IEEE TRANSACTIONS ON
EMERGING TOPICS
IN COMPUTING

Received 1 December 2017; revised 29 May 2018; accepted 9 July 2018.
Date of publication 26 July 2018; date of current version 3 March 2021.

Digital Object Identifier 10.1109/TETC.2018.2860051

Academic Influence Aware and Multidimensional Network Analysis for Research Collaboration Navigation Based on Scholarly Big Data

XIAOKANG ZHOU, (Member, IEEE), WEI LIANG, KEVIN I-KAI WANG, (Member, IEEE), RUNHE HUANG, (Senior Member, IEEE), AND QUN JIN, (Senior Member, IEEE)

X. Zhou is with the Faculty of Data Science, Shiga University, Hikone 522-8522, Japan and RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan
W. Liang is with the Key Laboratory of Hunan Province for Mobile Business Intelligence, Hunan University of Commerce, Changsha Hunan Sheng 410013, China
K. I.-K. Wang is with the Department of Electrical and Computer Engineering, The University of Auckland, Auckland, New Zealand
R. Huang is with the Faculty of Computer and Information Sciences, Hosei University, Tokyo 102-8160, Japan
Q. Jin is with the Faculty of Human Sciences, Waseda University, Tokorozawa, Japan
CORRESPONDING AUTHOR: Q. JIN (jin@waseda.jp)

ABSTRACT Scholarly big data, which is a large-scale collection of academic information, technical data, and collaboration relationships, has attracted increasing attentions, ranging from industries to academic communities. The widespread adoption of social computing paradigm has made it easier for researchers to join collaborative research activities and share academic data more extensively than ever before across the highly interlaced academic networks. In this study, we focus on the academic influence aware and multidimensional network analysis based on the integration of multi-source scholarly big data. Following three basic relations: Researcher-Researcher, Researcher-Article, and Article-Article, a set of measures is introduced and defined to quantify correlations in terms of activity-based collaboration relationship, specialty-aware connection, and topic-aware citation fitness among a series of academic entities (e.g., researchers and articles) within a constructed multidimensional network model. An improved Random Walk with Restart (RWR) based algorithm is developed, in which the time-varying academic influence is newly defined and measured in a certain social context, to provide researchers with research collaboration navigation for their future works. Experiments and evaluations are conducted to demonstrate the practicability and usefulness of our proposed method in scholarly big data analysis using DBLP and ResearchGate data.

INDEX TERMS Multidimensional network analysis, academic influence, scholarly big data, scholarly recommendation, research collaboration

I. INTRODUCTION

There is no doubt that the advancement of information and communication technologies has brought us into a new era of big data. We have been involved into the heterogeneous data environment and are dealing with various kinds of data everyday. Specifically, one kind of data published in various digital databases and shared on academic websites, has become increasingly popular, especially in academic and education institutions [1], which is called the scholarly big data. Scholarly big data is defined as the vast quantity of data related to scholarly undertaking and is produced from different scholarly sources with various formats, including journal articles, conference proceedings, theses, books, patents, presentation slides and experimental data [2]–[4]. Generally, it contains large scale of academic information (such as authors, papers, and citations), technical data (such as algorithms, figures, and tables), and collaboration relations across scholarly networks and digital libraries [5]. Recent investigations indicate that there are nearly 114 million recorded English-language scholarly documents accessible on the Web [6], and it is reported that a new scholarly document is published every 20 seconds [7], with a growth rate at tens of thousands per day [8]. This noticeable growth in electronic publishing with large-scale scholarly collections is attracting diversified research interests within broad research disciplines, aiming at making sense of scholarly data in a series of attractive aspects, such as research trend prediction, topic discovery, and expert finding, etc.

Meanwhile, with the rapid development of emerging computing paradigms (such as social computing), people are accustomed to conducting social activities with others online to share their learning and working experience. The high accessibility of Social Networking Service (SNS) has led to a new way to collect and share data from local environments and discover patterns in user associations over time [9]. Specifically, academic social networks (such as ResearchGate) have become more and more popular not only in academic communities, but also in industries due to the widespread adoption of Web 2.0 technology. For instance, ResearchGate integrates functions of publication dissemination and academic communication together, which effectively facilitates scientific information sharing and academic circle enlargement among researchers in social environments [10]. Generally, classic academic networks can be categorized as: co-word networks [11], co-author networks [12], citation networks [13], co-citation networks [14], and so on. In particular, these networks integrated together can provide us an opportunity to better utilize the heterogeneous datasets in socialized scientific systems based on network theories.

Recently, many research works have paid more attentions to advanced technologies on mining and analysis of multi-type relations among scholarly big data, resulting in numerous achievements in scholarly information extraction and integration [2], [15], [16], article recommendation and rating prediction [17]–[19], citation analysis and recommendation [20], [21], and research collaboration prediction and recommendation [22]–[24] etc. However, there are still several limitations when conducting heterogeneous network analysis, as follows. 1) Although lots of algorithms have been developed to measure the network-aware importance for academic recommendations, most recommendation systems only focus on a single kind of relation (e.g., co-author relationships or citation relationships) when constructing network models. A few works [25], [26] have considered two or three kinds of relations together, but the results can only be viewed as some hierarchical networks or multi-layer graphs. 2) Very often, when constructing a weighted graph to describe original relationships among authors or articles, for instance, the citation relationships, weights assigned to each edge are always treated as either 0 or 1. This is not sufficient to dynamically represent their real correlations. 3) Although most of the real-world scholarly networks can be viewed as heterogeneous and multi-typed objects, such as authors, papers, venues, and year of publication, have been utilized for network analyses, the datasets for experiments are mostly collected and extracted from one single data resourse (e.g., DBLP). The analysis of heterogeneity hidden in multiple scholarly networks has not been well conducted.

In this study, we focus on modeling and analysis of the multidimensional academic network, in which the time-varying academic influence is measured according to academic activities across social networks, in order to facilitate the research collaboration navigation in scholarly big data environments. Comparing with other related works, improvements are made with respect to the following aspects: 1) Researchers and articles are associated together to simultaneously represent their multi-type relations in a multidimensional network model. 2) In the constructed graph model, the weight assigned to each edge is quantified ranging from 0 to 1, which can more accurately describe their dynamic correlations with contextual academic information. 3) Both the publication data and academic activity data collected from different sources (i.e., DBLP and ResearchGate), are integrated together to conduct the evaluation experiment.

In particular, an academic influence aware and multidimensional network (AIMN) analysis method is proposed, in which three basic relations, namely Researcher-Researcher, Researcher-Article, and Article-Article relations, are considered together to construct the multidimensional network model from multi-source scholarly data. A set of measures is defined to quantitatively describe and analyze the multi-type correlatioins between different academic entities (e.g., researchers and articles) in the context of academic social networks. A recommendation algorithm is then developed based on the analysis of time-varying academic influence, to support researchers' collaboration works. Specifically, the major contributions of this paper are summarized as follows.

- A computational method to construct the multidimensional academic network, which can represent different correlations in terms of activity-based collaboration relationship, specialty-aware connection, and topic-aware citation fitness, among a series of academic entities with a set of measures.
- An improved algorithm based on Random Walk with Restart (RWR), in which the time-varying academic influence is newly introduced and measured in terms of the activeness of researchers and popularity of articles in a certain academic social context.
- A recommendation mechanism for research collaboration navigation based on the integration of multi-source scholarly data, which can efficiently provide researchers with scholarly recommendations to facilitate their collaboration works.

The remainder of this paper is organized as follows. Section II presents an overview of related works. In Section III, we introduce the structure of the multidimensional academic network and propose a set of measures to quantify the multi-type correlations within the constructed network model. In Section IV, after introducing the basic model of RWR, the academic influence of both researchers and articles are measured, and an algorithm is developed for the research collaboration navigation across academic social networks. The experiment and evaluation results using the DBLP and ResearchGate data are demonstrated in Section V. We conclude this study and give our promising perspectives regarding future research in Section VI.

