Advisor-Advisee Relationship Mining Based on Co-author Network

Shaohong Zhang, Zongbao Yang*, Jianyu Liu, Zhiqian Zhang, Xiaofei Xing, Ying Gao
Department of Computer Science, GuangZhou University, Guangzhou, P.R. China.

* Corresponding author. Tel.: 17718233721; email: wiseben@qq.com.
Manuscript submitted May 4, 2019; accepted July 20, 2019.
doi: 10.17706/jsw.14.9.388-399

Abstract: Advisor-advisee relationship among scholars is important in the academia circle. It contains abundant information about the academic inheritance, advisor recommendation and the forming of research communities, etc. The advisor-advisee relationship is always hiding behind the co-author network, however, there are some challenges when mining this kind of relationship. This relationship is always changing with time, the size of labeled data is limited and the authors' name ambiguity, etc. Previous works are focused on various aspects, including the citation network, the publication network and the co-author network, etc. To our best knowledge, all of these works are focused on the whole network, and none of them considered the credit allocation of the authors in each paper. Therefore, the relationship mining results may be influenced greatly by some high degree nodes. In this paper, we proposed a new method to solve this problem with the scholar data in DBLP. The credit allocation of each author is calculated, and the co-author network of DBLP is cut into smaller networks based on the characteristic. Then, the advisor-advisee relationship among researchers is mined based on these smaller co-author network. The results show that, the accuracy of this model is about 62.5%, however, this is an unsupervised method, which could save the time of training model and will not be influenced by the uncompleted training data set.

Key words: Advisor-advisee relationship, author credit allocation, co-author network, relationship mining.

1. Introduction

With the rapid growth of Internet and other communication platform, there are so many data were created every year. We are arriving at the era of big data. Information network contains lots of knowledge about relationships among people or other entities. For example, the advisor-advisee relationship is hidden in the co-author network; the friendship and family relationship are hidden in the social media network; the manager and worker relationship are hidden in the company management network.

Several projects aim to collect and show these kinds of relationships, such as LinkedIn[1], NeuroTree[2], Academic Family Tree[3], Mathematics Genealogy Project[4], etc. During which, users have to label their professional relationships, such as colleagues' relationship, in the LinkedIn themselves. The NeuroTree, Academic Family Tree and the Mathematics Genealogy Project require users to add the advisor-advisee relationship, too. In these projects, if someone you are interested in are not in the database, you might

[1] https://www.linkedin.com/
[2] https://neurotree.org/neurotree/
[3] https://academictree.org/
[4] https://www.genealogy.math.ndsu.nodak.edu/

cannot gain his academic tree, unless his information had been inserted in the database. In other words, all the information in these projects are labeled manual in a great extent. Therefore, it is necessary to design a method to uncover the hidden relationship types from the network automatically. In fact, there are some existing works aiming at discovering the relationship from complex information network [1]-[6], which provide a solid support for the analysis of co-author network.

The co-author network could be obtained from the published papers. Scientific papers are good way to share the achievements and innovation ideas among researchers. In recent years, there are more and more papers were published in different journals and conferences. Therefore, a great number of scholar data was produced. It is widely known that, scholar data contains the journal and conference papers, published books and patents, which have millions of authors, citations, etc. The massive scale related data is also including the co-author network, citation network, etc. We are facing the big scholar data [7], which is becoming a great challenge for the academic and industry to manage and analysis.

There is so many useful information hidden in big scholar data. So many researchers have analyzed it from different point of views. Such as, the ranking of journals and conferences [8]-[14], the credit allocation of each author [15]-[18], the influence of different scholars [19]-[22] and the finding of rising star in academic [23]-[26], etc. Besides, some popular and efficient academic research engines are created, including the Google Scholar[5], Microsoft Academic Search[6], Digital Bibliography & Library Project[7], Baidu Scholar[8], and Amier[9], etc. From which, we could gain almost all the useful information conveniently, however, none of them provide the relationship about advisor-advisee.

