. PubRank and StarRank. At first, descending order sorted list of top-30 authors is presented based on WMIRank score. The list also contains the current status of researcher, his organization, awards and top cited paper's citations. Then performance of three methods (WMIRank, PubRank and StarRank) is analyzed by computing average and standard deviation values of total papers and total citations for top-30 authors.

4.3 Performance Comparison

In this section, our proposed method is compared with baseline methods for performance analysis and then the results are demonstrated.

4.3.1 Ranking of Top-30 Authors by WMIRank

In this section, a ranking list of top-30 authors is presented using WMIRank by sorting them in descending order. Most of the authors have at least 100 citations for top cited paper and have brilliant google scholar profiles now. They are fellows of prestigious publication groups e.g. ACM, IEEE, AAAS, and NSERCC SIAM. They also have remarkable achievements such as IBM Outstanding Technical Achievement Awards, IBM Canada Research Impact of the Year Award, IBM Outstanding Innovation Award. Such achievements of our predicted authors are clear indicators that they were prospective rising stats during the period 1995-2000 and have now become experts.

One can see some of the authors in top 30 list were not able to become experts but are found as rising stars by WMIRank. There can several reasons, such as, they worked for their PhDs with distinguished professors or in top notch research labs and later were not able to continue same standards of research. As the started university careers with high workloads of teaching and less time for research and dint find similar motivation and environment to do research. Some of them may left research and got involved more in commercial projects. Although it is interesting to empirically investigate and analyze the reasons which is our other ongoing work "Standing on the Shoulders of Giants".

4.3.2 Comparison by Papers and Citations

Here, performance of proposed method is compared with baselines in terms of papers and citations. The comparison is performed by two metrics, Average and Standard Deviation. First three ranking lists of top-30 authors are computed using WMIRank, StarRank and PubRank, respectively. Second, average number of papers and average number of citations are computed for top-30 authors. The process of computing average and standard deviation for total number of papers and total number of citations is presented in Table 7.

For comparison, the performance of proposed method with two baselines is analyzed in terms of average and standard deviation metrics. As PubRank method only incorporated number of papers' information and static lists of venues therefore it output highest value of average number of papers as compared to WMIRank and StarRank as shown in Fig. 1. The WMIRank have a bit more average number of papers than that of StarRank.

The standard deviation values of number of papers' data presented by each method tell us the dispersion of data from average values. However, papers data presented by WMIRank have lowest standard deviation value as compared to StarRank and PubRank method as shown in Fig. 2. The lowest standard deviation value for WMIRank listing indicates that papers data tend to be very close to average value. For citations analysis, average and standard deviation are calculated for the total number of citations for top-30 authors for WMIRank, StarRank and PubRank. WMIRank outperforms and presents highest average number of citations as compared to StarRank and PubRank due to incorporation of weighted mutual influence of co-author citations, co-author order of appearance and co-author venue's citations.

StarRank only incorporated co-authors' order of appearance and dynamic publication venue score. The performance of StarRank is better than PubRank because PubRank only considers static venue score and papers information. For WMIRank, it has smaller standard deviation as compared to StarRank and more than PubRank. As the number of citations for PubRank top-30 authors is too few that's why it got less standard deviation here. However, in case of WMIRank has more number of average citations as compared to StarRank still has less average standard deviation. Standard deviation is an important factor to check the accuracy of method. Smaller standard deviation means it has stable rank which proves the effectiveness of WMIRank.

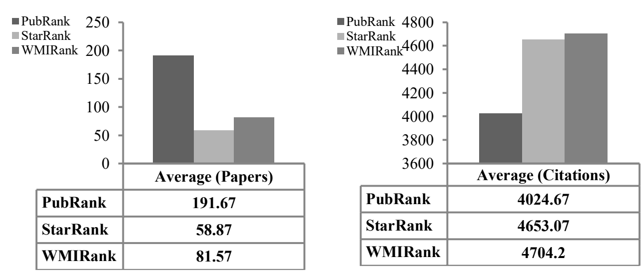

Figure 1: Average based performance comparison in terms of Papers & Citations of Top-30 Authors.

