Researchpedia
Open Access Journals

Received: 28 May 2020; **Revised:** 29 September 2021; **Accepted:** 19 December 2021; **Published:** 26 December 2021
Journal Name [Online ISSN Coming Soon], Volume 2, Issue 1, Article 10, Pages 111-116, Dec 2021

Digital Object Identifier 10.1111/RpJC.2020.DOINumber

All that Glitters is not Gold: Falsely Predicted Rising Stars

Ali Daud[1], Tehmina Amjad[2], Tayyaba Khaliq[2], and Malik Khizar Hayat[3]
[1]Department of Computer Science and Artificial Intelligence, University of Jeddah, Jeddah, Saudi Arabia
[2]Department of Computer Science and Software Engineering, International Islamic University, Islamabad, Pakistan
[3]Department of Information Technology, University of Haripur, Pakistan

Corresponding author: Tehmina Amjad (e-mail: tehminaamjad@iiu.edu.pk).

ABSTRACT Finding the rising stars in any domain is an attention-grabbing research topic these days with its significant application in various domains. Existing literature covers some methods that find the possible rising stars in domains including community question answering networks, bibliographic networks, sports networks, and telecommunication networks. Results of these methods are then used to select the competent personals for the corresponding field. Thus, these methods have great significance. Unfortunately, due to the unavailability of ground truth, the accuracy of the method and goodness to results is a challenging task. As all that glitters is not always gold, so identifying the false positive cases is equally essential. In this study, we analyze the bibliometric networks as a case study and investigated the reasons behind the falsely predicted rising stars. Less productivity and low-impact venue publications are the significant reasons of false predictions.

Keywords Finding Rising Stars, Bibliometric Networks, Machine Learning, Classification, Prediction, Academic Social Networks.

I. INTRODUCTION

Rising Stars are the individuals who might be at the start of their career but have the potential to reach high ranks in very short duration. The task of Finding Rising Stars (FRS) has a vital role in engaging the right young individuals on available junior positions, to escalate the performance and productivity of an organization and to exploit the advantages of their dynamic and vibrant comportment.

FRS methods can be very helpful in many fields like bibliometric networks, social networks like the forums and blogs and the production companies. Considerable work is done in the field of the ranking of academic entities in several types of bibliometric networks [1] including ranking of authors [2]–[9], ranking of journals [10]–[13] and publications [14]–[17]. FRS methods are also mostly explored for the bibliographic networks; however, their implications can be very wide. Consider an example of Human Recourse section, when a new employee needs to be appointed, the organization would like to select the best suitable candidate for the position available. If the FRS method predicts the talent of the applicant of becoming an expert shortly, it will be highly beneficial and in the best interest of the organization. Similarly, in the marketing and business sector, organizations used to investigate a new product with various business-oriented strategies before its official launch to ascertain that it will be able to gain the attention of many consumers soon or not. Thus, it is substantial for decision-making authorities to distinguish a nominee who has a high aptitude for becoming a rising star [18].
Significant work has been done for FRS in Bibliometric Networks (BNs) [19]–[24]. Some work was also done in the domains of Community Question Answer (CQA) networks [25], and sports networks [26]. Tsatsaronis et al. presented the typical evolution patterns for a person's career over time [27]. These are represented in Figure 1 as risings stars, well-established, stable entities, and declining entities.

Unfortunately, the ground truth measures results are not obtainable for the performance evaluation of the FRS methods. There can be some authors who were predicted as rising stars but failed to prove their predicted potential (false positive). Organizations have a great deal of interest in the accuracy of results of FRS methods as the future productivity of their businesses highly depends on the selections that are made based on these methods. Different methods and enhancements have been performed for improving the efficiency of finding rising star methods. In this research, we perform an analytical study to find the reasons behind falsely predicted rising stars. We choose the academic social network extracted from DBLP for the experimentation as most of the work for the FRS problem was done using bibliographic networks. To the best of our

knowledge, this is the first attempt to find the reasons behind the falsely predicted rising stars. The major contributions of this work are as follows:

