IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 23, 2017, accepted October 16, 2017, date of publication October 24, 2017, date of current version November 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2765363

Rising Star Forecasting Based on Social Network Analysis

ZHAOLONG NING, (Member, IEEE), YUQING LIU, JUN ZHANG, AND XIAOJIE WANG, (Student Member, IEEE)
School of Software, Dalian University of Technology, Dalian 116023, China

Corresponding authors: Jun Zhang (junezhang8900@gmail.com) and Xiaojie Wang (wangxj1988@mail.dlut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61502075; in part by the Fundamental Research Funds for the Central Universities under Grant DUT17RC(4)49; in part by the China Postdoctoral Science Foundation Funded Project under Grant 2015M580224; and in part by the Liaoning Province Doctor Startup Fund under Grant 201501166.

ABSTRACT With the development of cyber-physical-social networks, researches on evaluating social influence have drawn increasing interests. Social influence indicates the importance of people in social networks. As a typical type of social networks, how to evaluate scholar influence in the academic social network has been a practical issue for research institutions. In this paper, we aim at evaluating the latent influence of scholars to find academic rising stars, which refer to scholars that may have few papers and little impact currently, but he or she will become influential scholars in the future. Most of the current works focus on the assessment of rising stars. However, there exists a growth period for each scholar. It is unfair for young scholars with limited resources, who will make acquaintance at conferences and recommendations and who will learn from senior scholars. In this paper, we primarily propose StarRank, which is an improved PageRank method to calculate the initial values of rising stars, construct the social network via explicit and implicit links, and apply the neural network to predict scholars' rankings in the future. The experimental results on real data set demonstrate that our method has a better performance than the-state-of-the-art methods on the count of hitting rising stars and the spearman correlation coefficient.

INDEX TERMS Social influence, neural network, random walk algorithm, rising star.

I. INTRODUCTION

Nowadays, social influence prediction has becoming a popular research issue. With the developments of cyber-physical-social computing and network technologies, more and more communications are established via social networks [1]. There are many kind of cyber-physical-social networks, such as wireless sensor network [2] and vehicular social network [3]. With these we can find a social network and analysis the social influence. Social influence is a typical attribute of social network individual, and how to define and calculate it is a concerned problem [4]. Evaluating the influence of scholars can shed light on many problems, such as providing basis for award or funding application and job employment [5]. There are mature methods for finding network feature [6]. Due to the rich data and relationships in academic social networks, in this paper, we concentrate on finding rising stars in academic social networks.

A scholar publishes a paper for the first time, marking his (or her) formal entry into the academic career, in which he will collaborate with many other scholars and publish papers once and again [7]. Each scholar has a certain degree of excellence, which is reflected by the level of his collaborators and papers comprehensively. With the accumulation of papers and the expansion of his cooperative group, a scholar's academic career can be enhanced. Every year, many new scholars start their academic career and make their own contribution to the scientific undertaking. At the same time, each scholar may be subject to one or more academic institutions, which will try to attract outstanding scholars to join and expand their influence for their research strength improvement. The research strength of an institution depends on the number of excellent scholars and the high quality of papers. Excellent scholar means he or she has a significant impact in a specific academic area. In fact, most people they employ are those young scholars with a short academic career or those who have just started their academic career. As a result, it has become an important problem faced by the administration personnel on how to employ excellent scholars or the young scholars with potential who will become excellent in the future, i.e., the rising stars [8]. Excellent scholars are easily to judge by h-index, g-index and other methods. However rising star is hard to evaluate because of few evalucating indicators. In this paper we will forecast who is the rising star, and the scholar becoming excellent one finally.

Besides employment, how to identify a rising star is also a meaningful problem when choosing collaborators. According to the characteristics of social interaction, an excellent scholar will communicate and cooperate more with the scholars as excellent as them to publish more papers with high quality. Therefore, it is an important question for junior scholars, i.e., how to judge rising stars so as to select better collaborators, communicate and study with them [9]. According to the work of Albert *et al.* published in Science [10], each scholar's most influential papers are randomly distributed in the academic career. A rising star has a higher probability of publishing papers with high impact, and the publication of high-impact papers is not limited by the number of published papers, the length of career, and the number of citations. If a scholar can predict the future of other rising stars and cooperate with them, the probability of publishing high-impact papers will increase [11].

In addition, it is also of great importance to forecast the rising stars for post-graduate students. With little information of supervisors, students must make best decision for themselves to forecast who is the rising star. Each post-graduate student hopes to select excellent scholars as his of her mentor, who will offer more guidance and help. However, the number of excellent scholars in the field is limited, and it is unrealistic that the mentor of each graduate student is an excellent scholar. On the other hand, as the limited energy of excellent scholars, maybe the student's mentor is an excellent scholar nominally, but in fact the students are conducted by junior scholars in the same institution. These junior scholars are also excellent, but because of the shorter academic career and the smaller number of papers published and cited, their excellent degrees may be underestimated. Meanwhile, these junior scholars with shorter academic careers have more energy to conduct the students, so it is sometimes better to choose these scholars as mentors, and students will acquire more knowledge and experience than selecting senior scholars. Therefore, it is also vital for post-graduate students on how to predict the future rising stars and choose mentors.

In real life, the networks of scholars and papers are heterogeneous. Scholars, as nodes in the academic social network, will be affected by other nodes [12]. There are some kinds of links among scholars. Adamic *et al.* [13] regard Email as the communication way and make analyses. Twitter and Facebook are also typical communications for social analysis [14]. In fact, the relationship among scholars is embodied in the paper and the citations among papers reflect the degree of excellence of the paper, which is a significant judgment standard of excellent degree of scholars. The reality citation network is a complex network with the characteristics of scale-free, small-world and high concentration. In this network topology, nodes are closely linked. A paper has many references and is cited by others at the same time, which is quite similar to the Internet page links. Therefore, we can estimate the importance of the nodes in the citation network with existing algorithms of link analysis. The PageRank algorithm [15], as an excellent link analysis algorithm, is applied to the site rankings and works very well. In the Internet, the relationship between nodes is calculated by the hyper link, which is a one-way connection. A page will point to other pages, and it will be pointed by others, which is very similar to the citation network. The PageRank algorithm assumes that the probability of each page's rank flow is same for each reference. Therefore, the dynamic model of the random walk process can be adopted. The degree of page's excellence depends on the referenced pages according to the directions. The more times the page is cited, the higher excellence rank is, and the more important the page is.