4.4 Effect of Damping Factor

In this section, the effect of damping factor is analyzed to see its effect on rising stars finding by calculating ranking scores of our proposed method and two baseline methods. For this purpose, average citations of top-30 authors are calculated by each method for several values of damping factors. For WMIRank, the citations of authors remain stable on all the damping factor values and it is observed that maximum average citations are gained on all values of damping factor as shown in Fig. 3. However, it is seen that the average citations for top-30 authors calculated by StarRank do not remain stable for different values of damping factor. We get minimum average citations for StarRank at damping factor value of 0.15 then average citations are increased and values are continuously oscillating for other values of damping factor. For PubRank, maximum and minimum average citations are obtained at damping factor values of 0.45 and 0.35 and average citations obtained by PubRank are also not stable in comparison to WMIRank.

In the next subsections, comparison of authors' positions predicted by three methods using several values of damping factor are critically analyzed and ranking relocations are discussed.

4.4.1 Comparative Analysis at Damping factor 0.85 and 0.50

Here, comparison of WMIRank and baseline methods are presented for authors' positions obtained at damping factor values of 0.85 and 0.50. For ranking web pages, damping factor of 0.85 is usually used and it gives better results whereas 0.5 damping factor may be used for ranking publications. We observed that author's positions are varied by applying damping factor values of 0.85 and 0.50 as shown in Fig. 4 & 5.

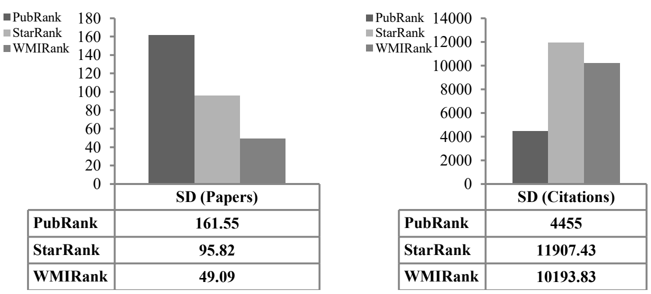

Figure 2: Standard Deviation performance comparison in terms of Papers & Citations, where SD is standard deviation.

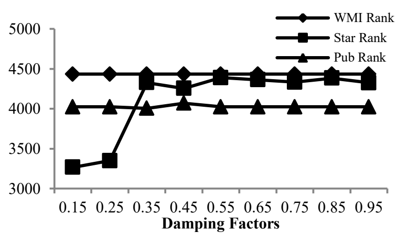

Figure 3: Average number of citations of top-30 authors, y-axis is average citations.

We can analyze that by using damping factor of 0.5, performance of WMIRank is comparatively improved w.r.t baseline methods as shown in Fig. 4. A comparison of authors' position and relocation by WMIRank w.r.t StarRank at damping factor of 0.5 is presented in next section.

4.4.2 Ranking Relocation for damping factor = 0.50

In this section, we examine the position relocation of authors' ranking proposed by WMIRank in comparison with StarRank and PubRank at damping factor value of 0.50. Authors with higher number of citations of publications will get higher positions as compared to authors with low number of citations. The position of authors proposed by WMIRank are compared with position of authors proposed by StarRank and position rise is identified to analyze the effectiveness of WMIRank over StarRank for rising star finding.

The lists of authors whose positions are upgraded are presented in Table 8 with WMIRank score. In Table 8, Bertram Ludascher was at position 10 and Marc P. C. Fossorier was at position 20 by StarRank due to high number of publications in high ranking venues. By WMIRank, Bertram Ludascher is switched at rank 2 that is 8 positions higher and Marc P. C. Fossorier is switched to rank 5 that is 15 positions higher than StarRank due to large number of citations. Bertram Ludascher and Marc P. C. Fossorier have 4943 and 6647 citations. The last column of table indicates the number of position upgraded by each author using WMIRank.

Next, we analyzed the fall in ranks of authors proposed by WMIRank method with respect to Star Rank method as shown in Table 9. The reason behind this position fall is due to fewer number of citations. As we can see that author Grace Ngai lost position down by 2 points from position 13 to position 15 and Dirk Jonscher lost position down by 5 point from position 12 to position 17 due to lower number of citations. However, these authors have higher number of publications in high ranking venues therefore they are ranked at higher positions by StarRank. The last column of table represents the number of positions down by each author.

We also compared ranking positions of authors proposed by WMIRank with positions of authors proposed by PubRank. The rise in authors' ranks is identified and results are presented in Table 10. The rise in ranks of authors is due to acquisition of large citations information based influence. E.g. author Yossi Azher switched at position 3 by 3 points from position 6 and Geoffrey Zweig switched at position 4 by 1 point from position 5.