II. RELATED WORK

Several issues relating to this study are discussed in this section. Foremost, issues of scholarly network analysis, and studies on academic literature and collaboration recommendation, are addressed respectively.

A. SCHOLARLY NETWORK ANALYSIS

Researchers have increasingly paid attentions on analyzing a variety of academic relationships in scholarly big data environments. Xia *et al.* [19] explored common author relations among pairwise articles based on two features: the ratio of pairwise articles with common author relations and the ratio of the most frequently appeared author. The result was utilized to improve the scientific article recommedndation according to author-based search patterns. Liu *et al.* [20] presented a recommendation method only using citation relations, in which papers were represented by each citing paper to calculate their similarities. In particular, Bai *et al.* [21] analyzed the so-called positive and negative citation relationships, which could be used to evaluate the impact of papers.

To extract the valuable information hidden across scholarly networks, several well-known algorithms have been applied, and proved to be effective. Sugiyama and Kan [17] utilized the traditional collaborative filtering algorithm to evaluate the significance of different sections of papers, which could be utilized to discover potential citation papers. West *et al.* [18] constructed an article-level citation network, and employed the hierarchical MapEquation method to generate clusters from the citation graph, which could provide researchers with the so-called "expert recommendation" and "classic recommendation". Xia *et al.* [24] improved the random walk based algorithm in a co-author network, to indentify valuable collaborators by considering the context of scholarly big data.

In addition to single-typed networks, multiple features are taken into account to model relationships among researchers and articles. Kong *et al.* [25] mined the research domains using a topic clustering model, and combined the results into a co-author based collaboration network to identify the more relevant collaborators. Guo *et al.* [26] designed a three-layer based recommendation model to describe multiple relations among authors and papers, aiming to provide the query-related recommendation using a simple random walk algorithm.

B. ACADEMIC LITERATURE AND COLLABORATION RECOMMENDATION

Generally, academic recommendations can provide a series of interesting results as mentioned above, including citation recommendation, collaborator recommendation, and conference recommendation etc. In particular, the citation recommendation can be classified into two categories: local citation recommendation and global citation recommendation [19], [20]. Huang *et al.* [27] built a neural probabilistic model to analyze semantic representations among cited papers, which could improve the recommendation accuracy according to the citation context. Ren *et al.* [28] constructed a cluster-based citation recommendation framework, to predict query-related ciations based on relations of clustered interest groups within heterogeneous bibliographic networks.

Collaborator recommendation is another significant issue to enhance academic associations for both researchers and institutions. Li *et al.* [29] analyzed the co-author network, and proposed three academic metrics to improve the link importance between two authors, in order to find new collaborators and facilitate collaboration research works. Considering several social principles (such as homophily and proximity), Brandao *et al.* [30] presented two metrics to analyze the academic context in academic social networks, and evaluated how these measures could influence collaboration recommendations.

As for the conference-related recommendation, Asabere *et al.* [23] developed a venue recommendation algorithm, in which conference participants' social capital, correlations, and research interests were utilized to analyze the context within a smart conference community. Experiments using real-world data demonstrated that this method could be used to suggest presentation session venues and improve the conference participation. Xia *et al.* [31] proposed a folksonomy-based recommendation mechanism to calculate conference participants' research interests, which could provide the socially aware recommendation of relevant scientific papers to other participants within the same smart conference.

In addition, studies on academic social recommendation systems have become increasingy popular [32]. Lopes *et al.* [33] presented an architecture for collaboration recommendation within academic social networks, in which metrics was engaged to measure the level of cooperation between pairs of researchers. Lee *et al.* [34] proposed a content-based recommendation algorithm using the metadata from authors' publications. Authors' research expertise, associated with their professional social networks, was utilized to predict co-author relationships. Chin *et al.* [35] built a system based on the analysis of online and offline interactions at the conference level, aiming to study how the O2O interactions could improve the friend recommendation and link more researchers together.

C. SUMMARY

Scholarly network analysis has been proved as an efficient way to deal with the explosive academic information along with diversified associations in scholarly big data environments. In particular, as a special social network, academic social network is playing a more and more important role in scholarly data dissemination and academic collaboration enlargement. Thus, to construct such scholarly network model, four major factors, namely, citation, co-author, research content, and academic activity, can be considered, which are shown in Table 1, comparing with other existing works.

As the comparison result shown in Table 1, in this study, the four important factors are taken into account to comprehensively describe and measure the academic collaboration related features in terms of three basic relations within our proposed network model. More precisely, the scientific publication data and academic activity data are seamlessly integrated together to analyze and quantify multi-type correlations between different academic entities.

Besides, comparing with traditional graph-based recommendation methods, the time-varying academic influence extracted from academic social networks, is newly introduced and analyzed for both researchers and articles in academic social recommendation systems. Furthermore, an improved

TABLE 1. Comparison of research works on scholarly network analysis.

Research	Citation	Co-Author	Research Content	Academic Activity
Sugiyama [17], West [18]	✓		✓	
Xia [19]		✓		✓
Liu [20], Bai [21]	✓			
Xia [24]		✓		
Kong [25]		✓	✓	
Guo [26]	✓	✓	✓	
AIMN (this work)	✓	✓	✓	✓

recommendation algorithm based on RWR model is developed to provide researchers with the reseach collaboration navigation in the socialized scholarly big data environments.

III. MODELING OF MULTIDIMENSIONAL ACADEMIC NETWORK

In this section, we introduce and define three basic relations to construct a multidimensional academic network in scholarly big data environments. A set of measures is proposed to represent and quantify multi-type correlations between different academic entities.

A. BASIC MODEL DESCRIPTION

The basic academic entities in scholarly data environments include two core factors, namely the researchers and articles. Figure 1 demonstrates the diversified relationships associated with various academic activities in scholarly big data environments, including following relationships and article recommendation behaviors existing in academic social networks, co-author and citation relationships from scientific publication data, etc.

Thus, three fundamental relations: Researcher-Researcher, Researcher-Article, and Article-Article, are taken into account to construct the multidimensional academic network. In particular, the definition is expressed as follows.

$$G_{AIMN}(V, E, C_T), \tag{1}$$

$V = R \cup A$ is a non-empty set of vertexes in the multidimensional network model.

$R = \{r_1, r_2, \ldots, r_m\}$ indicates a sub-set of vertexes in the constructed network, in which each r_i denotes a unique researcher. Specifically, $r_i = (RID_i, F_{r_i}, Act_i)_T$, in which RID_i indicates the researcher ID to identify a unique researcher; F_{r_i} is a vector with a series of extracted keywords to indicate the current research topics of researcher r_i during a selected time period T; Act_i indicates the so-called activeness factor of researcher r_i, and can be quantified based on the combination of impact factor (denoted from ResearchGate) and acamemic social activities during T.

$A = \{a_1, a_2, \ldots, a_n\}$ is a sub-set of vertexes in the constructed network, in which each a_i denotes a unique article published in the academic dataset. Specifically, $a_i = (AID_i, F_{a_i}, Pop_i)_T$, in which AID_i is the article ID to identify a unique article; F_{a_i} is a vector with a series of extracted keywords that represents the research topics included in article a_i; Pop_i indicates the so-called popularity factor of article a_i, and is quantified based on a combination of the frequencies of reads (denoted from ResearchGate) and citations during T.