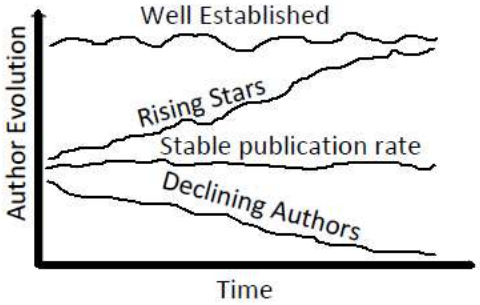

FIGURE 1. The Evolution of a Person's Career over Time [27]

Unfortunately, the ground truth measures results are not obtainable for the performance evaluation of the FRS methods. There can be some authors who were predicted as rising stars but failed to prove their predicted potential (false positive). Organizations have a great deal of interest in the accuracy of results of FRS methods as the future productivity of their businesses highly depends on the selections that are made based on these methods. Different methods and enhancements have been performed for improving the efficiency of finding rising star methods. In this research, we perform an analytical study to find the reasons behind falsely predicted rising stars. We choose the academic social network extracted from DBLP for the experimentation as most of the work for the FRS problem was done using bibliographic networks. To the best of our knowledge, this is the first attempt to find the reasons behind the falsely predicted rising stars. The major contributions of this work are as follows:

- Identifying falsely predicted rising stars
- Analyzing the reasons for the author's decline
- Improving existing FRS techniques
- The remainder of the paper is structured as follows: Section II reviews the literature, Section III provides the insights of the proposed method, Section IV discusses the experimentation and results and lastly, Section V concludes the findings.

II. RELATED WORK

The idea of FRS was first introduced by Li et al. as PubRank [18] in 2009. PubRank algorithm finds the evolving links in the social network of researchers by considering the mutual influence nodes including authors and papers as well as the dynamic features of the network over time. Afterwards, many people explored the idea in different dimensions [28]–[31]. Topic modeling methods have usually been applied in the past to identify the research interests of researchers. Observing the scientific growth, the trending topics can be identified as Stable, Hot, or Cold. Finding rising stars (junior researchers, who are at the start of their career) from a bibliometric network is a challenging task, specifically if the researchers have an interest in multiple sub-domains or are working on diverse topics. Existing methods for finding rising stars explore the co-author networks or citation networks, and ignore the textual content, which may help in finding rising stars through hot topics detection over time.

A publication contributing to a hot topic can be an indication that the author of that publication may be a rising star and can become an expert in that domain in the future. This study proposes the Hot Topics Rising Star Rank (HTRS-Rank) method for finding rising stars by detecting hot topics. HTRS-Rank finds the junior scholars, who contribute to hot topics at the start of their career and ranks them based on the presence of hot topics in their publications. AMiner five years dataset ranging from 2005–2009 is selected for experimentation. Top 10 researchers are considered to measure the association strength using rank correlation among HTRS-Rank and baseline methods. Experimental results show the efficiency of HTRS-Rank in comparison to the baseline methods. The proposed HTRS Rank (TF–IDF) provides low standard deviation for productivity, citations and sociality as compared to baseline methods for more social and highly cited authors. It is identified that HTRS-Rank (WordNet) emphasizes the semantic similarity of two sentences, whereas HTRS-Rank (TF–IDF) scheme emphasizes the uniqueness or importance of each term, therefore TF–IDF approach performs better than WordNet approach due to having higher correlation with StarRank and WMIRank [28]–[32] . Tsatsaronis et al. examined the changing aspects of the research profiles of authors [27]. With the help of unsupervised learning methods, they grouped the researchers into four categories including rising stars, well-established, stable, and declining. Two significant features were incorporated in PubRank and StarRank [22], which were

proposed that retains the features used by PubRank and involves the involvement of the researchers based on the mutual influence of their co-authors on each other and also handles the number of publications dynamically.

Daud et al. proposed a prediction-based method for FRS by examining co-author networks [20]. In order to forecast evolving scientists, they employed machine learning methods for FRS using publications, co-author, and venue-based features and their combinations. Later, WMIRank [32] was proposed which involves the weighted mutual influence of co-author's citations. They considered the order in which the author's name appears in a published article and the venue of publication. The results show that WMIRank can be successfully used to find future experts.

A machine learning-based study was conducted that focuses on the Pakistani scholarly society for forecasting evolving researchers in the scholarly bodies of Pakistan [24]. Classification models were applied using co-authors, authors, and venues as classification features, and the effect of these features was empirically investigated. The trialing was conducted on data of scholars and academicians of Pakistan crawled from the Web of Sciences.