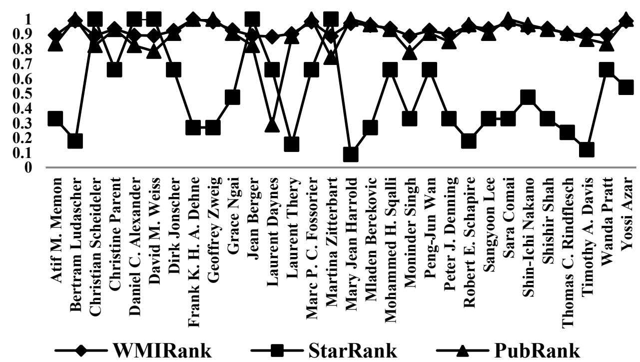

Figure 4: Effect of Damping Factor (0.5) on Authors Score (y-axis is author rank).

In Table 11 Shin-Ichi Nakano lost position by 2 points from position 8 to position 10 and Peng-Jun Wan lost position by 1 point from position 15 to position 16. Both these authors received less author order and citation based influence due to which decrease in their ranks is observed.

The PubRank and StarRank do not incorporate order and citations based mutual influence for finding rising stars. Therefore, we analyzed the performance of WMIRank and found it better than two baseline methods.

5. CONCLUSIONS

In this research, a new method is proposed for finding rising stars in co-author networks. Three types of attributes of co-author are hybridized for the formulation of WMIRank. The attributes are co-author citations, co-author order of appearance and co-author venue's citations. 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 the investigated task and can be useful for finding rising stars in other social networks. Although most of the top-30 authors are rising stars, but few of them were not able to become experts according

Table 6: Current Position of Authors by WMIRank where TCPC is Top Cited Paper Citations.

