Understanding Scholar Social Networks: Taking SCHOLAT as An Example

Min Gao[1,2], Yang Chen[1,2,*], Qingyuan Gong[1,2], Xin Wang[1,2], Pan Hui[3,4]
[1] School of Computer Science, Fudan University, China
[2] Shanghai Key Lab of Intelligent Information Processing, Fudan University, China
[3] Department of Computer Science, University of Helsinki, Finland
[4] Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong
mgao21@m.fudan.edu.cn, {chenyang, gongqingyuan, xinw}@fudan.edu.cn, panhui@cs.helsinki.fi

Abstract. Scholar social networks are composed of scholars and social connections among them. Studying such social networks can help promote academic exchanges and cooperation, and predict future trends in research. In this paper, we analyze SCHOLAT, a representative scholar social network in China, from three perspectives. First, we explore SCHOLAT's social graph, and we find this graph has a smaller average shortest-path length and a higher clustering coefficient than other social networks, for example, the collaboration network of Google Scholar and the Flickr social network. Moreover, we leverage the structural hole theory to identify important users on SCHOLAT. By comparing the top-500 structural hole spanners with 500 randomly selected users, we have found that the former has the higher values of several graph-based metrics, and they also connect more communities. Finally, we also undertake user group-based analysis, and we discover that the users belonging to Guangdong province, and the users from the top universities in China are well-connected and occupy important positions in the network.

Keywords: SCHOLAT Social Network, Social Graph Analysis, Structural Hole Theory, User Group-Based Analysis

1 Introduction

Scholar social networks [1–6] are a kind of social networks made up of scholars and their social connections. Such social networks have two main functions. On one hand, these networks allow researchers to create their homepage, upload their papers, maintain their information, and advertise their research outcomes to the public. On the other hand, these networks provide a platform for scholars to communicate and collaborate on scientific research topics, and build social connections. Plenty of scholar social networks are emerging, such as HASTAC [7], Academia [8], ResearchGate [10,11], and SCHOLAT [12]. SCHOLAT social network is a representative scholar social network, serving a large number of Chinese scholars. It provides services such as scholar information management and scholarly interaction, through which one can easily build her scholar homepage, and find scholars with the same research interests. And it also supports various types of interactions, such as making comments, posting “likes”, and reposting articles.

Scholar social networks are a particular type of social networks used for the purpose of scholarly communication. Researchers and scholars act as nodes and their relationships serve as edges, while for the general online social networks, the registered users and social connections between them correspond to the nodes and edges in the network, respectively. For example, Flickr [13] is a photo-sharing social network with millions of users, where users and their friendships can be modeled to nodes and edges. There are also certain differences between scholar social networks and collaboration networks [14,15]. Although both have researchers acting as nodes, the edges of the former are generated by online interactions (making friends, following or other forms of interactions) between scholars, while the edges of the latter are generated by joint publications between scholars. These differences in network composition motivate us to further explore scholar social networks. We aim to figure out whether there are differences among them and whether there are particularity of the user groups [17] in scholar social networks. However, existing work [18] have not carried these differences or give an analysis of the characteristics of scholar social networks. In this paper, we start with SCHOLAT social network to explore these differences and carry out a comprehensive analysis.

First, we conduct an informative analysis on the social graph of SCHOLAT. Specifically, we use social graph to model the users and their social connections and analyze this graph with representative graph metrics. We also compare the differences between SCHOLAT's social graph and other social graphs, for example, the collaboration social graph of Google Scholar and Flickr's social graph. We have found lots of unique features of SCHOLAT's social graph, such as a lower average shortest-path length and higher clustering coefficient, which reflects the feature of small-world networks [19]. By examining the distributions of some graph metrics, we discover that SCHOLAT's social graph is significantly different from the collaboration social graph of Google Scholar and Flickr's social graph.

Second, we identify important users [20] based on the structural hole theory [14, 22, 37] on SCHOLAT. Important users refer to the scholars bridging scholar communities and occupying key positions in the network. Due to the differences in academic impact of scholars, mining and tracking influential scholars in interdisciplinary research areas can help to identify research hotspots, and predict future trends. However, existing studies do not carry out the mining of structural hole spanners (SH spanners) in scholar social networks. To fill this gap, we dive into this problem by abundant structural hole theory-based analysis [21, 22] with pivotal metrics, for example, effective size. We first use the Louvain algorithm [16] to detect communities and study the sizes of the top-10 communities. Through comparative analysis between the top-500 SH spanners and the randomly selected 500 users, we can identify important users by effective size, who occupy more critical positions and resources than other users within the network.

