Scientometrics (2020) 123:429–449
https://doi.org/10.1007/s11192-020-03374-z

HNRWalker: recommending academic collaborators with dynamic transition probabilities in heterogeneous networks

Chen Yang[1] · Tingting Liu[1] · Xiaohong Chen[1] · Yiyang Bian[2] · Yuewen Liu[3]

Received: 12 October 2019 / Published online: 8 February 2020

Abstract
Multi-source information not only helps to solve the problem of sparse data but also improves recommendation performance in terms of personalization and accuracy. However, how to utilize it for facilitating academic collaboration effectively has been little studied in previous studies. Traditional mechanisms such as random walk algorithms are often assumed to be static which ignores crucial features of the linkages among various nodes in multi-source information networks. Therefore, this paper builds a heterogeneous network constructed by institution network and co-author network and proposes a novel random walk model for academic collaborator recommendation. Specifically, four neighbor relationships and the corresponding similarity assessment measures are identified according to the characteristics of different relationships in the heterogeneous network. Further, an improved random walk algorithm known as "Heterogeneous Network-based Random Walk" (HNRWalker) with dynamic transition probability and a new rule for selecting candidates are proposed. According to our validation results, the proposed method performs better than the benchmarks in improving recommendation performances.

Keywords Collaborator recommendation services · Heterogeneous networks · Random walk algorithms · Link prediction · Academic social platforms

Introduction

Academic collaboration is becoming increasingly prevalent due to the rapid development of knowledge and academic exchanges. Therefore, various academic social platforms such as ResearchGate from Germany, Academia from the United States of America, and ScholarMate from China are designed to strengthen these exchanges among

✉ Yiyang Bian
bianyiyang@nju.edu.cn

[1] College of Management, Shenzhen University, Shenzhen, Guangdong, People's Republic of China

[2] School of Information Management, Nanjing University, No. 163 Xianlin Rd, Qixia District, Nanjing, Jiangsu, People's Republic of China

[3] School of Management, Xi'an Jiaotong University, Xi'an, Shaanxi, People's Republic of China

researchers and help them find collaboration partners. These platforms provide users with information about relevant research fields and researchers' achievements (Lee et al. 2019) and create borderless and accessible opportunities for round-the-clock academic exchange. Specifically, academic collaborator recommendation is becoming increasingly important since it can automatically help users find appropriate academic collaborators (Khan et al. 2017).

Collaborator recommendation has greatly promoted the development of international disciplines by adopting various classical algorithms for academic collaborator recommendation (Montefusco et al. 2019). Among the methods, graph-based models with collaboration network features are widely used for recommending collaborators. These methods are classified into node-based methods and path-based methods. Node-based methods create a co-author network in which authors are denoted as nodes and their co-authors are regarded as neighbor nodes (Xie et al. 2018), adopting the recommendation strategy that the more similar two nodes are, the higher the probability of linkages. The basic similarity of the improved walking strategy proposed in this paper tests the structural features. It belongs to node-based methods. These methods are famous for their simple calculation, strong universal applicability, and ability to mine users' collaboration preferences with topological features. Different from the node-based methods, path-based methods consider global features rather than local features. In this paper, a novel path-based solution named "Heterogeneous Network-based Random Walk" (HNRWalker for short) is proposed. Random walk methods are the mainstream algorithms for collaborator recommendation, and they utilize the structure of networks, especially heterogeneous networks, to iterate and transfer the similarities of nodes.

Both node-based methods and path-based methods perform poorly on sparse data, and some of the existing studies about collaborator recommendation are based on homogeneous networks resulting the reduction of usable features (Zhou et al. 2017). One feasible solution is to use the multi-source information in heterogeneous networks, which can not only solve the problem of data sparsity but also improve the personalization and accuracy of recommending results (Valdeolivas et al. 2018; Zhou et al. 2017). In the long term, the use of heterogeneous networks will be necessary and meaningful in the diversified development of internet platforms. Therefore, this paper proposes a novel recommendation approach (HNRWalker) that performs well with heterogeneous networks and sparse data. Some traditional algorithms likes PageRank (Jung et al. 2016) and PersonalRank (Zhao et al. 2014) are not used in this paper, because their transition probability is static and too indistinct to reflect the real cooperative preferences of the target users. Our approach guarantees the randomness of the walking and positive guidance during the walking.

To sum up, this paper builds a heterogeneous network using institution and co-author networks. This heterogeneous network contains four different neighbor relationships based on the different institution affiliations and co-authorships. Then, several measures based on these four different neighbor relationships are provided to incorporate the characteristics of heterogeneous networks. Importantly, this set of similarity representation converts the static transition probabilities into dynamic transition probabilities, guides the random walking process, and retains the target users' past cooperation habits and preferences.

To test the effect of the improvements, three stages of comparative experiments were conducted. The proposed HNRWalker was compared with three types of benchmarks (Co-author Network-based Random Walk (CNRWalker) algorithm, five node-based link prediction methods, and two classical random walk algorithms). Results of the 3-stage experiments show that: (1) incorporating the institution network improves collaborator recommendation performance; (2) the new algorithm achieves higher precision and AUC

(short for Area under Curve) than the typical node-based link prediction methods; (3) the new mechanism performs better than the typical random walk algorithms.

There are three main contributions in this research. First, co-author and institution networks are combined to build heterogeneous networks to enable comprehensive profiling. In the heterogeneous networks, a new assessment mechanism is developed to generate candidate nodes. Second, the attributes of the institution and co-author networks are reasonably integrated to better represent the link formation scenes. Compared with PersonalRank and Random Walk with Restart algorithms, the improved walking strategy that applies dynamic transition probability better characterizes the transfer process of similarities among nodes. Third, all the recommended sequences are constructed based on the counted times that the starting nodes reach other candidate nodes. This is different from prior random walk algorithms which use the similarity between starting nodes and candidate nodes.

The rest of the paper is structured as follows. Section "Related work" details academic collaboration networks and collaborator recommendation methods including the link prediction and random walk algorithms. Then, the mechanism of our proposed random walk algorithm and its differences compared to the typical Random Walk with Restart algorithm are analyzed in Section "An improved random walk mechanism". In Section "The proposed model", we explain the details of our approach, including building the heterogeneous networks and formalization of migration measures that combine multiple attributes. Some validation experiments are designed and analyzed in Section "Experiments" and "Validation Results and Analyses". Section "Discussion" concludes this research and discusses future work.

Related work

In this section, we review the research related to this paper in four subsections. In Section "Research on collaborator recommender systems", we sort out the relevant research on collaborator recommendation and discuss previous related research from two aspects: feature extraction of researchers and modeling methods. We find that single and heterogeneous features are both leveraged, but heterogeneous cooperative networks are relatively more comprehensive choices. Thus, we compare the network features and specifically discuss the characteristics of heterogeneous networks in Section "Collaboration networks and heterogeneous networks", which provides the research basis for the heterogeneous networks selected in this paper. Because the graph-based methods are more suitable for processing heterogeneous network features, we discuss the graph-based models and two typical types of algorithms in "Graph-based models for collaborator recommendation". Section "Summary" is a summary of existing gaps in previous studies, which foreshadows the model proposed in this paper.

Research on collaborator recommender systems

How to find more relevant and appropriate academic collaborators is an essential academic question. Manual search is simple and direct, but it mainly relies on interpersonal relationships in reality, so it may fall into "local optimum" and miss the information in cooperative networks. In contrast, recommender systems can help scholars find appropriate collaborators with less cost. Therefore, scholars pay attention to the collaborator recommender systems research. Most previous research focuses on the two main aspects of collaborator

recommendation, selection of collaboration features and the recommendation modeling methods.