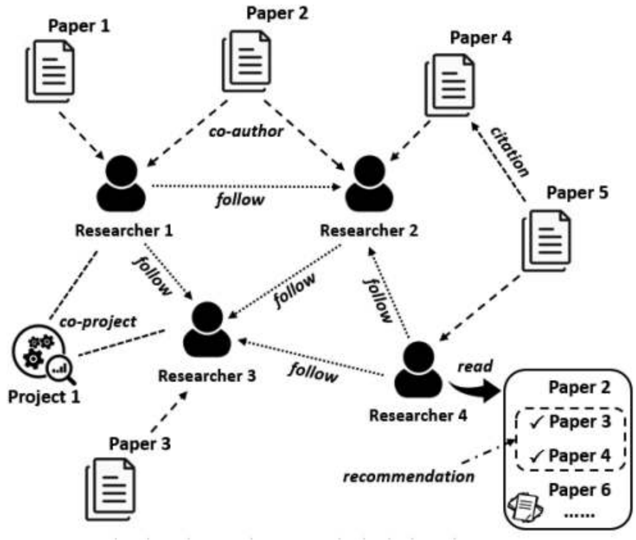


FIGURE 1. Multiple relationships in scholarly big data environments.

$E = \{e_{ij} \in V \times V \mid$ if a relationship exists between v_i and v_j, v_i, $v_j \in R \cup A\}$, is a collection of edges that connect the vertexes in V. Specifically, it contains three basic relations as mentioned above: 1) E_{RR} indicates the Researcher-Researcher relation, where the edge $\overrightarrow{r_i r_j}$, extending from vertex r_i to r_j, denotes an activity-based collaboration relationship from researcher r_i to r_j; 2) E_{RA}/E_{AR} indicates the Researcher-Article relation, where the edge $\overrightarrow{r_i a_j}/\overrightarrow{a_j r_i}$, extending from vertex r_i to a_j (or a_j to r_i), denotes a so-called specialty-aware connection from researcher r_i to article a_j (or a_j to r_i); and 3) E_{AA} indicates the Article-Article relation, where the edge $\overrightarrow{a_i a_j}$, extending from vertex a_i to a_j, denotes a topic-aware citation from article a_i to a_j.

$C_T = \{C_{T_{ij}} | \, if \, \exists \, e_{ij} \in E\}$ is a multi-tuple to describe the correlation on the corresponding edge, which includes a set of measures to describe and quantify these relationships within the multidimensional network. Each measure is defined to calculate the strength of a specific correlation between two connected academic entities during T.

Following these definitions discussed above, Figure 2 illustrates a conceptual example of the multidimensional academic network which is constructed based on the relationships shown in Figure 1. Details of the multi-type correlations between different researchers and articles are introduced and discussed in the following sections.

B. ANALYSIS OF RESEARCHER-RESEARCHER RELATIONS

The Researcher-Researcher relation demonstrates collaborations between two conncected researchers based on their academic associations. Specifically, the activity-based

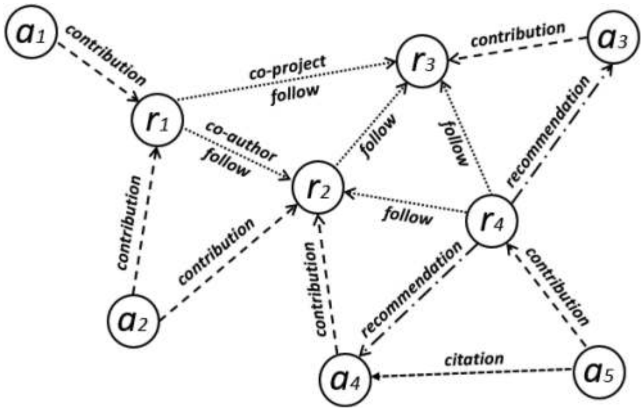


FIGURE 2. Conceptual image of a multidimensional academic network.

collaboration relationship is analyzed to characterize this kind of relation, which can be defined as follows.

$E_{RR} \subset R \times R$: This describes one kind of collaboration relations. Specifically, $E_{RR} = \{<r_i,\ ACR_{ij},\ r_j> |r_i,\ r_j \in R\}$, can be identified and analyzed by incorporating scientific publication data and academic activity data. ACR_{ij} indicates the weight assigned to edge $\overrightarrow{r_i r_j}$, to quantify the correlation in terms of the activity-based collaboration relationship from researcher r_i to r_j.

More precisely, co-author relationships from researchers' scientific publication data (e.g., DBLP), following relationships, and co-project-related relationships in their academic social networks (e.g., ResearchGate), are extracted to describe their explicit and implicit collaborations. In particular, the following relationship is employed to determine whether there exists an edge $\overrightarrow{r_i r_j}$, from researcher r_i to r_j, while the co-author relationship and co-project-related relationship between r_i and r_j are employed as additional weight assigned to this edge.

To measure this kind of activity-based collaboration relationship, the frequency of co-author and co-project-related collaborations, and the collaboration time (i.e., year), are utilized for calculation. Specifically, given two researchers r_i and r_j, the quantification of activity-based collaboration relationship ACR_{ij} within a time period T can be expressed as follows.

$$ACR_{ij} = \alpha FR_{ij} + \beta \sum_{x=1}^{\left|T_{A_{ij}}\right|} \frac{1}{t_c - t_{a_x} + 1} + \gamma \sum_{y=1}^{\left|T_{P_{ij}}\right|} \frac{1}{t_c - t_{p_y} + 1}, \quad (2)$$

where $\alpha,\ \beta,\ \gamma$ are the equilibrium coefficients, $\alpha + \beta + \gamma = 1$. FR_{ij} indicates the existing following relationship from r_i to r_j, and the value is equal to 1. $T_{A_{ij}} = \{t_{a_1}, \{t_{a_2}, t_{a_3}, \ldots\}$, lists the publication time of all the co-authored articles by researchers r_i and r_j during T, $t_{a_x} \in T_{A_{ij}}$. t_c indicates the current time (i.e., 2017), and $|T_{A_{ij}}|$ indicates the number of co-authored articles from researchers r_i and r_j during T. Likewise, $T_{P_{ij}} = \{t_{p_1}, t_{p_2}, t_{p_3}, \ldots\}$, lists the update time of all the co-projects joined by researchers r_i and r_j during T, $t_{p_y} \in T_{P_{ij}}$. $|T_{P_{ij}}|$ indicates the number of co-projects joined by researchers r_i and r_j during T.

C. ANALYSIS OF RESEARCHER-ARTICLE RELATIONS

Obviously, the Researcher-Article relation can be classified into two sub-types: from articles to researchers, and from researchers to articles. Specifically, the specialty-aware connection is analyzed to represent and characterize this kind of relation, which demonstrates the field-specific contribution of a researcher in an article, or the expertise-related recommendation from a researcher to an article. Firstly, this kind of connections from articles to researchers can be defined as follows.

$E_{AR} \subset A \times R$: This describes one kind of contribution relations. Specifically, $E_{AR} = \{<a_i,\ SAC_{ij}, r_j> \mid a_i \in A \wedge r_j \in R\}$, can be identified and analyzed from the scientific publication data. SAC_{ij} indicates the weight assigned to edge $\overrightarrow{a_i r_j}$, to quantify the correlation in terms of the field-specific contribution of researcher r_j in article a_i.

More precisely, the writtenby relationship with the co-author order in a specific article, is extracted to describe the corresponding contribution. In particular, the writenby relationship is employed to determine whether there exists an edge $\overrightarrow{a_i r_j}$, from article a_i to researcher r_j, while the co-author order in a_i is employed to calculate the weight assigned on this edge.

To measure this kind of field-specific contribution between researchers and articles hidden in co-author orders, it is assumed that the order, from the first author to the last, indicates the contributions of different authors to one specific article, from the strong to the weak. Moreover, it is noted that some special contributions (e.g., advice or suggestion from the supervisor) have not been considered in this situation (because in most of the case, the supervisor is always listed as the last author). Specifically, given a specific article a_i with the corresponding author list L_{a_i}, the quantification of field-specific contribution FSC_{ij} can be expressed as follows.

$$FSC_{ij} = \frac{\left|L_{a_i}\right| - s_j + 1}{\sum_{n=1}^{\left|L_{a_i}\right|} s_n}, \quad (3)$$

where L_{a_i} indicates the author list of $a_i.s_j = 1, 2, \ldots, |L_{a_i}|$, indicates the co-author order of r_j in the list.