Finally, we pay attention to the attributes of users and conduct a user group-based study on SCHOLAT. The same or similar user attributes often promote the establishment of direct or indirect associations among scholars, which are embodied in the forms of friendships, followings, collaborations, and other interactions. First, we extract users with attributes of research interests and affiliations. Then, we analyze user groups according to these two attributes. We have found the research interest with the largest number of users is "computer science". By comparing the distribution of user groups with different affiliations, we find that both geographic location and the ranking of universities in China can be used to divide users, and the former has a greater impact. We also discover that the users in Guangdong province or from top universities have a significant structural advantage on SCHOALT, which indicates these users are more well-connected and span over important positions in the social graph.

The paper structure is illustrated as follows. In Section 2, we provide some related works. Then we analyze the SCHOLAT social network from three perspectives and uncover some characteristics of SCHOLAT social network in Section 3. Conclusion and future work are discussed in Section 4.

2 Related Work

HASTAC [7] was commented by a National Science Foundation review as the world's first and oldest scholar social network in 2014[1]. Users in this network include researchers, scholars, and the general public with various interests. HASTAC contains more than 14,000 members, and offers them a free and open access community to teach and learn.

AMiner [23, 24], a big data mining and service system platform for scientific and technological information. It has collected more than 130,000,000 scholar profiles and 100,000,000 papers from different publication datasets. It offers some useful services, such as profile extraction, scholar ranking, and name disambiguation.

Google Scholar offers a wealth of information about authors and papers. Chen et al. [14] built a collaboration network of Google Scholar by referring to author profiles. They conducted a demographic analysis and provided an informative overview of scholars. In addition, they used social network analysis methods to analyze the collaboration network and explored several citation indicators.

Academia [8] is a typical scholar social network, which offers access to the users who tend to share their works and videos with others. Users can edit their personal information and upload papers to their profiles. In addition, it supports message sending between users. In the social graph of Academia, nodes are researchers, and edges are friendships among them. Their friendships are formed by paper sharing. Niyazov et al. [9] found that papers shared to Academia receive a 69 percent boost in citation over five years.

[1] https://www.hastac.org/about/history

ResearchGate [10,11] is a scientific social networking website that was launched in May 2008. The site is designed to promote scientific collaboration on a global scale. Users can maintain their profiles, and connect with colleagues, share research methods, and exchange ideas. In ResearchGate's social graph, nodes are researchers, and edges are their soical connections, such as sharing publications, collaborating with scholars, and communicating with researchers.

3 Data Analysis

In this section, we analyze the SCHOLAT social network from several perspectives. Firstly, we conduct a social graph-based analysis to observe the features of SCHOLAT's social graph (subsection 3.1). Then, we employ the structural hole theory to identify important users within this network (subsection 3.2). Finally, we carry out an analysis of the attribute-based user groups. We discover that the number of users with research interest "computer science" is the largest within this network, and we conclude that scholars from high-ranking universities are influential and occupy important positions within the network (subsection 3.3).

SCHOLAT offers an open-access dataset[2] of a scholar social network, which includes 16,007 nodes and 202,248 edges. Nodes in this network refer to the scholars within this network, and the edges correspond to the friendships between scholars.

3.1 Social Graph Analysis

We do social graph analysis by *igraph* package[3] for SCHOLAT and a series of findings are presented in this subsection. Specifically, we examine the difference between SCHOLAT's social graph and social graphs of two other networks, i.e., the collaboration network of Google Scholar [14] and the friendship network crawled from Flickr [36]. The information of these graphs is listed as Table 1.

Table 1: The description of three social graphs

Information	SCHOLAT	Google Scholar	Flickr
nodes	16,007	402,392	80,513
edges	202,248	1,234,019	5,899,882
type	indirect	indirect	indirect

We select four representative metrics to conduct social graph analysis as follows.

Degree [28]: in a network, the degree of a node refers to the number of its neighbors. In SCHOLAT's social graph, a scholar's degree equals the number of her connected scholars.

[2] https://www.scholat.com/research/opendata/#social_network
[3] https://igraph.org/python/

Eigenvector centrality [29]: this metric of a node holds that its centrality depends on the centrality of its neighbors. Bihari et al. [30] believed that the eigenvector centrality is more suitable for the discovery of influential researchers, and they used this metric to measure the influences of researchers. Similarly, in SCHOLAT's social graph, a scholar's influence can be calculated based on the influences of her friends.

Clustering coefficient [31]: for a node in a network, the definition of this indicator is the ratio of the number of edges between its neighbors divided by the number of possible edges between them. This metric quantifies the extent to which its neighbors form a cluster. In SCHOLAT's social graph, this metric reflects that the extent to which a scholar's friends are also friends.

