This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Social Recommendation With Large-Scale Group Decision-Making for Cyber-Enabled Online Service

Xiaokang Zhou, *Member, IEEE*, Wei Liang, *Member, IEEE*, Suzhen Huang, and Miao Fu

***Abstract*—Along with the development of several emerging computing paradigms and information communication technologies, it is said that cyber computing technology is playing an increasingly important role across cyber-related systems and applications. In this article, we focus on cyber–social computing and propose a computational model that integrates large-scale group decision-making (LSGDM) into social recommendations for cyber-enabled online services. As a concrete application example, a graph model is built to describe the LSGDM problem among researchers in scholarly big data environments. Following the basic profiling to describe decision-makers within scholarly networks, measures are defined to evaluate one researcher's academic performance and research outcome and further quantify correlations between them based on their collaboration relationships in a constructed network model. A two-stage large-scale decision-making solution is then proposed for social recommendations: A network partition algorithm is developed based on the identification of experts along with their influence extending to a group of researchers, and a random walk with the restart-based algorithm is improved to calculate the weighted decisions for group decision aggregation and alternative ranking. Experiments using the real-world data demonstrate the usefulness and effectiveness of our proposed model and method, which can provide the target researcher with more reliable recommendations.**

***Index Terms*—Cyber-enabled service, large-scale group decision-making (LSGDM), social influence, social network analysis, social recommendation.**

I. Introduction

RECENTLY, cyber computing technology is playing an increasingly important role in the personalized recommendation, viral marketing, resource management, and pattern recognition across cyber-related systems and applications. The highly developed cyber–social networks have seamlessly integrated people's routine life and social activities together into the cyberspace, which results in a strong need to develop systematic data models and computational intelligence algorithms, to represent and explain the new phenomena, behaviors, properties, and practices in cyber–social computing environments.

Social recommendation, which mainly focuses on the utilization of user-generated information and potential value of various relationships among them [1], has brought us lots of convenience in real-world online applications. However, it still suffers from data sparsity and cold start problems, especially when people need more reliable recommendations to deal with their time-varying requirements in big data environments. On the other hand, with the popularity of online social networks, large-scale group decision-making (LSGDM) becomes very important in practice, especially when strangers are involved together in a virtual large community during the socialization process [2]. It is necessary to better utilize the collective intelligence from group decision-making processes to provide people with more reliable recommendations.

As a special example, the prevalence of scholarly big data brings with a series of academic services, such as Google Scholar, Microsoft Academic, and CiteSeer, to access and manage the scientific findings. These provide opportunities for people to access human knowledge more widely but also lead to the commonly known information overload problem. Identifying researchers' current research areas across interdisciplinary fields, or even finding key articles in different domains becomes an extremely time-consuming task when facing with a large-scale and continuously growing number of research articles [3]. In particular, situations become worse and more complex when researchers are turning into a new research field. Similar to traditional recommendation systems, which aim at suggesting items of potential interest to solve information overload [4], academic recommendation systems focus on a variety of recommendation tasks, including article recommendation, reviewer assignment, and collaboration suggestion, in order to overcome the information overload problem in scholarly big data. However, it is still far away from satisfaction. Key challenges in dealing with the dynamics of scholarly data include: how to semantically characterize the time-varying research topics and trends in terms of authors' common interests; how to dynamically identify collaboration relationships and track the evolution of the corresponding research communities over time; and how to provide accurate and sufficient scientific article list or fine-grained expert recommendation to efficiently promote scientists' research [5]. In a word, it is essential to find an effective way to process, interpret, and analyze the scholarly data systematically, to

Manuscript received February 24, 2019; revised June 22, 2019; accepted July 14, 2019. The work was supported in part by the National Key R&D Program of China under Grant 2017YFE0117500, in part by the Natural Science Foundation of Hunan Province of China under Grant 2019JJ40150, in part by the Hunan Provincial Education Department Foundation for Excellent Youth Scholars under Grant 17B146, and in part by the Key Project of Hunan Provincial Education Department under Grant 17A113. *(Corresponding author: Wei Liang.)*

X. Zhou is with the Faculty of Data Science, Shiga University, Hikone 522-8522, Japan, and also with the RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan (e-mail: zhou@biwako.shiga-u.ac.jp).

W. Liang and M. Fu are with the Key Laboratory of Hunan Province for New Retail Virtual Reality Technology, Hunan University of Technology and Business, Changsha 410205, China (e-mail: weiliang@csu.edu.cn; miaofu1997@gmail.com).

S. Huang is with the Information and Network Center, Central South University, Changsha 410083, China (e-mail: huangsuzhen@csu.edu.cn).

Digital Object Identifier 10.1109/TCSS.2019.2932288

discover the hidden knowledge and potential research patterns among diversified academic activities.

Particularly, systematic integration of academic social networks can result in information-enriched graphs with high-context-based similarities and strengthened social relations. In addition to the early definition in [6], which argued that the large group decision-making usually consisted of more than 20 decision-makers, current studies pointed out that it became a common issue if the scale of decision-makers was between 20 and 50 when dealing with large group decision-making problems [7]–[9]. However, the situation becomes far more complex and uncertain when even more decision-makers are involved in big data environments. In-depth analysis strategy needs to be developed to deal with the heterogeneous nature of big data, in order to better understand dynamic collaborations from the user-generated data along with socialized academic activities. Accordingly, it is significant to develop an integrated solution with different functionalities mixed together in a coherent manner, which can provide efficient navigation to serve researchers at different levels and further support collaborative and interdisciplinary works across academics institutions. In this article, we design a computational model to integrate LSGDM into social recommendations, which can be utilized to improve the cyber-enabled online service. Specifically, we apply our approach in scholarly big data environments and propose a two-stage large-scale decision-making-based social recommendation (DMSR), including an expert-based network partition algorithm and a random walk with restart (RWR)-based decision weighting algorithm, to better support researchers' academic activities and research collaborations. The major contributions of this article are summarized as follows.

1) A graph model is presented to describe LSGDM issues within scholarly networks, in which formal descriptions are proposed to build profiles for scholarly decision-makers, and measures are defined to analyze correlations among them based on the evaluation of their academic performances and research outcomes along with collaboration relationships.
2) An expert-based network partition algorithm is developed to deal with the situation when a considerable number of decision-makers are engaged in big data environments, in which experts are identified according to the analysis of expertise from the academic profile and influence hidden in scholarly networks.
3) An RWR-based algorithm is improved for decision weighting in each divided sub-network, results of which are utilized in further group decision aggregation and alternative ranking to achieve the final decision-making solution.
4) The designed model and two-stage solution are applied in scholarly big data environments to provide the target researcher with more reliable social recommendations for academic collaboration support.