Sinatra *et al.* indicate that each scholar has a talent level, which will affect the excellence of career. There exist a lot of works on the scholars' assessment. Most of them are mainly based on the heterogeneous academic network topology, and the assessment of the importance in the network. The representative work is the PubRank algorithm proposed by Li *et al.* in 2009 [16]. PubRank algorithm is based on PageRank algorithm with random walk application in the scholar network. Each scholar node has a certain degree of excellence and the degree of excellence passes through the cooperations. The excellent degree of each scholar is assessed by the papers published and the journals in which they are published. In addition, the PubRank algorithm adds weights to the links of the scholar's network. The more times two scholars cooperate, the more chances of their outstanding levels flow. PubRank is a good algorithm to measure scholar's excellence. They believe that the top-ranking scholars contribute more to the paper. The closer a journal's contents are, the higher the journal rank is. StarRank uses a heterogeneous network of scholars - journals - papers and the information entropy method to calculate the degree of journal's excellence. LeaderRank proposed by Lü *et al.* [17] is based on PageRank and presents a new approach to find important nodes in social networks, which can also be used to estimate excellent nodes in scholar networks. Meanwhile, there is QRank [18] for searching rising stars. This method has a good performance with less time cost.

However, there exists little work on the prediction of scholar excellence. In our work, we simulate and analyze the spread of the excellent degree of scholars through the propagation model of complex networks. In the dissemination of information, a node spreads information at a stage, and in the next stage, it will receive information. After the continuous dissemination and deduction, information can be spread in the entire network. The dissemination of scholar academic level is similar to information. In a certain period, the academic level of a scholar can be improved when receiving, and the academic level of neighbors will be promoted in another period when spreading. Therefore, we can apply the propagation model of complex networks to analyze the spread of excellent degree.

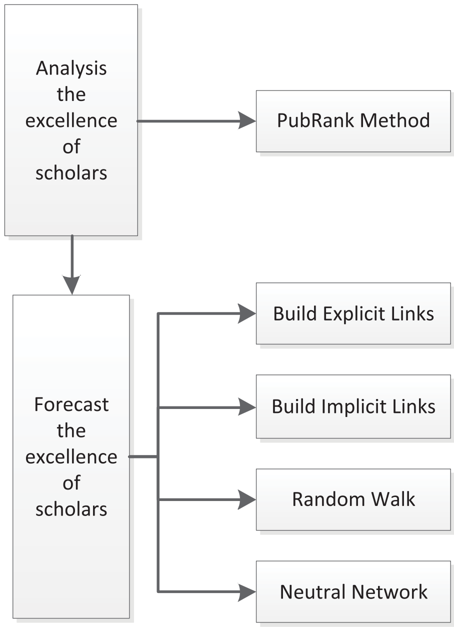


FIGURE 1. Illustration of our method procedure.

Most of the network analysis method are based on machine learning [19]. In this paper, we first process the scholar-paper heterogeneity network based on the PubRank algorithm and assess the excellent degrees of scholars by analyzing the quality of papers and the cooperative relationships. Next, we establish the academic network with explicit links. Then we assess the degrees of the scholarsąŕ activity and research extension, calculate the homogeneities among scholars, and establish the implicit scholar network links through the ternary relation and social activity. After establishing the links, the simulation is carried out by the random walk process. Finally, the calculated ranks and other features of scholars are trained by a neural network. The procedure of our method is shown in Fig. 1. We use the Support Vector Machine (SVM) method as a baseline method, which is a traditional regression method [20], and the result shows our method has a better performance.

II. ASSESSMENT OF SCHOLAR'S IMPACT

In this section, we improve the PubRank algorithm, and estimate the excellent degree of scholar. In the PubRank algorithm, some papers are published for just a short time and cited little. The assessment of each paper is mainly based on journals and the papers published in the same journal are identified as with same academic level, which is actually unfair. Some excellent papers sometimes may be published in ordinary journals due to anxiety and other factors. In addition, the quality of papers varies even if they are in the same journal. Therefore, the quality of a paper judged by its importance in the paper network will be more reasonable. It is true that some of the later published journals have a lower number of citations, even if they are of high quality. It will lower the author's excellent degree. However, it should be noted that what we estimate is the degree of scholar excellence at the moment. A less-cited paper has less attention at the moment, so the excellent degree of scholar at the moment does not depend on the paper. In other words, a less-cited paper does not have a significant impact on the scholar's current level of excellence.

In our approach, a heterogeneous network of scholar-paper is established and the excellent degree of scholar paper is evaluated. Then according to the excellent degree of paper and the cooperation between scholars, we evaluate the outstanding degree of scholars. We first estimate the excellent degree of paper as follows:

$$Rank(p) = \frac{1-d}{N} + d \times \sum_{i}^{N_p} \frac{Rank(i)}{N_i} \tag{1}$$

where $Rank(p)$ is the outstanding degree of the paper p. N_i is the number of papers cited in the i-th paper. An excellent paper will cite other excellent papers definitely, and be cited by other excellent papers. Therefore, the more excellence a paper is, the higher rank it is.

After calculating the excellent degree of each paper, we estimate the excellence of each scholar. Let each scholar's collection of papers be P, and the quality of each paper in P constitutes the publication level of the scholar. The scholar's publication level means the scholar's average paper quality, which roughly represents scholar's academic level. The calculation is as follows:

$$\lambda(a_i) = \frac{1}{|P|} \sum_{i=1}^{|P|} \frac{1}{\alpha^{r(pub_i)-1}} \tag{2}$$

where $\lambda(a_i)$ is the publication level of scholar a_i and $r(pub_i)$ is the excellent degree of scholar paper. α is an attenuation factor, and $0 < \alpha < 1$. The larger the α is, the smaller the value of λ is. The higher the value of λ is, the higher the scholar's publication level becomes.

When establishing the network of scholars, the communication between two scholars is an important judgment factor of the weight of network edge. The closer the communication between the two scholars is, the greater the probability of academic impact dissemination is. The cooperative relationship between the two scholars is reflected by cooperative papers. The authors of the same paper collaborate on the same work. The closer the cooperation is, the higher the weight of cooperation becomes. The weight of cooperation between two scholars is calculated as follows:

$$w(a_i, a_j) = \frac{(a_i, a_j)}{\sum_{k=1}^{|V|}(a_i, a_k)} \tag{3}$$

where (a_i, a_j) is the number of cooperative papers of scholar a_i and scholar a_j. The more times of a_i and a_j cooperation and the greater the proportion are, the higher the edge weight of the two nodes becomes. This communication relationship is an one-way link, meaning the degree of a_i concerns on a_j.