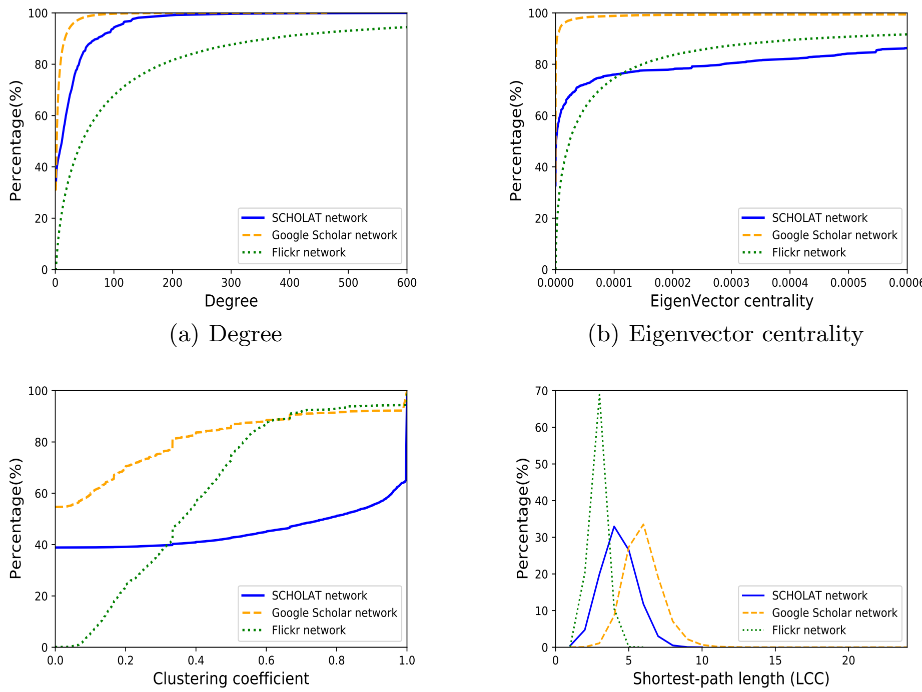
Shortest-path length [31]: this indicator refers to the length (or distance) of the shortest path between two nodes in a network. In SCHOLAT's social graph, the value of this metric can be used to indicate the smallest hops between two scholars establishing a connection.

We have made comparisons mainly on the following representative graph-based metrics including average degree, average clustering coefficient, network diameter of the largest connected component (LCC), and average shortest-path length of the LCC. The results are listed as Table 2.

Table 2: Comparison between social graphs of three social networks

Metrics	SCHOLAT	Google Scholar	Flickr
Average degree	25.27	6.13	146.56
Average clustering coefficient	0.55	0.20	0.17
Network diameter (LCC)	10	24	27
Average shortest-path length (LCC)	4.31	5.96	2.90

The average degree of SCHOLAT's social graph is 25.27, which is much larger than that of the Google Scholar's collaboration social graph. The reason for the large difference is that in SCHOLAT's social graph, edges can be generated as long as there is a friendship, while in the collaboration social graph of Google Scholar, edges are generated on the condition that there is a co-authorship. Therefore, forming an edge in SCHOLAT's social graph is much convenient than that in the collaboration social graph of Google Scholar. The average degree of SCHOLAT's social graph is less than that in the Flickr's social graph. This is because the users interact more intensively on Flickr. The average clustering coefficient of SCHOLAT's social graph is more than twice as large as that of the collaboration social graph of Google Scholar. This indicator of a scholar quantifies the extent to which her neighbors aggregate to form clusters with each other. The average clustering coefficient of SCHOLAT's social graph is also greater than that of Flickr. The network diameter of the largest connected component (LCC) of SCHOLAT's social graph is 10, which is smaller than that of LCC of the collaboration social graph of Google Scholar and greater than that of the LCC of Flickr's social graph. The SCHOLAT's social graph has the feature of small average shortest-path length and high clustering coefficient, which conforms to the criteria of small-world networks [19].