The remainder of this article is organized as follows. Section II presents an overview of related works. In Section III, we introduce the description of scholarly LSGDM problem within a structured network model. We discuss the two developed algorithms in a two-stage large-scale decision-making solution for social recommendations in Section IV. The experiment and evaluation results using the "ResearchGate" data are demonstrated in Section V. We conclude this article and give our promising perspectives regarding future research in Section VI.

II. Related Works

Several issues relating to this article are addressed in this section. Specifically, analysis of social network and influence, LSGDM in social environments, and recommendations across social media are discussed, respectively.

A. Social Network and Influence Analysis

Social network analysis has become a hotly discussed topic during these years. More significantly, social influence identification and measurement have been widely explored, not only for big data mining and analytics but also for intelligent system developments. Vatrapu *et al.* [10] presented a so-called social set analysis as a new approach to big data analytics, in which they discussed a basic framework for philosophies of computational social science, social data theory, and conceptual and formal models of social data. An analytical framework was then constructed to combine the big social data sets with organizational and societal data sets. Zhang *et al.* [11] built a statistical model, called socioscope, in which several functional components, including zoom, scale, and analysis tools, were developed to analyze the network structure, human social behavior, and further identify social relationships, groups, and community patterns in mobile computing environments. In particular, Song *et al.* [12] proposed a probabilistic method to model users' check-in movement behaviors with their social relationships, in order to analyze the social-spatial and social-temporal influence in location-based social network environments. Zhao *et al.* [13] analyzed the network structure within three special kinds of social relationships, namely, face-to-face social relationship, online social relationship, and self-report social relationship. Supervised classifiers were developed for the corresponding feature extraction and relationship prediction from the mobile trace data and online social data. Zhang *et al.* [14] designed an influence network model for public opinion propagation. A so-called public opinion control point selection algorithm was developed with a positive guidance technology to analyze the influence network controllability and public opinion diffusion. Wang *et al.* [15] presented a feature-based social influence evaluation model according to two factors: the importance of the user himself and the possibility of impacting others. A PageRank-based algorithm was developed with a social influence adjustment model to identify influence contributions. Jiang *et al.* [16] compared the traditional influence diffusion model, PageRank evaluation model, and use behavior model for user influence. Based on these, they built a user influence computing model considering both the user relationship and activity in their online social circles. Zhu *et al.* [17] focused on the influence maximization problem in social networks. They presented a design of the framework for price-related

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

propagation with two pricing strategies that considered the price into the influence spread model. Pan *et al.* [18] improved the dynamical parameters in traditional influence model and estimated how the influence changed over time with three examples using both the simulated and real data. Taking the social influence between users and their influence over the whole social network into account, Li and Xiong [19] proposed a social recommendation method based on users' direct connections and interactions, which could be used to improve the accuracy of recommendation results.

B. Large-Scale Group Decision-Making

In social network environments, LSGDM issues have become very common, strategies and mechanisms based on decision-making results are widely applied in practice. Urena *et al.* [20] reviewed the existent mechanisms regarding opinion dynamics and influence assessment in social networks. They discussed the challenges and research opportunities to utilize trust, reputation, and influence to facilitate decision-making processes and recommendation strategies, which could be leveraged in complex social networks scenarios. Wu *et al.* [21] defined the four tuple information to model and describe preferences of experts based on their indirect trust relationships, which could contribute to achieving a final solution for the group decision-making within social networks. To deal with the inconsistency issue in the group decision-making process, Liu *et al.* [22] developed a trust-based recommendation mechanism to provide personalized advice. The result could help group experts to achieve a higher consensus degree. Wu *et al.* [23] built a visual interaction framework for the consensus model in social network group decision-making processes. The trust relationship was analyzed to determine the experts' trust degrees, and a recommendation algorithm was then proposed to assist the consensus reaching process. Dong *et al.* [24] conducted a survey on consensus reaching processes in the social network group decision-making. They presented a basic framework with several social network concepts and identified the trust relationship and opinion evolution as the two important factors in consensus reaching process paradigms. A so-called two-stage solution was proposed in [9], which aimed to reduce the complexity of LSGDM problems and aggregate the comprehensive decision information based on a network partition algorithm and the shortest path algorithm. In particular, the trust relationship with linguistic information was modeled in [25]. Based on three levels of consensus degree and linguistic trust functions, a feedback mechanism with a theoretical framework was developed to provide the recommendation advice and increase the group consensus degree within a networked group. In [26], an estimation method was proposed based on the defined trust score for the multiple criteria group decision-making. A visual consensus aggregation model was constructed to facilitate the achievement of satisfied consensus level and further recommend experts with personalized advice.

C. Recommendations Across Social Media

Recently, social network-based recommendation approaches have drawn more and more attention in both academe and industry, which can provide more reliable recommendation results. Gao *et al.* [27] focused on the location-based social network and constructed a point-of-interest-based recommendation model. They analyzed users' geo-social correlations and influence and integrated them with a three-level joint pairwise ranking scheme to improve the recommendation accuracy. Lai *et al.* [28] considered users' trust relationships, product popularity, and social interactions, such as their rating, sharing, and posting behaviors in a social recommendation method, to provide them with relevant products according to their predicted preferences in social networks. Deng *et al.* [29] proposed a matrix factorization-based method for the social recommendation. The deep learning technology was utilized to deal with the initialization issue, and a so-called social trust ensemble learning model was built for community detections according to trust relations. Cui *et al.* [30] employed a matrix factorization method to integrate the deeper membership and the deeper friendship between users for social recommendations. In particular, the deeper membership similarity was calculated based on the Jaccard similarity coefficient, and the deeper friendship similarity was analyzed based on the random walk algorithm. To deal with real-world online applications, Zhao *et al.* [31] discussed a social recommendation framework based on the learning of user preference. The collaborative user–item relationship and item content features were integrated into a unified preference learning process, and the Frank–Wolfe algorithm was used to improve the result in an iterative procedure. Jiang *et al.* [32] developed a hybrid random walk mechanism, which incorporated the item transferability, such as popularity and behavioral consistency, to transfer the knowledge from multiple relational domains in graph-based applications. The results could be utilized to predict user–item links and provide recommendations to cold-start users. Stepan *et al.* [33] improved the traditional collaborative filtering algorithm, which took the spatial, temporal, and social context data into account together, to provide the optimized recommendations in location-based social networks. Yang *et al.* [34] exploited a nearest-neighbor-based recommender system with a set of matrix factorizations, which employed users' group affiliation information and social networks to improve the popularity-based social voting recommendations. Wang *et al.* [35] designed a joint social-content recommendation framework, which measured the relevance between users and items, the similarity between items, and influence between connected users, to improve the accuracy for re-sharing recommendations in online social networks. Sun *et al.* [36] presented a social-aware group recommendation framework, in which experts' relationships with group members were taken into account with their social behaviors. The group preference was then modeled based on users' tolerance and altruism features, and an algorithm was developed to provide recommendations under different social contexts.