On the other hand, connections from researchers to articles can be defined as follows.

$E_{RA} \subset R \times A$: This describes one kind of recommendation relations. Specifically, $E_{RA} = \{<r_i,\ ERR_{ij},\ a_j> |r_i \in R \wedge a_j \in A\}$, can be identified and analyzed from the academic activity data. ERR_{ij} indicates the weight assigned to edge $\overrightarrow{r_i a_j}$, to quantify the correlation in terms of the expertise-related recommendation from researcher r_i to article a_j.

More precisely, the recommendation-related activity of researcher r_i to article a_j with the similarity of research interest/topic between them, is extracted to describe the corresponding recommendation. In particular, the recommendation action from the researcher's academic activity is employed to determine whether there exists an edge $\overrightarrow{r_i a_j}$, from researcher r_i to article a_j, while the research similarity is employed to calculate the weight assigned to this edge.

To measure this kind of expertise-related recommendation between researchers and articles, the recommendation action can be identified from the academic activity data (e.g., from ResearchGate). A specific article may receive several recommendations from different researchers. However, the recommendation strength may be different due to the diverse research backgrounds of these recommenders. Thus, the research interest/topic similarty is utilized to quantify the correlation between the recommender and the recommended article. Furthermore, it is assumed that a closer similarity will lead to a stronger recommendation strength. A LDA based topic modeling method [36] is employed to identify and extract the topics from a certain document corpus. For the recommended article, the abstract and keywords will be extracted as the necessary document corpus, and abstracts and keywords in the recommender's current publications (for example, in the most recent year) will be selected to compose the document corpus for his/her research interest analysis. Finally, two vectors, F_{r_i} and F_{a_j}, with the extracted feature keywords, will be used to represent the research interest/topic of the researcher and article respectively, and further employed to calculate the similarity between them. Specifically, given a specific researcher r_i with his/her recommended article a_j, the measure of Cosine similarity is utilized to quantify the expertise-related recommendation ERR_{ij} as follows.

$$ERR_{ij} = \frac{\sum_{x=1}^{N}\left(F_{r_ix} * F_{a_jx}\right)}{\sqrt{\sum_{x=1}^{N} {F_{r_ix}}^2} * \sqrt{\sum_{x=1}^{N} {F_{a_jx}}^2}}, \tag{4}$$

where F_{r_i} and F_{a_j} are the N-dimensional vectors to describe the research interest/topic of the researcher r_i and article a_j respectively, $F_{r_ix} \in F_{r_i}$, $F_{a_jx} \in F_{a_j}$, $x \in N$.

D. ANALYSIS OF ARTICLE-ARTICLE RELATIONS

The Article-Article relation demonstrates the citation relationship between two connected articles. Specifically, the topic-aware citation fitness is analyzed to characterize this kind of relation, which can be defined as follows.

$E_{AA} \subset A \times A$: This describes one kind of citation relations. Specifically, $E_{AA} = \{<a_i, TCS_{ij}, a_j>|a_i, a_j \in A\}$, can be identified and analyzed from the scientific publication data. TCS_{ij} indicates the weight assigned to edge $\overrightarrow{a_ia_j}$, to quantify the correlation in terms of the topic-aware citation fitness from article a_i to a_j.

More precisely, the citation relationship with the research topic similarity between two articles, is extracted from the scientific publication data to describe their topic-aware citations. In particular, the citation relationship is employed to determine whether there exists an edge $\overrightarrow{a_ia_j}$, from article a_i to a_j, while the similarity of research topics is employed to calculate the weight assigned to this edge.

Generally, the edge $\overrightarrow{a_ia_j}$ extending from article a_i to a_j will be weighted as 0 or 1 as a_i cited a_j or not. However, in most of the cases, merely using the value of 0 or 1 cannot sufficiently capture the real correlation between two articles. Because when an article a_i is cited, the relevance of this article to the citing article a_j may be differed according to the purpose to refer to this paper. For instance, if a_i is cited in the introduction part, it may be more relevant to the background of a_j, rather than the similarity of the method discussed in a_i. Thus, it is necessary to dynamically change the value between 0 and 1, in order to distinguish the relevances of a series of cited articles to the citing article. To this purpose, the topic similarity is utilized to improve this kind of correlations between two connected articles in our model. Similar to the calculation process based on the topic modeling method [36] discussed in the previous section, two vectors, F_{a_i} and F_{a_j}, with the extracted feature keywords, will be used to represent the research topics of two articles respectively, and further employed to calculate the similarity between them. Specifically, given two specific articles a_i and a_j, the measure of Cosine similarity is utilized to quantify the topic-aware citation fitness TCF_{ij} as follows.

$$TCF_{ij} = \frac{\sum_{x=1}^{N}\left(F_{a_ix} * F_{a_jx}\right)}{\sqrt{\sum_{x=1}^{N} {F_{a_ix}}^2} * \sqrt{\sum_{x=1}^{N} {F_{a_jx}}^2}}, \tag{5}$$

where F_{a_i} and F_{a_j} are the N-dimensional vectors to describe the research topics of article a_i and a_j respectively, $F_{a_ix} \in F_{a_i}$, $F_{a_jx} \in F_{a_j}$, $x \in N$.

IV. ACADEMIC INFLUENCE AWARE RECOMMENDATION FOR RESEARCH COLLABORATION NAVIGATION

In this section, after introducing the basic RWR model, two measures are proposed to analyze the academic influence for both researchers and articles. An improved algorithm based on RWR is then developed for the research collaboration navigatioin.

A. THE RANDOM WALK WITH RESTART MODEL

The RWR model has been widely used to capture structure-aware features in a weighted graph. It has been proved to be an efficient way to calculate the similarity in terms of the relevance between two nodes in a constructed network model, which can achieve useful results in numerous social recommendation systems.

Thus, the RWR is used as the bacis model to generate recommendations from our constructed multidimensional academic network. In general, the well-known expression of RWR model can be described as follows.

$$HR^{(t+1)} = \lambda M * HR^{(t)} + (1-\lambda)q, \tag{6}$$

where λ, ranging from 0 to 1, is a damping coefficient. $HR^{(t)}$ indicates a ranking score vector at the iteration step t. q is the initial vector when starting the RWR model, which means $HR^{(0)} = q$. Typically, q is initialized as $[0, 0, \ldots, 1, \ldots, 0, 0]$, in which "1" indicates the target vertex v_i at the beginning. M is a transfer matrix, which indicates the probability of each vertex to transfer to the others.

Generally, given a specific vertex v_i as the starting node, the RWR model iteratively transmits to its neighborhood node based on the corresponding probability in M, calculated as λW_{ij}. On the contrary, it has the probability of $(1-\lambda)q_i$ to get back to itself.

Traditional RWR model usually treats each element W_{ij} as the same in the matrix M, which means the weight assigned to each edge is set to the same value. This is obviously not suitable when dealing with the multidimensional network model. Thus, it becomes an essential and necessary issue to make the transfer matrix to be more robust in multidimensional networks, which can benefit the improvement of RWR algorithm as well.

B. MEASURE FOR ACADEMIC INFLUENCE ANALYSIS

When running the RWR model, the rich information of both nodes and links will be integrated to generate a relevance score (or called ranking score) in a global way. Basically, given a target node beforehand, this kind of score depends on two important factors: the number of neighborhood nodes that are connected to a target node, and the importance of these nodes in the network model. Thus, in addition to the diverse correlations between two connected nodes, we introduce the academic influence, to measure the importance of each researcher or article in the constructed multidimensional network model.

To describe the academic influence for both researchers and articles in our multidimensional network model, given a specific vertex v_i, the academic influence AI_i can be expressed as follows.

$$AI_i = \begin{cases} Act_i, \textit{if } v_i \textit{ indicates a researcher} \\ Pop_i, \textit{if } v_i \textit{ indicates an article}, \end{cases} \tag{7}$$

where Act_i indicates the activeness of a specific researcher r_i, and Pop_i indicates the popularity of a specific article a_i respectively, during the current time period.