Algorithm 1 Forecast Method

Input: Dataset and paramters
Output: Excellent rank of scholars
Build explicit links with citations and cooperations
Set the random walk times T
for i = 0 to T **do** ąąąą
    Build implicit lines ąąąą
    Apply the random walk method
    Add the delta ranks
end for
Train data by neural network
return Ranks

The higher the edge weight is, the more chance a_i has to communicate to a_j.

After obtaining the scholar cooperation network and the publication level, we estimate the excellent degree of scholars. The evaluation method is as follows:

$$Rank(a_i) = \frac{1-d}{N} + d \times \sum_{j=1}^{|V|} \frac{w(a_i, a_j) \times \lambda(a_i) \times Rank(a_j)}{\sum_{k=1}^{|V|} w(a_k, a_j) \times \lambda(a_k)} \tag{4}$$

where $Rank(a_i)$ is the excellent degree of scholar a_i's, N is the total number of scholars, and d is a variable between 0 and 1 to control the network transmission ratio. The larger the value of d is, the higher the scholar position influence in the network becomes. Thus, we obtain the excellent degree of scholars at the initial moment and regard it as the initial weight of each scholar node.

III. RANK FORECASTING

In the scholar network, the excellent degree of scholar is a self-owned property of each node. The communication and influence of scholars have a significant impact on their interactions. A scholar who is fresh in the field of study will achieve a significant improvement if he is able to communicate frequently with excellent scholars in relevant fields. Establishing the scholar network links is an important step to predict the excellent degree of scholar. The scholar network structure is divided into two parts: explicit links and implicit links. The explicit links illustrate the author's cooperative relationship from the point of the paper, i.e., they are direct links. The implicit links are described by the activity degree and the triple of recommendation by the acquaintance. Explicit and implicit links reflect the scholar's learning relationship from two aspects respectively, and indicate the flow direction of the excellent degree. The dynamic model of the random walk has simple computational complexity and efficient computational results. The steps of rank forecasting is shown in Algorithm 1.

A. EXPLICIT LINK

An explicit link is a directed network construction according to the current paper cooperation and reference network. It expresses the communication relationship between the scholars. An ordinary scholar has established a one-way communication with a high-level scholar, and the ordinary scholar has a learning link to the high-level scholar. His academic level may be improved. Scholarsąŕ references and cooperative relations are mainly embodied by papers. In order to reflect these two relations, we need to establish a heterogeneous network, consisting of scholars and papers and regard papers as links of scholars. Two kinds of the heterogeneous network relations are shown in Figure 2.

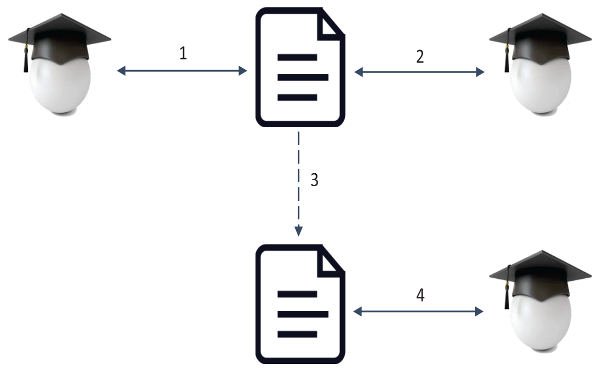


FIGURE 2. Illustration of heterogeneous network relations.

There are two kinds of edges in the network, including the cooperation relationship of scholar-paper and the reference relationship of paper-paper. The cooperation relationship of scholar-paper is a two-way edge because a scholar contributes to the paper in the cooperation and other scholars learn from their work. Similarly, the scholar can also learn from other scholars on his work. The reference relationship of paper-paper is a one-way edge because when a paper cites another, the paper learns from other papers, but the other papers do not learn from this paper. There are two kinds of links in the heterogeneous network represented by link 1-4 and link 1-2. Link 1-4 represents references and partnerships. Link 1-2 indicates that two authors have collaborated on a paper and have learned from each other in cooperation, so link 1-2 is bidirectional. Link 1-3-4 represents a reference learning relationship, and edge 3 is a reference relationship, so it is an unidirectional link. Through link 1-3-4, one scholar can learn from another scholar. Each scholar learns through his own link connected with other scholars in the heterogeneous network, each link has its own weight, and the weight represents the learning level of the two nodes. The higher the weight is, the closer the two nodes are connected and the higher the possibility of learning is. Here we assume that each scholar's learning ability is the same, that is, when two links have the same weight, the probabilities of the two scholar nodes to learn are the same. There are two kinds of links between the two nodes in the heterogeneous network. As the cooperative link and the reference link have different learning degrees, the weights of the two links are calculated differently. Suppose the link weight is S, then for the bidirectional cooperative link S_{12}, the calculation method is:

$$S_{12} = \overline{\sum_{\beta} a^{\alpha}} \tag{5}$$

where β is the paper collection of all the papers by cooperation, α is the paper's excellent degree, and a is an experimental constant. It is true that the more excellent a paper is, the more efforts are made by the collaborators, the more interaction and discussion they have, and the more opportunities for mutual learning. For the one-way reference link S_{134}, if a scholar cites more papers of another scholar, the scholar is more likely to have a greater amount of reading for the cited paper, and to learn and think more deeply about his work. Therefore, the greater proportion of cited papers exists, the more chances to learn and the larger weight of the link is. Here we assume that the contributions of each reference and each scholar to the paper are the same, then the unidirectional reference link is calculated as follows:

$$S_{134} = \overline{\sum_{\beta} (a^{\alpha} \frac{cita_{au}}{Citation})} \tag{6}$$

where $Cita_{au}$ is the number of references to the author's papers, and $Citation$ is the number of citations of this paper.

After computing the explicit link, we need to normalize it. Because we assume that the learning ability of each scholar is the same when constructing the link, the weight of each pointed edge represents the probability of the scholar's learning from other scholars. The normalization can not only facilitate the calculation but also make it possible to compare the relations between different scholars, which can be used for the later strength assessment.

B. IMPLICIT LINK

Next, we need to compute implicit links. Implicit links indicate the potential linkages between scholars, which are mainly reflected in acquaintances' mutual recommendation and acquaintance. The implicit link is an abstract link, which does not exist obviously in the society, but at a certain time two nodes will be constructed by the link with the some probability. To a great extent, the construction of implicit links is an abstract and fuzzy construction. The fuzziness, on one hand, is reflected by the probability of construction. On the other hand, it is reflected by the evaluation of network parameters.