III. Modeling of Scholarly Large-Scale Group Decision-Making Problem

In this section, we introduce and define the basic model to describe relationships among a series of decision-makers in a scholarly LSGDM problem. Formal descriptions are then given to represent the profile of and the correlation between decision-makers within scholarly networks.

A. Basic Model Description

To deal with the LSGDM problems issued within scholarly networks, two basic academic entities, namely, the researchers and the articles, are considered in scholarly environments. In particular, considering the basic principle of LSGDM problems, the co-author relationships between researchers and citations among articles are employed to construct the network model. Given a scholarly LSGDM problem, the basic model to describe the relationships among a series of decision-makers can be expressed as follows.

$$G(R, E, W). \tag{1}$$

$R = \{r_1, r_2, \ldots, r_n\}$ is a non-empty set of decision-makers participating in the decision-making process, in which each r_i denotes a specific researcher in the academic social network. Specifically, $r_i = (uid_i,\ P_i,\ A_i)$, in which uid$_i$ denotes the researcher ID to identify a unique researcher r_i, P_i is a five tuple with a series of attributes to describe the scholarly profile of r_i, and A_i denotes the set of articles published by r_i.

$E = \{e_{ij} = < r_i, w_{ij}, r_j > |$ if a relationship exists between researcher r_i and $r_j\}$ is a collection of edges that connect those researchers in R. Specifically, e_{ij} denotes the connection between two researchers r_i and r_j based on the co-author relationships extracted from their scientific publications and academic activities.

$W = \{w_{ij} |\text{if } \exists\ e_{ij} \in E\}$ is a set of measures to describe and quantify the relationship on the corresponding edge.

Following these above-addressed definitions, the initial indirect scholarly network *G (R, E, W)* is basically constructed according to co-author relationships, in order to deal with the LSGDM problem and better leverage the collective intelligence in scholarly big data environments.

B. Profiling of Decision-Maker in Scholarly Environments

Initially, each individual researcher, who can be viewed as the decision-maker in scholarly environments, is connected with others based on similar research interests across the large-scale scholarly network. Researchers with the similar specialty fields, research background, and knowledge framework will be easier to make the same decision. Therefore, for each researcher r_i, the profile P_i is designed to include the researcher's title, affiliation, years of scientific experience, h-index, and specialty fields, which can be expressed as follows.

$$P_i = (tit_i, aff_i, yse_i, hidx_i, sf_i). \tag{2}$$

Specifically, the specialty fields $\text{sf}_i = \{(wd_{il}, w_{wd_{il}})\}$, $l \in \{1, 2, \ldots, L$ can be represented as a L length set of keywords, which indicates the latest research topics that the researcher r_i focuses on. $w_{wd_{il}}$ denotes the weight of keyword wd_{il}, which can be calculated based on the TF-IDF method. Following this way, the profile generated based on the above formulation can be used to describe one researcher's latest academic status and further support the scholarly network partition process.

Another important issue is to measure a researcher's academic performance and research outcome within his/her research topics, which will be helpful to reveal his/her special role and status (e.g., an expert or not) in group decision-making problem. Typically, we define $A_i = \{a_{ik}\}, k \in \{1, 2, \ldots, K\}$ as a set of articles published by researcher r_i to measure his/her research outcome. In particular, each article is formalized as $a_{ik} = (AID_{ik}, F_{ik}, IF_{ik})$, in which AID_{ik} denotes the article DOI code to identify a unique article, F_{ik} denotes a set of keywords to represent research topics extracted from this article's title and abstract, and IF_{ik} is the value of the current impact factor of this article according to the journal it published in.

Accordingly, both P_i and A_i of a specific researcher r_i will be utilized in the decision-maker's role measurement and scholarly network partition, in order to identify and analyze his/her role and assigned into a sub-network when he/she is considered as a decision-maker in an LSGDM problem.

C. Analysis of Decision-Makers' Relationships

When handling the LSGDM problem, it is a key step to divide the whole large-scale scholarly network into a series of sub-networks based on the analysis of relationships among decision-makers. Generally, the diversified collaboration relationships among researchers are essential to identify their similar interests, opinions, and how the decision-makers can be influenced by each other during the consensus process across online academic social networks. In this article, the co-author collaboration relationships among researchers and citation relationships among articles are taken into account among decision-makers to deal with the correlation analysis and expert identification for the LSGDM problem.

Given a pair of researchers r_i and r_j, we define $A_i \cap A_j$ as the intersection of A_i and A_j, which indicates the co-author articles published by r_i and r_j. In particular, each article in the intersection is formalized as $a_x = (AID_x, F_x, IF_x)$, $x \in \{0, 1, \ldots, |A_i \cap A_j|\}$. Similarly, we define $A_i \cup A_j$ as the union of A_i and A_j, which indicates all the articles published by r_i and r_j. In particular, each article in the union is formalized as $a_y = (AID_y, F_y, IF_y)$, $y \in \{1, 2, \ldots, |A_i \cup A_j|\}$. Additionally, considering the different impact factors of each article, the correlation between r_i and r_j can be expressed as follows:

$$w_{ij} = \frac{\sum_{a_x \in A_i \cap A_j} IF_x}{\sum_{a_y \in A_i \cup A_j} IF_y}. \tag{3}$$

IV. Two-Stage Large-Scale Decision-Making Solution for Social Recommendation in Scholarly Network

In this section, following the preliminaries of the introduced basic model, a two-stage large-scale decision-making solution is proposed for social recommendations within scholarly networks. A network partition algorithm based on expert identification in the first stage and an RWR-based decision weighting algorithm in the second stage are developed for the LSGDM problem.