More precisely, the activeness is introduced to represent the current status on academic influence for each researcher. ResearchGate has given us a good example to interpret researchers' academic influence in terms of their specific impact factor. However, merely this factor is relatively static, thus is not suitable to represent the time-varying influence of each researcher. Specifically, researchers' update frequency of their academic activities is employed together with their impact factors to dynamically demonstrate their time-varying academic influence. Given a specific researcher r_i, the calculation of activeness of r_i can be expressed as follows.

$$Act_i = \frac{FA_{t_s r_i}}{FA_{T r_i}} * IF_{r_i}, \tag{8}$$

where $FA_{t_s r_i}$ indicates the frequency of updated activities of researcher r_i during the current time period t_s, while $FA_{T r_i}$ indicates the frequency of updated activities of r_i during the whole selected time period T. IF_{r_i} indicates the impact factor of r_i.

On the other hand, the popularity is introduced to represent the current status on academic influence for each article. The academic statistics based on both publication data and activitiy data in social environment is employed to represent this kind of influence of articles. Specifically, frequencies of reads and citations of an article published in social media (e.g., ResearchGate) are employed to dynamically capture the time-varying academic influence of each article. Given a specific article a_i, the calculation of popularity of a_i can be expressed as follows.

$$Pop_i = \nu \frac{FR_{t_s a_i}}{FR_{T a_i}} + (1-\nu) \frac{FC_{t_s a_i}}{FC_{T a_i}}, \tag{9}$$

where ν is an equilibrium coefficient, ranging from 0 to 1. $FR_{t_s a_i}$ indicates the frequency of reads of article a_i during t_s, while $FR_{T a_i}$ indicates the frequency of reads of a_i during T. Likewise, $FC_{t_s a_i}$ indicates the frequency of citations of article a_i during t_s, while $FC_{T a_i}$ indicates the frequency of citations of a_i during T.

C. MECHANISM FOR RESEARCH COLLABORATION NAVIGATION

As the core issue in the RWR model, the transfer matrix will directly influence the ranking score, and the recommendation result as well. In this study, both the correlation between the connected academic entities (i.e., researcher or article) and the academic influence on each vertex, are integrated together to construct the transfer matrix M. Specifically, given a multidimensional academic network with all researchers and articles connected, the element W_{ij} in the transfer matrix M can be expressed as follows.

$$W_{ij} = C_{T_{ij}} * AI_j, \tag{10}$$

where $C_{T_{ij}}$ indicates the value of a specific correlation between vertex v_i and v_j. AI_j indicates the corresponding academic influence of v_j. For instance, if the edge exists from article a_i to researcher r_j, the corresponding element W_{ij} in M will be calculated using FSC_{ij} in Eq. (3) and Act_j in Eq. (8).

Therefore, the RWR-based algorithm for research collaboration navigation is shown in Figure 3, which can more effectively transfer to more positive nodes based on the improved transfer matrix M.

Based on the algorithm shown in Figure 3, the ranked top-n nodes can be obtained. Consequently, given a specific researcher as the target node, the proposed multidimensional network model involves both the actual and potential collaboration-related interconnections associated around him/her. According to the improved RWR-based algorithm, a series of valuable nodes with high academic influence in the corresponding fields will be extracted and recommended. These recommended results can be provided as the research collaboration navigation to assist the target researcher to find more active researchers and more popular articles in more relecant research fields.

V. EXPERIMENT AND ANALYSIS

In this section, evaluation experiments are conducted to demonstrate the practicability and usefulness of our proposed model and method.

Input: The weighted graph G_{AIMN} (V, E, C_T) based on the multidimensional network model; a given target node vx
Output: A set of top-n nodes based on the ranking score to vx

Initialize q according to the given target node vx;
Initialize $HR^{(0)} = q$;
Initialize *NumOfIteration* ;
Initialize *MinDelta* for break ;
diff = 0 ;
for each W_{ij} in M
  if the edge is $\overrightarrow{r_i r_j}$
    $W_{ij} = ACR_{ij} * Act_j$;
  end if
  if the edge is $\overrightarrow{r_i a_j}$
    $W_{ij} = ERR_{ij} * Pop_j$;
  end if
  if the edge is $\overrightarrow{a_i r_j}$
    $W_{ij} = FSC_{ij} * Act_j$;
  end if
  if the edge is $\overrightarrow{a_i a_j}$
    $W_{ij} = TCF_{ij} * Pop_j$;
  end if
end for
for $i = 1$ to *NumOfIteration*
  $HR^{(t+1)} = \lambda M * HR^{(t)} + (1 - \lambda)q$;
  diff = $HR^{(t+1)} - HR^{(t)}$;
  if *diff* < *MinDelta*
    break ;
  end if
end for
Rank all the nodes based on their ranking scores ;
Return the top-n nodes for the target node vx ;

FIGURE 3. Algorithm for research collaboration navigation.

A. DATA SET

Both the original DBLP data and metadata crawled from ResearchGate are employed to conduct the evaluation experiments. More precisely, the DBLP data is collected to describe the basic publication record of each article, such as title, co-authors, crossref, publication years, etc. The data from ResearchGate is crawled to describe their diversified academic activities, including "co-project", "recommendation", "reads", "citations", etc. Finally, we collected 379,456 articles from 13,100 researchers, with the dataset period ranging from 1989 to 2017. In particular, the dataset was further divided into two subsets: we selected 10,000 researchers with their 80% articles as the training set, while the remaining data was used as the testing set. The detailed summary of the dataset is shown in Table 2.

B. EXPERIMENT DESIGN

To demonstrate the effectiveness of our AIMN method, three usually used evaluation metrics: precision, recall, and F1-measure, are employed to conduct the evaluation experiments according to the collaborations (e.g., existing co-author works in publications), which contains the following elements.

i). *NumTP*: having collaborations with the target node and recommended;

TABLE 2. Summary of the dataset.

	Statistics
Researchers	13,100
Articles	379,456
Projects	51,497
Reads	2,806,392
Citations	2,333,636
Recommednations	15,811

ii). *NumFP*: not having collaborations with the target node but recommended;

iii). *NumFN*: having collaborations with the target node but not recommended;

iv). *NumTN*: not having collaborations with the target node and not recommended;

With these elements, the precision, recall, and F1-measure metrics can be calculated as follows.

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

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

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

Accordingly, the following three methods, which consider multiple factors for research collaboration recommendations in scholarly big data envrionments, are chosen for comparison.

i). The basic RWR recommendation: This is the baseline method, which provides recommendations by running a basic RWR algorithm in the conducted multidimensional academic network.

ii). MVCWalker-based recommendation [24]: This is a RWR-based recommendation model, which considers the co-author order, latest collaboration time, and times of collaboration within a co-author network, to recommend the most valuable collaborators.

iii). CCRec-based recommendation [25]: This is a hybrid recommendation model, which considers publication contents and collaboration networks together, to provide personalized collaborator recommendations.

As for the settings of equilibrium coefficients used in each equation in our AIMN method, α was set as 0.3, while β and γ was set as 0.35 in Eq. (2), which means the importance of these three relationships was treated as equal as possible in this experiment. Likewise, the coefficient ν in Eq. (9) was set as 0.5. The default number of iterations in the algorithm was set as 30. In particular, following the test results discussed in [29], the damping coefficient λ was set as 0.8, which was used to determine the probability of RWR algorithm to jump back to the target vertex at the beginning.