(a) Degree

(b) Eigenvector centrality

(c) Clustering coefficient

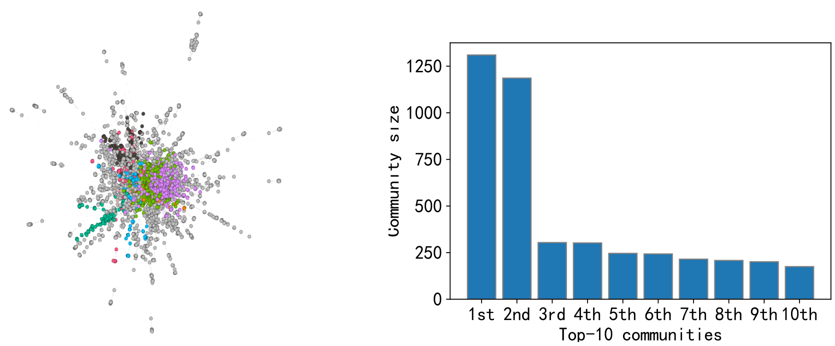
(d) Shortest-path length (LCC)

Fig. 1: Comparison on representative graph metrics among three social graphs

To analyze the SCHOLAT social graph, we compare the differences between this graph with the other two social graphs of the collaboration network of Google Scholar and Flickr network. Fig. 1(a)–Fig. 1(d) depicts the cumulative distribution function (CDF) of the metric of degree, eigenvector centrality, clustering coefficient, and the LCC's shortest-path length of these three graphs. In Fig. 1(a), for the SCHOLAT's social graph, the median value of the degrees is 11, which indicates that half of the users in this graph have a degree value greater than 11. While for the Flickr social graph, more than half of the users have over 150 neighbors. This is because friendships among users are relatively denser on Flickr than the relationship of connections between scholars. The distribution of the metric of eigenvector centrality is revealed in Fig. 1(b). Apparently, in SCHOLAT's social graph, the high values of this metric are noticeably more than that in the other two social graphs. That is to say, in SCHOLAT's social graph, there are high-influence scholars with influential friends. Fig. 1(c) corresponds to the CDF of the clustering coefficient. We find that the users in SCHOLAT's social graph have larger values of clustering coefficient than those in the social graphs of Google Scholar and Flickr. Fig. 1(d) plots the distribution of the shortest-path lengths (LCC) in each of the three social graphs. It can be observed that the average shortest-path length between users in SCHOLAT's social graph is smaller than that in the social graph of Google Scholar.

3.2 Structural Hole Theory-Based Important User Analysis

In this section, we perform structural hole theory-based analysis on important users in SCHOLAT network. Firstly, we use Louvain [16], a modularity-based algorithm, to detect communities in SCHOLAT network. As a result, the entire network is divided into 295 communities. Fig. 2(a) visualizes the result of the community division of SCHOLAT network. We use the modularity metric [32, 33] to measure the result of the community division. This metric is a value between -1 and 1. It measures the density of connections within communities versus the density of connections between communities. Pursuant to [34], the modularity value of the division on SCHOLAT is 0.856, much larger than 0.3, which signifies that the SCHOLAT network has a viable community structure. It can be found that the scholars within each community are more closely connected than scholars from different communities. Fig. 2(b) depicts the sizes of the top-10 communities within this network. The largest community has 1,310 users.



(a) Community division visualization (b) The sizes of the top-10 communities

Fig. 2: Analysis of communities within SCHOLAT social network

Based on the data contained in this network, we investigate the problem of important user identification by a structural hole theory-based analysis. In a social network, if there is no direct connection between several communities, it seems that there is a hole in the network structure. Structural hole (SH) refers to the gap between user communities. The users, occupying these special positions and playing a bridging role between different communities, are known as SH spanners. Burt [37] put forward the structural hole theory, which studies the structure formation of interpersonal networks and analyzes occupying what kind of positions in the network can bring more benefits or rewards to the nodes. The structural hole theory emphasizes that users occupying the positions of structural holes in networks can bring advantages in information dissemination and other resources to organizations and individuals [14,22,37]. As an important part of the structural hole theory, a series of SH metrics, including effective size, constraint, efficiency and hierarchy, are proposed to evaluate whether one node is probably serving as an SH spanner [22,37]. There is a wide range of applications of the structural hole theory, such as identification of high potential talents from newly-enrolled employees of a company [38], influence maximization in social networks [39], and proposing new metrics for measuring the importance of nodes [40]. Efforts have also been made to solve the problem of effectively mining the top-k SH spanners in social networks [41, 42]. We utilize SH theory to study the top-k important users within SCHOLAT network. Specifically, we use effective size [43, 44], a representative metric measuring the non-redundant links, to discover the SH spanners. The effective size of the ego network of node i is denoted as $e(i)$, and its definition is

