



# Graph Embedding for Scholar Recommendation in Academic Social Networks

Chengzhe Yuan<sup>1</sup>, Yi He<sup>2</sup>, Ronghua Lin<sup>2</sup> and Yong Tang<sup>2\*</sup>

<sup>1</sup>School of Electronics and Information, Guangdong Polytechnic Normal University, Guangzhou, China, <sup>2</sup>School of Computer Science, South China Normal University, Guangzhou, China

The academic social networks (ASNs) play an important role in promoting scientific collaboration and innovation in academic society. Accompanying the tremendous growth of scholarly big data, finding suitable scholars on ASNs for collaboration has become more difficult. Different from friend recommendation in conventional social networks, scholar recommendation in ASNs usually involves different academic entities (e.g., scholars, scientific publications, and status updates) and various relationships (e.g., collaboration relationship between team members, citations, and co-authorships), which forms a complex heterogeneous academic network. Our goal is to recommend potential similar scholars for users in ASNs. In this article, we propose to design a graph embedding-based scholar recommendation system by leveraging academic auxiliary information. First, we construct enhanced ASNs by integrating two types of academic features extracted from scholars' academic information with original network topology. Then, the refined feature representations of the scholars are obtained by a graph embedding framework, which helps the system measure the similarity between scholars based on their representation vectors. Finally, the system generates potential similar scholars for users in ASNs for the final recommendation. We evaluate the effectiveness of our model on five real-world datasets: SCHOLAT, Zhihu, APS, Yelp and Gowalla. The experimental results demonstrate that our model is effective and achieves promising improvements than the other competitive baselines.

**Keywords:** academic social networks, recommendation system, scholarly data, graph embedding, academic features

## 1 INTRODUCTION

Recent years have witnessed the fast-growing scholarly big data [1,2]. Against this background, academic social networks (ASNs) systems have aroused widespread attention; these systems provide scholars with an integrated platform to share their academic achievements and interact and collaborate with other scholars [3,4]. As a particular type of social networking, ASNs usually involve different academic entities and relationships. In ASNs, scientific collaboration plays an important role in promoting research and innovation. Scholar recommendation aims to help scholars in ASNs discover potential collaborators by measuring the correlation between scholars. Some research shows that collaboration is more likely to be undertaken between similar scholars [5].

Some studies explore scholarly recommendations in various real-world ASNs. For example, in [6], a user-based experimental approach to find experts on ResearchGate was proposed. Collaborator

## OPEN ACCESS

### Edited by:

Shudong Li,  
Guangzhou University, China

### Reviewed by:

Satyam Mukherjee,  
Shiv Nadar University, India  
Yun Yang,  
Swinburne University of Technology,  
Australia

### \*Correspondence:

Yong Tang  
ytang@m.scnu.edu.cn

### Specialty section:

This article was submitted to  
Social Physics,  
a section of the journal  
Frontiers in Physics

**Received:** 31 August 2021

**Accepted:** 18 October 2021

**Published:** 29 November 2021

### Citation:

Yuan C, He Y, Lin R and Tang Y (2021)  
Graph Embedding for Scholar  
Recommendation in Academic  
Social Networks.  
Front. Phys. 9:768006.  
doi: 10.3389/fphy.2021.768006

recommendation on ScholarMate is based on network topology [7,8] or academic information [9,10]. However, recommending suitable scholars in ASNs is not a trivial task. As pointed out by [2], ASNs are complex heterogeneous networks, which contain multiple types of nodes (e.g., scholars and papers) and links (e.g., citations and co-authorships). Therefore, how to characterize this heterogeneous academic information in scholarly recommendation system becomes very important.

Previous works have tried to integrate various academic information (e.g., different entity and relationship data) in conventional recommendation systems. However, these methods usually lead to biased recommendation results and fail to apply in large-scale networks [11]. Recently, studies [12,13] have proved the success of graph embedding models in heterogeneous network-based recommendation systems, as they are able to learn the latent features of nodes in large-scale networks, which in turn facilitates recommendations. The effectiveness of graph embedding-based models has been applied in various recommendation scenarios, such as movie recommendations [14] and POI recommendations [15].

Different from friend recommendations in conventional social networks, we aim to recommend potential similar scholars to users in heterogeneous academic social networks. In this article, we propose a graph embedding-based scholar recommendation approach in ASNs called GESRec. We break down the scholar recommendation process of our model into three stages. First, we construct the enhanced academic social networks by combining two types of academic features (e.g., attributes features and textual features) from scholar's academic information with user-user relationships, and the correlations between the scholar's academic features are calculated by the similarity of the corresponding embedding vectors. Then, the refined feature representations of the scholars are obtained from the enhanced ASNs by the graph embedding framework. Finally, top- $n$  potential scholars are generated based on the final recommendation. The main contributions of this work can be summarized as follows:

- We study an important yet challenging problem of recommending scholars in heterogeneous academic social networks that pose challenges beyond existing friend recommendation systems.
- We propose a graph embedding-based scholar recommendation system in ASNs by taking advantage of academic auxiliary information. The enhanced ASNs module integrates two types of academic features from scholars' academic information with user-user relationships, and the refined feature representations of the scholars are further obtained by graph embedding framework for scholar recommendation task.
- We conduct extensive experiments on three real-world datasets: SCHOLAT, Zhihu, and APS. The experimental results demonstrate that the proposed GESRec is effective and achieves promising improvements than the other competitive baselines.

The remainder of this article is organized as follows. **Section 2** introduces the related work. **Section 3** provides the problem

formulation and components of our model in detail. Then, experimental evaluations and detailed analysis are discussed in **Sections 4, 5**. Finally, the conclusions and the future work are presented in **Section 6**.

## 2 RELATED WORK

In this section, we review several existing studies related to our work. They are generally classified into three types: 1) academic social networks, 2) scholar recommendation, and 3) graph embedding-based recommendation.