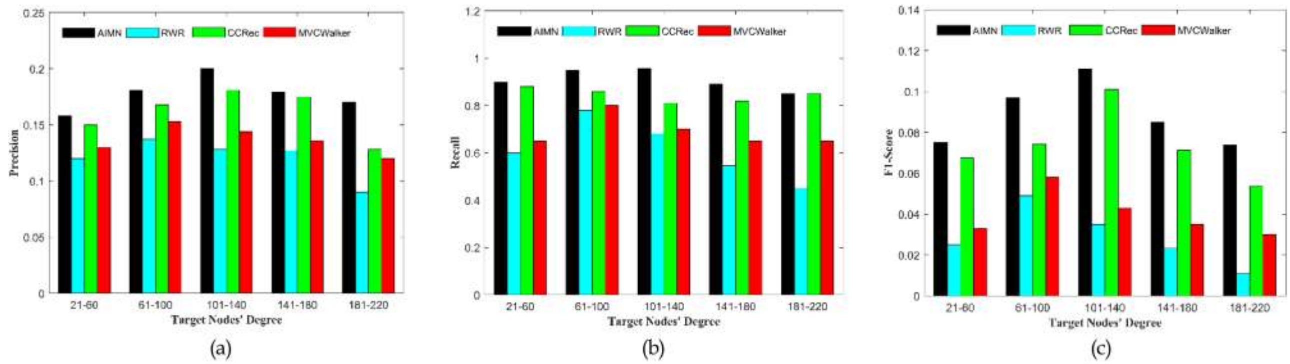


FIGURE 4. Performances based on different target nodes' degree. (a) Precision. (b) Recall. (c) F1.

C. PERFORMANCE EVALUATION

Considering the influence of different degrees of the selected target nodes, we demonstrate the performances of our AIMN method according to the different ranges of target nodes' degree. Specifically, the ranges of target nodes' degree were divided as: 21∼60, 61∼100, 101∼140, 141∼180, and 181∼220. Specifically, in each range, we selected 30 target nodes to conduct the experiment. The results are evaluated based on the precision, recall, and F1 metrics, and compared with the three mentioned recommendation methods, which are shown in Figures 4(a), 4(b), and 4(c) respectively. We give our observations and discussions based on these evaluation results as follows.

i) As a whole, according to the precision, recall, and F1 metrics, the performance of our AIMN method outperforms the other three recommendation methods in each range of target nodes' degree. This indicates the good applicability of our proposed method, especially when dealing with the target nodes in the degree range of 101∼140.

ii) It seems that our AIMN method and the CCRec-based recommendation method are more sensitive to the change of target nodes' degree, comparing with the other two methods. In particular, the CCRec-based method and our method could perform better in the degree range of 101∼140, while the MVCWalker-based method and the basic RWR method achieved their best results in the degree range of 61∼100. These results indicate that our method and the CCRec-based method, which consider multiple academic factors for research collaboration recommendations, can be more efficient in scholarly big data environments.

iii) In general, the results become better along with the increase of target nodes' degree (up to the degree range of 101∼140). This indicates the fact that given a target node, stronger connecctions in scholarly networks with richer collaboration information, will lead to a better recommendation result. On the other hand, the results become worse especially when the node degree is bigger than 200. This is to be expected as the scholarly network becomes more and more complex, it is more chanllenging and difficult to find optimized suggestions. Thus, it is necessary to involve the idea of academic influence to facilitate the research collaboration navigation.

D. COMPARISON ANALYSIS

To demonstrate the overall performance of the four methods for comparison, we randomly selected 100 target nodes, and calculated their average values in terms of each metric.

The evaluation results based on four metricsnamely, precision and recall rate, PR-Curve, and F1, are shown in Figures 5(a), 5(b), 5(c), and 5(d) respectively. We give our observations and discussions based on these comparison results as follows.

i) Basically, Figure 5(a) demonstates a general downward trend according to the precision metric. Logically, the precision rate decreases along with the increase of the length of recommendation list. Performances of all four methods become similar when the recommendation lists reach to 25 nodes. These results indicate that our AIMN method outperforms other three methods when recommending less than 25 nodes from our proposed model in terms of the precision metric.

ii) Similarly, Figure 5(b) demonstrates a general upward trend according to the recall metric. In particular, different from the performances of other three methods which increased gradually during the whole process, our AIMN method can reach the peak efficiently around recommending 25 to 30 nodes. This indicates the stability of our proposed model when the recommendation list is longer than 30 nodes in terms of the recall metric.

iii) As a whole, Figure 5(c) demonstrates the overall performances of the four methods according to the PR-curve. The area under curve illustrates the effectiveness of our AIMN method comparing with the others. Consequently, our method and the CCRec-based method perform better than the basic RWR and MVCWalker-based method, and our method achieves the best result according to the precision and recall metrics. This result indicates the necessity to take the multiple academic factors into account, in order to facilitate the research collaboration navigation.

iv) Finally, the F1-measure based results are shown in Figure 5(d). It is obvious that the recommendations based

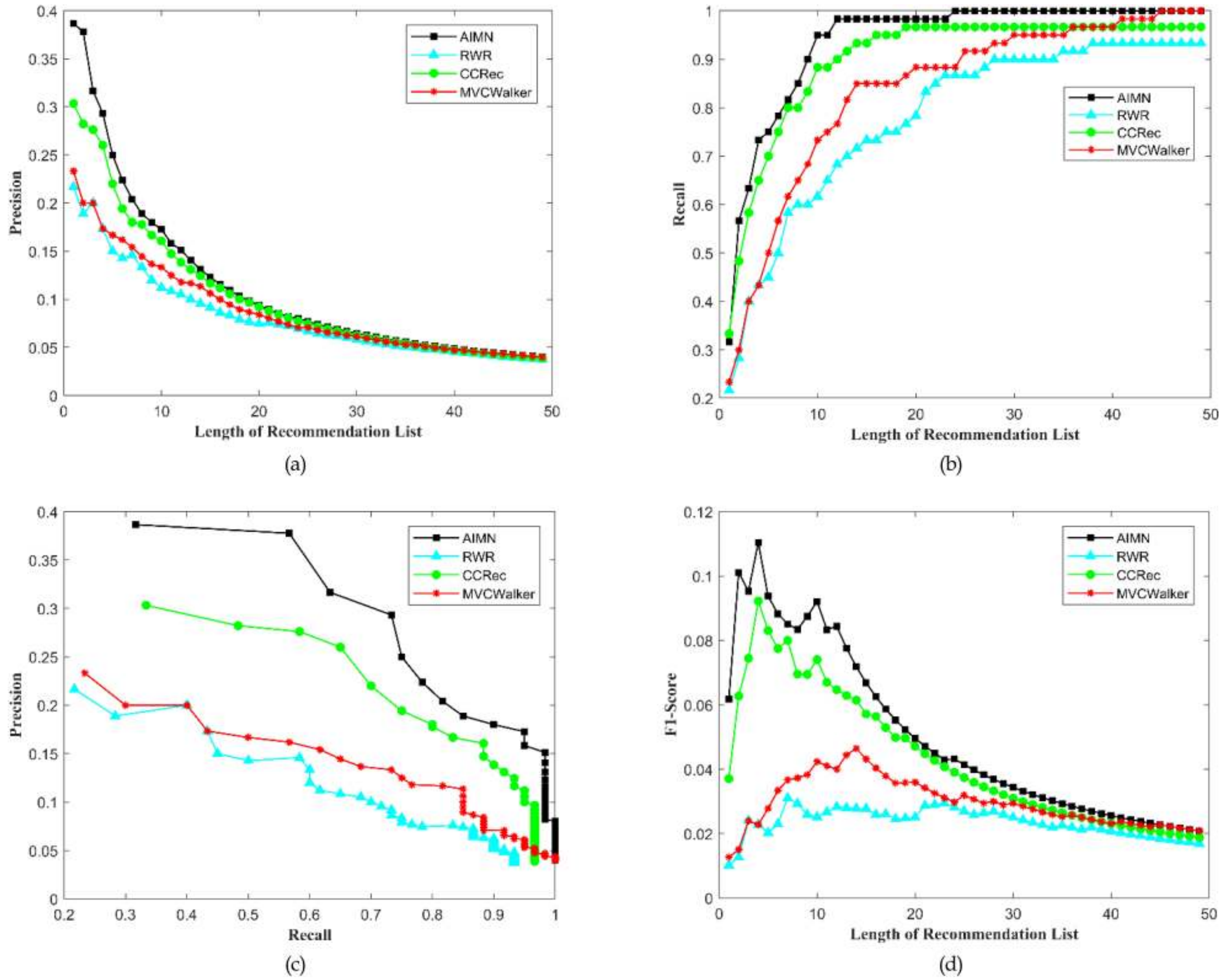


FIGURE 5. Comparison results among four methods based on different metrics. (a) Precision. (b) Recall. (c) PR-Curve. (d) F1.

on the CCRec-based and our method achieve better results comparing with the basic RWR and MVCWalker-based method, when the recommendation list is no more than 25 nodes. Especially, our AIMN method outperforms other three methods and reaches a peak of 0.11, when recommending around 5 nodes from the proposed model. These results indicate the importance that the analysis of multi-type correlations in a multidimensional network model can effectively benefit recommendations in scholarly big data environments.