A. Expert-Based Network Partition

The scale of a group decision-making problem is the most critical factor that will directly affect the efficiency of the consensus process. To deal with the complexity of LSGDM problems, the key technique is reducing the dimension of the scale of decision-makers during the decision-making process. Especially, influenced by some expert researchers, people aggregated in some smaller groups will reach the consensus more efficiently, which represents the group decision among a crowd of researchers. Therefore, the whole large-scale scholarly network can be partitioned into a series of sub-networks based on expert identification and relationship analysis among researchers.

1) Expert Identification Among Researchers: Similar to online social networks (e.g., Facebook and Twitter), scientific researchers usually share and acquire academic opinions and research outcomes across academic social networks (e.g., ResearchGate) as well. Typically, the outstanding publications and opinions from some experts in the specific field are more likely to inspire and influence the other researchers. These researchers who are influenced by the same experts may share the similar research interests and are more likely to achieve the compatible decision in LSGDM problems. Therefore, it is essential to identify those experts with their influenced groups of other researchers, which may resolve the LSGDM problem effectively and further promote their collaboration research works.

In general, an expert refers to a high-reputation researcher who owns outstanding expertise and far-reaching influence in his/her research field. The expertise of the researcher r_i based on his/her basic profile can be expressed as follows;

$$Pro_i = hidx_i + tit_i * \log(yse_i). \tag{4}$$

In addition, the researchers' published articles, including their co-author and citation relationships, are involved to measure the influence for an expert extending to a part of scholarly networks, which can be formalized as follows:

$$Inf_i = \sum_{a_{ik} \in A_i} IF_{ik} * (Coa_{ik} + \log(Cit_{ik})) \tag{5}$$

where Coa_{ik} and Cit_{ik} denote the number of co-authors and citations of article a_{ik} from researcher r_i, respectively.

Considering that more citations and higher reputation of a researcher may lead to more extensive influence and better collaborations, both the expertise based on the academic profile and influence hidden within scholarly networks are taken into account to identify if a given researcher can be viewed as an expert. Summarily, the identification of an expert among a series of researchers can be quantified as follows:

$$Expt_i = \alpha * Pro_i + (1 - \alpha) * Inf_i \tag{6}$$

where α is empirical coefficient to weight the importance of expertise and influence when identifying the expert.

Algorithm 1. Expert-based network partition

Input: The constructed network model $G(R, E, W)$;
The correlation threshold δ_{cor} .
Output: The expect-based researcher set = $\{N_1, \ldots, N_m\}$.

Initialize set N_e, and set $n = |R|$;
for i =1 to n
Calculate the expertise $expt_i$ by Eq. (4);
  for $j = i$ to n
    Calculate the correlation weight w_{ij} by Eq. (7);
  end for
calculate af_i by Eq. (5);
end for
Sort R to R' according to both $expt_i$ and af_i by Eq. (6) in descending order;
while $R' \neq \emptyset$ **do**
  Create a new N ;
  Fetch the top researcher r_i from R' ;
  $n = |R'|$;
  for k =1 to n
    if $w_{ik} \geq \delta_{cor}$
      $N = N \cup \{r_k\}$;
    end if
  end for
  $N = N \cup \{r_i\}$; $R' = R' - N$; $N_e = N_e \cup \{N\}$;
end while
Return $N_e = \{N_1, N_2, \ldots, N_m\}$

Fig. 1. Algorithm for large-scale scholarly network partition.

2) Network Partition Based on Expert Identification: It is noted that researchers are more likely to be influenced and inspired by experts with similar interests and opinions, which means that these researchers may achieve the consensus easily if they can be identified and grouped together efficiently. Therefore, comparing with traditional clustering methods, we take the identified expert as the initial center, then find and cluster the other related researchers who may be influenced by him/her, and finally generate the corresponding sub-networks. This clustering process is efficient and significant in scholarly network partition, especially when dealing with LSGDM problems. The expert-based network partition algorithm is shown in Fig. 1.

As described in Algorithm 1 in Fig. 1, we first identify experts among all the researchers according to the expertise from their academic profiles and influence across scholarly networks. Furthermore, we calculate correlations between each researcher and the identified expert and then group them into the corresponding cluster if the measured correlation is greater than the threshold δ_{cor}. Following this way, the whole scholarly network can be divided into a series of sub-networks $N_1, N_2, \ldots, N_m$, each of which will include the identified expert in the center together with a group of influenced researchers.

B. Decision Weighting Based on Random Walk With Restart

The RWR technique has been proven to be an important method to measure the relevance between two nodes based on the structure-aware feature in a weighted network model.

Algorithm 2. RWR-based decision weighting within a sub-network

Input: A sub-network divided from the constructed network model $G\ (R,\ E,\ W)$ with the researcher set N_p ;
A given target node r_{tgt} ;
The iteration threshold δ_{ite}.
Output: List $\overrightarrow{L_p}$ of weighted decisions for researchers in N_p.

```
Initialize q according to the given target node r_tgt;
Initialize HR^(0) = q ;
Initialize NumOfIteration ;
Initialize threshold δ_min for iteration break;
diff = 0 ;
for i = 1 to NumOfIteration
    HR^(t+1) = λM * HR^(t) + (1 − λ)q ;
    diff = HR^(t+1) − HR^(t) ;
    if diff < δ_ite;
        break ;
    end if
end for
Sort all the nodes based on their ranking scores;
Return the weighted decisions according to ranking scores as
    list L_p = < s^p_1, s^p_2, ..., s^p_|N_p| > for the target node r_tgt
```

Fig. 2. Algorithm for weighted decision computation.

To solve the LSGDM problem, the RWR-based method is developed to figure out researchers' weighted decisions within each sub-network during the decision-making process.

More precisely, to calculate each researcher's weighted decision based on their correlations in a given sub-network, the basic RWR model can be expressed as follows:

$$HR^{(t+1)} = \lambda M * HR^{(t)} + (1-\lambda)q \tag{7}$$

where HR^t denotes a vector of weighted decisions at the iteration step t. M is a transfer matrix describing the probability of each vertex to transfer to the others in a structured scholarly network graph G (R, E, W). λ, ranging from 0 to 1, is a damping coefficient. q is the initial vector when starting the RWR model, $HR^0 = q$.