$$e(i) = n - \frac{2t}{n} \tag{1}$$

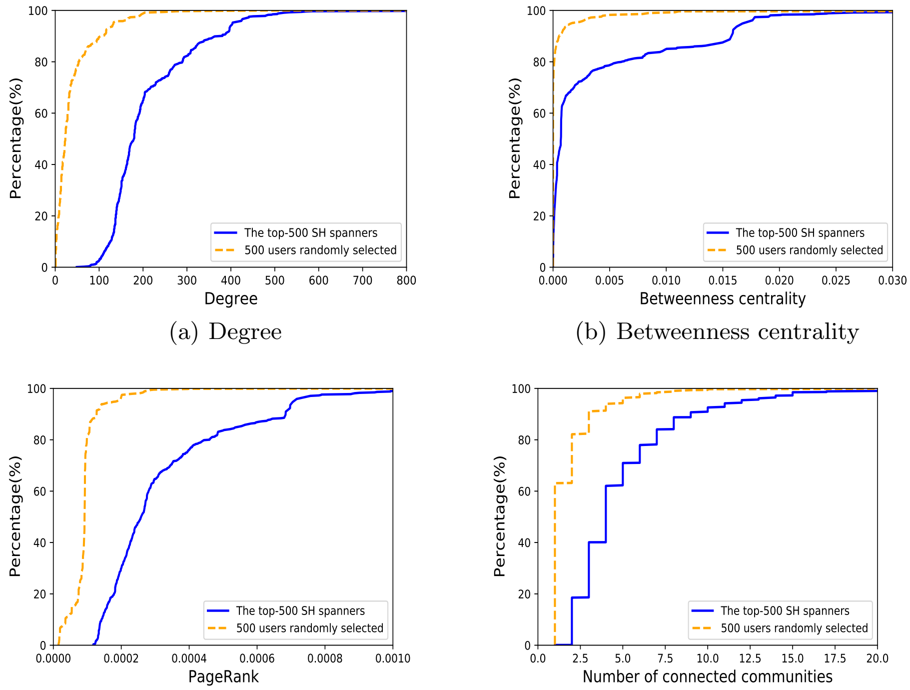
where n and t are the number of nodes and edges within the ego network of node i, respectively.

On the basis of this metric, we study the feature of the top-500 SH spanners from several perspectives, including the degrees, the betweenness centrality values, the PageRank values and the number of communities (N_{com}) they bridge. In order to illustrate the difference between the top-500 SH spanners and the randomly selected 500 users, we use the CDF to characterize the distribution of the values on the above metrics of the two user groups. The results are shown in Fig. 3(a) – Fig. 3(d).

According to Fig. 3, we find some significant differences between the top-500 SH spanners (represented as the solid lines in the figure) and the randomly selected 500 users (represented as the dotted lines in the figure). In Fig. 3(a), we can find that half of the top-500 SH spanners have more than 200 neighbors, while about 50% of the 500 randomly selected users have degree values lower than 100. Similarly, compared with the 500 randomly selected users, the top-500 SH spanners whose betweenness centrality values are more than 0.0023 account for nearly 50%, which means that these SH spanners play a decisive role in the process of information dissemination on the entire network. Fig. 3(c) shows the difference on the metric of PageRank, and we can find that half of the top-500 SH spanners have PageRank values larger than 0.00025, which are much higher than those of the 500 randomly selected users. As shown in Fig. 3(d), we discover that the top-500 SH spanners bridge more communities than other users. And we also find these important nodes with higher values of graph-based metrics, such as the degrees, betweenness centrality values, which further validates that the top-500 SH spanners are more important in the whole network. For example, node #4809, a scholar from South China Normal University, has the largest value of effective size and connects 22 communities. What is more, it is the user with the greatest values of degree, betweenness centrality in the network. This means that user #4809 is able to play a critical role in the information propagation and resource control on the whole network.

3.3 User Group-Based Analysis

This subsection mainly analyzes different groups of users according to two important attributes, i.e., research interest and affiliation. We first conduct the



(a) Degree (b) Betweenness centrality

(c) PageRank (d) Number of connected communities

Fig. 3: Difference between the top-500 SH spanners and the randomly selected 500 users

data preprocessing and the extraction on users providing the attributes of research interest and affiliation. As a result, we get 11,996 users providing the research interest information and 8,970 users providing the affiliation information. Then, through statistical calculation and comparative analysis, we have explored the differences between groups of users with different research interests and affiliation information, respectively.

Due to the unbalanced distribution of the attributes of users, which may lead to inaccurate analysis, we need to make some data preprocessing. First, we filter out incorrect values and null values that violate common sense. Second, by observing the dataset, we find that some attributes of users are ambiguous. One possible reason is that users are not willing to let others know the precise information about themselves. Such as, "university of science and technology" is a vague expression. So we remove such data entries for user group-based analysis.

We first divide the users according to their research interests. Table 3 lists the top-10 keywords of research interests of users. The keyword of "computer science" has the highest proportion of users, indicating that there are about 6,084 (50.7%) scholars interested in computer science among all users of SCHOLAT. And the average degree (Avg. Degree) of users with 7 out of the top-10 keywords is higher than the average degree of the entire graph. In general, it can be recognized that the top-10 keywords include most of the research hot-spots in computer science and engineering. There are also research directions in other disciplines, such as culture industry and law.