In fact, mining the advisor-advisee relationship have a significant meaning for the academic circle. For example, it can improve the effect of double-blind peer review, construct the academic graph, and help to understand the academic inheritance. Besides, if we know the advisor-advisee relationship between researchers, we could discover how the research communities have been formed and how many talents the advisor has fostered easily. Therefore, we could do more analysis on how a researcher influenced the academic research community and how the research topics have been emerged and evolved. The advisor-advisee relationship can also provide a good reference for people to find an appropriate scholar, because people might pay attention on not only the personal academic achievement but also how many related experts one has forested [27].

As is known to all, the advisor-advisee relationship among the researchers is hidden in the co-author network and no co-author pair could be neglected, since every collaboration could potentially be a context of nurturing [28]. There are some open web sites are available to retrieve the advisor(s) of a researcher. Such as the NeuroTree, Mathematics Genealogy Project, The Academic Family Tree. During which, the Mathematics Genealogy Project is more related to the subject of Math. The Academic Family Tree contained great number of fields, such as the Education, Philosophy, History and computer science, etc.

However, because of the limiting of data set and the characteristic of the advisor-advisee relationship (it will change with the time), it is a great challenge to deal with this problem in a very high degree of accuracy.

The rest part of this paper is organized as fellows. Some related works are introduced in section 2; In section 3, we formulate the problem and introduce the approach of this paper detailed; In section 4, we show the experiment results; Finally, several conclusions and the future work are given according to our research results in section 5.

[5] http://scholar.google.com/
[6] https://academic.microsoft.com/
[7] https://dblp.uni-trier.de/
[8] http://xueshu.baidu.com/
[9] https://www.aminer.cn/

2. Related Works

In recent years, there are more papers are focused on describing the mentor-ship, including the behaviors between supervisors and students, the characteristic of cooperation in the mentor-ship, and so on. Besides, there are some previous works focused on relationship mining. For example, the discovering of family relationship [29], the detecting of social relationships [30], the detection of communication relationship [31], etc. All of these works provide a solid basement for this paper.

However, there are only a few works focused on the advisor-advisee relationship mining. Han, etc. had proposed a method to detect the advisor-advisee based on the frequency terms mining [35], which may miss some important information that hidden behind the co-author network. Wang, etc. proposed a time-constrained probabilistic factor graph model (TPFG) to mining the advisor-advisee relationship between researchers [27]. It is based on the citation network and the co-author network; however, it did not consider the credit allocation of each author in the paper. Therefore, the result may be influenced by the high degree authors in the co-author network. Wang, etc. proposed a deep learning framework to mining the advisor-advisee relationship, based on the co-author network. It did not consider the credit allocation of authors, either [36]. Zhao, etc. proposed a deep model, equipped with improved Refresh Gate Recurrent Units, to discovery the hidden advisor-advisee relationship from the co-author network [37]. This work also neglects the credits allocation of authors.

3. Problem Describing

In order to make the problem clearer, we formulated the problem. Some abbreviates are introduced as following:

1) P_a stands for the set of papers, which author a published. We can denote it as $P_a = \left\{ P \mid (P_a)^t \le T \right\}$, where the T stands for the year when we obtained the data set (2018), $(P_a)^t$ means author \$a\$ published paper in year t.

2) $coA(a)$ stands for the co-authors of author a, where the $coA(a) = \left\{ (a,b) \mid a,b \in P_a \right\}$.

3) $First(a)$ stands for the year when author a published his first paper.

4) Y_a^b stands for the time span between author a and b on the first published paper.

5) $Credit(P_a)$ stands for the crediting allocation of author a in paper P.

6) $Ad(a)$ stands for the advisor of author a.

3.1. Assumptions

To make our approach worked, we should make some assumptions based on the background trues.

Assumption1: The first published paper of the advisor is early than the advisee. We could show this relationship as $First(advisor) < First(advisee)$.

Assumption2: The papers published by the advisees must co-author with his advisor, during the year he was supervised by his tutor. We could denote it as $(a, Ad(a)) \in P_a^t$, where $t \in \left[T - First(a), T \right]$ and $t \in Z^+$.

Assumption3: The advisor of author a must in the set of his co-coauthors. We denote it as $Ad(a) \in coA(a)$.