Specifically, q is initialized as [0, 0, …, 1, …, 0, 0], in which "1" denotes the target vertex r_i at the beginning. M is initialized based on w_{ij} appended on each edge e_{ij}, which is quantified according to the existing co-author collaboration relationships among researchers within the divided sub-network. The RWR-based algorithm to calculate the weighted decision for each researcher within a given sub-network is shown in Fig. 2.

Based on the algorithm shown in Fig. 2, a given researcher is considered as the target node when generating the weighted decision vector. The iteration will keep running untill the difference between $HR^{(t+1)}$ and $HR^{(t)}$ is less than threshold δ_{ite} (here, we set the threshold as 10^{-7}), which means that the results are converged and the weighted decisions become stable finally. The sorted list $\vec{L}_p$, including the ranking scores of the chosen researchers (i.e., alternatives), will be recognized as their weighted decision information for group N_p and contribute to the group decision aggregation process.

TABLE I
SUMMARY OF THE DATA SET

	Statistics
Researchers	13,100
Articles	379,456
Citations	2,333,636

C. Group Decision Aggregation and Alternative Ranking

To achieve the final decision solution for the LSGDM problem, the last step is to aggregate the decision information from each sub-network, including the group weighting based on the divided sub-networks and alternatives ranking according to the aggregated decision information. For each sub-network N_p in set N derived from Algorithm 1 in Fig. 1, the weight ω_{N_p} for the divided group in each sub-network can be measured based on the number of DMs (i.e., researchers) included in this group, which is expressed as follows:

$$\omega_{N_p} = \frac{Num_{N_p}}{\sum_{N_a \in N} Num_{N_a}}. \tag{8}$$

Given a target researcher r_{tgt} with the corresponding list $\vec{L}_p$ for each N_p, the alternative list from m groups can be represented as $L_{alt} = \{S_q^p | q \in \{1, 2, \ldots, |N_p|\}, p \in \{1, 2, \ldots, m\}\}$. The calculation for the group decision information aggregation can be quantified as follows:

$$D_{S_q^p} = \sum_{p=1}^{m} \sum_{q=1}^{|N_p|} \omega_{N_p} * S_q^p \tag{9}$$

where S_q^p denotes the group decision information of r_q in the divided researcher group in sub-network N_p.

Finally, the sorted alternatives with their corresponding decision information will be recommended to the target researcher r_i as the optimal alternatives for the LSGDM problem solutions.

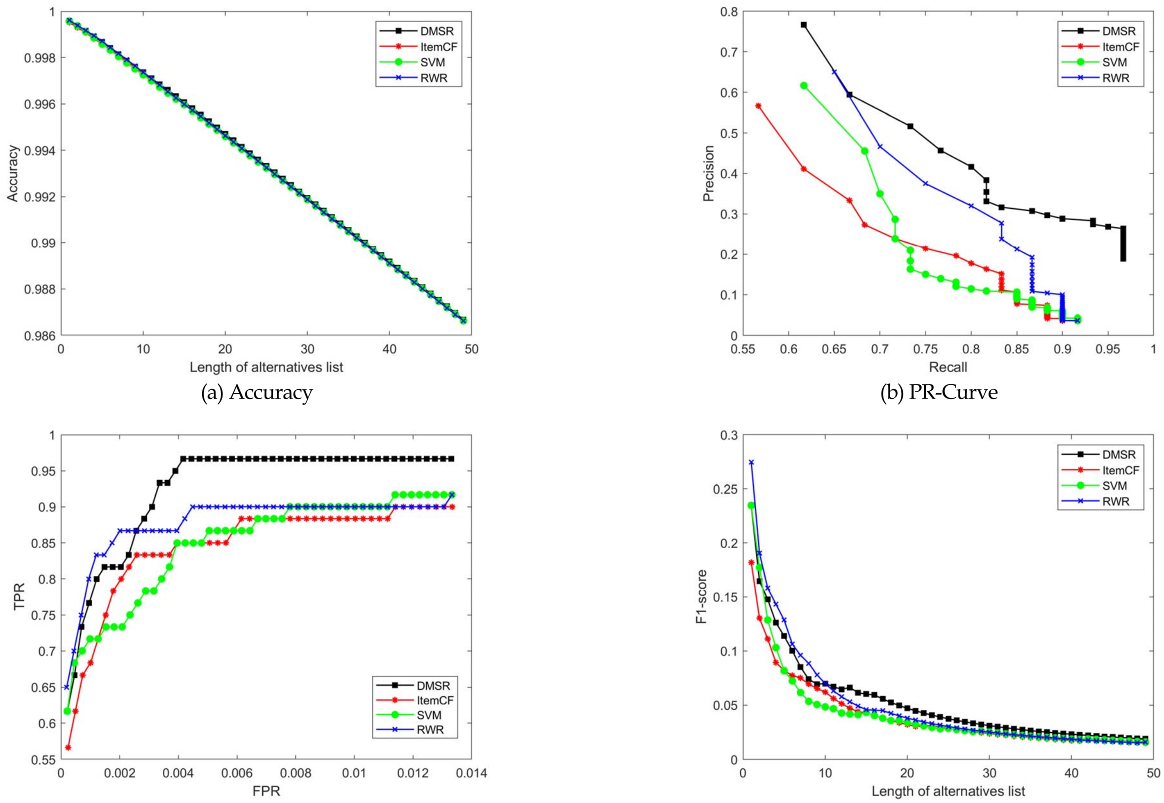
V. EXPERIMENT AND ANALYSIS

In this section, to show the practicability of the proposed solution for LSGDM problems, evaluation experiments are implemented with real-world scholarly data. Comparisons with other related methods are conducted to demonstrate the effectiveness of our proposed method.

A. Data Set

The scholarly data crawled from the online academic social network "ResearchGate," which is a global online academic service that contains scientists associated with their academic activities, are employed to conduct evaluation experiments. Specifically, two kinds of data provided by ResearchGate, namely, the researcher profile data, which includes one researcher's title, affiliation, years of scientific experience, h-index, and specialty fields, and the published article data,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) Accuracy (b) PR-Curve (c) ROC-Curve (d) F1

Fig. 3. Comparison results among four methods based on different metrics. (a) Accuracy. (b) PR-Curve. (c) ROC-Curve. (d) F1.

which indicates researchers' co-author collaboration and citation relationships, are applied to evaluate the performance of our proposed solution in scholarly LSGDM problems.

Finally, we collected more than 350 000 articles from 13 000 researchers. The detailed summary of the data set is shown in Table I. Considering the scale of decision-makers that we discussed earlier, we evaluated our method with 10 000 researchers as decision-makers to simulate the LSGDM problem in big data environments.