Early research focused on content features and network features. Content features include citation articles (Alshareef et al. 2018), keywords (Cohen and Ebel 2013), titles (Luong et al. 2015), and so on. Network features include co-author networks (Zhang 2017) and venue networks (Zhou et al. 2017), etc. Some content features such as citation articles and titles are translated into semantic network features (Chaiwanarom and Lursinsap 2015; Davoodi et al. 2013). Single and heterogeneous network features are both leveraged in previous studies. However, multi-dimensional networks provide more comprehensive information than single networks, so heterogeneous networks constructed by multi-dimensional networks are advantageous for recommending academic collaborators. The discussion on collaboration networks can be found in Section "Collaboration networks and heterogeneous networks".

The mainstream processing method is to represent and calculate these features in the network structure, and the network-based processing methods are more appropriate for feature modeling, particularly graph-based models. They can be further classified into node-based methods and path-based methods. To facilitate readers' understanding, in Table 1 we classify the relevant literature from two different feature selection strategies and two different graph-based modeling methods. Node-based methods can measure the social proximity within single networks (Mahapatra et al. 2019); and the path-based methods are commonly used by scholars in heterogeneous networks (Mahapatra et al. 2019; Sun et al. 2019). Graph-based models are further discussed in Section "Graph-based models for collaborator recommendation".

Collaboration networks and heterogeneous networks

Establishing collaborations depends heavily on past cooperative relationships. The significance of collaboration networks in collaboration formation has been proved by Hoang et al. (2019), Song et al. (2019) and Bergé (2017). So, recommending collaborators based on features of collaboration networks is reasonable and effective.

Since heterogeneous networks contain more collaboration information than single homogeneous networks and can better model real-world collaboration systems (Sun and Han 2013), heterogeneous networks constructed by combining homogeneous networks are more beneficial for collaborator recommendation. As Table 1 shows, Zhou et al. (2017) and Guo and Chen (2014) utilized multiple networks in their collaborator recommendation research. Zhou et al. (2017) built their heterogeneous network with multiple types of objects (papers, terms, and venues). Guo and Chen (2014) built their heterogeneous network with a co-author network and a citation network.

Different from prior works, our research builds a heterogeneous network with the institution network and the co-author network. Co-author networks are commonly used in collaborator recommender systems (Chuan et al. 2018; Makarov et al. 2016), while the institution information has not been widely used even though it has been proved to be efficient information for collaborator recommendation in academic social networks (Brandao and Moro 2012) and can be used to predict the possibility of cooperation between uncooperative institutions (Bornmann and Leydesdorff 2015). Results of experiments in stage 1 shown in Sects. "Experiments" and "Validation results and analyses" also verify the role of institution networks.

Table 1 Classification of previous collaborator recommendation studies

Graph-based models	Homogeneous networks		Heterogeneous networks	
	Network types	Studies	Network types	Studies
Node-based methods	Co-author network	(Chuan et al. 2018; Huynh et al. 2013; Makarov et al. 2016)		
Path-based methods	Co-author network	(Li et al. 2014; Xia et al. 2014)	Co-author network and citation network, term network and venue network	(Zhou et al. 2017)
			Co-author network and citation network	(Guo and Chen 2014)

Graph-based models for collaborator recommendation

Based on the heterogeneous network and its characteristics, this paper uses graph-based models for collaborator recommendation, similar to Chaiwanarom and Lursinsap (2015), Zhou et al. (2017), and Sun et al. (2019). According to the emphasis on different network attributes, graph-based link prediction studies can be classified into node-based methods and path-based methods, also known as local indexes and global indexes respectively. Most of them have the characteristics of simple calculation and high recommendation accuracy in dense data-sets. Therefore, many researches try to obtain better recommendation performance by combining them with other methods, such as in the studies of Li et al. (2020), Xiao et al. (2018), and Wu et al. (2017). Next, node-based methods and path-based methods are elaborated in detail.

Node-based methods for collaborator recommendation

Some common and popular local indicators are Adamic–Adar Index (Adamic and Adar 2003), Jaccard Coefficient (Fang et al. 2018), Common Neighbors (Yao et al. 2016), Resource Allocation (Zhou et al. 2009), and Preferential Attachment (Weaver 2015). These indexes have the advantage of minor calculation compared with other algorithms in recommender systems. They can maintain their high calculation speed even in large and complex networks. However, this high-speed calculation is achieved by sacrificing other features since such algorithms only consider the neighbor features of the topology. Therefore, they have poor performances when the data is sparse, and it is hard to use such algorithms individually. However, when combined with other algorithms, these methods can maintain their high-speed calculation and perform well even with sparse data. In this paper, we combine these node-based methods with HNRWalker and compare their performances. Formulas of all the node-based methods are shown in Table 2. In the equations, $|\Gamma_i|$ and $|\Gamma_j|$ refer to the degrees of node i and j. $|\Gamma_i \cap \Gamma_j|$ represents the co-neighbor quantity of i and j. z is the co-neighbor of j and h. k_z is equal to $|\Gamma_i \cap \Gamma_j|$.

Path-based methods for collaborator recommendation

Research in collaborator recommendation usually employs path-based methods with common indicators such as the Katz Index (Li et al. 2016) and SimRank (Du et al. 2015). For example, by combining the friendship network and the auxiliary information from the user-item network, Symeonidis and Perentis (2014) converted the Katz

Table 2 Node-based methods

Node-based method	Abbreviation	Equation
Adamic–Adar index	AA	$\mathrm{AA}_{jh} = \sum_{z \in \Gamma_j \cap \Gamma_h} \frac{1}{\log(k_z)}$
Jaccard coefficient	JA	$\mathrm{JA}_{jh} = \frac{\lvert\Gamma_j \cap \Gamma_h\rvert}{\lvert\Gamma_j \cup \Gamma_h\rvert}$
Common neighbors	CN	$\mathrm{CN}_{jh} = \lvert\Gamma_j \cap \Gamma_h$
Preferential attachment	PA	$\mathrm{PA}_{jh} = \lvert\Gamma_j\rvert * \lvert\Gamma_h$
Resource allocation	RA	$\mathrm{RA}_{jh} = \sum_{z \in \Gamma_j \cap \Gamma_h} \frac{1}{k_z}$

Index to fit the multi-modal network. It has been proved that the introduction of heterogeneous information facilitates object search and increases recommendation accuracy (Zarrinkalam et al. 2018; Zhang 2017).

The random walk algorithm has been widely used in academic collaborator recommendation. Luong et al. (2015) used the co-author order, latest collaboration time point and frequency of collaboration to compute the walk edges and proposed their Academic Random Walk with Restart algorithm based on the author–author graph. Similarly, frequency of collaboration is also used in this paper to show the strength of co-author relationships. Zhou et al. (2017) used two importance measures (sequence importance and freshness importance) to weight edges in a heterogeneous network constructed with multiple types of objects (papers, items, and venues) and proposed the RWR-CR (Random Walk with Restart-based Collaborator Recommendation) algorithm. These studies have shown that the random walk algorithm is suitable for recommending collaborators.

The random walk strategy can also be regarded as a special link prediction method. The core principle of the strategy is to start from an initial node and randomly walk according to the neighbor relationship between nodes (Ostroumova Prokhorenkova and Samosvat 2016). When the target node arrives at its neighbor nodes, a new network node is formed. The random walk strategy emphasizes the connections between nodes, and the selection of the next node (hop) is closely related to the neighborhood characteristics of the current node. After constant rounds of iterations, the target node's own relational networks can be formed.

Summary

One serious deficiency in some random walk algorithms like PageRank, PersonalRank and Random Walk with Restart is that they do not combine the heterogeneous network structure with the node features, or just treat the nodes with equal importance as their transition probabilities are static (Valdeolivas et al. 2018). It is clearly that each node and linkage owns individual characteristics such as the strength of the relationships between nodes. AA, JA, and other node-based methods have proved that they can well characterize the relationship strength between neighbors. This relationship strength is a representation of the transitivity for the random walk algorithm. So, characteristics of the nodes should be considered while random walking. Moreover, academic collaboration is strongly dependent on relationship networks, and the collaboration intensity of scholars affects the possibility of the next collaboration. Liu and Jansen (2017) once used the number of social activity features such as retweets and likes to express one's popularity among his/her followers. Similarly, the number of papers that two researchers or institutions have published together can reflect the strength of their cooperation; thus, we also use the number of collaborations to express the importance of neighbor nodes to target nodes in the networks.