#	Name	Position	Organization	Awards	TCPC
1	Frank K. H. A. Dehne	Professor	Research Lab, Carleton University School of Computer Science, Ottawa, Canada	IBM Canada Research Impact Of The Year Award, 2012. Carleton University Research Achievement Award, 1993, 1999, 2012.	188
2	Bertram Ludascher	Professor	University of California, Davis		1331
3	Yossi Azar	Professor & Head	Tel-Aviv University, Israel	Bruno Memorial Award in 2001 Erdős Prize in 1989. Feher prize in 1991. Pólya Prize in 2000. Landau Prize in 2005. Gödel Prize in 2005; Israel Prize for mathematics in 2008 EMET Prize for mathematics in 2011.	383
4	Geoffrey Zweig	Research Manager & Principal Researcher	Microsoft	IBM Outstanding Innovation Award, 2005	364
5	Marc P. C. Fossorier	Professor	University of California, Davis	NSF Career Award form Promising Young Faculty, 1998, 2001. Medal for Excellence Research by Regents	296
6	Sara Comai	Professor	Department of Electrical & Information Technology, Milano, Italy	Chorafas award, 2000	297
7	Mary Jean Harrold	Professor	Georgia Institute of Technology	IEEE Fellow "for contributions to software systems, 2011. Microsoft Third Top Author in Software engineering of All Time, 2013. ACM Named Top Ranking Software Engineering Researcher in World, 2007.	294
8	Mladen Berekovic	Professor	IDA Institute TU Braunschweig		71
9	Robert E. Schapire	Professor	Princeton University	Paris Kanellakis Theory & Practice Award Received Gold Prize in 2003	398
10	Shin-Ichi Nakano	Professor	Gunma University, Kiryu, Japan	Imperial Prize of Japan academy Imperial Academy Prize Duke of Edinburgh Prize	155
11	Shishir Shah	Associate Professor	University of Houston Houston, Texas Area,	Technical Excellence Award, 1994	145
12	Mohammed H. Sqalli	Assistant Professor	KFUPM, Dhahran, Saudi Arabia.	KFUPM Excellence in Teaching Award, 2013 KFUPM Excellence in Academic Advising Award, 2012 Second Prize Student Paper Award, 1996	61
13	Christine E. Parent	Assistant Professor & Post-Doctoral Researcher	University of Texas, Austin	NSERCC Postdoctoral Fellowship, 2010 SFR Excellence Teaching Award, 2008 SFU Teaching Assistant Award, 2008 SFU Graduate Fellowship, 2005 CU Graduate Fellowship, 1998-2000 German Embassy Prize, 1999	115
14	Sangyoon Lee	Assistant Professor	Department of Economics University of Wisconsin-Madison	Korean Culture and Entertainment Awards, 2012 MBC Drama Award, 2010 Korean Broadcasting Awards, 2009	10
15	Grace Ngai	Assistant Professor	Polytechnic University Hong Kong	Royal Bank of Canada Award, 1999, 2000 McDonald Scholarship, 1994	330
16	PengJun Wan	Associate Professor	Illinois Institute of Technology, Chicago	ACM Sigmobile Distinguhshed Service Award, 2008	860
17	Dirk Jonscher	Professor	University of Geneva		56
18	Thomas C. Rindflesch	Information Research Specialist	National Library of Medicine, Washington D.C.	Fellow American College of Medical Informatics, 2005 NLM Group awards, 2006 MFGRD award, 2009	135
19	Laurent Thery	Project Manager	Valeo Engine& Electrical Systems, Paris, France		105
20	Timothy A. Davis	Professor	University of Florida	Fellow Society of Industrial and Applied Mathematics, 2013	937
21	Peter J. Denning	Distinguished Professor	GMU E-Center for E-Business, Monterey	IEEE Fellow,1982 AAAS Fellow, 1984 ACM Fellow, 1993 NSF Distinguished Education Fellow, 2007 ISOC Jon Postel Award for CSNET, 2009 CRA Computing Research Award,1989 ACM Distinguished Service Award,1989 Centennial Engineering Award, 1992 ACM Outstanding Contribution Award, 1999 ACM SIGCSE Outstanding CS Educator Award, 1999 ACM Karl Karlstrom Outstanding Educator Award, 1996 ACM SIGCSE Lifetime Achievement Award, 2009	2223
22	Wanda Pratt	Assistant Professor	University of California, Irvine.	SIGCHI Best of CHI honor Award	204
23	Atif M. Memon	Associate Professor	Department of Computer Science University of Maryland	Fellowships Andrew Mellon Foundation Award, 1995 NSF CAREER award, Gold Medal Award	256
24	Daniel C. Alexander	Professor	Centre for Medical Image Computing, Department of Computer Science	IEEE Transaction on Medical Imaging	504
25	Christian Scheideler	Assistant Professor	Department of Computer Science Johns Hopkins University	ACM Symposium Award, 2012	461
26	Jean Berger	Business Development Manager	Express Division at SDL plc, Chicago		179
27	David M. Weiss	Professor	Iowa State University	IEEE Software 25th Anniversary Top Pick Selection ACM SIGSOFT Impact Paper Award Best Retrospective Paper Award from 7th International Conference on Software Engineering	965
28	Moninder Singh	Research Staff Member	Thomas J. Watson Research Center, Yorktown Heights, NY USA	IBM Eminence & Excellence Award, 2012 IBM Outstanding Technical Achievement Award, 2012 IBM Research Accomplishment Awards 2003, 2004, 2008, 2011	931
29	Martina Zitterbart	Professor	Karisruhe Institute of Technology	Alcatel SEL Research Award, 2002	135
30	Laurent Daynes	Senior Staff Engineer	Sun Microsystems Laboratories	OOPSLA 10-Year Most Influential Paper Award, 2011	161

to their profiles in 2014. 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. Additionally, 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 and finding rising cloud services provider in cloud environment.

Table 7: Authors' ranking by score (WMIRank) with total papers and citations.

#	Authors	Scores (WMIRank)	Recent Status	
			# of Papers up to 2014	# of Citations upto 2014
1	Frank K. H. A. Dehne	0.470	154	2643
2	Bertram Ludascher	0.462	101	4943
3	Yossi Azar	0.460	165	6745
4	Geoffrey Zweig	0.458	63	2595
5	Marc P. C. Fossorier	0.457	151	6647
6	Sara Comai	0.453	71	1839
7	Mary Jean Harrold	0.451	131	11195
8	Mladen Berekovic	0.441	58	445
9	Robert E. Schapire	0.438	139	56194
10	Shin-Ichi Nakano	0.432	64	895
11	Shishir Shah	0.427	58	497
12	Mohammed H. Sqalli	0.427	21	127
13	Christine Parent	0.425	60	2487
14	Sangyoon Lee	0.425	42	223
15	Grace Ngai	0.420	36	791
16	Peng-Jun Wan	0.418	155	7366
17	Dirk Jonscher	0.416	9	185
18	David E. Wilkins	0.402	9	1205
19	Knerstin Dautenhahn	0.402	127	4140
20	Sreejit Chakravarty	0.400	96	1155
21	Peter J. Denning	0.399	160	10609
22	Torsten Grust	0.397	52	1922
23	Nevenka Dimitrova	0.397	77	1577
24	Laura M. Haas	0.395	81	6627
25	Christian Scheideler	0.393	117	2222
26	Jean Berger	0.393	16	293
27	Mikio Aoyama	0.393	43	548
28	Daniel Hoffman	0.392	35	1052
29	Dominique Michelucci	0.392	42	595
30	Franck Cappello	0.389	114	3364
		Sum	2447	141126
		Average	81.57	4704.2
		Standard Deviation	49.09	10193.83