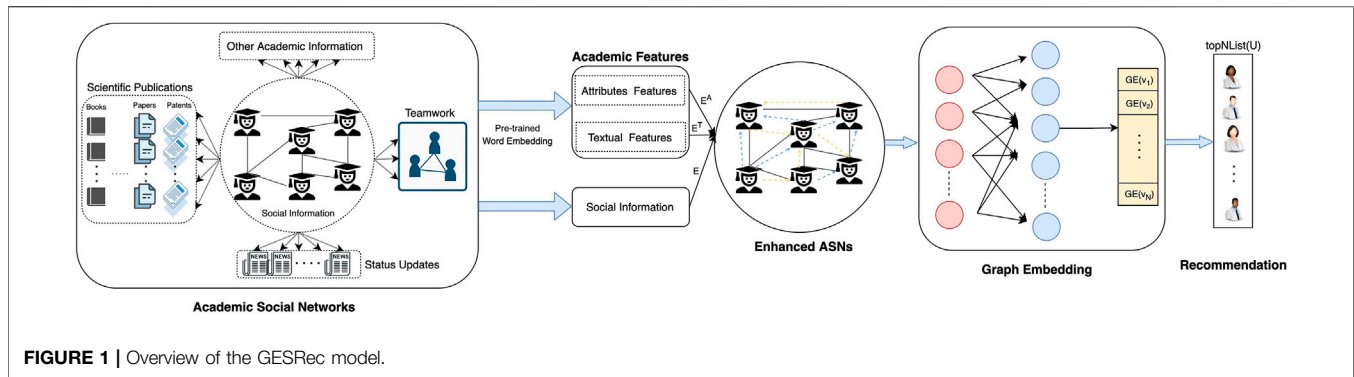
### 2.1 Academic Social Networks

In recent years, academic social networks have received widespread attention due to the large volume of scholarly big data, e.g., entities (publications, scholars, etc.) and their relationships (citations, co-authorships, etc.) [2]. ASNs provide various research topics, such as community detection [16,17], knowledge sharing [18,19], and recommendation systems [20,21]. A line of research focuses on making use of ASNs for particular recommendations [21]. A course recommendation model in ASNs was proposed, which combines association rules algorithm with an improved multi-similarity algorithm of multi-source information. By constructing a weighted co-authorship network [22], a method called VOPRec is proposed, which utilizes the node embedding technology to explore text information and network structure information for papers recommendation. Overall, most of these studies focus on analyzing ASNs for the specific type of recommendation services.

### 2.2 Scholar Recommendation

Scholar recommendation is one of the main tasks in the scholarly recommendation, which can be useful in finding similar scholars for potential collaborations. A related research topic is the expert recommendation, which means trying to detect the most knowledgeable people in some specific topics [23]. For example, in [24], an expert recommendation system based on the Pearson correlation coefficient and the FP-growth association rule was proposed. A multilevel profile-based expert finding method for expert recommendation in online scientific communities is proposed in [25]. Some works focus on constructing expert recommendation systems based on social networks [26–28]. Besides, in [9], a context-aware researcher recommendation system to encourage university-industry collaboration on industrial R&D projects was introduced [6]. Moreover, a pathway-based approach was proposed that takes into consideration both pages and navigations, which aims to identify the right experts with relevant expertise or experience for a given topic.

There are some previous works that are similar to our study. For example, in [29], a research-fields-based graph in ASNs was built, and a community-based scholar recommendation model was proposed [30]. The previous study was extended by processing the problem of subset community through the GraphChi framework in parallel and recommending the



scholars within the community according to the relevant recommendation rules [10]. Moreover, a heterogeneous network-based scholar recommendation approach was designed by integrating researchers' characteristic and relationship information [11]. Recommending potential friends for scholars in ASNs was proposed by considering both network topology and scholars' academic information.

As mentioned above, previous scholar recommendation methods usually fail to fully explore the auxiliary information in ASNs. In this article, we enhance the academic social networks with full auxiliary information for scholar recommendations.

## 2.3 Graph Embedding-Based Recommendation

Graph embedding techniques have attracted a great deal of attention. A line of research attempts to apply graph embedding in recommendation systems. One of the advantages of graph embedding is the ability to learn the low-dimensional representations of nodes in large-scale networks [10]. Graph embedding has been exploited in heterogeneous information networks for various recommendation scenarios, such as “co-author recommendation,” “social recommendation,” and “movie recommendation” [13]. A heterogeneous co-occurrence network was constructed to learn the latent representations of users and items. Some works focus on using meta-path-aware similarity between user and item for recommending items [31–33]. In [15], a graph embedding-based model for POI recommendation was proposed, by decomposing the high-order interrelationships among users, POIs and items to a set of pairwise interaction relationships to address data sparsity and cold start problems.

Different from the above methods, our study focuses on scholar recommendation in ASNs and we build enhanced ASNs by considering two types of academic features of scholars (attributes features and textual features), which in turn can be fully explored by graph embedding. Moreover, our work is complementary to the scholar recommendation task as our model can be extended to other graph embedding-based frameworks to recommend scholars in different scenarios.

## 3 METHODS

In this section, first, we present the formulation of the scholar problem in ASNs. Then, we introduce the details of the proposed model: Graph Embedding for Scholar Recommendation (GESRec) model in **Figure 1**, including enhancing academic social networks with auxiliary information modules and graph embedding for scholar recommendation framework. Finally, we show the overall algorithm and details.

### 3.1 Problem Formulation

We denote the academic social networks as  $G = (V, E, F)$ , where  $V = \{v_1, v_2, \dots, v_{|V|}\}$  denotes the set of scholars,  $E = \{e_{ij} | v_i \in V, v_j \in V\}$ , and the edge  $e_{ij}$  between nodes  $v_i$  and  $v_j$  indicates that there are academic connections between these scholars.  $f_i = \{x_1, x_2, \dots, x_k\}$  represents the scholar  $v_i$  that contains a set of  $k$  academic features. We define the set of academic features shared by all scholars together constituted as  $F = \bigcup_{v_i \in V} f_i$ . We define the correlation between scholars  $v_i$  and  $v_j$  as  $p(v_i, v_j) = \text{sim}(w_i, w_j)$ , where  $w_i$  denotes the embedding vector of scholar  $v_i$  learned by graph embedding technique and  $\text{sim}(\bullet)$  means an assigned similarity calculation method.

The scholar recommendation in academic social networks problem takes the scholars  $U = \{u_1, u_2, \dots, u_{|U|}\}$  and academic social networks  $G = (V, E, F)$  as input and then calculates the correlation  $p(u_i, u_j | u_i \in U, u_j \in U, i \neq j)$  by iterations through all scholars in  $U$ . It outputs top- $n$  similar scholars list for each scholar, where each list is ranked by the calculated correlation  $p(\bullet)$ .

### 3.2 Enhancing Academic Social Networks With Auxiliary Information

Different from other social networks platforms (e.g., Twitter and Facebook), academic social networks usually contain various types of academic data. Some recent researches have pointed out that additional academic data can be used as auxiliary information to extract better academic relationships between scholars and improve scholar recommendation performances [18]. In academic social networks, we divide the academic features into attributes features and textual features. In particular, attribute features include scholar's biographical information, research interests, and collaborative team;

these features can usually be expressed as lists of academic terminology. Textural features include a group of the scholar's academic achievements, such as scientific papers, patents, and research projects, and these features usually contain the title, abstract and other textual information of the academic achievement. In this study, we process the two features in separate ways, and the extracted academic relationships networks  $E'$  are added to the original user-user networks  $E$  to jointly construct enhanced academic social networks  $E^C = E' \cup E$ , and  $G' = (V, E^C, F)$  is used for subsequent scholar recommendations.