Table 3: The top-10 keywords of research interests of SCHOLAT users

Research interest	#Users	Avg. Degree
computer science	6,084	70.5
artificial intelligence	1,284	33.7
automation	1,116	30.3
big data	636	36.2
machine learning	636	26.9
network security	516	38.5
privacy preservation	478	25.0
Internet of things	504	26.7
culture industry	422	24.9
law	320	23.6

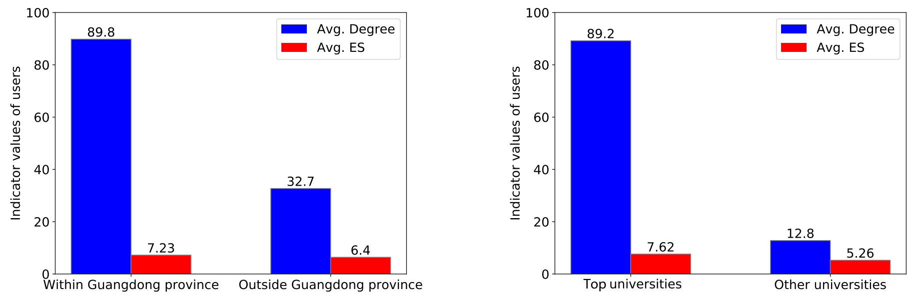
In addition, we investigate the users' affiliation based on the geographic locations and the university ranking. Since the SCHOLAT website is located in Guangdong province, we count the number of users whose affiliations belong to Guangdong province or not. We find that the number of users in Guangdong province is 5,808, higher than that out of Guangdong province, which is 3,912. This reveals the impact of geographic location on attracting users. We compare the users from Guangdong province and out of Guangdong province in terms of the average degree (Avg. Degree) and average effective size (Avg. ES), as shown in Fig. 4(a). It can be spotted that the Avg. Degree and the Avg. ES of users from Guangdong province are also higher than those out of Guangdong province. This also indicates that these users are more well-connected and occupy more important positions.

In order to further analyze user groups according to university ranking, we exclude the users with the affiliations of universities outside China and vague meanings, and the final number of domestic universities is 208. We consider one widely adopted standard in China, i.e., we refer to the "Project 211[4]", which covers 116 top universities in China. In this paper, we define the universities belonging to "Project 211" as top universities in China. We compare the users with different ranked universities in terms of the Avg. Degree and Avg. ES, and the results are depicted in Fig. 4(b). The numbers of users belonging to the top universities and other universities are 6,118 and 2,852, respectively. In other words, users from top-ranked universities account for a larger proportion of SCHOLAT users. Similarly, the Avg. Degree and the Avg. ES of users from the top universities are higher than those from other universities. This is because

[4] Project 211, commonly known as 211 universities, is a project of the Education Ministry of China to build about 100 key disciplines universities for the 21st century, including 116 universities.

the top universities have more educational resources and scientific research advantages than other universities, which have a greater influence on the academic ecosystem in China.

By comparing the two subgraphs in Fig. 4, the differences reflect that geographic location and university ranking are important to distinguish users, and the former takes a more significant role. Moreover, compared with other users, the users in Guangdong province or from the top-ranked universities occupy more structural advantages and are important in SCHOLAT.



(a) User groups based on geographic location

(b) User groups based on university ranking

Fig. 4: Differences between user groups on two metrics

4 Conclusion and Future Work

In this paper, we conduct a series of analysis on SCHOLAT social network, including social graph analysis, structural hole theory-based important user analysis, and user group-based analysis. We have found that SCHOLAT's social graph has obvious small-world features with a large average clustering coefficient and a small average shortest-path length, which are different from other social graphs of social networks, i.e., the collaboration network of Google Scholar (collaboration network) and the Flickr network (online social network). Based on the structural hole theory, we have observed the top-500 SH spanners are more important with the network structure than the randomly selected 500 users, and the SH spanners have a higher level of connectivity, and span over critical positions within the network. We further study the research interests of the scholars and find that a large portion of the scholars are working in the computer science and engineering-related field of research. We also find that the users from top universities in China, and the users from Guangdong province, play a more important role in SCHOLAT network.

We plan to further enhance our work in scholar social networks from the following three aspects. First, we would like to explore the composition and characteristics of user communities. Second, we plan to study the dynamics and evolution of the social connections. Last but not least, we would like to investigate the identification of potential fake accounts [45–47].