Assumption4: Each co-authors of author a could become a's advisor with a certain probability. In this paper, we only chose those authors who have 90% chance of becoming the advisor of a.

3.2. Approach

Previous works are most based on the citation network and co-author network, which contained the whole nodes in the network. Therefore, it will be influenced greatly by the author who have high degree in the co-author network. For example, author a has co-authored three papers with his advisor $Ad(a)$ in his student life. However, these three papers are also co-authored with author B, who has published 100 papers already and has great academic influence in the related field. Therefore, the whole network related methods may have a great probability to predict the advisor of author a is B. In this case, author B is the node with high degree, if we take the whole network into account, the result may be influenced by this kind of author greatly. In order to avoid this problem, we should know the distribution of the degree in this co-author network, and cut this co-author network into smaller ones based on the degree correlation.

Assortativity coefficient is a measure to evaluate the degree correlation of a network. Therefore, we calculated the assortativity of this co-author network. We first defined some symbols as following:

$$\begin{cases} q_k = P_n(k) \\ e_{ij} = P(i,j) = \dfrac{m(i,j)\mu(i,j)}{2M} \\ \sigma_q^2 = \sum_k k^2 q_k^2 - \left[\sum_k k q_k\right]^2 \end{cases}$$

During which, k stands for the degree of nodes, $P_n(k)$ stands for the probability of a randomly selected neighboring node with a degree of k under a randomly selected node in the network. $m(j,k)$ stands for the number of edges between the nodes with a degree of j and the nodes with a degree of k. If $j=k$, then $\mu(i,j)=1$, else $\mu(i,j)=2$. σ_q^2 stands for the variance of q_k, M stands for the count of edge in the network.

Then, we gain the formulation to calculate the assortativity of this co-author network.

$$r = \frac{1}{\sigma_q^2}\sum_{j,k} jk(e_{jk} - q_j q_k) \tag{1}$$

In Eq. (1), r stands for the assortativity coefficient and $r \in [-1,1]$. If $r > 0$, this network will be an assortative one, which means the nodes with high degree will be more likely to connect with the high degree nodes. If the $r < 0$, the network will be disassoratative, which means the nodes with low degree will be more likely to connect with the high degree nodes.

To make the calculation simpler for this co-author network, we changed the formation of the Eq. (2), and gain the new assortativity coefficient calculate equation as following:

$$\begin{aligned} \mathrm{r} &= \frac{cov(k_i,k_j)}{\sigma_x^2} \\ &= \frac{S_1 S_e - (S_2)^2}{S_1 S_3 - (S_2)^2} \end{aligned} \tag{2}$$

where:

$$\begin{cases} cov(k_i,k_j)=\dfrac{\sum\limits_{i,j}(a_{ij}-\dfrac{k_ik_j}{2M})k_ik_j}{\sum\limits_{i,j}(k_i\delta_{ij}-\dfrac{k_ik_j}{2M})k_ik_j} \\ \sigma_k^2=\dfrac{\sum\limits_{i,j}(a_{ij}-\dfrac{k_ik_j}{2M})k_ik_j}{\sum\limits_{i,j}(k_i\delta_{ij}-\dfrac{k_ik_j}{2M})k_ik_j} \\ S_e=\sum\limits_{i,j}a_{ij}k_ik_j=2\sum\limits_{i,j\in E}k_ik_j \\ S_1=\sum\limits_i k_i \\ S_2=\sum\limits_i (k_i)^2 \\ S_3=\sum\limits_i (k_i)^3 \end{cases}$$

During which, a_{ij} is the adjacent matrix of the co-author network. k_i is the degree of node i. E is the set of nodes, which contains all the authors.