Academic features can be expressed as the intuitive representation of the scholar's academic information, which is represented in the form of terminology vocabulary. To better mining the underlying semantics of these academic features, we utilize a pre-trained word embedding model [34] to get the academic features embedding vectors, which are further processed to calculate the similarity among scholars. The set of academic features  $f_i = \{a_1, a_2, \dots, a_k\}$  means that the scholar  $u_i$  contains  $k$  academic features, which are divided into attribute features  $f_i^a$  and text features  $f_i^t$ .

For the attribute features, we get the representation vectors by the pre-trained word embedding model and then calculate the correlation between these vectors by the cosine similarity. Meanwhile, in order to avoid heavy computational burdens, we set a threshold  $\theta_t$  to filter low correlations, and for reducing the centrality of high-frequency scholars, we also set a top- $n$  list to balance the number of scholars. Besides, as to prevent the effect of differences in the number of scholar's attribute features, we define the correlation between scholar's attribute features as follows:

$$attrSim(u_i, u_j) = \max(\cosSim(emb_{x_m}, emb_{x_n}) \mid x_m \in f_i^a, x_n \in f_j^a) \tag{1}$$

$$e_{ij}^a = \begin{cases} attrSim(u_i, u_j), & \text{if } attrSim(u_i, u_j) \geq \theta_a \text{ and } attrSim(u_i, u_j) \text{ in } topk_a \\ 0, & \text{else} \end{cases} \tag{2}$$

where  $f_i^a$  and  $f_j^a$  denote the set of attributes features for scholars  $u_i$  and  $u_j$ ,  $emb_{x_m}$  is the embedding vector representation of attribute feature  $x_m$  obtained by the pre-trained word vector model. By computing the similarity values of the embedding vectors between two sets of attribute features ( $f_i^a$  and  $f_j^a$ ), we take the maximum value as the similarity between attribute features  $attrSim(u_i, u_j)$ . If the value of  $attrSim(u_i, u_j)$  is greater than the threshold  $\theta_a$ , and in the top- $k_a$  list of similarity between attribute features, we add a new edge  $e_{ij}^a$  between scholar's attribute features correlation.

For the textual features, we process them by cleaning, tokenization, and stemming and combine all the processed text features. Similar to the attribute features, we also get the correlation between textual features by the cosine similarity of the corresponding representation vectors. We define the correlation between scholar's textual features as follows:

$$textSim(u_i, u_j) = \cosSim(emb_{t_i}, emb_{t_j}), t_i = concat(f_i^t) \tag{3}$$

$$e_{ij}^t = \begin{cases} textSim(u_i, u_j), & \text{if } textSim(u_i, u_j) \geq \theta_t \text{ and } textSim(u_i, u_j) \text{ in } topk_t \\ 0, & \text{else} \end{cases} \tag{4}$$

where  $f_i^t$  and  $f_j^t$  denote the set of textual features for scholars  $u_i$  and  $u_j$ . By combining all the text features  $f_i^t$  of the scholar  $u_i$ , we get a document-level textual features  $t_i$ ,  $textSim(u_i, u_j)$  means the cosine similarity between attribute features. We also use thresholds  $\theta_t$  and  $k_t$  to determine whether to add an edge  $e_{ij}^t$  between scholar's textual features correlation.

By combining the multiple academic relationships between scholars with user-user connection, we define the enhanced academic social networks as follows:

$$E^C = E \cup E^A \cup E^T \tag{5}$$

$$G' = (V, E^C, W) \tag{6}$$

where  $E^A = \{e_{ij}^a \mid u_i \in U, u_j \in U\}$  represents the set of scholar's attribute features correlations,  $E^T$  denotes the set scholar's textual features correlations, and  $G'$  is a directed weighted academic social networks.

### 3.3 Graph Embedding for Scholar Recommendation

In order to utilize multiple academic relationships for better scholar recommendation performance, we try to adopt the graph embedding method to learn the latent feature representations of the scholars for scholar recommendation. As we get enhanced academic social networks  $G' = (V, E^C, F)$  from previous section, we aim to learn a low-dimensional embedding  $\mathbf{v} \in \mathbb{R}^d$  for each node  $v \in V$ , where  $d \ll |V|$  is the dimension of the representation space. The learned embeddings are considered to contain informative academic characteristics, which are useful in the scholar recommendation task. The whole algorithm framework is shown in Algorithm 1. First, we obtain two types of scholar's academic features (attributes features  $f_i^a$  and textual features  $f_i^t$ ). Second, we calculate the corresponding correlation between scholar's attribute features  $attrSim(u_i, u_j)$  and textual features  $textSim(u_i, u_j)$ . Third, we construct enhanced ASNs by combining the multiple academic relationships between scholars with user-user connections in **Eqs 5, 6**. Then, we get the refined feature representations  $\mathbf{v}_i$  of the scholars by the graph embedding framework. Finally, the top- $n$  list of ( $sorSimList(U)$ ) similar scholars is recommended to the user  $u_i$  as the potential recommendation list.

For each node  $v \in V$ , we construct the neighborhood  $\mathcal{N}_v$  by the skip-gram with negative sampling strategy. We formulate the overall objective of representation learning in our network as follows:

$$\max_f \sum_{v \in V} \log \Pr(\mathcal{N}_v \mid f(v)) \tag{7}$$

where  $f: V \rightarrow \mathbb{R}^d$  is a mapping function from node to  $d$ -dimensional representation space and  $\mathcal{N}_v \subset V$  is the neighborhood of node  $v$ . Following LINE [35], we define  $\Pr(\mathcal{N}_v \mid f(v))$  by the softmax function as follows.