For implicit link construction, we use parameters of scholar research extensiveness and scholar activity. A scholar's research extensiveness and research activity have a great influence on his excellent degree. The more extensive and active the research is, the more academic conferences he will participate in and publish in a relatively large number of journals; then the greater possibility of contact, learning, and cooperation with other excellent scholars exists, which has a great effect on the mutual learning to promote themselves. At the same time, we assume that each scholar has the same effort in each direction. The more extensive a scholar has studied, the less effort is made in one direction, and it has restricted communication. The extensiveness of scholar research is expressed as follows:

$$Wide_{Au} = \frac{|N_{Au}|}{|N|} \tag{7}$$

where N_{Au} is the set of title words of all the papers published by Au, and N is the set of title words in the whole paper data set. The title of an article can fully summarize the paper. After elimination of the attributives with no practical significance, the proportion of the number of title words in the whole paper can be used to show the extensiveness of scholar research in a certain degree.

Scholar activity is also one of the main parameters of prediction, which has a great influence on the degree of scholar excellence. In various fields, scholars will participate in many academic conferences and forums. In the meeting, scholars can make acquaintances and have a discussion with each other. In face of the same research directions and research points, two scholars are likely to have a cooperative relationship for future mutual learning. Therefore, it is an important part of constructing implicit links to analyze the learning relations of two scholars through social activity. The more active the scholar is, the more communication he will attend, and he will have more chances of meeting other outstanding scholars. With more communication, he will make greater academic advancement. An active scholar is more inclined to communicate with other active scholars. Therefore, the link analysis method can be used to estimate the degree of scholar activity, which can be calculated by:

$$Active_{Au} = \frac{1-d}{N} + d\sum_{j=1}^{m} \frac{Active_{Au_j}}{L(Au_j)} \tag{8}$$

where $L(Au_j)$ is the number of scholar's collaborators, that is, the out-degree of the node in the network. d is the adjustment parameter. N is the number of scholars in the network.

How to quantify activity has become an important problem to construct implicit links. "Small World Network" is an important phenomenon in social networks. In the general social network, sometimes we will find that two people who do not know each other can get to know via few edges, which is a very important phenomenon in the real network. In the scholar network, this phenomenon is still obvious. As a result, a scholar can connect with another scholar by only a few edges. Therefore, the link between scholars is close. In a real academic network, scholars have a very wide range of contacts, the edges in the network are also extensive. At the same time, the real scholar network has a large number of nodes, and presents a close feature. In such a network, the number of nodes is extremely large, but the distance between two points is limited. Therefore, we can view it as an infinite network.

In this infinite-scale network, a scholar can spread academic knowledge to others and obtain knowledge from others as well. We can apply the SIS model to abstract this relationship. SIS is a typical network model to simulate the interaction of people and make immunization [21]. The spread of knowledge is similar to the infectious disease transmission, which makes people turn to a special status and return to initial status after some time. In the SIS model of a typical infinite-scale network, there are two kinds of nodes: the susceptible node S and the infected node I. At a moment, susceptible nodes will be infected by surrounding infected nodes in a certain probability and become infected nodes. Meanwhile, infected nodes will recover from the infected state and become susceptible nodes. The threshold in the SIS model is defined as:

$$\lambda_c = \frac{< k >}{< k^2 >} \tag{9}$$

When the probability of infection in the network exceeds λ_c, the nodes of the entire network will be infected over time. We can use this threshold to characterize the activity of the scholar. The more active the scholar is, the higher the probability of spreading the academic level exists, and the greater the probability of infecting others in the SIS model is. Note that in the SIS model, the weight of each edge is the same, that is, the probability of mutual infection is the same. In the real network, each scholar has a different degree of activity. Thus, in the evaluation it needs to be scaled to the activity interval. When $Active_A > Activet$, the scholar is regarded as active. $Active_t$ is defined as:

$$Active_t = Active_{min} + (Active_{max} - Active_{min}) \times \lambda_c \tag{10}$$

When two scholars are similar, they are more likely to establish a learning relationship. Here we use the coefficient of homogeneity to compare the similarity of two scholars. Homogeneity is an important feature in social network [22], which affects the behavior of scholars. In our work, the higher the homogeneity is, the higher the degree of similarity between the two scholars is, and the greater the probability of mutual learning exists. The homogeneity is denoted by ε, i.e.,

$$\varepsilon(A, B) = \frac{|N_A \cap N_B|}{|N_A \cup N_B|} e^{-|\frac{Rank_A}{Wide_A} - \frac{Rank_B}{Wide_B}|} \tag{11}$$

where N_A and N_B are published papers' title sets of scholars respectively. Rank and wide show the scholarsą́ energy distribution on the research field. The higher the ε is, the more similar two scholars are. In the calculation of homogeneity, we consider the differences of research fields and energy distribution of the two scholars. The greater the difference between the research fields is, the more different the energy distributes, then the lower the homogeneity becomes, and the less opportunity to form a learning relationship.

Meanwhile, although there are many conferences in academic area, one scholar has little spirit to join a conference less similar with his research. We use a parameter $Simi$ to calculate the similar research ares. If two scholars have similar research area, they will have more chances to join the same conferences and make acquaintance. $Simi_{AB}$ is calculated as:

$$Simi_{AB} = \frac{|N_A \cap N_B|}{|N_A \cup N_B|} \tag{12}$$

If $Simi_{AB}$ is less than the similarity threshold $T_{similar}$, then a link will not exist between A and B. If both A and B are active scholars, and they have similar research areas, then a link with the strength of ε will be established between them in the probability of P, where

$$P = \min\{(Active_A + Active_B)\varepsilon(A, B), Active_t\} \tag{13}$$

In addition, acquaintance recommendation is a common kind of implicit relationship. If the relationship between scholar A and B is close and the relationship between scholar B and scholar C is also close, A, B and C can constitute a ternary relationship group, A and C can also have great possibility to get to know each other and cooperate, which is a common situation in social life. The ternary relation group is based on the explicit link and deduced in the normalized explicit link. A typical triple relationship group ABC is shown in the Fig. 3.