B. Experiment Setup

The data set was further divided into two subsets: researchers and their 80% articles were selected to construct the scholarly network model, while the remaining data were used as the testing set. To demonstrate the effectiveness of the proposed method, six usually used evaluation metrics, namely, precision, recall, F1, true positive rate (TRP), false positive rate (FPR), and accuracy, are employed to conduct the evaluation and comparison, which contains the following elements.

1) *NumTP*: The alternatives having collaborations with the target node and selected.
2) *NumFP*: The alternatives having no collaboration with the target article but selected.
3) *NumFN*: The alternatives having collaborations with the target article but not selected.
4) *NumTN*: The alternatives having no collaboration with the target article and not selected.

Based on the above-mentioned definitions, accuracy, precision, recall, F1, TRP, and FPR metrics can be calculated as follows:

$$Accuracy = \frac{NumTP + NumTN}{NumTP + NumTN + NumFP + NumFN} \tag{10}$$

$$Precision = \frac{NumTP}{NumTP + NumFP} \tag{11}$$

$$Recall = \frac{NumTP}{NumTP + NumFN} \tag{12}$$

$$F1 = \frac{2Precision{*}Recall}{Precision + Recall} \tag{13}$$

$$TPR = \frac{NumTP}{NumTP + NumFN} \tag{14}$$

$$FPR = \frac{NumFP}{NumFP + NumTN}. \tag{15}$$

The following three methods, which are usually used for social recommendations in big data environments, are chosen for comparisons.

1) *The Basic RWR Recommendation [37]:* This is the baseline method that provides recommendation results by running a basic RWR algorithm in a conducted network model.
2) *ItemCF-Based Recommendation [38]:* This is an item-based collaborative filtering recommendation method that is widely used in social recommendations especially when dealing with information explosion issues.

3) *SVM-Based Recommendation [39]:* This is a support vector machine-based recommendation method that uses a supervised learning model to improve the recommendation result based on the classification analysis.

As for the settings of equilibrium coefficients used for calculations in our method, α was set as 0.5 in (6), which indicates the same importance of expertise and influence when identifying the expert. In addition, the default number of iterations in Algorithm 2 in Fig. 2 was set as 100, and the damping coefficient λ was set to 0.8 to determine the probability of RWR to return back to the target node at the beginning.

C. Performance Evaluation

To evaluate the effectiveness of the proposed method when handling the large-scale decision-making problem, we conducted experiments under a recommendation scenario that letting 10 000 researchers select a maximum of 50 alternatives to support their academic collaborations.

The performance of the proposed method was evaluated based on the academic recommendation scenario and compared with the three methods according to the six metrics introduced earlier, which are shown in Fig. 3(a)–(d), respectively. It is noted that the receiver operating characteristic (ROC) curve shown in Fig. 3(c) is created by plotting the TPR against the FPR at various threshold settings. We give our observations and discussions based on these evaluation results as follows.

1) Fig. 3(a) shows that all the four methods obey nearly the same accuracy curves that decrease as the length of alternatives increases. The overlapping curves indicate that the four methods are capable of providing valid decision information when there are not so many candidate alternatives.
2) Fig. 3(b) demonstrates the overall performances of the four methods according to the precision–recall (PR)-curve. Obviously, the proposed method and the RWR method perform better than the basic ItemCF and SVM method according to the area under curve. Our method achieves the best result according to the PR-Curve, which illustrates the effectiveness of our method comparing with the others. This result indicates the importance to calculate the weighted decision based on the analysis of different researchers' collaboration relationships within a constructed network model.
3) Fig. 3(c) demonstrates a general upward trend based on the ROC-Curve. Basically, all the four curves stay above the diagonal line from the left bottom to the top right corner, which represents a relatively good classification result comparing with the random selection of alternatives. Considering that the best possible recommendation result would be in the upper left corner, especially the coordinate (0, 1) of the ROC space, which means that the curve closer to the upper left corner demonstrates the better performance, the proposed method outperforms the other three methods because its curve is the closest one to the upper left corner and away from the diagonal line.

TABLE II

Top-10 Alternative Selections Based on Different Methods

Reference node	TSDM	RWR	ItemCF	SVM
no. 26	no. 26	no. 26	no. 26	no. 26
no. 05	no. 86	no. 48	no. 47	no. 61
no. 47	no. 47	no. 07	no. 19	no. 05
no. 86	no. 05	no. 16	no. 52	no. 52
no. 34	no. 32	no. 34	no. 86	no. 86
no. 32	no. 18	no. 32	no. 32	no. 32
no. 18	no. 30	no. 18	no. 30	no. 47
no. 75	no. 34	no. 05	no. 05	no. 19
no. 30	no. 28	no. 52	no. 75	no. 31
no. 52	no. 75	no. 30	no. 18	no. 65

4) According to the results of the F1 metric shown in Fig. 5(d), the proposed method outperforms the other three methods and reaches a peak of 0.34. This result demonstrates the necessity to reduce the dimension of an LSGDM problem based on a comprehensive consideration of decision-makers' academic expertise and reciprocal influence in scholarly big data environments.

To evaluate the effectiveness in selecting the valuable collaborative alternatives, we compared our proposed method with the other three methods according to the ranking of the top-10 alternative selections. The experiment was conducted under the scenario that letting 500 researchers select 50 alternatives. The Delphi method [40] was employed as the traditional group decision-making to conduct the baseline result for reference.

As the results shown in Table II, the basic RWR method is the least effective one among all the decision-making solutions. This can be explained, as it simply sets the same weight of each element in the transfer matrix, thus cannot adapt to the complex situation in LSGDM problems. On the other hand, our proposed method outperforms the other three methods in terms of the effectiveness of alternative selection. This is because our method takes multiple factors into account across scholarly networks, which can efficiently contribute to the sub-network partition based on the expert identification and decision weighting for the group decision information aggregation in scholarly big data environments.

VI. Conclusion

In this article, we presented a computational method to integrate LSGDM into social recommendations. A two-stage large-scale decision-making solution was proposed to provide users with more reliable recommendations, which could enhance the cyber-enabled online service in cyber–social computing environments.

First, we designed and constructed a graph model to describe the LSGDM problem in scholarly big data environments. Specifically, based on the evaluation of one researcher's academic performance and research outcome, a series of formal descriptions were introduced to build the profile of each decision-maker within a constructed scholarly network model. Measures were defined to quantify the correlation between two associated researchers according to their collaboration relationships.