**Algorithm 1.** Graph embedding for scholar recommendation model.

```

Input : the scholars  $U$ ; the academic social networks  $G = (V, E, A)$ ; the attribute features threshold  $\theta_a, K_a$ ; the textual features threshold  $\theta_t, K_t$ 
Output :  $topNList(u_i) = u_{i_1}, u_{i_2}, \dots, u_{i_n}, u_i \in U$ 
1 for  $u_i \in U$  do
2    $f_i = \{a_1, a_2, \dots, a_k | a_i \in A\} \leftarrow u_i$ ;
3    $\{f_i^a, f_i^t\} \leftarrow f_i$ ;
4   for  $x_m \in f_i^a$  do
5      $emb_{x_m} \leftarrow x_m$ ;
6   end
7   for  $x_n \in f_i^t$  do
8      $emb_{x_n} \leftarrow x_n$ ;
9   end
10 end
11 for  $u_i \in U$  do
12   for  $u_j \in U$  do
13      $attrSim(u_i, u_j) \leftarrow \langle u_i, u_j \rangle$ ;
14     if  $attrSim(u_i, u_j) \geq \theta_a$  &  $attrSim(u_i, u_j)$  in  $topK_a$  then
15        $e_{ij}^a \leftarrow \langle v_i, v_j \rangle, w_{ij}^a = attrSim(u_i, u_j)$ ;
16     end
17      $textSim(u_i, u_j) \leftarrow \langle u_i, u_j \rangle$ ;
18     if  $textSim(u_i, u_j) \geq \theta_t$  &  $textSim(u_i, u_j)$  in  $topK_t$  then
19        $e_{ij}^t \leftarrow \langle v_i, v_j \rangle, w_{ij}^t = textSim(u_i, u_j)$ ;
20     end
21   end
22 end
23  $E^C = \{E \cup E^A \cup E^T\}$ ;
24  $w_{ij}^c = w_{ij}^a + w_{ij}^t + w_{ij}^f$ ;
25  $G^c = \{V, E^C, W\}$ ;
26 for  $v_i \in V$  do
27    $v_i = GE(v_i)$ ; // Graph embedding for vector representation of  $v_i$ 
28    $v_i \leftarrow u_i$ ;
29 end
30 for  $u_i \in U$  do
31   for  $u_j \in U$  do
32      $Sim(u_i, u_j) = Sim(vec_i, vec_j)$ ;
33   end
34    $sortSimList(u_i) \leftarrow Sim(u_i, u_j | u_j \in U, i \neq j)$ ;
35 end
36 return  $topN(sortSimList(U))$ ;

```

$$\Pr(\mathcal{N}_v | f(v)) = \frac{\exp(f(\mathcal{N}_v) \cdot f(v))}{\sum_{c \in V} \exp(f(c) \cdot f(v))} \quad (8)$$

Considering the heavy cost of computing when the number of nodes in the network is relatively large, we employ negative sampling [36] to maximize the log probability of the softmax as follows:

$$\log \sigma(f(\mathcal{N}_v) \cdot f(v)) + \sum_{m=1}^M \mathbb{E}_{v_m \sim P(v)} [\log \sigma(-f(v_m) \cdot f(v))] \quad (9)$$

where  $M$  is the number of negative samples and  $\sigma(\bullet)$  denotes the sigmoid function.  $P(V)$  is the noise distribution of nodes. We apply stochastic gradient descent (SGD) to optimize this objective (9).

Based on the graph embedding module, we get the low-dimensional embedding of all scholars  $U$ . We calculate the similarity between two scholars based on the cosine similarity of these two low-dimensional embedding vectors as follows:

$$Sim(u_i, u_j) = \frac{\mathbf{v}_i \times \mathbf{v}_j}{\|\mathbf{v}_i\| \|\mathbf{v}_j\|} \quad (10)$$

Finally, the top- $n$  similar scholars are recommended to the selected scholar as the potential recommendation list.

### 4 EXPERIMENT SETUP

In this section, we first describe the statistics of five real-world datasets. Then, we describe the evaluation metrics used for experimental results. Finally, we conduct detailed experiments and a case study to demonstrate the effectiveness of the proposed

model. The following subsections introduce the design of the experiment setup.

### 4.1 Datasets

We conduct experiments on five widely used real-world datasets: SCHOLAT, Zhihu, APS, Yelp, and Gowalla. These datasets are publicly accessible, real-world data with various sizes, sparsity and domains (e.g., academic social networks domain: SCHOLAT, Zhihu, and APS; popular social networks domain: Yelp and Gowalla). The statistics of the datasets are listed in Table 1.

- SCHOLAT<sup>1</sup> is an emerging vertical academic social networking system designed and built specifically for researchers in China. The main goal of SCHOLAT is to enhance collaboration and social interactions focused on scholarly and learning discourses among the community of scholars. In addition to social networking capabilities, SCHOLAT also incorporates various modules to encourage collaborative and interactive discussions, for example, chat, email, events, and news posts. We constructed the SCHOLAT dataset based on [37]. More specifically, first, we only select scholars with more than ten friends to ensure that the relationship network is not too sparse. Then, we clean and desensitize these data. Finally, we collected 19,841 scholars and 63,686 friends relations between them. We also extracted 3,309 academic attributes and 67,462 textual features for the scholars.
- Zhihu<sup>2</sup> is China’s most popular Q&A platform. Users ask questions, express their opinions on different issues, and share, exchange, and discuss knowledge. We use the same version of the Zhihu dataset from [38], which contains 10,000 active users and 43,894 friends relations between them, and the context describing the user’s topic of interest is used as the textual features. Since Zhihu is mainly a platform for users to ask and answer questions, users do not usually have academic attributes features.
- APS<sup>3</sup> (American Physical Society, APS) is a non-profit academic organization, which aims to promote the development of research in physics through academic journals, scientific conferences, and exhibitions. The APS database contains over 600,000 papers from 18 core physics journals, including paper metadata and citations. We select PRA (Physical Review A: Atomic, Molecular, and Optical Physics) journals to construct the APS dataset. First, we process the corresponding JSON metadata to extract co-author relationships between the authors as friend relationships. Then, we extract the authors’ academic attributes by the glove.6B.200d<sup>4</sup> pre-trained word vector, which is obtained from the English corpus using the GloVe model [39]. In order to ensure that the relationship network is not too sparse, we also only select scholars with more than

<sup>1</sup><https://www.scholat.com/research/opendata/>

<sup>2</sup><https://www.zhihu.com/>

<sup>3</sup><https://www.aps.org/index.cfm>

<sup>4</sup><https://www.kaggle.com/incorpes/glove6b200d>

**TABLE 1** | Statistics of the datasets.

|                      | SCHOLAT | Zhihu  | APS     | —             | Yelp      | Gowalla   |
|----------------------|---------|--------|---------|---------------|-----------|-----------|
| #Users               | 19,841  | 10,000 | 14,279  | #Users        | 45,919    | 29,858    |
| #Relationships       | 63,686  | 43,894 | 446,685 | #Items        | 45,538    | 40,981    |
| #Attributes features | 3,309   | \      | 6,221   | #Interactions | 1,185,065 | 1,027,370 |
| #Textual features    | 67,462  | 10,000 | 81,088  | Density       | 0.056%    | 0.084%    |

ten friends. Overall, we collect 14,279 scholars and 446,685 co-authorship relations between them. We also extract 6,221 academic attributes and 81,088 textual features for the scholars.