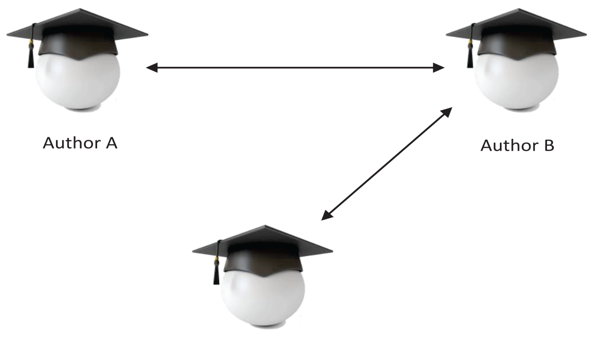


FIGURE 3. Illustration of triple relationship.

In Fig. 3, A and B constitute a two-way link, B and C constitute a two-way link. If the edge weight of A and B is high and the edge weight of B and C is high, A and C will form a link in the future with strong possibility. It is noted that the links are unidirectional for the scholar nodes and the direction represents the transfer of learning, thus the constructed implicit links are also unidirectional, so that the weights of link S_{AC} and link S_{CA} are different. We need to give the definition of strong links before computing implicit links. When the weight S of an unidirectional link is larger than the threshold S_t, we call the link S as a strong link. The strength of the link determines the tightness of the learning relationship between the two nodes. If the relationship is strong, they will pay more attention to each other and the recommendation relationship between them will be strong. For instance, the relationship between A and B is close as well as the relationship between B and C, then A is likely to have a close learning tendency towards C, and a learning link is constructed.

If link S_{AB} and link S_{BC} are bidirectional strong links, then,

$$S_{AC} = \begin{cases} \sqrt{\sum S_{AB} \cdot S_{BC}} & S_{AC} > S_t \\ 0 & else \end{cases} \tag{14}$$

where $S_t = S_{min} + (S_{max} - S_{min}) \times \lambda_c$.

After constructing the implicit links, we need to process them so that the sum of the out-degree of each node is 1. On the one hand, the calculation scale will be consistent with the explicit links and the data can be processed uniformly. On the other hand, it will be possible to compare the link relations of different scholar nodes.

C. IMPACT EVALUATION

After constructing the implicit and explicit links, the next step is to evaluate the dissemination of excellent degree on the links. The rank of scholar excellence spreads across the entire network of scholars, and the random walk process is employed for evaluation. We average and analyze them based on statistical data to evaluate the spread of scholars in the real environment. The random walk is a very classical method of dynamic evolution. Based on the previous implicit and explicit link networks, we deduce for M generations with the following derivation:

$$\Delta Rank_A = Rate \times \sum_k \max\{\frac{S_{Ak}}{S_A}(Rank_k - Rank_A) \times P_{Ak}, 0\} \tag{15}$$

where $Rate$ is the learning rate, S_{Ak} is the weight of link Ak, S_A is the sum of all out-degrees of A, and P_{Ak} is the probability of receiving impact, which can be calculated as follows:

$$P_{Ak} = \begin{cases} 1, & random(0,1) > S_{Ak} \\ 0, & else \end{cases} \tag{16}$$

It can be seen that the greater the difference between the scholarsąŕ academic levels is and the higher the weight of learning becomes, the more the level is promoted. An active scholar has higher activity and social degrees, so relatively their social relations based on social activities have played a significant role in improving their academic levels. Therefore, we deduce the explicit and implicit links separately, and the differences calculated by each deduction are denoted as $\delta Rank_E$ and $\delta Rank_I$ respectively. For scholars, the final improvement of after a deduction is:

$$\Delta Rank_a = (Active_a \times \Delta Rank_H + (1 - Active_a) \times \Delta Rank_E) \times Rate \tag{17}$$

where $Rate$ is a constant value to control the speed of spread.

After deduction, the implicit links will be recalculated to evaluate the social uncertainty. In addition, after promoting academic levels, the homogeneity between scholars will change, resulting in changes of learning weights. Therefore, it is necessary to recalculate and construct the implicit links. After the simulation of K times, the excellent degrees of scholar are the results of our experiments.

D. RANK FORECASTING

To ensure the forecast result, we should consider the impact factors as many as possible. Talent rank is one of the features of achievements. If a scholar with high talent has few spirits on research, the count of citations will be not large. Thus, finding outstanding scholars only based on rank is not much reliable. To solve this problem, we use a neural network to deal with the relationship between features and the achievements.

The neural network is trained with outstanding rank, count of co-authors, count of publications, count of citations, and the citation number per year. The output of the neural network is the count of citations in the future. All the data should be normalized.

In this neutral network, we use sigmod as the active function. For sigmod function has a restrict output range, it is hard to diffuse. Meanwhile, sigmod function has a output range between [0,1], which is suitable for the training data. The structure of neutral network is designed by 3 layers, which is illustrated as Fig. 4. Between every 2 layers we apply full-connection. After training the network, it is used to predict the outstanding scholars.

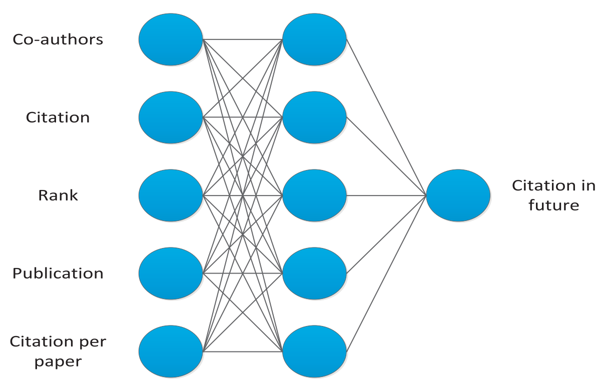


FIGURE 4. Illustration of structure of neural network.

IV. EXPERIMENTS & RESULTS

The scholar data set of American Physical Society (APS) is used as the experimental data set. APS data set includes the data set of more than 10 kinds of its publications. In APS data set, we can get many indexes, such as DOI, collaborators, titles, and references. To furthermore conduct the experiments and analyses, we extract the data of 5 years for experiments and comparison, count the number of the scholarąŕs cited papers each year, and take it as the evaluating indicator. Different evaluating indicator will lead to different performance [23]. In our work, we process the sub data set from 1975 to 1980 and the set from 1980 to 1985 for prediction, and use citations from 1985 to 1995 and citations from 1990 to 2000 to verify the results. Details can be found in Tables 1 and 2.