Second, an expert-based network partition algorithm was developed, which could be used to deal with a considerable number of decision-makers in big data environments, comparing traditional group decision-making solutions. More precisely, both the expertise based on academic profiles and influence hidden within scholarly networks were taken into account in an integrative strategy for expert identifications. The whole scholarly network was then divided into several sub-networks based on the extending of influence from the identified expert to a group of researchers using an improved clustering algorithm.

Third, an RWR-based algorithm was improved for the weighted decision calculation in each divided sub-network. In particular, the initial node in the algorithm was recognized as the target researcher, and a vector of the weighted decision was generated for the further group decision aggregation and alternative ranking processes. The sorted results could be provided to the target researcher as the optimal alternatives to achieve the final LSGDM solution.

Finally, our designed two-stage DMSR solution was applied in the scholarly big data environment. Experiments and evaluations were conducted based on the real-world data crawled from "ResearchGate," which demonstrated the effectiveness and usefulness of our proposed model and method in supporting researchers' academic collaborations with more reliable social recommendations.

In future studies, we will focus more on the dynamics and heterogeneity in big data environments. Algorithms will be improved to deal with more complex situations for more flexible cyber-enabled services.

REFERENCES

[1] J. Tang, X. Hu, and H. Liu, "Social recommendation: A review," *Social Netw. Anal. Mining*, vol. 3, no. 4, pp. 1113–1133, 2013.

[2] T. Wu, X. W. Liu, and F. Liu, "An interval type-2 fuzzy TOPSIS model for large scale group decision making problems with social network information," *Inform. Sci.*, vol. 432, pp. 392–410, Mar. 2018.

[3] H. Saggion and F. Ronzano, "Scholarly data mining: Making sense of scientific literature," *Proc. ACM/IEEE Joint Conf. Digit. Libraries (JCDL)*, Toronto, ON, Canada, Jun. 2017, pp. 1–2.

[4] H. Liu, X. Kong, X. Bai, W. Wang, T. M. Bekele, and F. Xia, "Context-based collaborative filtering for citation recommendation," *IEEE Access*, vol. 3, pp. 1695–1703, Oct. 2015.

[5] F. Osborne and E. Motta, "Rexplore: Unveiling the dynamics of scholarly data," in *Proc. IEEE/ACM Joint Conf. Digit. Libraries*, London, U.K., Sep. 2014, pp. 415–416.

[6] X. H. Chen and R. Liu, "Improved clustering algorithm and its application in complex huge group decision-making," *Syst. Eng. Electron.*, vol. 28, no. 11, pp. 1695–1699, 2006.

[7] I. Palomares, L. Martínez, and F. Herrera, "A consensus model to detect and manage noncooperative behaviors in large-scale group decision making," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 3, pp. 516–530, Jun. 2014.

[8] B. Liu, Y. Shen, Y. Chen, X. Chen, and Y. Wang, "A two-layer weight determination method for complex multi-attribute large-group decision-making experts in a linguistic environment," *Inf. Fusion*, vol. 23, pp. 156–165, May 2015.

[9] T. Wu, K. Zhang, X. Liu, and C. Cao, "A two-stage social trust network partition model for large-scale group decision-making problems," *Knowl.-Based Syst.*, vol. 163, pp. 632–643, Jan. 2019.

[10] R. Vatrapu, R. R. Mukkamala, A. Hussain, and B. Flesch, "Social set analysis: A set theoretical approach to big data analytics," *IEEE Access*, vol. 4, pp. 2542–2571, 2016.

[11] H. Zhang, R. Dantu, and J. W. Cangussu, "Socioscope: Human relationship and behavior analysis in social networks," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 6, pp. 1122–1143, Nov. 2011.

[12] Y. Song, Z. Hu, X. Leng, H. Tian, K. Yang, and X. Ke, "Friendship influence on mobile behavior of location based social network users," *J. Commun. Netw.*, vol. 17, no. 2, pp. 126–132, Apr. 2015.

[13] H. Zhao, H. Zhou, C. Yuan, Y. Huang, and J. Chen, "Social discovery: Exploring the correlation among three-dimensional social relationships," *IEEE Trans. Comput. Social Syst.*, vol. 2, no. 3, pp. 77–87, Sep. 2015.

[14] L. Zhang, T. Wang, Z. Jin, N. Su, C. Zhao, and Y. He, "The research on social networks public opinion propagation influence models and its controllability," *China Commun.*, vol. 15, no. 7, pp. 98–110, Jul. 2018.

[15] G. Wang, W. Jiang, J. Wu, and Z. Xiong, "Fine-grained feature-based social influence evaluation in online social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 9, pp. 2286–2296, Sep. 2014.

[16] J. Wei, G. Mengdi, W. Xiaoxi, and W. Xianda, "A new evaluation algorithm for the influence of user in social network," *China Commun.*, vol. 13, no. 2, pp. 200–206, Feb. 2016.

[17] Y. Zhu, D. Li, R. Yan, W. Wu, and Y. Bi, "Maximizing the influence and profit in social networks," *IEEE Trans. Computat. Social Syst.*, vol. 4, no. 3, pp. 54–64, Sep. 2017.

[18] W. Pan, W. Dong, M. Cebrian, T. Kim, J. H. Fowler, and A. S. Pentland, "Modeling dynamical influence in human interaction: Using data to make better inferences about influence within social systems," *IEEE Signal Process. Mag.*, vol. 29, no. 2, pp. 77–86, Mar. 2012.

[19] C. Li and F. Xiong, "Social recommendation with multiple influence from direct user interactions," *IEEE Access*, vol. 5, pp. 16288–16296, 2017.

[20] R. Ureña, G. Kou, Y. Dong, F. Chiclana, and E. Herrera-Viedma, "A review on trust propagation and opinion dynamics in social networks and group decision making frameworks," *Inf. Sci.*, vol. 478, pp. 461–475, Apr. 2019.

[21] J. Wu, R. Xiong, and F. Chiclana, "Uninorm trust propagation and aggregation methods for group decision making in social network with four tuple information," *Knowl.-Based Syst.*, vol. 96, pp. 29–39, Mar. 2016.

[22] Y. Liu, C. Liang, F. Chiclana, and J. Wu, "A trust induced recommendation mechanism for reaching consensus in group decision making," *Knowl.-Based Syst.*, vol. 119, pp. 221–231, Mar. 2017.