- Yelp<sup>5</sup> is a popular online directory for local business reviewing and social networking sites. The Yelp dataset is a subset of Yelp's businesses, reviews, and user data. This dataset consists of 16,239 users, 14,282 businesses, and 198,397 ratings ranging from 1 to 5. In addition, this dataset contains social relations and attribute information of businesses.
- Gowalla<sup>6</sup> is a location-based social networking website where users share their locations by checking in. Same as in [40], this dataset consists of 29,858 users, 40,981 items, and 1,027,370 interaction.

For each dataset, we train the academic attributes by pre-trained word vector models to obtain the corresponding embedding vectors. We pre-process the textual features by tokenization, lower-casing, and stemming, and then we transform the pre-processed textual features into the corresponding feature vectors by the pre-trained word vector models. We randomly split all the datasets into training, validation, and testing sets by the partition 80%:10%:10%, respectively.

## 4.2 Evaluation Metrics

In this researcher recommendation task, given a scholar, a practical recommendation model generates a top- $n$  ranked list of researchers. We evaluate the performance of our model and other baseline models by four evaluation metrics: Precision@N, Recall@N, F1 score, and NDCG@N. The first three evaluation metrics are formulated as follows:

$$\text{Precision@N} = \frac{1}{m} \sum_{i=1}^m \frac{R_a \cap T_a}{R_a} \quad (11)$$

$$\text{Recall@N} = \frac{1}{m} \sum_{i=1}^m \frac{R_a \cap T_a}{T_a} \quad (12)$$

$$\text{F1@N} = \frac{2 * \text{Precision@N} * \text{Recall@N}}{\text{Precision@N} + \text{Recall@N}} \quad (13)$$

where  $m$ ,  $R_a$ , and  $T_a$  represent the number of scholars, the predicted list of items, and the ground-truth items associated with scholar  $i$ , respectively. F1 score is the harmonic mean of

Precision and Recall. Besides, we also adopt NDCG@N (Normalized Discounted Cumulative Gain) as the evaluation metric to judge the quality of the top- $n$  ranking list.

## 4.3 Comparison Methods

To verify the effectiveness of our proposed model, we employ the following methods as baselines:

- ItemPop: ItemPop is a simple method that recommends the most popular items. Items are ranked based on the observed frequency of each aspect in this item's historical reviews. All scholars are recommended with the same top- $n$  item lists.
- ALS [41]: ALS (Alternating Least Squares) is a matrix factorization model that attempts to estimate the user-item rating matrix as the product of two lower-rank matrices. ALS minimizes two loss functions alternatively.
- BPR [42]: BPR (Bayesian Personalized Ranking) is a matrix factorization-based model that utilizes pairwise ranking loss, which is tailored to learn from implicit feedback.
- DeepWalk [43]: DeepWalk is a word2vec based network embedding model, which learns latent representations by predicting the local neighborhood of nodes sampled from random walks on the graph.
- HERec [31]: HERec is a heterogeneous information network embedding-based approach that utilizes auxiliary information for recommendation. This model designs a random walk strategy to filter the node sequence for network embedding.
- Multi-GCCF [40]: Multi-GCCF is a graph convolution-based recommendation framework, which employs a multi-graph encoding layer to integrate the information provided by the user-item, user-user, and item-item graphs.

GESRec\_NW, GESRec\_A, and GESRec\_T are three variants of our proposed model GESRec. To verify the importance of academic relationships between scholars, GESRec\_NW treats all relationships between scholars equally important. GESRec\_A and GESRec\_T only use attribute information and textual feature, respectively, as supplementary information to extend the interaction between scholars.

## 5 RESULTS AND DISCUSSIONS

In this section, we report and analyze the results of our experiments with Precision, Recall, NDCG, and F1 metrics on five real-world datasets in various domains. In particular, we first

<sup>5</sup><https://www.yelp.com/dataset/download>

<sup>6</sup><https://snap.stanford.edu/data/loc-gowalla.html>

**TABLE 2** | Experimental results on SCHOALT dataset with Precision@N and Recall@N metrics.

| Metrics   | N  | ItemPop | ALS   | BPR   | DeepWalk | HERec | Multi-GCCF | GESRec_NW | GESRec_A | GESRec_T | GESRec       |
|-----------|----|---------|-------|-------|----------|-------|------------|-----------|----------|----------|--------------|
| Precision | 10 | 0.055   | 0.006 | 0.135 | 0.142    | 0.188 | 0.190      | 0.076     | 0.077    | 0.087    | <b>0.192</b> |
|           | 20 | 0.046   | 0.007 | 0.115 | 0.120    | 0.140 | 0.158      | 0.060     | 0.059    | 0.066    | <b>0.162</b> |
|           | 30 | 0.041   | 0.006 | 0.096 | 0.105    | 0.125 | 0.135      | 0.048     | 0.048    | 0.053    | <b>0.143</b> |
|           | 40 | 0.039   | 0.006 | 0.062 | 0.087    | 0.096 | 0.105      | 0.041     | 0.041    | 0.044    | <b>0.114</b> |
|           | 50 | 0.036   | 0.006 | 0.053 | 0.065    | 0.067 | 0.073      | 0.035     | 0.035    | 0.038    | <b>0.088</b> |
| Recall    | 10 | 0.102   | 0.005 | 0.226 | 0.264    | 0.289 | 0.307      | 0.125     | 0.129    | 0.138    | <b>0.318</b> |
|           | 20 | 0.145   | 0.013 | 0.297 | 0.325    | 0.364 | 0.374      | 0.185     | 0.188    | 0.198    | <b>0.407</b> |
|           | 30 | 0.213   | 0.018 | 0.331 | 0.394    | 0.433 | 0.440      | 0.214     | 0.220    | 0.226    | <b>0.442</b> |
|           | 40 | 0.240   | 0.023 | 0.363 | 0.426    | 0.465 | 0.47       | 0.233     | 0.241    | 0.242    | <b>0.487</b> |
|           | 50 | 0.289   | 0.024 | 0.409 | 0.462    | 0.483 | 0.495      | 0.245     | 0.253    | 0.252    | <b>0.501</b> |

The best performance in each case is highlighted.