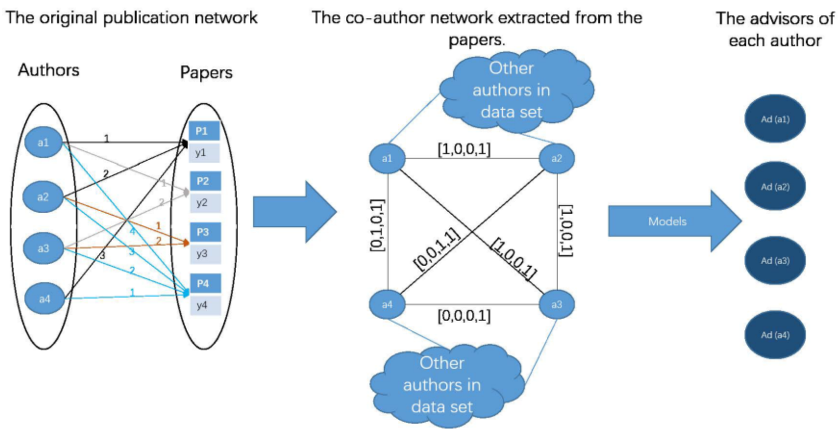


Fig. 1. Example of the relationship mining of the co-author network.

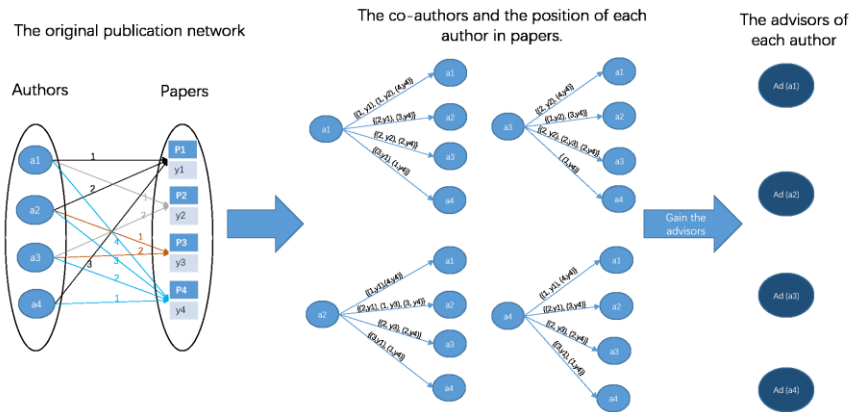


Fig. 2. Example of the relationship mining of the splited co-author network.

With the Eq. (2), we could calculate the assortativity coefficient of the co-author network, and the result is 1, which means the co-author network of DBLP is assortative greatly. Therefore, we cut the co-author network into some smaller ones. Different from the other methods, the method in this paper mines the advisor-advisee relationship in many smaller networks, each network just contains the author himself and his co-authors. We could compare these two kinds of models though the following simple figures.

In Fig. 1, it considers the whole co-author network together. The vector $[1,0,0,1]$ on the edge between a1 and a2 means that a1 has collaborated with a2 in year y1 and y2 (there is none limitations on y1 and y2, they could even be the same year). In this kind of model, the co-authors would be influenced by each other and every authors in the data set would also have a great influence in the result, especially these high degree authors. In Fig. 2, the dictionary $\{(1,y1),(1,y2),(4,y4)\}$ stands for the position of the co-author in one paper and the published year of one paper. In this model, every author and their co-authors are extracted in a single network. Therefore, the result only influenced by the collaborate patterns, such as, the collaborate time span, the number of collaborate papers and the time when the first paper published, etc. The other authors who are not in the smaller network will not influence the results any more. It could also make sure all the advisors of author a are in P_a.

3.3. Credit Allocation of Each Author

Credit allocation plays an important role in multidisciplinary research, especially for these works done by more than one author. How to determine the credit allocation of the authors has an important application in a wide range area, including the hiring, funding and the salary, etc. There are some methods to deal with credit allocation, such as, 1) For the single-author papers, the sole author gets all the credit. This rule is accepted around the scientific field all the time, while, it will be not effective in the papers with multi-authors. 2) For the multi-authors papers, there are three mainstream methods to solve this problem. First, views all the papers with multi-authors as the sole author [40], [41]. Second, views all the authors in the published papers has the equal credit allocation [42], [43]. Third, allocates the scientific credit according to the order or role of co-author in the published paper. There are some previous works to solve the credit allocation, based on the order of authors in the paper [15]-[17]. In this paper, we used the following Eq. (3) to calculate the credit allocation of each author in a multi-authors paper.

$$C^{p}(a)=\frac{\frac{1}{i}}{\sum_{i=1}^{n}(\frac{1}{i})} \tag{3}$$

In the Seq (3), $C^{p}(a)$ stands for the allocation credit of author a in the paper p, which the number of co-authors in paper p is n.