III. PROPOSED METHOD

This is an analytical study in which we applied the StarRank [22] method on a dataset extracted from DBLP ranging from 1996 to 2014. The data from 1996 to 2000 is pre-processed and the performance of all predicted rising stars is analysed in years ranging from 2001 to 2014. The predicted rising stars are analysed and verified either they have proved themselves as a future expert or not.

The steps performed in the methodology are as follows:

- Data from DBLP is divided on a year-wise basis into three equal-size chunks,
- Number of citations and publications are considered
- Authors with low citations and publications are identified as false positives
- Timewise analysis is performed and six major reasons for falsely predicted rising stars are identified
- Profiles of authors are visited to find the possible reasons for false predictions

Figure 2 shows the details of the steps performed.

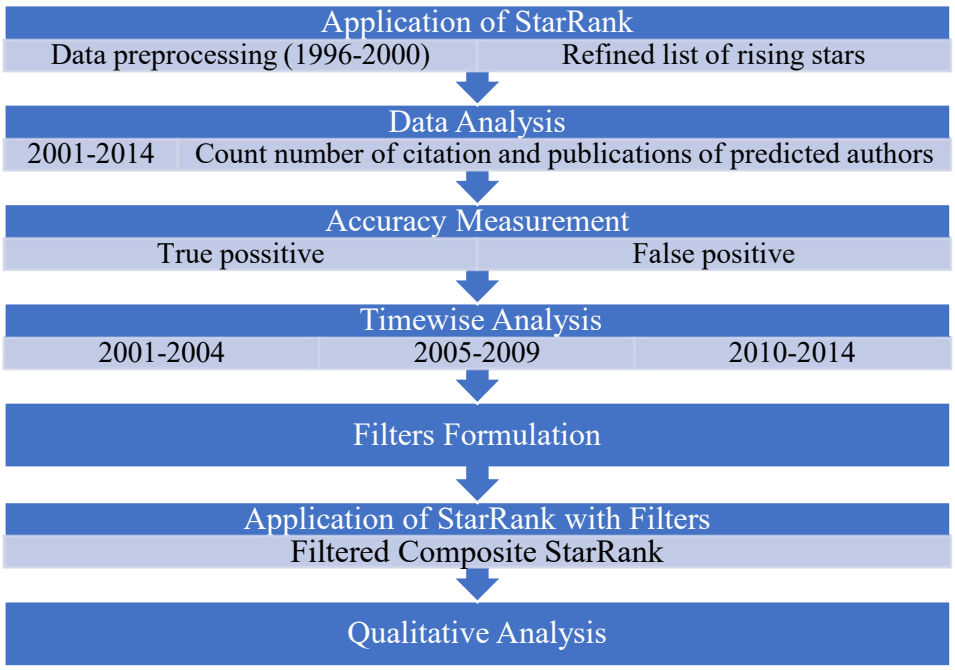

FIGURE 2. The Methodology

To get the rising stars list to analyze their further performance, the StarRank method is implemented. Data obtained from DBLP is preprocessed and is divided into three parts of five years each, as mentioned in Figure 2. Statistics of the dataset are shown in Table 1. StarRank is applied on divided chunks of the dataset and got a list of authors with StarRank scores. We sorted the authors according to their score value and selected the top 50 authors to analyze their performance. We formulated the filters in such a way that expected false-positive cases can be removed to improve the performance of the FRS method.

TABLE 1. Dataset Statistics (1996-2014)

Variables	Count
Authors	194181
Publications	230495
Venues	1977

IV. RESULTS AND DISCUSSION

To analyze the performance and output of authors ranked by StarRank, we calculated the number of citations and publications of the top fifty authors. Authors who are predicted as a future rising star and achieved their expectations are categorized as true positives. Authors who are predicted as future rising stars but failed to achieve their expectations are categorized as false positives. Among the top fifty authors, we found fifteen authors as false positives, the remaining thirty-five listed as true positives as per their performance found on the web. We considered the authors as true positives if they have several citations greater than or equal to 500. If an author has citations less than 500, she is considered as a false positive.