Based on the points above, this paper builds a heterogeneous network with the institution network and co-author network. Then, several measures are provided to quantify the institution network features and the co-author network features and evaluate four kinds of neighbor relationships. These measures succeed in converting the static transition probabilities into dynamic transition probabilities. More details about the new random walk algorithm (HNRWalker) are discussed in Sects. "An improved random walk mechanism" and "The proposed model".

An improved random walk mechanism

Finding new collaborators through previous collaborators can be seen as an extension of relationships. From the perspective of network structure, this extension phenomenon is a process in which a node relies on its neighbor nodes to walk outward. The condition for the node to walk is the strong relationships between nodes. For the walking algorithm, it is the transitivity. In this way, random walk algorithm which emphasizes transitivity is the typical algorithm for recommending academic collaborators. Following the basic idea of random walk, this paper proposes a random walk algorithm whose mechanism is wholly different from others. For ease of understanding, we take Random Walk with Restart (RWR for short) (Valdeolivas et al. 2018) as an example. In RWR, the current node keeps walking to its neighbor nodes in the graph model. When a neighbor node is reached, it will continue to walk with a static probability, or abandon walking and return back to the target node. After a finite number of iterations, the probabilities of reaching a vertex in the graph model gradually converge to a series of static values. These values are regarded as the importance degree associated with the starting node. It is generally believed that the higher the degree of similarity, the more likely it is to be a potential collaborator. The RWR formula is:

$$p^{(t+1)} = (1-d)Sp^{(t)} + \mathrm{d}q. \tag{1}$$

In Eq. 1, $p^{(t+1)}, p^{(t)}$ and q are column vectors. $p^{(t)}$ means the probability distribution of the step t, and $p_i^{(t)}$ is the probability that the starting node arrives at node i after t steps. The column vector q is a restart vector indicating the starting state, and q_i represents the probability of the starting node in node i. In the initialization phase, the value of the starting node (the target user) is set to 1, while other nodes are set to 0. S is the transition probability matrix and $S_{i,j}$ represents the current transition probability from node i to j. d is the probability that the starting node turns back to the target node.

Equation 1 makes it clear that this iterative process disobeys the Markov Chain. It will accumulate hop by hop and iteration by iteration until it eventually converges. However, our proposed HNRWalker would obey the Markov Chain. There are two main differences between HNRWalker and RWR. First, RWR has a static transition probability, but HNRWalker dynamically calculates the transition probability based on the attributes of nodes in heterogeneous networks. The specific calculation method of the dynamical probability will be explained in Section "The proposed model". Another difference is that RWR uses iterated and cumulative calculating to measure the importance of each node toward the target nodes. Instead of focusing on the importance degree of each candidate node, our algorithm counts the number of times that the target node reaches each candidate node during thousands of iterations. In each of the iterations, the target node will randomly move forward three hops along the network to determine if the visited node can get a recommended opportunity. After a sufficient quantity of iterations, the number of recommendations of each node are aggregated and counted. The more visits, the more similarity there is between the target node and the candidate node and the more likely they are to establish linkages. For a clearer presentation, the differences between HNRWalker and RWR are listed in Table 3.

The random walk process of our algorithm is given in Fig. 1. It is emphasized that the final recommendation set is obtained by the Top-N candidate nodes that accumulate the most recommendation opportunities. Here, each time the target node arrives at other candidate nodes, these candidate nodes will add one recommendation opportunity except for

Table 3 Differences between HNRWalker and RWR

Difference	HNRWalker	RWR
Transition probability	Dynamic	Static
Range of transition probability	Relative size	Absolute size
Markov chain	Obey	Disobey
Convergence	Non-convergence	Convergence
The possibility of restart	Dynamic	Static
Output	Visiting times	Similarity

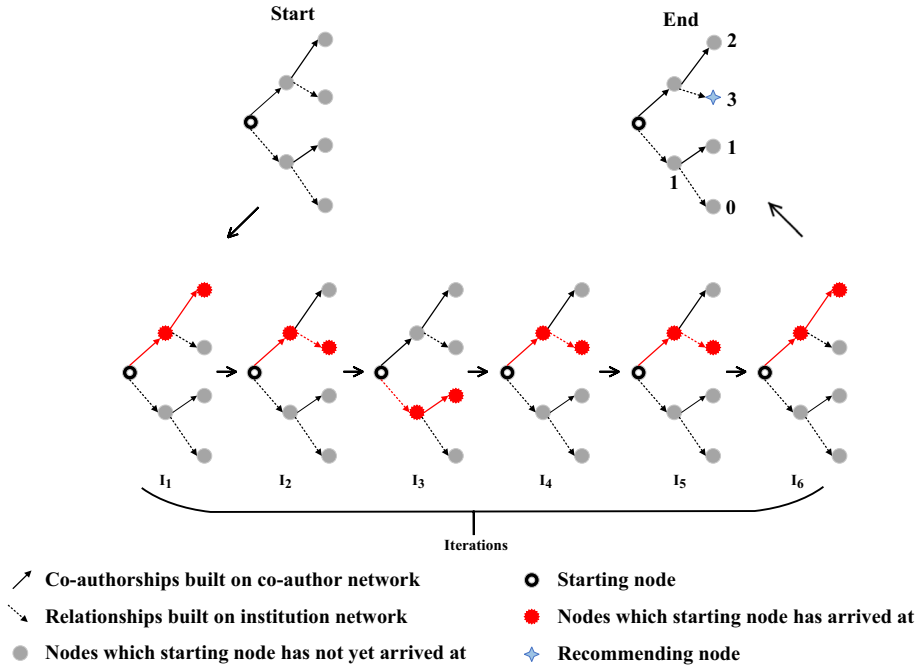


Fig. 1 An example of the HNRWalker mechanism. (After 6 rounds of iterations, the blue star node will be recommended since the starting node accesses it most frequently)

the nodes connected with the target node by a solid line. More details about heterogeneous networks, neighbor relationships, and transition probability are in the next section.

The proposed model

Based on the institution network and co-author network, we build a heterogeneous network and extract features from authors' co-authorship data and institutional cooperation data. A set of attribute similarity measures is defined to give target nodes positive guidance during random walking. Our approach is shown in Fig. 2.

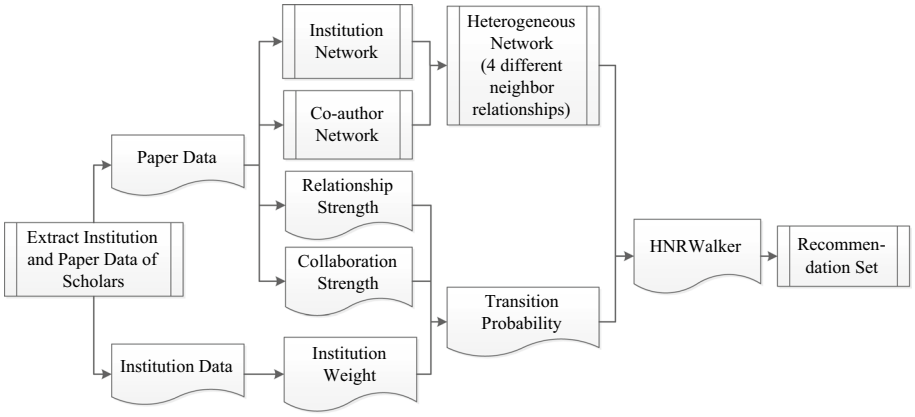


Fig. 2 HNRWalker recommendation process

Construction of the heterogeneous network

Existing studies have proved that institution information is important for academic collaboration research. For example, Hoekman et al. (2009) believed that institutional information can be used to study authors' collective cooperative behavior. Ortega and Aguillo (2013) proved that Google Scholar Citations (GSC) is a suitable tool for cooperation research only at the institution- and nation level. Bornmann and Leydesdorff (2015) found that institution information in co-authorship data can be used to predict the possibility of cooperation between uncooperative institutions. In fact, cooperation between institutions is mainly accomplished by inner members, so it is reasonable to believe that institution information can be used to predict the collaboration of members. Importantly, according to the research of Yan and Guns (2014), predicting collaborations with information at author and institution level is recommended.