**TABLE 3** | Experimental results on Zhihu dataset with Precision@N and Recall@N metrics.

| Metrics   | N  | ItemPop | ALS   | BPR   | DeepWalk | HERec | Multi-GCCF   | GESRec_NW | GESRec_A | GESRec_T | GESRec       |
|-----------|----|---------|-------|-------|----------|-------|--------------|-----------|----------|----------|--------------|
| Precision | 10 | 0.024   | 0.001 | 0.038 | 0.056    | 0.068 | 0.096        | 0.085     | —        | —        | <b>0.106</b> |
|           | 20 | 0.020   | 0.001 | 0.029 | 0.049    | 0.057 | 0.084        | 0.073     | —        | —        | <b>0.089</b> |
|           | 30 | 0.018   | 0.001 | 0.020 | 0.041    | 0.051 | 0.076        | 0.065     | —        | —        | <b>0.082</b> |
|           | 40 | 0.015   | 0.001 | 0.018 | 0.033    | 0.042 | <b>0.072</b> | 0.051     | —        | —        | 0.071        |
|           | 50 | 0.011   | 0.001 | 0.016 | 0.029    | 0.036 | <b>0.065</b> | 0.042     | —        | —        | 0.063        |
| Recall    | 10 | 0.130   | 0.003 | 0.159 | 0.213    | 0.296 | 0.347        | 0.245     | —        | —        | <b>0.365</b> |
|           | 20 | 0.203   | 0.014 | 0.201 | 0.266    | 0.324 | 0.378        | 0.283     | —        | —        | <b>0.398</b> |
|           | 30 | 0.244   | 0.017 | 0.247 | 0.315    | 0.372 | 0.411        | 0.311     | —        | —        | <b>0.423</b> |
|           | 40 | 0.263   | 0.022 | 0.278 | 0.378    | 0.402 | <b>0.462</b> | 0.338     | —        | —        | 0.454        |
|           | 50 | 0.271   | 0.027 | 0.286 | 0.401    | 0.427 | <b>0.474</b> | 0.347     | —        | —        | 0.469        |

The best performance in each case is highlighted.

**TABLE 4** | Experimental results on APS dataset with Precision@N and Recall@N metrics.

| Metrics   | N  | ItemPop | ALS   | BPR   | DeepWalk | HERec | Multi-GCCF | GESRec_NW | GESRec_A | GESRec_T | GESRec       |
|-----------|----|---------|-------|-------|----------|-------|------------|-----------|----------|----------|--------------|
| Precision | 10 | 0.136   | 0.002 | 0.327 | 0.333    | 0.341 | 0.356      | 0.326     | 0.316    | 0.340    | <b>0.362</b> |
|           | 20 | 0.088   | 0.002 | 0.242 | 0.246    | 0.243 | 0.251      | 0.224     | 0.217    | 0.232    | <b>0.257</b> |
|           | 30 | 0.070   | 0.002 | 0.193 | 0.182    | 0.185 | 0.189      | 0.170     | 0.165    | 0.175    | <b>0.201</b> |
|           | 40 | 0.055   | 0.002 | 0.151 | 0.145    | 0.154 | 0.156      | 0.137     | 0.132    | 0.141    | <b>0.168</b> |
|           | 50 | 0.042   | 0.002 | 0.133 | 0.130    | 0.133 | 0.133      | 0.114     | 0.111    | 0.117    | <b>0.135</b> |
| Recall    | 10 | 0.184   | 0.002 | 0.446 | 0.451    | 0.466 | 0.494      | 0.456     | 0.415    | 0.484    | <b>0.520</b> |
|           | 20 | 0.221   | 0.003 | 0.588 | 0.602    | 0.618 | 0.622      | 0.578     | 0.531    | 0.605    | <b>0.630</b> |
|           | 30 | 0.248   | 0.005 | 0.612 | 0.644    | 0.664 | 0.688      | 0.629     | 0.582    | 0.656    | <b>0.699</b> |
|           | 40 | 0.262   | 0.007 | 0.649 | 0.676    | 0.687 | 0.705      | 0.657     | 0.611    | 0.682    | <b>0.726</b> |
|           | 50 | 0.267   | 0.009 | 0.665 | 0.698    | 0.711 | 0.720      | 0.673     | 0.629    | 0.699    | <b>0.733</b> |

The best performance in each case is highlighted.

provide detailed performance of our proposed model with six different baselines (for Yelp and Gowalla datasets, we evaluate our experiments with the same setting in this work [40]). Then, we analyze the impact of different factors of our model on scholar recommendation in academic social networks. Finally, we provide a visualization of scholars embedding vectors on the SCHOLAT dataset.

## 5.1 Performance Comparison

Tables 2–5 report the results of our proposed model and baselines across SCHOLAT, Zhihu, APS, Yelp, and Gowalla datasets. We adopt Precision@N, Recall@N, and NDCG@N as

the evaluation metrics for these datasets. As can be seen, in academic social networks datasets, our model outperforms all of the baseline models on the Precision and Recall metrics on SCHOLAT and APS datasets and has comparable performance with the Multi-GCCF model [40] on the Zhihu dataset. In other popular social networks datasets. Our model has the best performance on the Recall and NDCG metrics on Yelp dataset and has comparable performance with the best baseline MultiGCCF model on the Gowalla dataset.

For the SCHOLAT dataset, in Table 2, it is clear that our GESRec model constantly outperforms all baselines with  $N \in \{10, 20, 30, 40, 50\}$  in Precision and Recall metrics,

**TABLE 5** | Experimental results on Yelp and Gowalla datasets with Recall@20 and NDCG@20 metrics.

|            | Yelp          |               | Gowalla       |               |
|------------|---------------|---------------|---------------|---------------|
|            | Recall@20     | NDCG@20       | Recall@20     | NDCG@20       |
| ItemPop    | 0.0086        | 0.0117        | 0.0362        | 0.1081        |
| ALS        | 0.0004        | 0.0007        | 0.0006        | 0.0008        |
| BPR        | 0.0494        | 0.0662        | 0.1291        | 0.1878        |
| DeepWalk   | 0.0534        | 0.0712        | 0.1395        | 0.1960        |
| HERec      | 0.0599        | 0.0732        | 0.1547        | 0.2037        |
| Multi-GCCF | 0.0667        | 0.0810        | <b>0.1595</b> | <b>0.2126</b> |
| GESRec     | <b>0.0669</b> | <b>0.0816</b> | 0.1587        | 0.2119        |

The best performance in each case is highlighted.

demonstrating the effectiveness of our model when applied to scholar recommendation task. According to the results, GESRec improves over the best baseline Multi-GCCF method by 0.2–1.5% in terms of Precision and 0.2–3.3% in terms of Recall. This suggests that by exploiting the auxiliary academic information with multi-relation graphs, our model can obtain higher performance in recommending scholars in ASNs. We notice that the GESRec model has better performance than other competitive deep graph embedding-based models (i.e., Multi-GCCF, HERec, and DeepWalk), which indicates that by fusing scholar's attributes features and textual features with user-user relationships, the GESRec model is able to have better academic learning representation vectors of the scholars. Besides, we also observe that deep graph embedding-based models are consistently superior to the traditional top- $n$  recommendation methods (ItemPop, ALS, and BPR). One of the reasons might be that the user-user relations in the SCHOLAT are relatively too sparse, so they fail to obtain scholars' academic features for recommendations. Meanwhile, significant improvements are also observed over three variants models: GESRec\_NW, GESRec\_A, and GESRec\_T. This demonstrates the effectiveness of utilizing multi-relation networks in graph embedding models. We further observe that the performance of GESRec\_T is slightly better than GESRec\_NW and GESRec\_A. The possible reason is that academic textual features are more suitable than attribute features and user-user connections in expressing friendship preferences among scholars.