The dataset is divided into further three chunks 2001-2004, 2005-2009, and 2010-2014, and counted the number of citations and publications for all predicted rising star authors as shown in the Table 2. Google is used as a search engine and various social and professional sites including their homepages, google scholar profiles, DBLP profiles, and similar ones are visited to track the performance of authors and also for analyzing the reasons of failure in showing predicted performance. Their number of citations, number of publications, venues of publications, the impact factor of venues, the effectiveness of co-authorship, and their order of appearances, the affiliation of authors during different periods, and the ranking of affiliated organizations are verified.

StarRank considers the dynamic ranking of venues based on entropies. After a detailed qualitative analysis of information available on the internet about the false-positive rising stars, we came up to find two major reasons that are (1) authors with a smaller number of publications and (2) venues with a high entropy value. Based on these reasons, we proposed that authors with many publications with less than a decided threshold value must be removed from the data before the application of the FRS method to improve the performance of these methods. Secondly, we propose that the dataset shall be refined based on the entropy of the publication's venues. Entropy is used for the performance evaluation of StarRank in the existing literature. The journals which are more topic-specific and are strict criteria for accepting a manuscript have a low entropy, while on the other hand, the journals that accept manuscripts from a wider range of fields have a higher entropy. We applied the formulated filters on StarRank and proposed a new method – Filtered Composite (FC) StarRank. By applying the FC StarRank, we generated an improved resultant list of rising stars.

TABLE 2. Timewise Analysis of False Positive Authors*

Year	2001-2005		2006-2010		2011-2014		2017	
Author name	**Cit**	**Pub**	**Cit**	**Pub**	**Cit**	**Pub**	**Cit**	**Pub**
W. Dean Bidgood Jr.	0	0	0	0	0	0	0	0
Steven Fraser	9	15	4	28	0	4	0	0
Debra A. Hensgen	180	2	9	1	0	0	0	0
Jerry Baulier	0	0	0	0	0	0	0	0
Michel Barreteau	0	0	0	2	1	1	0	0
RiittaSmeds	0	3	0	3	0	0	0	0
Carol Harger	0	0	0	0	0	0	0	0
Aijun Li	0	4	0	8	0	4	0	0
Maria GiuseppinaCampi	0	0	0	0	0	0	0	0
HariBalachandran	16	2	0	0	0	0	0	0
Uwe Engelmann	0	5	0	1	0	0	0	0
Lucimar F. de Carvalho	0	4	0	0	0	0	0	0
Chin-Hsiung Wu	8	9	1	3	0	1	0	0

*The values shown in the table are subject to the statistics of the dataset.

Further study of the publication rate, citations received, their affiliations, the prestige of the venues where they publish, we were able to categorize the authors into five categories.

The first category contains the authors who appeared for a short period. Their publications period started from 1996 to onwards and ends in the first decade of 2000. Some of them published 2 to 7 papers with good co-authors and received good citations. Their web pages are visited by searching their names in Google and we found the affiliations of these authors. When we visit the website of their affiliated institutes, IT companies and research centers their record is not found on concerned websites. There is no profile found neither on social sites nor on professional sites.

The second category contains the authors who were active in research for a short time. They published a limited number of publications, about 2 to 7 papers. They had good co-authorship and received a good number of citations against limited

publications. Later on, they changed their affiliation and started working in different fields rather than research. Some of them have a good professional record and some of them have not.

The third category contains the authors with a good professional record and working on key posts now infamous IT companies as well as top-ranked universities. They have fewer papers and received fewer citations but their work is continuous. It was slow but steady.

The fourth category contains the authors with a good professional record, working continuously in IT companies and institutes. Although their productivity was incessant, they published an average number of papers and received fewer citations because of low-ranked venues. The venues where papers are published do not have a high impact.

The fifth category contains the authors with low-ranked affiliations; they do not belong to high-ranked institutions or organizations. They produced continuous but slow pace hence resulting in fewer publications. Publication venues are mostly low ranked and ultimately received fewer citations. We further summarize the findings into five reasons as follows:

1. The author is not affiliated with the high-ranked institute.
2. The author stopped pursuing her career as a researcher and changed her interests. She diverted from research, got indulged in other activities, for example, she started working as a software engineer or any position in the industry.
3. The author stopped research work during her first decade and further we were not able to find any record of their services in any other field using the internet.
4. Author changed her field of interest over time and started research in other fields.
5. Author published her articles in low-impact venues.
6. Author published papers with less Author Contribution Weight (ACW). The author's contribution-based mutual influence is verified. Co-authorship is ineffective later in the first and second decade of 2000.