All the experiments are written by python. The machine learning library for neutral network is pybrain. Pybrain is an open-source library for building neutral network in python environment. With this environment we can concentrate more on the network structure than the implementation. The library for SVM is sklearn. Sklearn is a powerful machine learning library. It provides classification, regression, clustering and other functions. This library is supposed

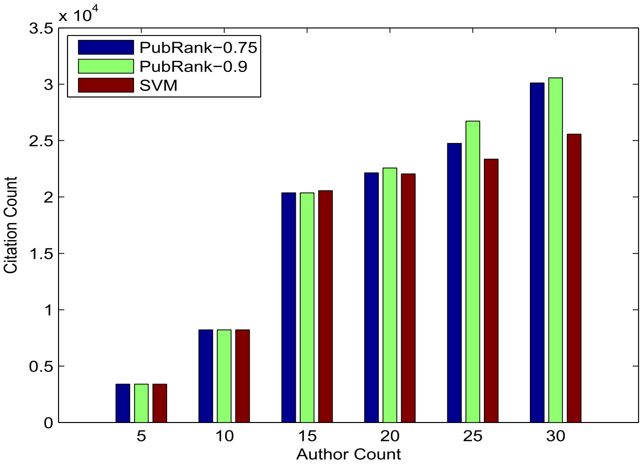


FIGURE 5. Performance of PubRank from 1975 to 1980.

TABLE 1. Data with PubRank from 1975 to 1980.

Author Count	PubRank-0.75	PubRank-0.90	SVM
5	3406	3402	3406
10	8212	8212	8212
15	20365	20365	20561
20	22138	22565	22050
25	24752	26730	23353
30	30109	30569	25564

TABLE 2. Data with PubRank from 1980 to 1985.

Author Count	PubRank-0.75	PubRank-0.90	SVM
5	12471	14545	12471
10	22993	22841	21727
15	33230	33230	31518
20	40465	39450	40013
25	54274	49545	42996
30	57835	57721	54435

by Google and becomes a mainstream library for researchers.

The baseline method is based on the SVM method. It is a traditional prediction model. There are some prediction works based on SVM [24]–[26] We train the SVM model with all processed data and the rank we used in the neural network. To show the influence of wide similarity threshold $T_{similar}$, we set $T_{similar}$ as 0.75 and 0.9 to conduct the experiments. The corresponding performance evaluations can be found in Figs. 5 and 6.

The SVM model is set with the following parameters. The kernel function of SVM is rbf kernel. The tolerance of model is 0.001. SVM model will not stop running until the tolerance is lower than 0.001.

From the results shown in Figs. 7 and 8, we can find that with the increase of author counts, our method hits more citations than SVM, i.e., more the counts, more significant the difference of performance is. Although different thresholds have different performances, both of them are better than SVM. To show the influence of initial rank, we use StarRank as our initial rank.

From the result in Tables III and IV, we can find that StarRank can also have a better performance than SVM when

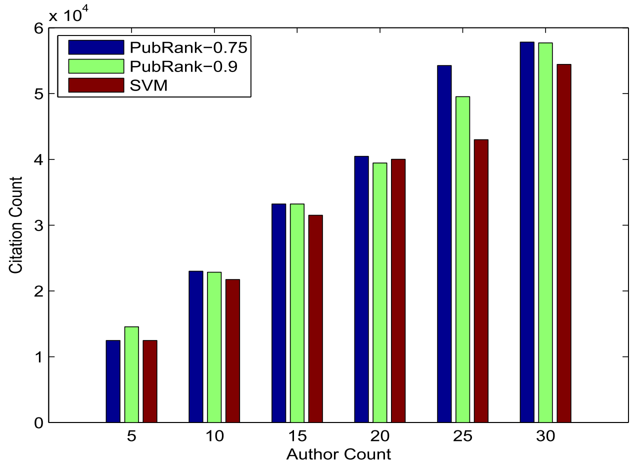


FIGURE 6. Performance of PubRank from 1980 to 1985.

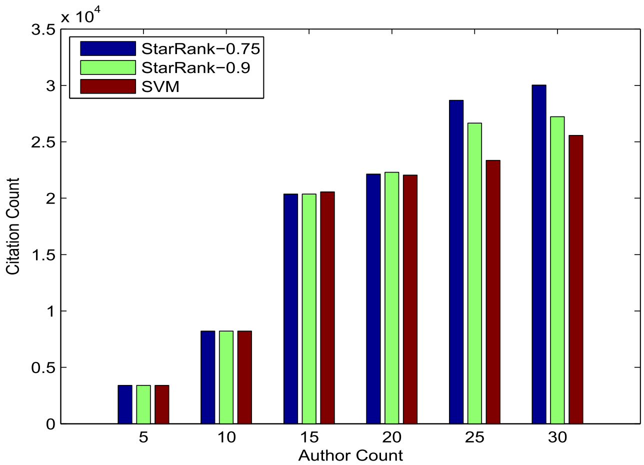


FIGURE 7. Performance of starRank from 1975 to 1980.

TABLE 3. Data with StarRank from 1975 to 1980.

Author Count	StarRank-0.75	StarRank-0.90	SVM
5	3406	3402	3406
10	8212	8212	8212
15	20365	20365	20561
20	22138	22302	22050
25	28681	26662	23353
30	30035	27226	25564

TABLE 4. Data with StarRank from 1980 to 1985.

Author Count	StarRank-0.75	StarRank-0.90	SVM
5	16126	14545	12471
10	19604	22841	21727
15	30796	33230	31518
20	40378	39450	40013
25	47940	49545	42996
30	54299	57721	54435

the count is large. Top author's hit rate is another factor to show the performance effectiveness. In fact, our task is to find future rising stars in academic social network as many as possible. Therefore, the performance for top author's hit rate is more important than the citation count. We set the

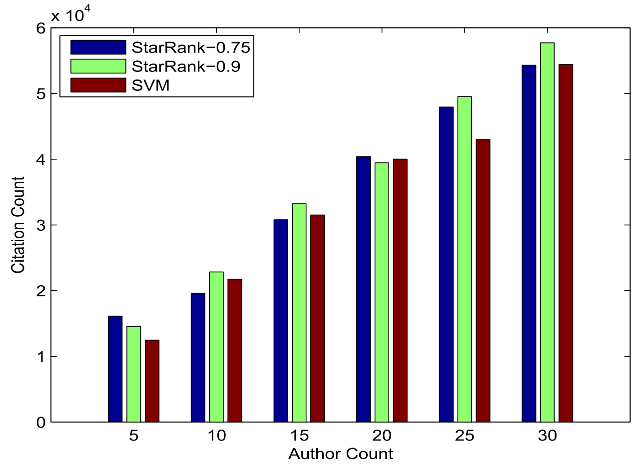


FIGURE 8. Performance of starRank from 1980 to 1985.