Based on the viewpoints above, two kinds of relationship networks can be established based on cataloguing information of research papers. One is the co-author network. Linkages between two authors indicate that they have been co-authors before. The other one is the institution network. Linkages between two institutions indicate that their scholars have once been co-authors in academic publications. We show the heterogeneous network structure in Fig. 3.

Random walk meta-paths

There is a fundamental problem in heterogeneous information networks whose network structure relates multiple types of nodes and linkages. In order to combine these different nodes with different linkages and utilize various attributes to achieve academic collaborator recommendation, we must tease out these meta-paths. In general, these paths are divided into three categories. The first category is based on the co-author network. $\text{scholar} \xrightarrow{\text{paper}} \text{scholar} \xrightarrow{\text{paper}} \text{scholar}$ is the simplest and most direct meta-path. Related recommendation is the same as ordinary friend recommendation. The second category is based on the institution network. If two institutions have a close cooperation relationship, the internal scholars can form meta-paths. A sample of this kind of meta-path is

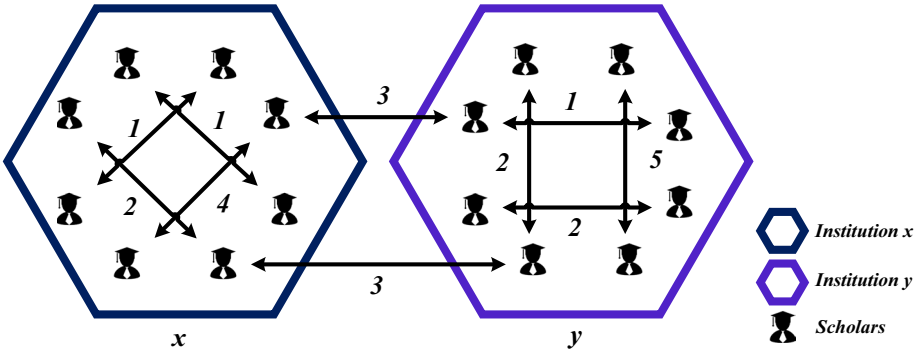


Fig. 3 A demonstration of heterogeneous network structure. (A line with a double-sided arrow indicates that two scholars have once published papers together, and the numbers beside the arrows denote the number of papers they have published together)

$\text{scholar} \xrightarrow{\text{institution}} \text{scholar} \xrightarrow{\text{institution}} \text{scholar}$. The third category is a mixture of the first and second categories, such as $\text{scholar} \xrightarrow{\text{institution}} \text{scholar} \xrightarrow{\text{paper}} \text{scholar}$. Linkage pairs in this kind of meta-path are formed by the co-authorships between scholars and partnerships between institutions. All kinds of meta-paths are shown in Fig. 4.

According to Figs. 3 and 4, starting node A has four kinds of "Neighbors". They are B, C, D and E. These neighbors have different institution and co-author relationships. B and C are A's real collaborators because they have published papers together. The difference between B and C is that B and A are in the same institution while C and A are not. D and E are A's potential collaborators. These potential relationships are built on the institution network. The difference between D and E is that D and A belong to the same institution while E and A do not.

HNRWalker will also regard D and E as A' s neighbors. When the starting node A reaches D or E, the recommender system will make a record because D and E have the qualification to be recommended as potential neighbors. Definitely, every visited node (whether a real neighbor or a potential neighbor) will then be the new starting node and keep walking to its neighbors as starting node A did. In our proposed walking mechanism, each iteration covers three hops, which is a common search space setting. Candidate nodes within the space can perform well according to the research of Chaiwanarom and Lursinsap (2015) and Cohen and Ebel (2013). Since more hops lowers efficiency, three hops

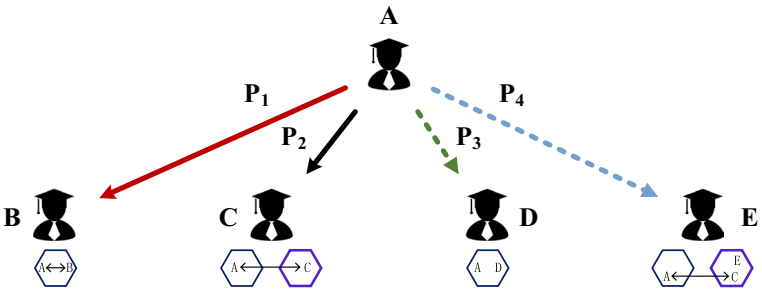


Fig. 4 Path graph of random walk in heterogeneous network

ensure both accuracy and efficiency. Then if the visited node is not the starting node's previous collaborators, it will be recorded.

Transition probability of HNRWalker

Methods based on node similarities Traditional link prediction methods can be divided into node-based similarity methods and path-based similarity methods. The former methods are more widely used because of their lower computational cost and stronger universality. Because HNRWalker needs a huge number of iterations, node-based similarity methods will be the base of the transition probability.

Node-similarity methods are based on a classic assumption: the larger the similarity between two nodes, the more possibility they will make a linkage. Previous improvements are mostly based on the number of co-neighbors, the degrees of nodes and the correlation between nodes. On one hand, the most extensive idea is that the more co-neighbors between two nodes, the more possibility that a potential relationship exists between them. On the other hand, the less degrees of two nodes, the more possibility they turn to establish a new linkage. Based on our experiments in stage 2 shown in Section "Experiments", we found that the Jaccard Coefficient performs best to measure the structural attributes between nodes. The formula is given in Eq. 2:

$$\mathrm{JA}_{ij} = \frac{|\Gamma_i \cap \Gamma_j|}{|\Gamma_i| + |\Gamma_j| - |\Gamma_i \cap \Gamma_j|}. \tag{2}$$

where $|\Gamma_i|$ and $|\Gamma_j|$ refer to the degrees of nodes i and j, or the numbers of their neighbors. $|\Gamma_i \cap \Gamma_j|$. represents the number of co-neighbors between i and j. This method is an indicator that comprehensively considers the number of co-neighbors and the degree of nodes. Past experiments show that it usually improves the efficiency without compromising accuracy much (Samanthula and Jiang 2015).

Collaboration ratio between scholars Xia et al. (2014) introduced the number of collaborations into their random walk model. However, absolute numbers cannot reflect the sence of the relationships, so we need to incorporate the total number of researchers' publications. We de t relationship strength between two scholars as the following formula:

$$\mathrm{Colfre}_{ij} = \frac{\mathrm{Paper}_{ij}}{\mathrm{Paper}_i + \mathrm{Paper}_j - \mathrm{Paper}_{ij}}. \tag{3}$$

We use $0 \leq \mathrm{Colfre}_{ij} \leq 1$ to indicate the strength of cooperation between two scholars and use Paper_{ij} to indicate the number of papers co-authored by scholars i and j. Paper_i and Paper_j are the number of papers published by scholars i and j respectively.