Table 3 shows the false positive authors and the reasons that were identified behind their false prediction and the category to which they belong to.

TABLE 3. List of False Positive Authors

Sr. No	Author Name	Category	Reason No
1	W. Dean Bidgood Jr.	2nd	2,6
2	Steven Fraser	4th	5
3	Debra A. Hensgen	2nd	2
4	Jerry Baulier	2nd	2
5	Michel Barreteau	4th	4,6
6	RiittaSmeds	3rd	6
7	Carol Harger	1st	3
8	Aijun Li	5th	1,6
9	Maria GiuseppinaCampi	1st	3
10	HariBalachandran	1st	3,6
11	Uwe Engelmann	2nd	2
12	Lucimar F. de Carvalho	4th	4,5,6
13	Chin-Hsiung Wu	5th	1

VII. CONCLUSION

Finding the rising stars is a very interesting research direction with its application in various domains. The problem of FRS has significant importance for the organizations that apply these methods while selecting suitable candidates for different positions. The falsely predicted rising stars if selected by an organization, can cause damage to the performance, and expected output of that organization. In this study, we considered the bibliometric networks as a field of study and analysed the reasons behind the false prediction of rising stars. The top fifty authors predicted by the StarRank method are analysed and we came up with significant reasons behind false predictions. From the analysis of results, it can be concluded that the authors can have different reasons for the decline majorly including less productivity and publication in low-impact journals. In the future, we are interested in exploring the reason for false prediction in other networks.

REFERENCES

[1] T. Amjad, A. Daud, and N. R. Aljohani, "Ranking authors in academic social networks: a survey," *Libr. Hi Tech*, vol. 36, no. 1, pp. 97–128, 2018.
[2] T. Amjad and A. Daud, "Indexing of authors according to their domain of expertise," *Malays. J. Libr. Inf. Sci.*, vol. 22, no. 1, pp. 69–82, 2017.
[3] T. Amjad, A. Daud, A. Akram, and F. Muhammed, "Impact of mutual influence while ranking authors in a co-authorship network," *Kuwait J. Sci.*, vol. 43, no. 3, pp. 101–109, 2016.