E. DISCUSSION

In summary, according to the precision, recall rate, and F1-score results, the performance of basic RWR method is the least effective for the research collaboration navigation among all four methods. This is because the conventional RWR method simply assigns all the elements in the transfer matrix with the same value, which is not suitable to analyze the multi-type correlations in a multidimensional network model. The situation becomes even worse when dealing with the sparsity issue in big data envrionments. In contrast, all the other three methods consider the different link importance when constructing the transfer matrix. For instance, the MVCWalker-based method defines that matrix based on several factors among co-author relationships, thus could achieve a better result comparing with the basic RWR recommendation.

Obviously, our AIMN method and the CCRec-based method perform better than the other two recommendation methods. This can be explained in two aspects. First, in addition to the co-author relationships considered in the MVCWalker-based method, at least two or three other relations, such as the similarity of publication content, citations among articles, and academic interactions in social media, are taken into account in the CCRec-based and our method. These underline the importance of utilizing multiple academic factors to improve the research collaboration navigation in scholarly big data environments. Second, since the MVCWalker-based recommendation considers solely the existing co-author relationships to construct the network model, it becomes more suitable to those researchers who have known each other, or already established their collaboration relatioships. Accordingly, when extending to wider application scenarios, our method and the CCRec-based method would achieve better results because they can provide not only recommendations based on the existing collaborations, but also recommend potential collaborations according to their indirect relationships.

Our AIMN method outperforms other models in terms of the more effective recommendation results as shown in Figure 5(a), 5(b), 5(c), and 5(d). This can be summarized as: our model has a more comprehensive consideration for both the scientific publication and academic social activity in scholarly big data environments. A variety of academic factors, including research content, co-author, citation, and academic activity, are integrated together to analyze their multi-type correlations. These provide more contextual information and enrich the multidimensional network model. In addition, the time-varying academic influence is extracted from academic social networks and analyzed for both researchers and articles in our recommendation algorithm. This can contribute to capturing those collaboration-related features and associations more accurately and timely, and further facilitate the research collaboration navigation in scholarly big data environments. Therefore, our method can provide the remarkably good scholarly recommendations especially with a relatively shorter recommendation list.

VI. CONCLUSIONS

In this study, we presented a multidimensional network model with a set of measures to describe and quantify multi-type correlations among a series of academic entities (i.e., researchers and articles) in scholarly big data environments. An improved RWR-based algorithm with newly defined academic influence was developed to provide researchers with research collaboration navigation to facilitate their future works. The main findings of this study are summarized as follows.

First, a multidimensional network model was constructed based on three basic relations mixed together. A set of measures was introduced and defined to represent and quantify multi-type correlations in terms of activity-based collaboration relationships, specialty-aware connections, and topic-aware citation fitness among a series of researchers and articles connected together.

Second, the time-varying academic influence was defined and measured for both researchers and articles in a certain academic social context. An improved algorithm based on RWR model was developed, in which both the multi-type correlations and time-varying academic influence were taken in to account, to facilitate the research collaboration navigation.

Third, experiments and evaluations were conducted using an integrated dataset from DBLP and ResearchGate. Analysis results demonstrated the effectiveness of our proposed model and method in providing researchers with scholarly recommendations for research collaboration navigation in scholarly big data envionments.

In future studies, we will focus on in-depth understanding of the dynamics of scholarly big data. Experiments will be designed considering the heterogeneous data environment. More evaluations will be conducted to optimize the coefficients in equations and improve the algorithm to adapt more complex situations using the multidimensional network model.

ACKNOWLEDGMENTS

The work has been partially supported by the National Science Foundation of China under Grant Nos. 61273232, 61772196 and 61472136, and the Hunan Provincial Education Department Foundation for Excellent Youth Scholars under Grant No. 17B146.

REFERENCES

[1] J. Priem, “Scholarship: Beyond the paper,” *Nature*, vol. 495, no. 7442, pp. 437–440, 2013.

[2] K. Williams, J. Wu, S. R. Choudhury, M. Khabsa, and C. L. Giles, “Scholarly big data information extraction and integration in the CiteSeer X digital library,” in *Proc. IEEE 30th Int. Conf. Data Eng. Workshops*, 2014, pp. 68–73.

[3] K. Williams, L. Li, M. Khabsa, J. Wu, P. C. Shih, and C. L. Giles, “A web service for scholarly big data information extraction,” in *Proc. IEEE Int. Conf. Web Serv.*, 2014, pp. 105–112.

[4] 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. 1, 2017.

[5] Y.-R. Lin, H. Tong, J. Tang, and K. S. Candan, “Guest editorial: Big scholar data discovery and collaboration,” *IEEE Trans. Big Data*, vol. 2, no. 1, pp. 1–2, Jan.-Mar. 2016.

[6] Z. Wu, J. Wu, M. Khabsa, K. Williams, H. Chen, W. Huang, S. Tuarob, S. R. Choudhury, A. Ororbia, P. Mitraz, and C. L. Giles, “Towards building a scholarly big data platform: Challenges, lessons and opportunities,” in *Proc. IEEE/ACM Joint Conf. Digital Libraries*, 2014, pp. 117–126.

[7] R. Munroe, “The rise of open access,” *Sci.*, vol. 342, no. 6154, pp. 58–59, 2013.

[8] M. Khabsa and C. L. Giles, “The number of scholarly documents on the public web,” *Plos One*, 2014.

[9] Y. Liu, J. Venkatanathan, J. Goncalves, E. Karapanos, and V. Kostakos, “Modeling what friendship patterns on facebook reveal about personality and social capital,” *ACM Trans. Comput.–Human Interact.*, vol. 21, no. 3, 2014, Art. no. 17.

[10] M. Thelwall and K. Kousha, “ResearchGate: Disseminating, communicating, and measuring scholarship?” *J. Assoc. Inf. Sci. Technol.*, vol. 66, no. 5, pp. 876–889, 2015.

[11] S. Milojevic, C. R. Sugimoto, E. Yan, and Y. Ding, “The cognitive structure of library and information science: Analysis of article title words,” *J. Amer. Soc. Inf. Sci. Technol.*, vol. 62, no. 10, pp. 1933–1953, 2011.

[12] F. Osareh, R. Khademi, M. K. Rostami, and M. S. Shirazi, “Coauthorship network structure analysis of iranian researchers’ scientific outputs from 1991 to 2013 based on the social science citation index (SSCI),” *Collnet J. Scientometrics Inf. Manage.*, vol. 8, no. 2, pp. 263–271, 2014.

[13] M. Newman, *Networks: An Introduction*. London, United Kingdom: Oxford Univ. Press, 2010.

[14] S. X. Zhao and F. Y. Ye, “Power-law link strength distribution in paper cocitation networks,” *J. Amer. Soc. Inf. Sci. Technol.*, vol. 64, no. 7, pp. 1480–1489, 2013.