For the Zhihu dataset, in **Table 3**, we can notice that our model achieves the best performance in Precision and Recall metrics with  $N \in \{10, 20, 30\}$  and is comparable to the Multi-GCCF model with  $N \in \{40, 50\}$ , which is one of the best baseline models in the SCHOLAT dataset. Since the Zhihu dataset has no attribute features, we can only use a small amount of textual information as auxiliary information to reconstruct the original user-user networks, and only one variant GESRec\_NW is applied in this dataset. Compared to the best baseline Multi-GCCF, our model has slightly worse performance with  $N \in \{40, 50\}$ . A possible reason is that Zhihu is not a true ASN, and the motivation for users to establish relationships in common social networks might be different from that for academic social networks (i.e., scholars tend to follow other scholars who have similar academic features).

For the APS dataset, we observe similar results of our model in **Table 4**, which are similar to those of the SCHOLAT dataset. Although Multi-GCCF is the strongest baseline on all datasets, our model outperforms Multi-GCCF by up to 1.2 and 2.6% in terms of Precision and Recall. ALS performs the worst among all the models, as ALS is a matrix factorization-based model, but the user-user relations matrix is generally too sparse. We also notice that the overall experimental results on APS are better than Zhihu and SCHOLAT, which prove the validity of improving the performance of scholar recommendations by leveraging the academic features and relationships among scholars in ASNs. Due to the similarity of the SCHOLAT and APS datasets, we do not provide qualitative analysis in this experiment.

To further investigate the effectiveness of our proposed model, we also investigate the performance for all the models in other popular social networks domain datasets: Yelp and Gowalla. In **Table 5**, we find similar results as we observed in academic social networks datasets. ItemPop and ALS models give the worst performance on all datasets since these traditional recommendation methods fail to consider auxiliary information (i.e., the information of user-item interactions). We further notice that BPR outperforms ItemPop and ALS models because BPR is able to learn individual users' preference information. However, ItemPop is a simple model and ALS fails to deal with the sparse users-items interaction matrix.

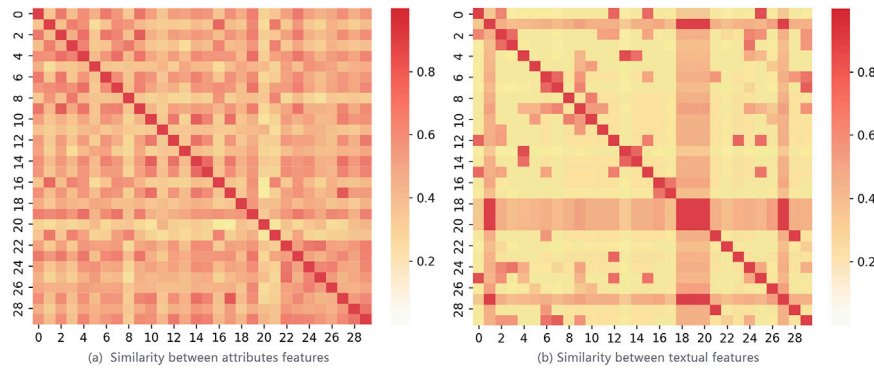
We observe that the deep graph embedding-based models (GESRec, Multi-GCCF, HERec, and DeepWalk) consistently outperform the general methods (ItemPop, ALS, and BPR), which indicates the advantages of deep graph embedding-based model in processing graph structure. Compared to the best baseline, Multi-GCCF, our model achieves better performance in the Yelp dataset but has slightly worse performance in the Gowalla dataset. One reason could be that the Multi-GCCF model provides better representations of the users when the Gowalla is relatively dense than Yelp. However, our model is able to capture better representations of the users when the Yelp dataset contains more auxiliary information of user-item interactions.

The advantages of Precision@N and Recall@N scores on all academic social networks datasets suggest that our model is able to make better scholar recommendations in academic social networks by building better academic learning representation vectors of the scholars. In addition, the comparison results of Recall@20 and NGCG@20 on two popular social networks datasets further confirm the effectiveness of our proposed model.

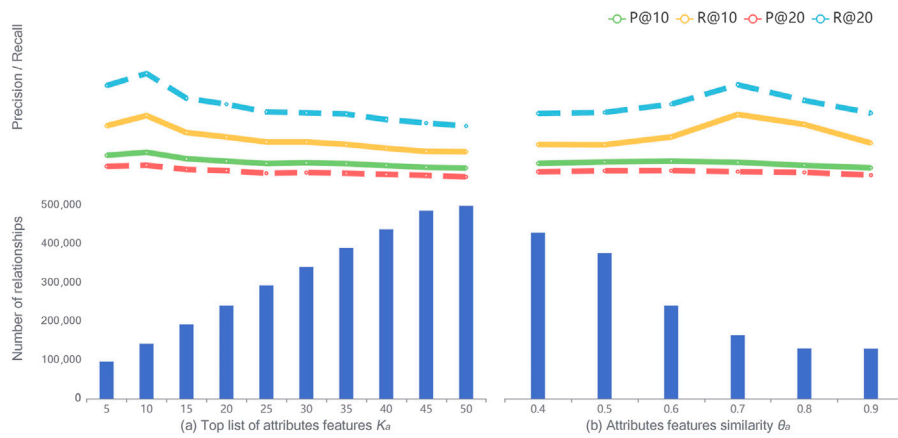
## 5.2 Impact of Different Factors

GESRec model converts scholars' multiple academic data (i.e., attributes features and textual features) into academic relationships between scholars, which are further utilized as auxiliary information to enhance the original user-user relationships. GESRec shows obvious superiority against all the baselines. However, the differences between the embedding vectors of attributes features and textual features may affect the quality of academic relationships between scholars, which in turn may affect the effectiveness of recommendations.