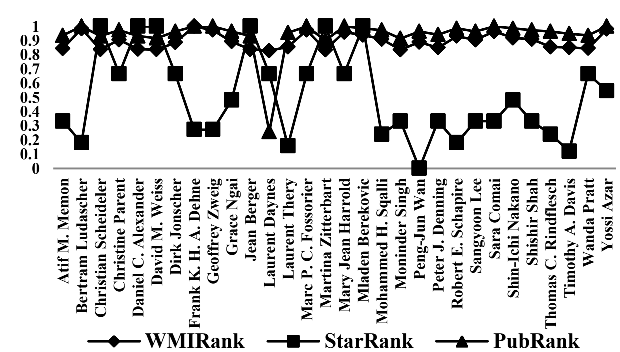

Figure 5: Effect of Damping Factor (0.85) on Authors Score (y-axis is author rank).

Table 8: Authors' position Up by WMIRank w.r.t StarRank.

Author	Score WMIRank	Position in WMIRank	Position in StarRank	Position Up
Frank K. H.A. D.	0.941	1	2	+1
Bertram Ludascher	0.930	2	10	+8
Yossi Azar	0.927	3	6	+3
Geoffrey Zweig	0.924	4	7	+3
Marc P. C. Fossorier	0.923	5	20	+15
Sara Comai	0.917	6	22	+18
Mary Jean Harrold	0.914	7	29	+22
Mladen Berekovic	0.901	8	25	+17
Shishir Shah	0.882	11	23	+12
Mohammed H. Sqalli	0.882	12	27	+15
Christine Parent	0.879	13	24	+11
Sangyoon Lee	0.879	14	26	+12
Peng-Jun Wan	0.869	16	30	+14
Christian Scheideler	0.835	25	28	+3

Table 9: Authors' position down by WMIRank w.r.t StarRank

Author	Score WMIRank	Position in WMIRank	Position in StarRank	Position Down
Robert E. Schapire	0.896	9	1	-8
Shin-Ichi Nakano	0.888	10	3	-7
Grace Ngai	0.871	15	13	-2
Dirk Jonscher	0.867	17	12	-5
Thomas C. Rindflesch	0.847	18	17	-1
Laurent Thery	0.847	19	14	-5
Timothy A. Davis	0.844	20	15	-5
Peter J. Denning	0.843	21	16	-5
Wanda Pratt	0.840	22	19	-3
Atif M. Memon	0.839	23	11	-12
Daniel C. Alexander	0.837	24	21	-3
Jean Berger	0.835	26	4	-22
David M. Weiss	0.835	27	18	-9
Moninder Singh	0.833	28	9	-19
Martina Zitterbart	0.833	29	5	-24
Laurent Daynes	0.829	30	8	-22

Table 10: Authors' position up by WMIRank w.r.t PubRank.

Author	Score WMIRank	Position in WMIRank	Position in PubRank	Position Up
Frank K. H. A. D.	0.941	1	2	+1
Yossi Azar	0.927	3	6	+3
Geoffrey Zweig	0.924	4	5	+1
Sara Comai	0.917	6	7	+1
Mladen Berekovic	0.901	8	10	+2
Sangyoon Lee	0.879	14	17	+3
Grace Ngai	0.871	15	16	+1
Laura M. Haas	0.837	24	25	+1

Table 11: Authors' position down by WMIRank w.r.t PubRank.

Author	Score WMIRank	Position in WMIRank	Position in PubRank	Position Down
Bertram Ludascher	0.930	2	1	-1
Marc P. C. Fossorier	0.923	5	3	-2
Mary Jean Harrold	0.914	7	4	-3
Shin-Ichi Nakano	0.888	10	8	-2
Peng-Jun Wan	0.869	16	15	-1
Dirk Jonscher	0.867	17	14	-3
Christian Scheideler	0.835	25	24	-1

REFERENCES

[1] Abbasi, R.A., Maqbool, O., Mushtaq, M., Aljohani, N.R., Daud, A., Alowibdi, J.S. and Shahzad, B., Saving lives using social media: Analysis of the role of twitter for personal blood donation requests and dissemination. Telematics and Informatics., ISSN 0736-5853, http://dx.doi.org/10.1016/j.tele.2017.01.010 (2017).