Acknowledgement

This work is sponsored by National Natural Science Foundation of China (No. 62072115, No. 71731004, No. 61602122, No. 61971145), China Postdoctoral Science Foundation (No. 2021M690667), the Research Grants Council of Hong Kong (No. 16214817), the 5GEAR project and FIT project from the Academy of Finland. Yang Chen is the corresponding author.

References

1. E. Giglia. Academic Social Networks: it's time to change the way we do research. *European Journal of Physical and Rehabilitation Medicine*, 47(2): 345–349, 2011.
2. X. Kong, Y. Shi, S. Yu, J. Liu, and F. Xia. Academic social networks: Modeling, analysis, mining and applications. *Journal of Network and Computer Applications*, 132: 86–103, 2019.
3. S. Mason. Adoption and usage of Academic Social Networks: a Japan case study. *Scientometrics*, 122(3): 1751–1767, 2020.
4. C. Huang. Social network site use and academic achievement: A meta-analysis. *Computers & Education*, 119: 76–83, 2018.
5. L. Subelj, D. Fiala, and M. Bajec. Network-based statistical comparison of citation topology of bibliographic databases. *Scientific Reports*, 4(1): 6496:1–6496:10, 2014.
6. Y. Tang. Scholar Social Networks and Big Data Research. In *Proceedings of the 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2021.
7. C. Puschmann, and M. Bastos. How digital are the digital humanities? An analysis of two scholarly blogging platforms. *PLOS ONE*, 10(2): e0115035, 2015.
8. C. Sugimoto, S. Work, V. Lariviere, and S. Haustein. Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology*, 68(9): 2037–2062, 2017.
9. Y. Niyazov, C. Vogel, R. Price, B. Lund, D. Judd, A. Akil, M. Mortonson, J. Schwartzman, and M. Shron. Open access meets discoverability: Citations to articles posted to Academia.edu. *PLOS ONE*, 11(2): e0148257, 2016.
10. M. Yu, Y. Wu, W. Alhalabi, H. Kao, and W. Wu. ResearchGate: An effective altmetric indicator for active researchers? *Computers in Human Behavior*, 55: 1001–1006, 2016.
11. M. Thelwall, and K. Kousha. ResearchGate articles: Age, discipline, audience size, and impact. *Journal of the Association for Information Science and Technology*, 68(2): 468–479, 2017.
12. Q. Xu, L. Qiu, R. Lin, Y. Tong, C. He, and C. Yuan. An Improved Community Detection Algorithm via Fusing Topology and Attribute Information. In *Proceedings of the 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 1069–1074, 2021.
13. M. Cha, A Mislove, and K. Gummadi. A measurement-driven analysis of information propagation in the Flickr social network. In *Proceedings of the 18th International Conference on World Wide Web (WWW)*, 721–730, 2009.

14. Y. Chen, C. Ding, J. Hu, R. Chen, P. Hui, and X. Fu. Building and Analyzing a Global Co-Authorship Network Using Google Scholar Data. In *Proceedings of the 26th International Conference on World Wide Web Companion (WWW Companion)*, 1219–1224, 2017.
15. A. Abbasi, K. Chung, and L. Hossain. Egocentric analysis of co-authorship network structure, position and performance. *Information Processing & Management*, 48(4): 671–679, 2012.
16. V. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P1000, 2008.
17. A. Peleshchyshyn, V. Vus, O. Markovets, and S. Albota. Identifying specific roles of users of social networks and their influence methods. In *Proceedings of the 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, 2: 39–42, 2018.
18. X. Wu, and C. Zhang. Finding high-impact interdisciplinary users based on friend discipline distribution in academic social networking sites. *Scientometrics*, 119(2): 1017–1035, 2019.
19. D. J. Watts. Networks, Dynamics, and the Small-World Phenomenon. *American Journal of Sociology*, 105(2): 493–527, 1999.
20. Q. Gong, Y. Chen, X. He, Y. Xiao, P. Hui, X. Wang, and X. Fu. Cross-site Prediction on Social Influence for Cold-start Users in Online Social Networks. In *ACM Transactions on the Web*, 15(2): 6:1–6:23, 2021.
21. M. Rezvani, W. Liang, W. Xu, and C. Xu. Identifying top-k structural hole spanners in large-scale social networks. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM)*, 263–272, 2015.
22. Z. Lin, Y. Zhang, Q. Gong, Y. Chen, A. Oksanen, A. Ding. Structural Hole Theory in Social Network Analysis: A Review. *IEEE Transactions on Computational Social Systems*, 1–16, 2021. doi:10.1109/TCSS.2021.3070321
23. J. Tang. Aminer: Toward understanding big scholar data. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM)*, 467–467, 2016.
24. H. Wan, Y. Zhang, and J. Zhang. AMiner: Search and mining of academic social networks. *Data Intelligence*, 1(1): 58–76, 2019.
25. M. Biryukov, and C. Dong. Analysis of computer science communities based on DBLP. In *Proceedings of International Conference on Theory and Practice of Digital Libraries (TPDL)*, 228–235, 2010.
26. C. Moreira, P. Calado, and B. Martins. Learning to rank academic experts in the DBLP dataset. *Expert Systems*, 2015, 32(4): 477–493.
27. N. Aggrawal, and A. Arora. Visualization, analysis and structural pattern infusion of DBLP co-authorship network using Gephi. In *Proceedings of the 2nd International Conference on Next Generation Computing Technologies (NGCT)*, 494–500, 2016.
28. L. C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3): 215–239, 1978.
29. B. Ruhnau. Eigenvector-centrality—a node-centrality?. *Social Networks*, 22(4): 357–365, 2000.
30. A. Bihari, and K. Pandia. Eigenvector centrality and its application in research professionals' relationship network. In *Proceedings of the 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*, 510–514, 2015