3.4. Disambiguating Authors Name

Author name disambiguation problem is to match the mentioned persons to the actual persons in the digital libraries. In fact, a proper evaluation of author name disambiguation for digital libraries becomes critical if the validity of research findings mined would be affected by the correctness of author identification in the data.

Researchers have been working on this matching problem for decades years. Some scholars have resolved name ambiguity using simple heuristics like simplifying name strings in a format of forename initials and surname [44]. Meanwhile, others have applied machine learning algorithms to disambiguate all names in digital libraries [45]. However, there is not a universal solution for this problem, which may be caused by several reasons, such as the limiting of data set, especially lacking of the test set. Therefore, there are some works are focused on creating the test data set [46]-[48]. However, these classic test collections are small and could not be used to study properties of defects, which would have great influence in the reveling of new approaches to match mentions and persons.

Among them, the DBLP's author name disambiguation performs well even on large ambiguous name blocks. When compared with other name disambiguation algorithms, the solution of DBLP always performances more compete, which possibly due to its hybrid disambiguation approach combining algorithmic and manual error correction [49]. What is more, the DBLP has a solution of the authors with the same name. For example, Jiawei Han 0001, Jiawei Han 0002, Jiawei Han 0003. We made the same assumption like many other scholars that, the DBLP's disambiguation is accurate [50].

Based on the analyzing above, we used the longest common substring to match the authors' name. For an author, if the name of a mentioned person full matched a name in the DBLP's name set, we would set the person in DBLP is the mentioned person. If the name could not full match the name in DBLP, we would set these authors, whose name have 95% same sequence substring, are the candidates of the mentioned person.

3.5. Gaining the Advisor of an Author

With the factors above, we could calculate the advisor for a given author with the following equations.

$$\begin{cases} Y_{a_c}^{a} = First(a_c) - First(a) \\ N(a_c) = count(\{P_a \mid Y_{a_c}^{a} > conT\}) \\ adV(a_c) = (\sum_{p \in p_a} C^p(a_c)) Y_{a_c}^{a} N(a_c) \end{cases} \tag{4}$$

In Eq. (4), a_c stands for one of the co-author of author a, $a_c \in coA(a)$. $N(a_c)$ stands for the number of papers that a collaborated with his co-author a_c. $count()$ is a function that could calculate the number of papers in a given set. $conT$ stands for the mean collaborate year between the advisors and advisees. According to [51], we could know that the average length of advisor–advisee relationship in academic networks is five years. Therefore, we set $conT = 5$.

3.6. Getting the Academic Tree

Academic tree is organized as a family tree of scientists and scholars according to mentoring relationships, often in the form of dissertation supervision relationships[10]. The visualized one could show the academic inheritance clearly and it will be a great convenience for people to find the quality and quantity of the students one has forested. It will be easy for people to find the expert in a special field.

To show the academic tree clearly, we used the modified deep first search algorithm to obtain all the possible advisor-advisee pairs and show them in the academic tree.

4. Experimental Results

The accuracy of this model is about 62.5%. However, this is an unsupervised model, which may not be influenced by the labeled data, and it could also save the time of model training in some extent.

The co-author network graph about an author, Zongbao Yang is shown in Fig. 3. From which we could find

[10] https://en.wikipedia.org/wiki/Academic_genealogy

that some nodes with larger degrees tend to connect nodes with larger degrees. In fact, this is a reasonable phenomenon in academia. The academic community is a group that places great emphasis on knowledge sharing and collaboration, and the influential people tend to work with experts who are equally influential in certain areas. In particular, the phenomenon of interdisciplinary cooperation is now widespread, and finding people who have influence in a certain field will promote more excellent scientific research results.