Institutional weight This study combines the institution network with the co-author network. Therefore, in the random walk process, it is possible to use the cooperation relationship between institutions to adjust the node relevance. Obviously, the strength of the cooperation between institutions will also affect the possibility of cooperation between their inner members. In particular, if one institution has a large quantity of research partners, how to distinguish the importance of partners is pivotal. A simple way is to count the

scholars in each institution and their collaboration frequency. Briefly, we define the weight between the institutions in Eqs. 4–6 as follows:

If both scholars i and j belong to institution $Insti_x$, then the institution weight SInst_Wei_{ij} will be:

$$\text{SInst\_Wei}_{ij} = \frac{1}{\text{Num}_x} = \frac{1}{\text{Num}_y}. \tag{4}$$

If scholars i and j belong to different institutions $Insti_x$ and $Insti_y$, then the institution weight DInst_Wei_{ij} will be:

$$\text{DInst\_Wei}_{ij} = \frac{\text{Colfre}_{xy}}{\text{Num}_x + \text{Num}_y} \tag{5}$$

$$\text{Colfre}_{xy} = \frac{\text{Paper}_{xy}}{\text{Paper}_x + \text{Paper}_y - \text{Paper}_{xy}} \tag{6}$$

Num_x and Num_y represent the total number of scholars within the institutions of i and j. $0 \leq \text{Colfre}_{xy} \leq 1$ is the collaboration ratio of institution x and y and can be calculated with the number of papers they published before. It can be proved that if two scholars come from the same institution, the weight of the institution will be larger and more advantageous than the institutional weight of two scholars from different institutions. The aim we set institutional weight is to highlight t importance of the institution when recommending collaborators.

Transition probability of HNRWalker: the combination of multiple attributes Previous literature shows that the recommendation effect of the node-similarity methods is susceptible to the close relationships between nodes. In general, the denser the relationships between nodes, the better the recommendation result. However, most social networks including the co-author network, are not dense in the real environment. Therefore, it is necessary to improve and innovate the structural similarity of network nodes. We tried to join the institution network and combine the attributes of the heterogeneous network to formalize new indicators.

As shown in Fig. 4:

P_1: If scholars i (A) and j (B) are in the same institution, and they once have published papers together, we can define the transition probability from i to j in Eq. 7:

$$\Pr o_{ij} = (\text{SInst\_Wei}_{ij} + \text{Colfre}_{ij}) * \text{JA}_{ij} = \left(\frac{1}{\text{Num}_x} + \frac{\text{Paper}_{ij}}{\text{Paper}_i + \text{Paper}_j - \text{Paper}_{ij}}\right) * 1 \tag{6}$$

P_2: If scholars i (A) and j (C) are not in the same institution, but they once have published papers together, we can define the transition probability from i to j in Eq. 8:

$$\text{Pro}_{ij} = \left(\text{DInst\_Wei}_{ij} + \text{Colfre}_{ij}\right) * \text{JA}_{ij} = \left(\frac{\text{Colfre}_{xy}}{\text{Num}_x + \text{Num}_y} + \frac{\text{Paper}_{ij}}{\text{Paper}_i + \text{Paper}_j - \text{Paper}_{ij}}\right) * 1 \tag{7}$$

P_3: If scholars i (A) and j (D) are in the same institution, but have published papers together (Colfre_{ij}=0), we can define the probability of walking from i to j in Eq. 9:

$$\text{Pro}_{ij}=(\text{SInst\_Wei}_{ij}+\text{Colfre}_{ij})*\text{JA}_{ij}=\frac{1}{\text{Num}_x}*\frac{\left|\Gamma_i\cap\Gamma_j\right|}{\left|\Gamma_i\right|+\left|\Gamma_j\right|-\left|\Gamma_i\cap\Gamma_j\right|} \tag{8}$$

P_4: If scholars i (A) and j (E) are not in the same institution and have n published papers together ($\text{Colfre}_{ij}=0$), we can define the probability of walking from i to j in Equation:

$$\text{Pro}_{ij}=(\text{DInst\_Wei}_{ij}+\text{Colfre}_{ij})*\text{JA}_{ij}=\left(\frac{1}{\text{Num}_x+\text{Num}_y}\right)*\frac{\left|\Gamma_i\cap\Gamma_j\right|}{\left|\Gamma_i\right|+\left|\Gamma_j\right|-\left|\Gamma_i\cap\Gamma_j\right|}. \tag{10}$$

It can be inferred from Eqs. 4 and 5 that SInst_Wei_{ij}. is always higher than DInst_Wei_{ij}, which means that two scholars in the same institution have higher institutional weight than two scholars in different institutions. Moreover, it is certain that JA_{ij} in P_3–P_4 ranges from 0 to 1 and could not be higher than 1. In P_1 and P_2, we set JA_{ij} to 1 because we assume that two scholars who have cooperated before own the highest JA value at 1. There is another difference between P_1 and P_3, P_2 and P_4. In P_1 and P_2, we add the collaboration frequency to Pro_{ij} so that we can respect scholars' past cooperation habits during random walking.

Pseudocode of the algorithm

In summary, the procedure of the proposed HNRWalker can be described in Table 4 as follows.

Experiments

To evaluate our algorithm, a 3-stage experiment is adopted in this study. The experiments were operated in the environment of an Inter(R) Core(tm) i5 processor, Windows10 system, 16 G memory computer, and Python 2.7. Some web crawler tools were used to acquire journal papers published in Library and Information Science (LIS) during 2005–2015 from ISI (Institute for Scientific Information) Web of Science (WoS), which is maintained by Thomson Reuters Scientific and widely used as the standard tool to generate citation data for research (Wang et al. 2016). Importantly, scholars were identified by both names and institutions. That is to say, scholars with the same name but in different institutions are regarded as different scholars. Meanwhile, if a scholar belongs to two different institutions, he or she is divided into two scholars. Finally, our data-set includes 72,315 papers published by 65,535 authors from 2005 to 2015. While processing the data, we found that the data is sparse. Regarding the scholar pool, 59,342 scholars had less than five collaborators and only 1.31% of the scholars had more than ten collaborators (or we could call them neighbor nodes in the networks).

We conducted one-to-one numbering for all papers and authors and used unique identifiers throughout the experiments. Then, we randomly selected 35 scholars with more than five collaborators as the target seeds. To divide the data into the test set and

Table 4 Procedure of the designed HNRWalker algorithm

Algorithm：HNRWALKER	
1:	Input: $user_{target}$ (user who need collaborator recommendation)
2:	$Collaboration\_set_{scholar}$ (cooperation situation of scholars)
3:	$Collaboration\_set_{institude}$(cooperation situation of institutions)
4:	Output: $recom\_set_{target}$ (recommendation set).
5:	Begin: $user_{target}$ iteration
6:	For i=1 to k do
7:	For j=1 to 3 do
8:	Find the node set $node\_set_{hoop_j}$ that current node can continue walking to
9:	Calculate the transition probability set $pro\_set_{hoop_j}$
10:	Generate a random number R based on $pro\_set_{hoop_j}$
11:	Get the probability range and find the land point $Land\_Point_j$
12:	If $Land\_Point_j$ not in $Collobration\_Set_{Scholar}$
13:	Add $Land\_Point_j$ to the $Candidate\_Set_{target}$
14:	End for
15:	End for
16:	Count the number of occurrences of each node in the $Candidate\_Set_{target}$
17:	Sort the number of times for building the $Recom\_Set_{target}$
18:	End

- **Note.** k is a pre-set number of iterations, which can be adjusted according to the size of the network.

training set, we randomly put one-fifth of the partnerships into the test set and put the remaining partnerships into the training set. This procedure ensures that the number of partnerships between the training set and the test set remains at a golden ratio of 4:1, which is an appropriate ratio chosen by many researchers. After cutting, the entire training set contains 154,562 linkage pairs, and the test set contains 39,237 linkage pairs.

Finally, three stages of comparative experiments were conducted to validate our previous conjectures.

Stage 1 To test whether introduction of the institution network can improve the recommendation results, we compared our proposed HNRWalker algorithm with the Co-author Network-based Random Walk (CNRWalker) algorithm. This type of benchmark is similar to HNRWalker. However, since CNRWalker is based on a homogeneous network (co-author network), all their linkages are built between co-authors, which also means that there are only two meta-paths (p_1 and p_2 in Fig. 4) while walking. The transition probabilities of path p_1 and p_2 are measured by Eqs. 7 and 8. In a word, HNRWalker_JA is compared with CNRWalker_JA in this stage of experiment.