31. T. Zhou, G. Yan, and B. Wang. Maximal planar networks with large clustering coefficient and power-law degree distribution. *Physical Review E*, 71(4): 046141, 2005.
32. M.-E. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*,103(23): 8577–8582, 2006.
33. Y. Chen, J. Hu, Y. Xiao, X. Li, and P. Hui. Understanding the User Behavior of Foursquare: A Data-Driven Study on a Global Scale. *IEEE Transactions on Computational Social Systems*, 7(4): 1019–1032, 2020.
34. H. Kwak, Y. Choi, Y.-H. Eom, H. Jenog, and S. Moon. Mining communities in networks: A solution for consistency and its evaluation. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference (IMC)*, 301–314, 2009.
35. E. Saccenti, M. Hendriks, and A. Smilde. Corruption of the Pearson correlation coefficient by measurement error and its estimation, bias, and correction under different error models. *Scientific Reports*, 10(1): 438:1–438:19, 2020.
36. L. Tang, and H. Liu. Scalable Learning of Collective Behavior based on Sparse Social Dimensions. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM)*, 2009.
37. R. S. Burt. Structural Holes: The Social Structure of Competition. *Harvard University Press*, 1992.
38. Y. Ye, H. Zhu, T. Xu, F. Zhuang, R. Yu, and H. Xiong. Identifying high potential talent: A neural network based dynamic social profiling approach. In *Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM)*, 718–727, 2019.
39. J. Zhu, Y. Liu, and X. Yin. A new structure-hole-based algorithm for influence maximization in large online social networks, *IEEE Access*, 5: 23405–23412, 2017.
40. M. Lu. Node importance evaluation based on neighborhood structure hole and improved topsis, *Computer Networks*, 178: 107336, 2020.
41. T. Lou, and J. Tang. Mining structural hole spanners through in-formation diffusion in social networks, in *Proceedings of the 22nd International Conference on World Wide Web (WWW)*, 825–836, 2013.
42. W. Xu, M. Rezvani, W. Liang, J. Yu, and C. Liu. Efficient algorithms for the identification of top-k structural hole spanners in large social networks. *IEEE Transactions on Knowledge and Data Engineering*, 29(5): 1017–1030, 2017.
43. Y. D. Kwon, R. H. Mogavi, E. U. Haq, Y. Kwon, X. Ma, and P. Hui. Effects of ego networks and communities on self-disclosure in an online social network. In *Proceedings of International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2019.
44. Q. Ying, D. Chiu, and X. Zhang. Diversity of a User's Friend Circle in OSNs and Its Use for Profiling. In *Proceedings of the 10th International Conference on Social Informatics (SocInfo)*, 2018.
45. Q. Gong, Y. Chen, X. He, Z. Zhuang, T. Wang, H. huang, X. Wang, and X. Fu. DeepScan: Exploiting deep learning for malicious account detection in location-based social networks. *IEEE Communications Magazine*, 56(11): 21–27, 2018.
46. Q. Gong, J. Zhang, Y. Chen, Q. Li, Y. Xiao, X. Wang, and P Hui. Detecting malicious accounts in online developer communities using deep learning. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM)*, 1251–1260, 2019.
47. X. He, Q. Gong, Y. Chen, Y. Zhang, X. Wang, and X. Fu. DatingSec: Detecting Malicious Accounts in Dating Apps Using a Content-Based Attention Network. *IEEE Transactions on Dependable and Secure Computing*, 18(5): 2193–2208, 2021.