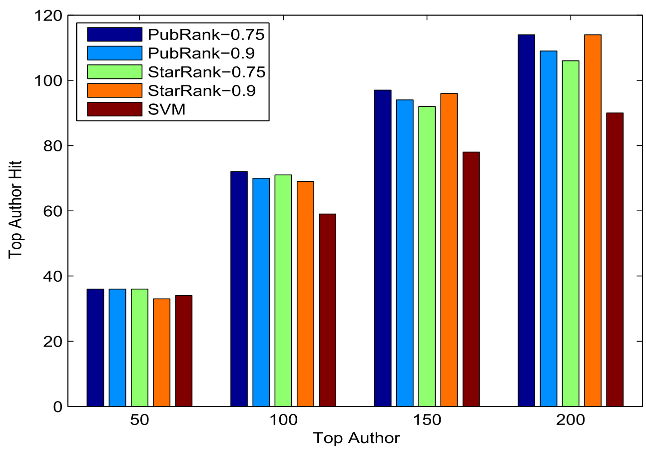


FIGURE 9. Top authors hit from 1975 to 1980.

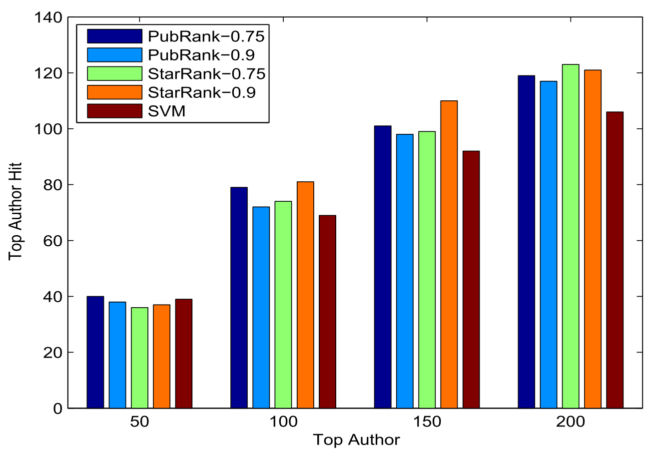


FIGURE 10. Top authors hit from 1980 to 1985.

threshold as 0.25, that is, we regard the top 25% scholars as top authors. We count the scholar's hit by algorithms. The results are shown in Fig. 9 and 10. From the results we can find that our method can hit more top authors in short scopes. Both PubRank and StarRank can have a better performance than SVM in most situations.

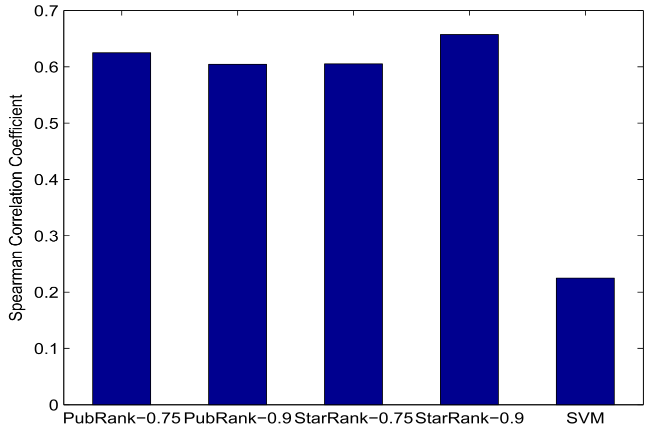


FIGURE 11. Comparison of spearman correlation coefficient from 1975 to 1980.

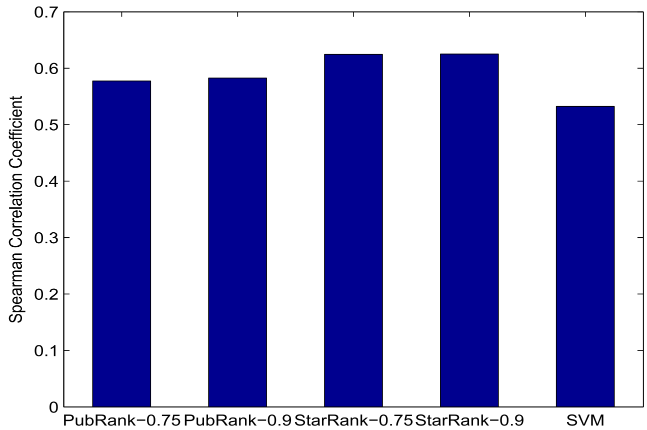


FIGURE 12. Comparison of spearman correlation coefficient from 1980 to 1985.

To compare the rationality, we regard spearman correlation coefficient as the evaluation index. Spearman correlation is a factor to estimate the correlation between 2 series of data, which is widely used in statistics. The larger the coefficient value is, the more rational the method is. The result of 2 time intervals are shown in Figs. 11 and 12. We can find the coefficient of SVM is always less than our method, which means our method has a better performance. In addition, From the experiment of StarRank, we can find the result is unstable. Sometimes the performance is not better than SVM. Initialization with PubRank may have a stable result. These experiments show our method is effective to assess the excellent rank and have a better performance.

V. CONCLUSION

In this paper, we propose the StarRank to find rising stars, and evaluate the diffusion of outstanding rank in heterogeneous networks. We consider the explicit link and implicit link in the network, and construct the transmission links of outstanding rank. Finally, we fit neural network with the rankings. We compare the performance of our method with the existing prediction method, and the results show our method has a better performance. In the future, more datasets will be processed and tested to explore the performance of our method. Meanwhile, we will consider a quantified method to decide the iteration time and the increasing rate. In addition, more methods will be employed for the initial excellent ranking.

REFERENCES

[1] Z. Ning, F. Xia, X. Hu, Z. Chen, and M. S. Obaidat, ''Social-oriented adaptive transmission in opportunistic Internet of smartphones,'' *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 810–820, Apr. 2017.

[2] Z. Ning, Q. Song, Y. Yu, Y. Lv, X. Wang, and X. Kong, ''Energy-aware cooperative and distributed channel estimation schemes for wireless sensor networks,'' *Int. J. Commun. Syst.*, vol. 30, no. 5, p. e3074, 2017.

[3] Z. Ning, F. Xia, N. Ullah, X. Kong, and X. Hu, ''Vehicular social networks: Enabling smart mobility,'' *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 16–55, May 2017.