The academic tree of him is shown in the Fig. 4. From which, we could find that the academic tree is a directed graph and the arrow points to the advisor, the start point of an arrow is the advisee. For example, the advisor of Zongbao Yang is Shaohong Zhang, the advisor of Shaohong Zhang is Hau-san Wong, and the rest can be done in the same manner.

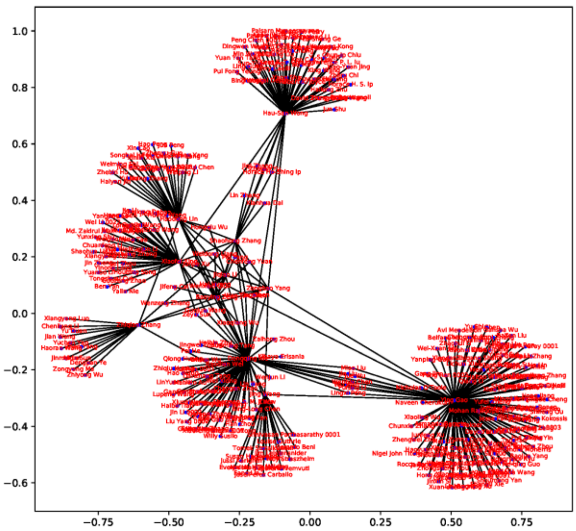


Fig. 3. The co-authors network of Zongbao Yang in two generations.

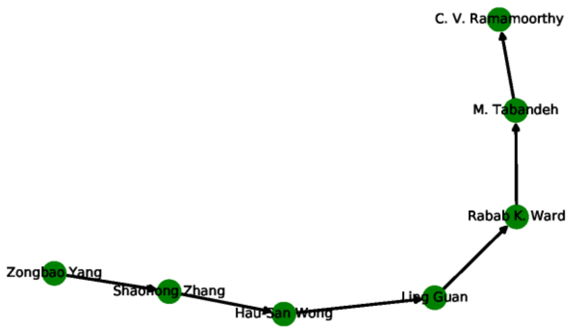


Fig. 4. The academic tree of Zongbao Yang.

5. Contribution

(1) We have created a data set of advisor-advisee relationship, which contained 14747 pairs of adviser-advisees. This data set could be used as the train set. To our best knowledge, this data set is the only one opening set of this kind of relationship.

(2) We used an unsupervised model to mining this kind of relationship, which could save the time of model training and unlimited by the size of the labeled data set.

6. Conclusion and Future Work

6.1. Conclusion

Advisor-advisee relationship is an important relationship in the academic circle. However, it is always changing with time and other challenges will arise when mining this relationship, including the limiting of data set, the ambiguous of authors name, the allocating of credit about co-authors, etc.

As to the ambiguous of authors' name, the method of longest common substring is used to find the candidate author. For the credit allocation, a weighted distribution method was used to allocate the credit of co-authors, which is based on the order of authors in the author list. Besides, previous works are all based on the whole co-author and citation network, which will be influenced greatly by the high degree authors in the network. Therefore, we proposed an innovation method by cutting the original co-author network into some smaller ones, and mining the advisor-advisee relationship in these small networks.

Finally, we designed an advisor score function to calculate the probability of an author who could be the advisor. As an unsupervised method, the accuracy of this method is about 62.5%. When given a determined author, the method in this paper could draw the academic tree and his co-author network graph. From the academic tree, we could find the academic inheritance of one author clearly.

6.2. Future Work

Similar with other social entities, scholars have their own communities. Mining the advisor-advisee relationship from the scholar communities may obtain a more accurate result. People in the same communities always have the same research interest and have a higher probability to cooperate with each other. Especially for a younger researcher, he may be more likely to work with their advisors or other experienced researchers. Therefore, there are more advisor-advisee relationship hidden behind the scholar communities.

However, identifying the communities is full of challenges, because of the complex influence factors. For example, when detect the scholar communities, it is an inevitable problem to measure the influence of an author, which is based on the quality and quantity of the papers he has published and the credits he has made for these papers. However, the quality of published paper has a great relevance with the influence of publications, which are most relate to the citation counts of their published papers and the quality of these citing papers, etc. It is a complex problem, however, meaningful.