[23] J. Wu, F. Chiclana, H. Fujita, and E. Herrera-Viedma, "A visual interaction consensus model for social network group decision making with trust propagation," *Knowl. Based Syst.*, vol. 122, pp. 39–50, Apr. 2017.

[24] Y. C. Dong, Q. Zha, H. Zhang, G. Kou, H. Fujita, F. Chiclana, and E. Herrera-Viedma, "Consensus reaching in social network group decision making: Research paradigms and challenges," *Knowl. Based Syst.*, vol. 162, pp. 3–13, Dec. 2018.

[25] J. Wu, L. Dai, F. Chiclana, H. Fujita, and E. Herrera-Viedma, "A minimum adjustment cost feedback mechanism based consensus model for group decision making under social network with distributed linguistic trust," *Inf. Fusion*, vol. 41, pp. 232–242, May 2018.

[26] J. Wu, F. Chiclana, and E. Herrera-Viedma, "Trust based consensus model for social network in an incomplete linguistic information context," *Appl. Soft Comput.*, vol. 35, pp. 827–839, Oct. 2015.

[27] R. Gao *et al.*, "Exploiting geo-social correlations to improve pairwise ranking for point-of-interest recommendation," *China Commun.*, vol. 15, no. 7, pp. 180–201, Jul. 2018.

[28] C.-H. Lai, S.-Y. Lee, and H.-L. Huang, "A social recommendation method based on the integration of social relationship and product popularity," *Int. J. Hum.-Comput. Stud.*, vol. 121, pp. 42–57, Jan. 2019.

[29] S. Deng *et al.*, "On deep learning for trust-aware recommendations in social networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 5, pp. 1164–1177, May 2017.

[30] L. Cui, D. Pi, and J. Zhang, "DMFA-SR: Deeper membership and friendship awareness for social recommendation," *IEEE Access*, vol. 5, pp. 8904–8915, 2017.

[31] Z. Zhao, H. Lu, D. Cai, X. He, and Y. Zhuang, "User preference learning for online social recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 9, pp. 2522–2534, Sep. 2016.

[32] M. Jiang *et al.*, "Social recommendation with cross-domain transferable knowledge," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 11, pp. 3084–3097, Nov. 2015.

[33] T. Stepan, J. M. Morawski, S. Dick, and J. Miller, "Incorporating spatial, temporal, and social context in recommendations for location-based social networks," *IEEE Trans. Comput. Social Syst.*, vol. 3, no. 4, pp. 164–175, Dec. 2016.

[34] X. Yang *et al.*, "Collaborative filtering-based recommendation of online social voting," *IEEE Trans. Computat. Social Syst.*, vol. 4, no. 1, pp. 1–13, Mar. 2017.

[35] Z. Wang, L. Sun, W. Zhu, S. Yang, H. Li, and D. Wu, "Joint social and content recommendation for user-generated videos in online social network," *IEEE Trans. Multimedia*, vol. 15, no. 3, pp. 698–709, Apr. 2013.

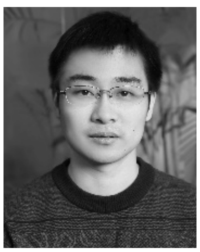
[36] L. Sun, X. Wang, Z. Wang, H. Zhao, and W. Zhu, "Social-aware video recommendation for online social groups," *IEEE Trans. Multimedia*, vol. 19, no. 3, pp. 609–618, Mar. 2017.

[37] H. Park, J. Jung, and U. Kang, "A comparative study of matrix factorization and random walk with restart in recommender systems," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2018, pp. 756–765.

[38] A. Salah and M. Nadif, "Social regularized von Mises–Fisher mixture model for item recommendation," *Data Mining Knowl. Discovery*, vol. 31, no. 2, pp. 1218–1241, 2017.

[39] F. Ali *et al.*, "Merged ontology and SVM-based information extraction and recommendation system for social robots," *IEEE Access*, vol. 5, pp. 12364–12379, 2017.

[40] B. Vahdani, S. M. Mousavi, R. Tavakkoli-Moghaddam, and H. Hashemi, "A new design of the elimination and choice translating reality method for multi-criteria group decision-making in an intuitionistic fuzzy environment," *Appl. Math. Model.*, vol. 37, no. 4, pp. 1781–1799, 2013.



Xiaokang Zhou (M'12) received the Ph.D. degree in human sciences from Waseda University, Tokyo, Japan, in 2014.

From 2012 to 2015, he was a Research Associate with the Department of Human Informatics and Cognitive Sciences, Faculty of Human Sciences, Waseda University. Since 2016, he has been a Lecturer with the Faculty of Data Science, Shiga University, Hikone, Japan. He has been a Visiting Researcher with the RIKEN Center for Advanced Intelligence Project (AIP), RIKEN, Tokyo, since 2017. He has been engaged in interdisciplinary research works in the fields of computer science and engineering, information systems, and social and human informatics. His current research interests include ubiquitous and social computing, big data mining and analytics, machine learning, behavior and cognitive informatics, cyber–physical–social systems, cyber intelligence, and cyber-enabled applications.

Dr. Zhou is also a member of the IEEE Computer Society, the Association for Computing Machinery, USA, the Information Processing Society of Japan, Japan, and The Japanese Society for Artificial Intelligence, Japan.

Wei Liang (M'19) received the M.S. and Ph.D. degrees in computer science from Central South University, Changsha, China, in 2005 and 2016, respectively.

He is currently with the Key Laboratory of Hunan Province for Mobile Business Intelligence, Hunan University of Technology and Business, Changsha. He has published more than ten papers at various conferences and journals, including *Future Generation Computing Systems*, the *Journal of Computer and System Sciences*, and *Personal and Ubiquitous Computing*. His current research interests include information retrieval, data mining, and artificial intelligence.

Suzhen Huang received the B.S. and M.S. degrees in engineering from Central South University, Changsha, China, in 2012 and 2015, respectively.

She is currently with the Information and Network Center, Central South University. Her current research interests include cloud computing and recommendation algorithms.

Miao Fu is currently a Senior Student major in information and computing science with the Hunan University of Technology and Business, Changsha, China. She also studied statistics and linear algebra at Brown University, Providence, RI, USA, in 2017. Her current research interests include recommendation algorithms and analysis of social or bioinformatics networks.

Dr. Fu is one of the National First Prize winners of Mathematical Contest in Modeling in 2017.