Stage 2 Following Eqs. 7–10, we combined our Heterogeneous Network-based Random Walk (HNRWalker) algorithm with node-based link prediction methods (JA, AA, RA, CN, and PA) as HNRWalker_JA, HNRWalker_AA, HNRWalker_RA, HNRWalker_CN, and HNRWalker_PA. Then, we compared their performance with JA, AA, RA, CN and PA (formulas are shown in Table 2). In this way, the effect of the HNRWalker algorithm with new similarity assessment measures can be verified. It is easy to find that the link weight of the real neighbors (co-authors) is 1 from Eqs. 7 and 8. This is because the similarity calculated by JA ranges from 0 to 1 and the maximum is 1. However, as the range of CN and PA is not limited in [0, 1], we need to normalize the values. The normalization strategy was adopted in three steps. First, for neighbors built by the co-author network, we found the maximum values of PA and CN. Then, we calculated the PA and CN for neighbors built by the institution network. Last, these values were divided by the maximum value to get their ratios.

Stage 3 Our proposed HNRWalker was compared with PersonalRank (PR for short) and RWR (Eq. 1) to verify the efficiency of the proposed random walk mechanism. The PR formula is shown in Eq. 11. d is the probability of stopping walking and turning back to the starting node. N is the number of nodes in the network. in(i) and out(j) refer to the in-degree of i and the out-degree of j.

$$\mathrm{PR}_i = \frac{d}{N} + (1-d)\sum_{j\in \mathrm{in}(i)} \frac{PR_j}{|\mathrm{out}(j)|} \tag{11}$$

Evaluation indexes of Precision (Shi et al. 2016) and AUC (short for Area Under Curve) (Zahr et al. 2010) are selected in this paper. Precision, shown in Eq. 12, is equal to the ratio of actual positive examples in the recommended positive examples. AUC measures the probability that the score of an actual existing linkage selected randomly will be higher than the score of non-existent linkage, which is also selected randomly. Assuming that we compare them N times independently, and there are n' existing links with higher scores and n'' linkages with the same scores, then the value of AUC can be calculated with Eq. 13.

$$\mathrm{Precision} = \frac{\text{the number of correct positive predictions}}{\text{the number of positive predictions}} \tag{12}$$

$$AUC = (n' + 0.5n'')/N \tag{13}$$

The value of AUC ranges from 0 to 1, and the higher the value, the better the recommendation performance of the algorithm. If all the prediction scores are made randomly, then AUC equals 0.5. Therefore, it is generally recognized that the AUC values of good recommendation algorithms should not be lower than 0.5, which means their performances are even worse than the random conjectures (Huang and Ling 2005).

Validation results and analyses

In Tables 5 and 6 and Fig. 5, the results of the three stages of experiments are shown respectively. Higher evaluation indicator values are highlighted in bold in tables, and figures utilize the color histogram to compare performance differences.

Table 5 Experiment performance (stage 1): comparison of HNRWalker_JA and CNRWalker_JA

	Precision (%)					AUC (%)				
	@1	@2	@3	@5	@10	@1	@2	@3	@5	@10
HNRWalker_JA	**28.13**	**26.56**	**19.79**	**16.25**	**9.69**	61.35	**71.34**	**74.37**	**74.83**	**78.47**
CNRWalker_JA	15.63	15.63	15.63	10.00	6.25	**62.96**	61.30	61.40	70.10	72.33

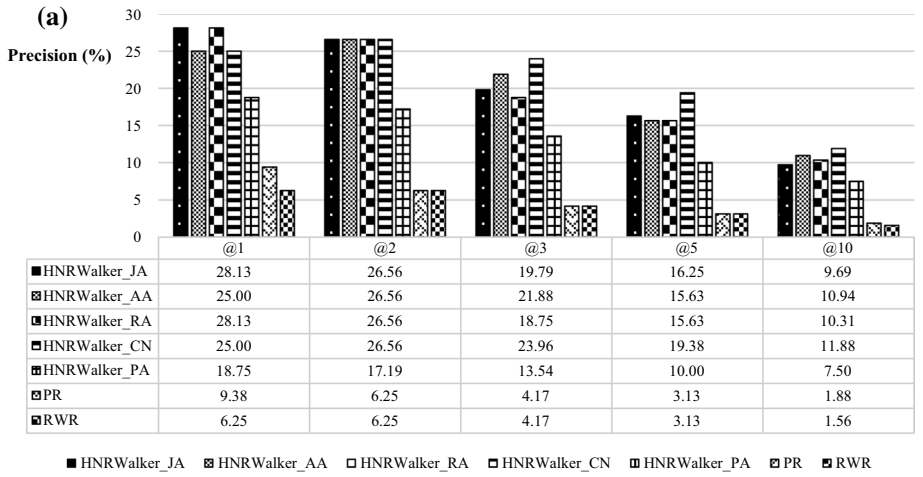
Table 6 Experiment performance (stage 2): comparison of HNRWalker and Node-based Methods

	Precision (%)					AUC (%)				
	@1	@2	@3	@5	@10	@1	@2	@3	@5	@10
HNRWalker_JA	**28.13**	**26.56**	**19.79**	**16.25**	**9.69**	**61.35**	**71.34**	**74.37**	**74.83**	**78.47**
JA	15.63	12.50	15.63	15.00	8.54	37.26	45.88	55.97	65.84	72.19
HNRWalker_AA	**25.00**	**26.56**	**21.88**	**15.63**	**10.94**	**71.88**	**71.96**	**72.76**	**74.99**	**70.45**
AA	9.38	9.38	14.58	13.75	9.18	58.26	48.61	46.64	55.14	63.45
HNRWalker_RA	**28.13**	**26.56**	**18.75**	**15.63**	**10.31**	**62.32**	**70.96**	**77.85**	**73.60**	**73.22**
RA	9.38	7.81	7.29	10.63	8.23	47.43	52.38	55.42	42.15	47.31
HNRWalker_CN	**25.00**	**26.56**	**23.96**	**19.38**	**11.88**	**52.08**	**55.44**	**65.63**	**71.27**	**75.47**
CN	12.50	14.06	15.63	15.00	9.81	41.46	50.20	51.23	58.38	62.48
HNRWalker_PA	**18.75**	**17.19**	**13.54**	**10.00**	**7.50**	50.00	47.51	49.77	55.12	**59.99**
PA	15.63	12.50	9.38	8.75	7.91	**62.81**	**67.51**	**66.63**	**58.73**	48.77

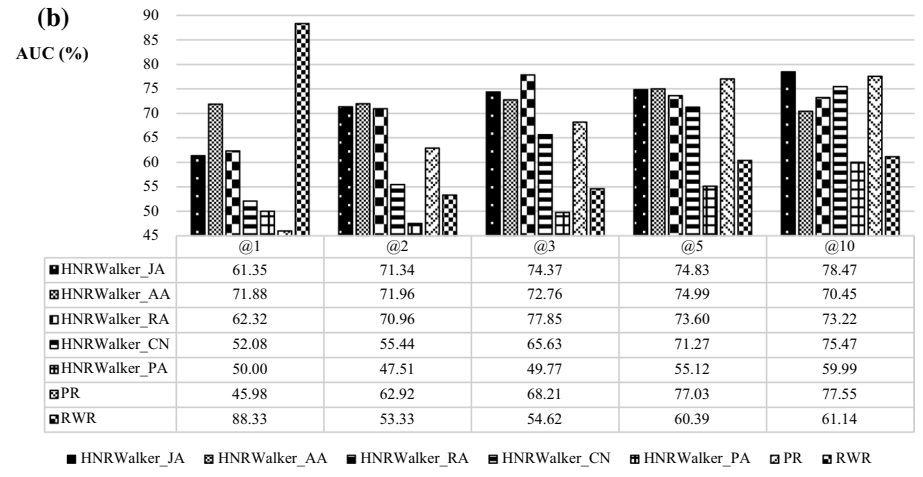
According to Table 5, the HNRWalker_JA algorithm obtains higher precision, which means that it performs significantly better than CNRWalker_JA regardless of the number of scholars recommended. Meanwhile, for the AUC index, HNRWalker_JA also has higher AUC than CNRWalker_JA in total. The results show that introducing the institution network is effective.