[4] Z. Ning, L. Liu, F. Xia, B. Jedari, I. Lee, and W. Zhang, ''CAIS: A copy adjustable incentive scheme in community-based socially aware networking,'' *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3406–3419, Apr. 2017.

[5] F. Xia, W. Wang, T. M. Bekele, and H. Liu, ''Big scholarly data: A survey,'' *IEEE Trans. Big Data*, vol. 3, no. 1, pp. 18–35, Mar. 2017.

[6] Q. Zhang, L. T. Yang, X. Liu, Z. Chen, and P. Li, ''A Tucker deep computation model for mobile multimedia feature learning,'' *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 13, no. 3S, pp. 1–18, 2017.

[7] Z. Ning *et al.*, ''A cooperative quality-aware service access system for social internet of vehicles,'' *IEEE Internet Things J.*, 2017, doi: 10.1109/JIOT.2017.2764259.

[8] J. Zhang, Z. Ning, X. Bai, X. Kong, J. Zhou, and F. Xia, ''Exploring time factors in measuring the scientific impact of scholars,'' *Scientometrics*, vol. 112, no. 3, pp. 1301–1321, 2017.

[9] X. Bai, I. Lee, Z. Ning, A. Tolba, and F. Xia, ''The role of positive and negative citations in scientific evaluation,'' *IEEE Access*, vol. 5, pp. 17607–17617, 2017.

[10] R. Sinatra, D. Wang, P. Deville, C. Song, and A.-L. Barabási, ''Quantifying the evolution of individual scientific impact,'' *Science*, vol. 354, no. 6312, p. aaf5239, 2016.

[11] J. Zhang, Z. Ning, X. Bai, W. Wang, S. Yu, and F. Xia, ''Who are the rising stars in academia?'' in *Proc. IEEE/ACM Joint Conf. Digit. Libraries*, Jun. 2016, pp. 211–212.

[12] J. Tang, J. Sun, C. Wang, and Z. Yang, ''Social influence analysis in large-scale networks,'' in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 807–816.

[13] L. Adamic and E. Adar, ''How to search a social network,'' *Social Netw.*, vol. 27, no. 3, pp. 187–203, 2005.

[14] F. Xia, X. Su, W. Wang, C. Zhang, Z. Ning, and I. Lee, ''Bibliographic analysis of nature based on Twitter and Facebook altmetrics data,'' *PLoS ONE*, vol. 11, no. 12, p. e0165997, 2016.

[15] L. Page, S. Brin, R. Motwani, and T. Winograd, ''The pagerank citation ranking: Bringing order to the Web,'' Stanford InfoLab, Tech. Rep., 1999, pp. 1–17.

[16] X.-L. Li, C. S. Foo, K. L. Tew, and S.-K. Ng, ''Searching for rising stars in bibliography networks,'' in *Proc. Int. Conf. Database Syst. Adv. Appl.*, 2009, pp. 288–292.

[17] L. Lü, Y.-C. Zhang, C. H. Yeung, and T. Zhou, ''Leaders in social networks, the delicious case,'' *PLoS ONE*, vol. 6, no. 6, p. e21202, 2011.

[18] L. Li, X. Wang, Q. Zhang, P. Lei, M. Ma, and X. Chen, ''A quick and effective method for ranking authors in academic social network,'' in *Multimedia and Ubiquitous Engineering*. New York, NY, USA: Springer-Verlag, 2014, pp. 179–185.

[19] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, ''An improved deep computation model based on canonical polyadic decomposition,'' *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published.

[20] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, ''PPHOPCM: Privacy-preserving high-order possibilistic c-Means algorithm for big data clustering with cloud computing,'' *IEEE Trans. Big Data*, to be published.

[21] S. Lee, L. E. Rocha, F. Liljeros, and P. Holme, ''Exploiting temporal network structures of human interaction to effectively immunize populations,'' *PLoS ONE*, vol. 7, no. 5, p. e36439, 2012.

[22] M. McPherson, L. Smith-Lovin, and J. M. Cook, ''Birds of a feather: Homophily in social networks,'' *Annu. Rev. Sociol.*, vol. 27, no. 1, pp. 415–444, 2001.

[23] Q. Zhang, L. T. Yang, Z. Chen, P. Li, and M. J. Deen, ''Privacy-preserving double-projection deep computation model with crowdsourcing on cloud for big data feature learning,'' *IEEE Internet Things J.*, to be published.

[24] T. Ebina, H. Toh, and Y. Kuroda, ''Loop-length-dependent svm prediction of domain linkers for high-throughput structural proteomics,'' *Peptide Sci.*, vol. 92, no. 1, pp. 1–8, 2009.

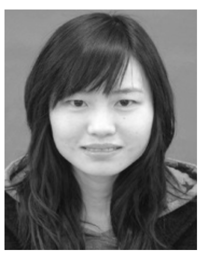
[25] L. J. Wee, D. Simarmata, Y.-W. Kam, L. F. Ng, and J. C. Tong, ''SVM-based prediction of linear b-cell epitopes using bayes feature extraction,'' *BMC Genomics*, vol. 11, no. 4, p. S21, 2010.

[26] S. Hirose, K. Shimizu, S. Kanai, Y. Kuroda, and T. Noguchi, ''POODLE-L: A two-level SVM prediction system for reliably predicting long disordered regions,'' *Bioinformatics*, vol. 23, no. 16, pp. 2046–2053, 2007.

ZHAOLONG NING (S'11–M'14) received the M.Sc. and Ph.D. degrees from Northeastern University, China, in 2011 and 2014, respectively. From 2013 to 2014, he was a Research Fellow with Kyushu University, Japan. He is currently an Assistant Professor at Dalian University of Technology. His research interests include big scholar data, network optimization, and social computing. He has authored over 60 papers in the above areas.

YUQING LIU is currently working toward the B.Sc. degree at the School of Software, Dalian University of Technology, Dalian, China. His current research interests include data mining, operation research, and image processing.

Jun Zhang received the B.Sc. and M.Sc. degrees in software engineering from Jilin University, Changchun, China, in 2010 and 2012, respectively. She is currently working toward the Ph.D. degree at the Mobile and Social Computing Laboratory, School of Software, Dalian University of Technology, Dalian, China. Her research interests include big scholarly data, evaluation of scientific impacts, and computational social science.



XIAOJIE WANG (S'16) received the M.S. degree from Northeastern University, China, in 2011. She is currently working toward the Ph.D. degree at the School of Software, Dalian University of Technology, Dalian, China. Her research interests are social computing and network security. From 2011 to 2015, she was a Software Engineer at NeuSoft Corporation.