[15] S. Tuarob, S. Bhatia, P. Mitra, and C. L. Giles, “AlgorithmSeer: A system for extracting and searching for algorithms in scholarly big data,” *IEEE Trans. Big Data*, vol. 2, no. 1, pp. 3–17, Mar. 1, 2016.

[16] S. Amjad, H. Mukhtar, and C. Dunne, “Automating scholarly article data collection with action science explorer,” in *Proc. Int. Conf. Open Source Syst. Technol.*, 2014, pp. 160–169.

[17] K. Sugiyama and M.-Y. Kan, “A comprehensive evaluation of scholarly paper recommendation using potential citation papers,” *Int. J. Digit. Libraries*, vol. 16, no. 2, pp. 91–109, Jun. 2015.

[18] J. D. West, I. Wesley-Smith, and C. T. Bergstrom, “A recommendation system based on hierarchical clustering of an article-level citation network,” *IEEE Trans. Big Data*, vol. 2, no. 2, pp. 113–123, Jun. 1, 2016.

[19] F. Xia, H. Liu, I. Lee, and L. Cao, “Scientific article recommendation: Exploiting common author relations and historical preferences,” *IEEE Trans. Big Data*, vol. 2, no. 2, pp. 101–112, Jun. 1, 2016.

[20] 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, 2015.

[21] X. Bai, I. Lee, Z. Ning, A. Tolba, and F. Xia, “The Role of Positive and Negative Citations in Scientific Evaluation,” in *IEEE Access*, vol. 5, pp. 17607–17617, 2017.

[22] Y. Garay, M. Akbar, and A. Q. Gates, "Towards identifying potential research collaborations from scientific research networks using scholarly data," *Proc. IEEE/ACM Joint Conf. Digital Libraries*, 2016, pp. 217–218.

[23] N. Y. Asabere, F. Xia, W. Wang, J. J. P. C. Rodrigues, F. Basso, and J. Ma, "Improving smart conference participation through socially aware recommendation," *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 5, pp. 689–700, Oct. 2014.

[24] F. Xia, Z. Chen, W. Wang, J. Li, and L. T. Yang, "MVCWalker: Random walk-based most valuable collaborators recommendation exploiting academic factors," *IEEE Trans. Emerging Topics Comput.*, vol. 2, no. 3, pp. 364–375, Sep. 2014.

[25] X. Kong, H. Jiang, Z. Yang, Z. Xu, F. Xia, and A. Tolba, "Exploiting publication contents and collaboration networks for collaborator recommendation," *PLoS One*, vol. 11, no. 2, 2016, Art. no. e0148492.

[26] L. Guo, X. Cai, F. Hao, D. Mu, C. Fang, and L. Yang, "Exploiting fine-grained co-authorship for personalized citation recommendation," *IEEE Access*, vol. 5, pp. 12714–12725, 2017.

[27] W. Huang, Z. Wu, C. Liang, P. Mitra, and C. Lee Giles, "A neural probabilistic model for context based citation recommendation," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 2404–2410.

[28] X. Ren, J. Liu, X. Yu, U. Khandelwal, Q. Gu, L. Wang, and J. Han, "ClusCite: Effective citation recommendation by information network-based clustering," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 821–830.

[29] J. Li, F. Xia, W. Wang, and H. Jiang, "Acrec: A co-authorship based random walk model for academic collaboration recommendation," in *Proc. Companion Publication 23rd Int. Conf. World Wide Web*, 2014, pp. 1209–1214.

[30] M. A. Brandão, M. M. Moro, G. R. Lopes, and J. P. M. Oliveira, "Using link semantics to recommend collaborations in academic social networks," in *Proc. 22nd Int. Conf. World Wide Web*, 2013, pp. 833–840.

[31] F. Xia, N. Y. Asabere, H. Liu, N. Deonauth, and F. Li, "Folksonomy based socially-aware recommendation of scholarly papers for conference participants," in *Proc. 23rd Int. Conf. World Wide Web*, 2014, pp. 781–786.

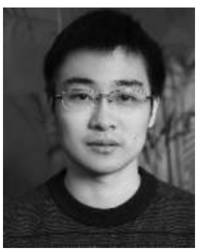
[32] R. Van Noorden, "Online collaboration: Scientists and the social network," *Nature*, vol. 512, no. 7513, pp. 126–129, 2014.

[33] G. R. Lopes, M. M. Moro, L. K. Wives, and J. P. M. de Oliveira, "Collaboration recommendation on academic social networks," in *Proc. Int. Conf. Conceptual Model.*, 2010, pp. 190–199.

[34] D. H. Lee, P. Brusilovsky, and T. Schleyer, "Recommending collaborators using social features and MeSH terms," in *Proc. Amer. Soc. Inf. Sci. Technol.*, 2011, pp. 1–10.

[35] A. Chin, B. Xu, H. Wang, and X. Wang, "Linking people through physical proximity in a conference," in *Proc. 3rd Int. Workshop Model. Soc. Media*, 2012, pp. 13–20.

[36] W. Liang, Z.X. Lu, Q. Jin, and M. Wu, "Analyzing of research patterns based on a temporal tracking and assessing model", *Personal Ubiquitous Comput.*, vol. 20, no. 6, pp. 933–946, 2016.



XIAOKANG ZHOU (M'12) received the PhD degree in human sciences from Waseda University, 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, Japan. From 2016, he has been a lecturer with the Faculty of Data Science, Shiga University, Japan. He also works as a visiting researcher in the RIKEN Center for Advanced Intelligence Project (AIP), RIKEN, Japan, from 2017. He has been engaged in interdisciplinary research works in the fields of computer science and engineering, information systems, and human informatics. His recent research interests include ubiquitous and social computing, data mining and analytics, behavior and cognitive informatics, user modeling, human-computer interaction, and information seeking and recommendation. He is a member of the IEEE CS, and ACM, USA, IPSJ, Japan, and JSAI, Japan.

WEI LIANG received the MS and PhD degrees in computer science from Central South University, in 2005 and 2016, respectively. He is currently working at the Key Laboratory of Hunan Province for Mobile Business Intelligence, Hunan University of Commerce, China. His research interests include information retrieval, data mining, and artificial intelligence. He has published more than 10 papers at various conferences and journals, including FGCS, JCSS, and PUC.

KEVIN I-KAI WANG (M'04) received the bachelor's of engineering (Hons.) degree in computer systems engineering and the PhD degree in electrical and electronics engineering from the Department of Electrical and Computer Engineering, the University of Auckland, New Zealand, in 2004 and 2009 respectively. He is currently a senior lecturer in the Department of Electrical and Computer Engineering, the University of Auckland. He was also a research engineer designing commercial home automation systems and traffic sensing systems from 2009 to 2011. His current research interests include wireless sensor network based ambient intelligence, pervasive healthcare systems, human activity recognition, behaviour data analytics and bio-cybernetic systems.

RUNHE HUANG (SM'17) received the PhD degree in computer science and mathematics from the University of the West of England, in 1993. She is a full professor in the Faculty of Computer and Information Sciences at Hosei University, Japan. She received a Sino-Britain Friendship Scholarship for her PhD study in UK. Her research fields include cognitive computing, brain modeling, computational intelligence computing, big data, machine learning. She is a senior member of the IEEE.

QUN JIN (M'95, SM'17) is a professor at the Networked Information Systems Laboratory, the Department of Human Informatics and Cognitive Sciences, Faculty of Human Sciences, Waseda University, Japan. He has been extensively engaged in research works in the fields of computer science, information systems, and social and human informatics. He seeks to exploit the rich interdependence between theory and practice in his work with interdisciplinary and integrated approaches. His recent research interests cover human-centric ubiquitous computing, behavior and cognitive informatics, big data, data quality assurance and sustainable use, personal analytics and individual modeling, intelligence computing, blockchain, cyber security, cyber-enabled applications in healthcare, and computing for well-being. He is a senior member of the Association of Computing Machinery (ACM), Institute of Electrical and Electronics Engineers (IEEE), and Information Processing Society of Japan (IPSJ).