According to Table 6, link prediction algorithms such as AA, CN, and RA have poor recommendation effects since their AUC values are sometimes lower than 0.5. However, if the numbers of the collaborators that systems recommended are less than 10, PA gets suitable AUC values. We have reason to believe that if the data is sparse, nodes with higher degrees are more likely to be recommended. It is easy to see that most of the combinational algorithms have better performance compared with the link prediction algorithm, especially if we just recommend a few (one to three) collaborators to each scholar. There are two possible reasons for this result. First, in a sparse data-set, the role of neighbors is more important than other features. Secondly, random walk is based on the topological graph. In this graph, the existence of common neighbors will expand the walking path. Therefore, nodes with more common neighbors have larger advantage during the random walk. Compared with AA, JA, etc., the improvement effect of PA is not as good as others since it considers the node degree but not the common neighbors.

We found that both RWR and PR are much lower in precision value than HNRWalker (Fig. 5). Traditional random walk algorithms cannot fully utilize the network information because of their random mechanism. Besides, they do not obey Markov Chain during the process of iteration and walking, which means that there are evolvements among the visited nodes. But in fact, further study is needed to prove whether this evolution is reasonable. More importantly, the AUC of HNRWalker_JA,



	@1	@2	@3	@5	@10
HNRWalker_JA	28.13	26.56	19.79	16.25	9.69
HNRWalker_AA	25.00	26.56	21.88	15.63	10.94
HNRWalker_RA	28.13	26.56	18.75	15.63	10.31
HNRWalker_CN	25.00	26.56	23.96	19.38	11.88
HNRWalker_PA	18.75	17.19	13.54	10.00	7.50
PR	9.38	6.25	4.17	3.13	1.88
RWR	6.25	6.25	4.17	3.13	1.56



	@1	@2	@3	@5	@10
HNRWalker_JA	61.35	71.34	74.37	74.83	78.47
HNRWalker_AA	71.88	71.96	72.76	74.99	70.45
HNRWalker_RA	62.32	70.96	77.85	73.60	73.22
HNRWalker_CN	52.08	55.44	65.63	71.27	75.47
HNRWalker_PA	50.00	47.51	49.77	55.12	59.99
PR	45.98	62.92	68.21	77.03	77.55
RWR	88.33	53.33	54.62	60.39	61.14

Fig. 5 Experiment performance (stage 3): comparison of HNRWalker, PR, and RWR

HNRWalker_AA, and HNRWalker_RA are higher than 60%, while RWR and PR are not.

Thus, we conclude that: (1) Compared with three types of benchmarks in terms of Precision and AUC indexes, the combination of HNRWalker and JA, AA, RA, and CN have better performance. (2) HNRWalker is suitable to combine with the index that considers common neighbors. For the index focused on the degree of nodes (PA), the combinative effect may not be good, especially in terms of AUC. (3) In general, our purposed HNRWalker_JA had higher effect in Precision and AUC indexes than the typical PR and RWR. This demonstrates that the designed transition probability which combined with the characteristics of heterogeneous networks and the novel walking mechanism plays a powerful role in improving the accuracy of recommendations.

Discussion

Conclusions

This paper makes innovations on both the selection of collaboration features and recommendation methods. Specifically, this paper builds a heterogeneous network with institution and co-author networks. It proves the importance of the institution characteristic for collaborator recommendation and provides a referable method to build heterogeneous networks in recommender systems research. This research identifies four types of neighbor relationships in the heterogeneous network and proposes a set of similarity assessment measures. This set of similarities includes the nodes' co-author information, neighbor information, and institution information. It can effectively measure the similarity between nodes and convert static transitivity into dynamic transitivity. In fact, transition probability is an important part of the random walk algorithm. Some effective improvement toward it can significantly improve the performance of the whole algorithm. Our experiment in stage 3 demonstrated the effectiveness of the improvement, so it can be applied to other walking algorithms. Furthermore, this paper proposes a novel method that obtains the recommendation set by sorting the recommendation opportunities candidates get. This is a new way of sorting candidates. All the validation results show that the proposed approach performs much better than the benchmarks and the fusion of cooperation elements and the improvements based on the random walk algorithm are effective. From the perspective of networks and methods, the proposed HNR-Walker is practical and could be extended to venues recommendation, object classification, and segmentation.

Limitations and future works

This study has several limitations that should be improved in the future. First, data we used are journal papers published in LIS, and two questions are worth further exploration. One is whether there are disciplinary differences in scientific collaboration, such as in biology, chemistry, and linguistics. The other one is that we treated the scholars who belong to multiple institutions as distinct users. Although there are no more than 15 scholars who belong to multiple institutions in our data, we still hope to find a strategy to deal with this problem in the future. Second, only institution and co-author networks are constructed into the heterogeneous network. Other networks such as geographic location have not been considered. How to make use of these relationships and build a more comprehensive and broader heterogeneous network will be the next focus of our research. Third, the method proposed in this paper is mainly based on the neighbor structure, while the semantic relevance of papers and scholars' interests have not been considered. As semantic analysis can mine the transition of scholars' academic interest and recognize their current interests, topic modeling methods and semantic language models will be utilized in future to construct semantic features.

Acknowledgements This research was supported by grants from National Natural Science Foundation of China [71701134], Humanity and Social Science Youth Foundation of Ministry of Education of China [16YJC630153], Guangdong Basic and Applied Basic Research Foundation [2019A1515011392] and Natural Science Foundation of Guangdong Province of China [2017A030310427].

References

Adamic, L. A., & Adar, E. (2003). Friends and neighbors on the web. *Social networks, 25*(3), 211–230.

Alshareef, A. M., Alhamid, M. F., & El Saddik, A. (2018). Recommending Scientific Collaboration Based on Topical, Authors and Venues Similarities. In *Paper presented at the 2018 IEEE international conference on information reuse and integration (IRI)* (pp. 55–61). IEEE.

Bergé, L. R. (2017). Network proximity in the geography of research collaboration. *Papers in Regional Science, 96*(4), 785–815.

Bornmann, L., & Leydesdorff, L. (2015). Topical connections between the institutions within an organisation (institutional co-authorships, direct citation links and co-citations). *Scientometrics, 102*(1), 455–463.

Brandao, M. A., & Moro, M. M. (2012). Affiliation influence on recommendation in academic social networks. In *Paper presented at the AMW* (pp. 230–234).

Chaiwanarom, P., & Lursinsap, C. (2015). Collaborator recommendation in interdisciplinary computer science using degrees of collaborative forces, temporal evolution of research interest, and comparative seniority status. *Knowledge-Based Systems, 75,* 161–172.

Chuan, P. M., Ali, M., Khang, T. D., & Dey, N. (2018). Link prediction in co-authorship networks based on hybrid content similarity metric. *Applied Intelligence, 48*(8), 2470–2486.

Cohen, S., & Ebel, L. (2013). Recommending collaborators using keywords. In *Paper presented at the proceedings of the 22nd international conference on World Wide Web* (pp. 959–962). ACM.

Davoodi, E., Kianmehr, K., & Afsharchi, M. (2013). A semantic social network-based expert recommender system. *Applied Intelligence, 39*(1), 1–13.

Du, L., Li, C., Chen, H., Tan, L., & Zhang, Y. (2015). Probabilistic SimRank computation over uncertain graphs. *Information Sciences, 295,* 521–535.

Fang, W., Yang, G., & Hu, Z. (2018). An improved DV-Hop algorithm with Jaccard coefficient based on optimization of distance correction. In *Paper presented at the international conference on bio-inspired computing: theories and applications* (pp. 457–465). Springer.