Acknowledgment

The work described in this paper was partially supported by grants from Scientific and Technological Project of Guangzhou [Project No. 201607010053, 201707010284], Guangdong Natural Science Foundation of China [Grant No. 2018A030313922], Science and Technology Projects of Guangdong Province, China [Grant No. 2016B010127001], the funding of Guangzhou education scientific research project [Project No. 1201730714], 2016 Education reform project of Guangdong Province [Project Name 'Research on Construction and Mining methods in Knowledge Graphs of Computer Sciences Courses', Project No. 426, File [2016]236]], the Postgraduate Educational Reform project of Guangdong Province [No. 2017JGXM-MS45].

References

[1] Li, W., Chen, Y., Hu, T., & Luo, J. (2018). Mining the relationship between emoji usage patterns and personality. *Proceedings of the Twelfth International Conference on Web and Social Media* (pp. 648–651), Stanford, California, USA.

[2] Yong, S., & Asano, Y. (2018). Purpose-feature relationship mining from online reviews towards

purpose-oriented recommendation. *IEICE Transactions 101-D*.

[3] Zhang, X., & Butts, C. T. (2017). Activity correlation spectroscopy: A novel method for inferring social relationships from activity data. *Social Netw. Analys*, *7(1)*.

[4] Huang, C., Wang, D., Zhu, S., & Mann, B. (2017). Toward local family relationship discovery in location-based social network. *Social Netw. Analys. Mining*, *7(1)*.

[5] Connes, V., Dugu´e, N., & Guille, A. (2018). Is community detection fully unsupervised? The case of weighted graphs. *Proceedings The 7th International Conference on Complex Networks and Their Applications Complex Net Works*, Cambridge, UK.

[6] Moctar, A. O. M., & Sarr, I. (2018). Survey on social ego-community detection. *Proceedings of the 7th International Conference on Complex Networks and their Applications Complex*.

[7] Xia, F., Wang, W., Bekele, T. M., & Liu, H. (2017). Big scholarly data: A survey. *IEEE Trans. Big Data*, *3(1)*.

[8] Yu, D., Wang, W., Zhang, S., Zhang, W., & Liu, R. (2017). A multiple-link, mutually reinforced journal-ranking model to measure the prestige of journals. *Sci-320 Scientometrics*, *111(1)*.

[9] Jiang, X., Sun, X., Yang, Z., Zhuge, H., & Yao, J. (2016). Exploiting hetero generous scientific literature networks to combat ranking bias: Evidence from the computational linguistics area. *JASIST*, *67(7)*.

[10] Valderrama, P., Escabias, M., & Jim´enez-Contreras, E. (2018). Proposal of a stochastic model to determine the bibliometric variables influencing the quality of a journal: Application to the field of dentistry. *Scientometrics*, *115(2)*.

[11] Milojevic, S., Radicchi, F., & Bar-Ilan, J. (2017). Citation success index − An intuitive pair-wise journal comparison metric. *J. Informetrics*, *11(1)*.

[12] Xu, F., Liu, W., & Rousseau, R. (2015). Introducing sub-impact factor (SIF-) sequences and an aggregated sif-indicator for journal ranking. *Scientometrics*, *102(2)*.

[13] Waltman, L., Eck, N. J. V., Leeuwen, T. N. V., & Visser, M. S. (2012). Some modifications to the SNIP journal impact indicator. *J. Informetrics*, *7(2)*.

[14] Yu, D., Wang, W., Zhang, S., Zhang, W., & Liu, R. (2017). A multiple-link, mutually340 reinforced journal-ranking model to measure the prestige of journals. *Scientometrics*, *111(1)*.

[15] Abbas, A. M. Polynomial weights or generalized geometric weights: Yet another scheme for assigning credits to multiple authors.

[16] Egghe, L., Rousseau, R., & Hooydonk, G. V. (2000). Methods for accrediting publications to authors or countries: Consequences for evaluation studies.

[17] Assimakis, N., & Adam, M. (2010). A new author's productivity index: p-index. *Scientometrics*, *85(2)*.