[4] T. Amjad, A. Daud, D. Che, and A. Akram, "MuICE: Mutual Influence and Citation Exclusivity Author Rank," *Inf. Process. Manag.*, vol. 52, no. 3, pp. 374–386, 2015.
[5] T. Amjad *et al.*, "Standing on the shoulders of giants," *J. Informetr.*, vol. 11, no. 1, pp. 307–323, 2017.
[6] M. Ausloos, "A scientometrics law about co-authors and their ranking: the co-author core," *Scientometrics*, vol. 95, no. 3, pp. 895–909, 2013.
[7] D. Bouyssou and T. Marchant, "Ranking authors using fractional counting of citations: An axiomatic approach," *J. Informetr.*, vol. 10, no. 1, pp. 183–199, 2016.
[8] M. Dunaiski, J. Geldenhuys, and W. Visser, "Author ranking evaluation at scale," *J. Informetr.*, vol. 12, no. 3, pp. 679–702, 2018.
[9] M. Dunaiski, J. Geldenhuys, and W. Visser, "How to evaluate rankings of academic entities using test data," *J. Informetr.*, vol. 12, no. 3, pp. 631–655, 2018.
[10] K. C. Chan, G. S. Seow, and K. Tam, "Ranking accounting journals using dissertation citation analysis: A research note," *Account. Organ. Soc.*, vol. 34, no. 6, pp. 875–885, 2009.
[11] F. L. DuBois and D. Reeb, "Ranking the international business journals," *J. Int. Bus. Stud.*, vol. 31, no. 4, pp. 689–704, 2000.
[12] D. Bouyssou and T. Marchant, "Consistent bibliometric rankings of authors and of journals," *J. Informetr.*, vol. 4, no. 3, pp. 365–378, 2010.
[13] C.-F. Tsai, "Citation impact analysis of top ranked computer science journals and their rankings," *J. Informetr.*, vol. 8, no. 2, pp. 318–328, 2014.
[14] D. Hong, F. Baccelli, and others, "On a joint Research Publications and Authors Ranking," 2012, Accessed: Jun. 24, 2014. [Online]. Available: http://hal.archives-ouvertes.fr/hal-00666405/
[15] D. Peiris and R. Weerasinghe, "Citation network based framework for ranking academic Publications and venues," in *Advances in ICT for Emerging Regions (ICTer), 2015 Fifteenth International Conference on*, 2015, pp. 146–151. Accessed: Jun. 23, 2016. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7377681
[16] A. Perianes-Rodríguez and J. Ruiz-Castillo, "Multiplicative versus fractional counting methods for co-authored publications. The case of the 500 universities in the Leiden Ranking," *J. Informetr.*, vol. 9, no. 4, pp. 974–989, 2015.
[17] M. Schreiber, "Fractionalized counting of publications for the g-Index," *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, no. 10, pp. 2145–2150, 2009.
[18] X.-L. Li, C. S. Foo, K. L. Tew, and S.-K. Ng, "Searching for rising stars in bibliography networks," in *Database Systems for Advanced Applications*, 2009, pp. 288–292. Accessed: Apr. 30, 2016. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-00887-0_25
[19] L. Li and H. Tong, "The child is father of the man: Foresee the success at the early stage," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 655–664. Accessed: Apr. 04, 2016. [Online]. Available: http://dl.acm.org/citation.cfm?id=2783340
[20] A. Daud, M. Ahmad, M. S. I. Malik, and D. Che, "Using machine learning techniques for rising star prediction in co-author network," *Scientometrics*, vol. 102, no. 2, pp. 1687–1711, 2015.
[21] G. Panagopoulos, G. Tsatsaronis, and I. Varlamis, "Detecting rising stars in dynamic collaborative networks," *J. Informetr.*, vol. 11, no. 1, pp. 198–222, 2017.
[22] A. Daud, R. Abbasi, and F. Muhammad, "Finding rising stars in social networks," in *Database Systems for Advanced Applications*, 2013, pp. 13–24. Accessed: Apr. 30, 2016. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-37487-6_4
[23] J. Zhang *et al.*, "Cocarank: A collaboration caliber-based method for finding academic rising stars," in *Proceedings of the 25th International Conference Companion on World Wide Web*, 2016, pp. 395–400.
[24] T. Amjad, A. Daud, S. Khan, R. A. Abbasi, and F. Imran, "Prediction of Rising Stars from Pakistani Research Communities," in *2018 14th International Conference on Emerging Technologies (ICET)*, 2018, pp. 1–6.
[25] L. T. Le and C. Shah, "Retrieving rising stars in focused community question-answering," in *Asian Conference on Intelligent Information and Database Systems*, 2016, pp. 25–36. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-662-49390-8_3
[26] H. Ahmad, A. Daud, L. Wang, H. Hong, H. Dawood, and Y. Yang, "Prediction of rising stars in the game of cricket," *IEEE Access*, vol. 5, pp. 4104–4124, 2017.
[27] G. Tsatsaronis *et al.*, "How to become a group leader? or modeling author types based on graph mining," in *Research and Advanced Technology for Digital Libraries*, Springer, 2011, pp. 15–26. Accessed: Apr. 09, 2016. [Online]. Available: http://link.springer.com/10.1007%2F978-3-642-24469-8_4
[28] X. Kong, Y. Shi, S. Yu, J. Liu, and F. Xia, "Academic social networks: Modeling, analysis, mining and applications," *J. Netw. Comput. Appl.*, vol. 132, pp. 86–103, 2019.
[29] A. Daud *et al.*, "Finding rising stars in bibliometric networks," *Scientometrics*, pp. 1–29, 2020.
[30] A. Daud, F. Abbas, T. Amjad, A. A. Alshdadi, and J. S. Alowibdi, "Finding rising stars through hot topics detection," *Future Gener. Comput. Syst.*, vol. 115, pp. 798–813, Feb. 2021, doi: 10.1016/j.future.2020.10.013.
[31] A. Nawaz and M. S. I. Malik, "Rising stars prediction in reviewer network," *Electron. Commer. Res.*, pp. 1–23, 2021.
[32] A. Daud *et al.*, "Finding Rising Stars in Co-Author Networks via Weighted Mutual Influence," in *Proceedings of the 26th International Conference on World Wide Web Companion*, 2017, pp. 33–41. Accessed: Apr. 22, 2017. [Online]. Available: http://dl.acm.org/citation.cfm?id=3054137