Guo, Y., & Chen, X. (2014). Cross-domain scientific collaborations prediction with citation information. In *Paper presented at the 2014 IEEE 38th international computer software and applications conference workshops* (pp. 229–233). IEEE.

Hoang, D. T., Nguyen, N. T., Tran, V. C., & Hwang, D. (2019). Research collaboration model in academic social networks. *Enterprise Information Systems, 13*(7–8), 1023–1045.

Hoekman, J., Frenken, K., & Van Oort, F. (2009). The geography of collaborative knowledge production in Europe. *The Annals of Regional Science, 43*(3), 721–738.

Huang, J., & Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on knowledge Data Engineering, 17*(3), 299–310.

Huynh, T., Hoang, K., & Lam, D. (2013). Trend based vertex similarity for academic collaboration recommendation. In *Paper presented at the international conference on computational collective intelligence* (pp. 11–20). Springer.

Jung, J., Shin, K., Sael, L., & Kang, U. (2016). Random walk with restart on large graphs using block elimination. *ACM Transactions on Database Systems, 41*(2), 1–43. https://doi.org/10.1145/2901736.

Khan, S., Liu, X., Shakil, K. A., & Alam, M. (2017). A survey on scholarly data: From big data perspective. *Information Processing Management, 53*(4), 923–944.

Lee, J., Oh, S., Dong, H., Wang, F., & Burnett, G. (2019). Motivations for self-archiving on an academic social networking site: A study on researchgate. *Journal of the Association for Information Science Technology, 70*(6), 563–574.

Li, Z., Liang, X., Zhou, X., Zhang, H., & Ma, Y. (2016). A link prediction method for large-scale networks. *Chinese Journal of Computers, 39*(42), 1–18.

Li, S., Song, X., Lu, H., Zeng, L., Shi, M., & Liu, F. (2020). Friend recommendation for cross marketing in online brand community based on intelligent attention allocation link prediction algorithm. *Expert Systems with Applications, 139,* 112839. https://doi.org/10.1016/j.eswa.2019.112839.

Li, J., Xia, F., Wang, W., Chen, Z., Asabere, N. Y., & Jiang, H. (2014). Acrec: a co-authorship based random walk model for academic collaboration recommendation. In *Paper presented at the proceedings of the 23rd international conference on World Wide Web* (pp. 1209–1214). ACM.

Liu, Z., & Jansen, B. J. (2017). Identifying and predicting the desire to help in social question and answering. *Information Processing Management, 53*(2), 490–504.

Luong, N. T., Nguyen, T. T., Jung, J. J., & Hwang, D. (2015). Discovering co-author relationship in bibliographic data using similarity measures and random walk model. In *Paper presented at the Asian conference on intelligent information and database systems* (pp. 127–136). Springer.

Mahapatra, R., Samanta, S., Pal, M., & Xin, Q. (2019). RSM index: A new way of link prediction in social networks. *Journal of Intelligent Fuzzy Systems* (Preprint), pp. 1–15.

Makarov, I., Bulanov, O., & Zhukov, L. E. (2016). Co-author recommender system. In *Paper presented at the international conference on network analysis* (pp. 251–257). Springer.

Montefusco, A. M., do Nascimento, F. P., Sennes, L. U., Bento, R. F., & Imamura, R. (2019). Influence of international authorship on citations in Brazilian medical journals: a bibliometric analysis. *Scientometrics, 119*(3), 1487–1496.

Ortega, J. L., & Aguillo, I. F. (2013). Institutional and country collaboration in an online service of scientific profiles: Google Scholar Citations. *Journal of Informetrics, 7*(2), 394–403.

Ostroumova Prokhorenkova, L., & Samosvat, E. (2016). Recency-based preferential attachment models. *Journal of Complex Networks, 4*(4), 475–499.

Samanthula, B. K., & Jiang, W. (2015). Secure multiset intersection cardinality and its application to jaccard coefficient. *IEEE Transactions on Dependable, 13*(5), 591–604.

Shi, B., Ifrim, G., & Hurley, N. (2016). Learning-to-rank for real-time high-precision hashtag recommendation for streaming news. In *Paper presented at the proceedings of the 25th international conference on World Wide Web* (pp. 1191–1202). International World Wide Web Conferences Steering Committee.

Song, R., Xu, H., & Cai, L. (2019). Academic collaboration in entrepreneurship research from 2009 to 2018: A multilevel collaboration network analysis. *Sustainability, 11*(19), 5172. https://doi.org/10.3390/su11195172.

Sun, Y., & Han, J. (2013). Meta-path-based search and mining in heterogeneous information networks. *Tsinghua Science Technology, 18*(4), 329–338.

Sun, N., Lu, Y., & Cao, Y. (2019). Career age-aware scientific collaborator recommendation in scholarly big data. *IEEE Access, 7,* 136036–136045.

Symeonidis, P., & Perentis, C. (2014). Link prediction in multi-modal social networks. In *Paper presented at the joint European conference on machine learning and knowledge discovery in databases* (pp. 147–162). Springer.

Valdeolivas, A., Tichit, L., Navarro, C., Perrin, S., Odelin, G., Levy, N., et al. (2018). Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics, 35*(3), 497–505.

Wang, X., Fang, Z., & Sun, X. (2016). Usage patterns of scholarly articles on web of science: A study on web of science usage count. *Scientometrics, 109*(2), 917–926.

Weaver, I. S. (2015). Preferential attachment in randomly grown networks. *Physica A: Statistical Mechanics its Applications, 439,* 85–92.

Wu, J., Zhang, G., & Ren, Y. (2017). A balanced modularity maximization link prediction model in social networks. *Information Processing Management, 53*(1), 295–307.

Xia, F., Chen, Z., Wang, W., Li, J., & Yang, L. T. (2014). MVCWalker: Random walk-based most valuable collaborators recommendation exploiting academic factors. *IEEE Transactions on Emerging Topics in Computing, 2*(3), 364–375.

Xiao, Y., Li, X., Wang, H., Xu, M., & Liu, Y. (2018). 3-HBP: A three-level hidden Bayesian link prediction model in social networks. *IEEE Transactions on Computational Social Systems, 5*(2), 430–443.

Xie, Z., Ouyang, Z., Li, J., Dong, E., & Yi, D. (2018). Modelling transition phenomena of scientific coauthorship networks. *Journal of the Association for Information Science Technology, 69*(2), 305–317.

Yan, E., & Guns, R. (2014). Predicting and recommending collaborations: An author-, institution-, and country-level analysis. *Journal of Informetrics, 8*(2), 295–309.

Yao, L., Wang, L., Pan, L., & Yao, K. (2016). Link prediction based on common-neighbors for dynamic social network. *Procedia Computer Science, 83,* 82–89.

Zahr, N., Arnaud, L., Marquet, P., Haroche, J., Costedoat-Chalumeau, N., Hulot, J. S., et al. (2010). Mycophenolic acid area under the curve correlates with disease activity in lupus patients treated with mycophenolate mofetil. *Arthritis Rheumatism, 62*(7), 2047–2054.

Zarrinkalam, F., Kahani, M., & Bagheri, E. (2018). Mining user interests over active topics on social networks. *Information Processing Management, 54*(2), 339–357.

Zhang, J. (2017). Uncovering mechanisms of co-authorship evolution by multirelations-based link prediction. *Information Processing Management, 53*(1), 42–51.

Zhao, T., Xiao, R., Sun, C., Chen, H., Li, Y., & Li, H. (2014). Personalized recommendation algorithm integrating roulette walk and combined time effect. *Journal of Computer Applications, 34*(4), 1114–1117.

Zhou, X., Ding, L., Li, Z., & Wan, R. (2017). Collaborator recommendation in heterogeneous bibliographic networks using random walks. *Information Retrieval Journal, 20*(4), 317–337.

Zhou, T., Lü, L., & Zhang, Y.-C. (2009). Predicting missing links via local information. *The European Physical Journal B, 71*(4), 623–630.