[18] Shen, H.-W., & Barab´asi, A.-L. (2014). Collective credit allocation in science. *Proceedings of the National Academy of Sciences*, *111(34)*.

[19] Donner, P. (2018). Effect of publication month on citation impact. *J. Informetrics*, *12(1)*.

[20] Zhao, F., Zhang, Y., Lu, J., & Shai, O. (2019). Measuring academic influence using heterogeneous author-citation networks. *Scientometrics*, *(1)*.

[21] Jiang, X., Sun, X., Yang, Z., Zhuge, H., & Yao, J. (2016). Exploiting heterogeneous scientific literature networks to combat ranking bias: Evidence from the17360 computational linguistics area. *JASIST*, *67(7)*.

[22] Liu, Z., Huang, H., Wei, X., & Mao, X. (2014). Tri-rank: An authority ranking framework in heterogeneous academic networks by mutual reinforce. *Proceedings of the 26thIEEE International Conference on Tools with Artificial Intelligence*.

[23] Zhang, C., Liu, C., Yu, L., Zhang, Z., & Zhou, T. (2017). Identifying the academic rising stars via pairwise citation increment ranking, in: Web and Big Data. *Proceedings, Part I*, 475–483.

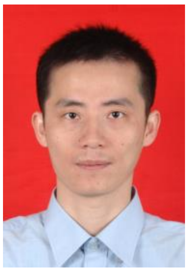
[24] Daud, A., Aljohani, N. R., Abbasi, R. A., Rafifique, Z., Amjad, T., Dawood, H., & Alyoubi, K. H. (2017).

Finding rising stars in co-author networks via weighted375 mutual influence. *Proceedings of the 26th International Conference on World Wide Web Companion*.

[25] Zhang, J., Xia, F., Wang, W., Bai, X., Yu, S., Bekele, T. M., & Peng, Z. (2016). Coca rank: A collaboration caliber-based method for finding academic rising stars. *Proceedings of the 25th International Conference on World Wide Web*.

[26] Zhang, J., Ning, Z., Bai, X., Wang, W., Yu, S., & Xia, F. (2016). Who are the rising stars in academia? *Proceedings of the 16th ACM/IEEE-CS on Joint385 Conference on Digital Libraries*.

[27] Wang, C., Han, J., Jia, Y., Tang, J., Zhang, D., Yu, Y., & Guo, J. (2010). Mining advisor-advisee relationships from research publication networks. *Proceedings of18the 16th ACM SIGKDD International Conference on Knowledge Discovery390 and Data Mining*.



Shaohong Zhang is an associate professor in the Department of Computer Science at Guangzhou University. He was a postdoctoral fellow in the Department of Computer Science, City University of Hong Kong. He received the PhD degree from Department of Computer Science, City University of Hong Kong. His research interests include pattern recognition, data mining, and bioinformatics.

Zongbao Yang was born in Anhui province. He received his B.S. degree from Guangdong Medical University, Dongguan, China, in 2016. He is currently perusing the M.S. degree in Guangzhou University, Guangzhou, China. His research interests include data mining, machine learning, social network and citation network. He is a member of China Computer Federation (CCF).

Liu Jianyu was born in Guangdong Province. He received his B.S. degree from Zhongshan University of Zhongshan University of Electronic Science and Technology, China in 2017. His currently pursuing his M.S. degree in Guangzhou University, Guangzhou, China. His research interests include data mining, machine learning, natural language processing, and deep learning.

Zhiqian Zhang was born in Guangdong province. She is majored in computer science and pursuing her bachelor degree in Guangzhou University. Her research interests includs data mining, machine learning.

Xiaofei Xing received PhD degree in computer science from Central South University, China, in 2012. He has been a visiting research fellow at University of Tsukuba, Japan. His research interests include modeling and performance evaluation in wireless sensor networks and mobile computing. He is a member of IEEE and China Computer Federation (CCF).

Ying Gao received his Ph.D. degree from South China University of Technology, Guangzhou, China, in 2002. He is currently a professor with the School of Computer Science and Educational Software, Guangzhou University, Guangzhou, China. His main research interests include intelligent optimization algorithms, pattern recognition, and signal processing.