[2] Amjad, T., Ding, Y., Daud, A., Xu, J., and Malic, V. Topic-based Heterogeneous Rank. Scientometrics, 104 ,1 (2015), 313-334.

[3] Amjad T, Ding Y, Xu J, Zhang C, Daud A, Tang J, Song M. Standing on the shoulders of giants. Journal of Informetrics. 11(1):307-23 (2017).

[4] Daud, A. Using Time Topic Modeling for Semantics-Based Dynamic Research Interest Finding. Knowledge-Based Systems (KBS). 26 (2012): 154-163.

[5] Daud, A., Li, J., Zhou, L., and Muhammad, F. Temporal Expert Finding through Generalized Time Topic Modeling. Knowledge-Based Systems (KBS), 23, 6(2010), 615-625.

[6] Daud, A., Li, J., Zhou, L., and Muhammad, F. A Generalized Topic Modeling Approach for Maven Search. In Proceedings of International Asia-Pacific Web Conference and Web-Age Information Management, 2009. 138-149.

[7] Daud, A., Abbasi, R., and Muhammad, F. Finding Rising Stars in Social Networks. Database Systems for Advanced Applications (LNCS), 2013, 7825, 13-24.

[8] Daud, A., Ahmad, M., Malik, M., and Che, D. Using machine learning techniques for rising star prediction in co-author network. Scientometrics, *102,*2 (2015), 1687-1711.

[9] Guns, R., and Rousseau, R. (2014). Recommending research collaborations using link prediction and random forest classifiers. Scientometrics. *101*, 2(2014), 1461-1473.

[10] Huang, S., Yang, B., Yan, S., and Rousseau, R. Institution name disambiguation for research assessment. Scientometrics, *99*,3(2014), 823-838.

[11] Khan, H.U., Daud, A., Ishfaq, U., Amjad, T., Aljohani, N., Abbasi, R.A. and Alowibdi, J.S., Modelling to identify influential bloggers in the blogosphere: A survey. Computers in Human Behavior, 68, pp.64-82 (2017).

[12] Li, X.K., Foo, C.S., Tew, K.L., and Ng, S.K. Searching for Rising Stars in Bibliography Networks. In Proceedings of the 14th International Conference on Database Systems for Advanced Applications 2009, 288-292.

[13] Long, P.M., Lee, T. K., & Jaffar, J. Benchmarking Research Performance in Department of Computer Science. School of Computing, National University of Singapore 1999, http://www.comp.nus.edu.sg/~tankl/bench.html.

[14] Sekercioglu, C. H. Quantifying Coauthor Contributions. Science, 322, 5900(2008), 371.

[15] Shoaib, M., Daud, A., Khiyal, M. S. H. Improving Similarity Measures for Publications with Special Focus on Author Name Disambiguation. Arabian Journal of Science and Engineering, 40, 6(2015), 1591-1605.

[16] Shin, D., Kim, T., Choi, J., and Kim, J. Author name disambiguation using a graph model with node splitting and merging based on bibliographic information. Scientometrics, 100,1(2014), 15-50.

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

[18] Tsatsaronis, G., Varlamis, I., and Norvag, k. How to Become a Group Leader? Or Modeling Author Types Based on Graph Mining. LNCS 2011,6966. 15-26.

[19] Wang, G.A., Jiao, J., Abrahams, A. S., Fan, W., and Zhang, Z. Expert Rank: A topic-aware expert finding algorithm for online knowledge communities. Decision Support Systems, 54,3 (2013), 1442-1451.

[20] Wu, H., Li, B., Pei, Y., and He, J. Unsupervised author disambiguation using Dempster–Shafer theory. Scientometrics, 101,3 (2014), 1955-1972

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

[22] Yan, R., Huang, C., Tang, J., Zhang, Y., and LI, X., To Better Stand on the Shoulder of Giants. JCDL '12 Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries New York, 2012, 51-60.

[23] Zhang, G., Ding, Y., and Milojevic, S. Citation content analysis (CCA): A method for syntactic and semantic analysis of citation content. Journal of the American Society for Information Science & Technology, 64,7(2013), 1490-1503.