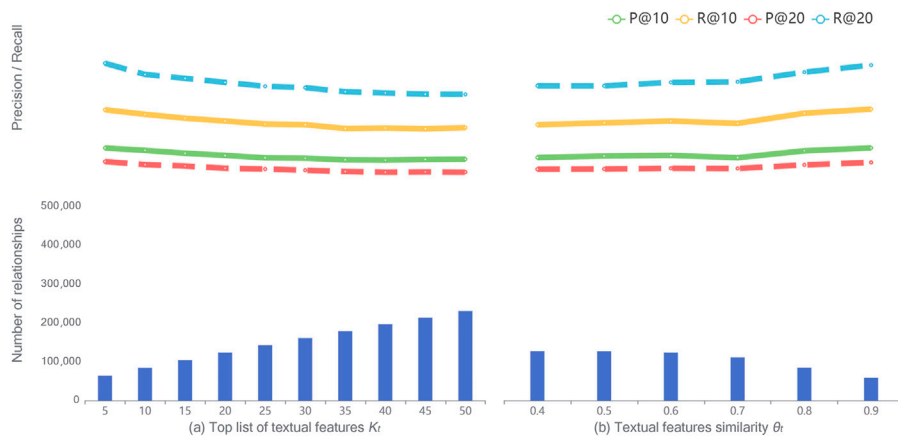




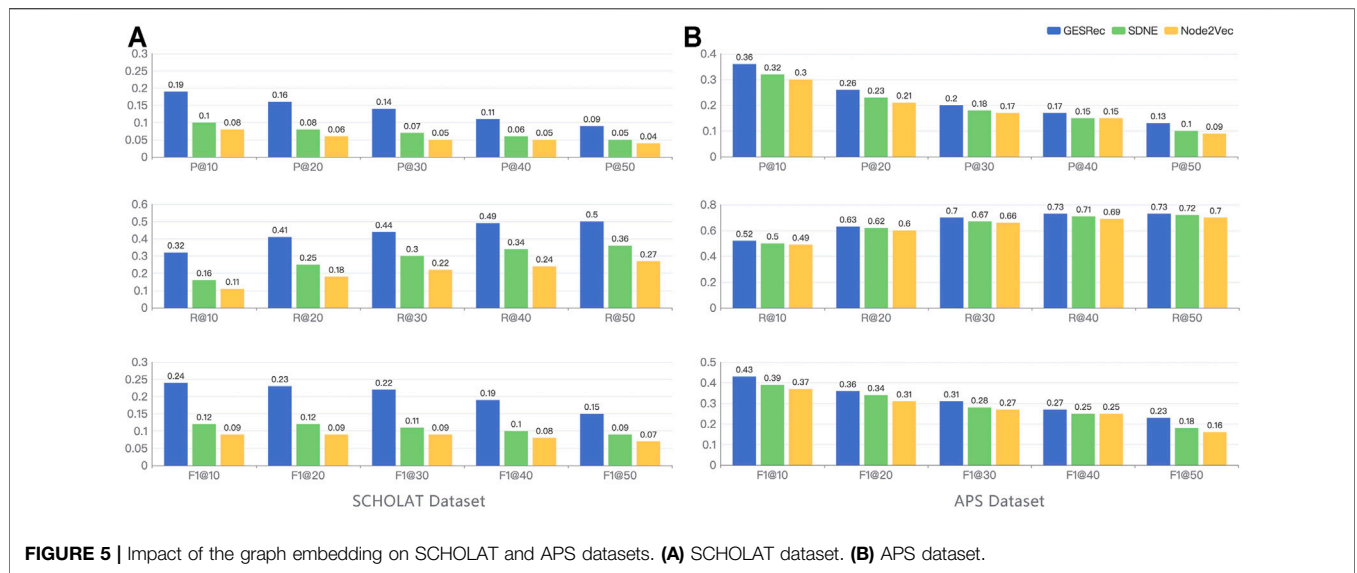
**FIGURE 2 |** Heatmap of the similarities between academic features from SCHOLAT. **(A)** Similarity between attributes features. **(B)** Similarity between textual features.



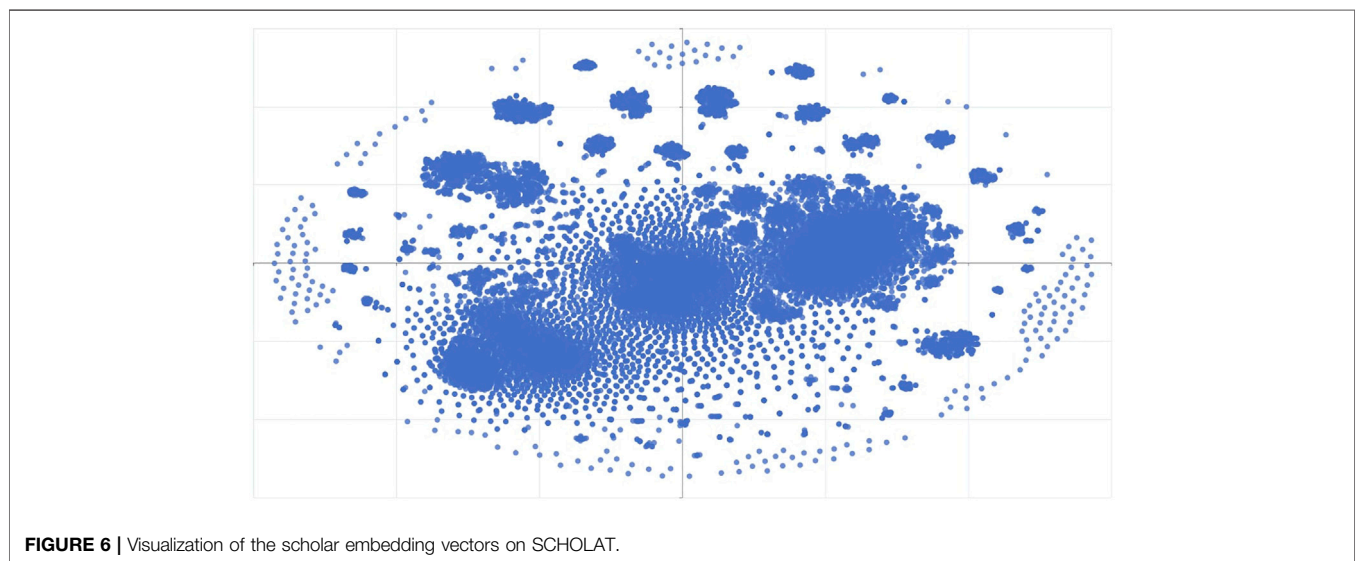
**FIGURE 3 |** Impact of the attributes features on SCHOLAT dataset. **(A)** Top list of attribute features  $K_a$ . **(B)** Attributes features of similarity  $\theta_a$ .



**FIGURE 4 |** Impact of the textual features on SCHOLAT dataset. **(A)** Top list textual features  $K_r$ . **(B)** Textual features of similarity  $\theta_r$ .



**FIGURE 5 |** Impact of the graph embedding on SCHOLAT and APS datasets. (A) SCHOLAT dataset. (B) APS dataset.

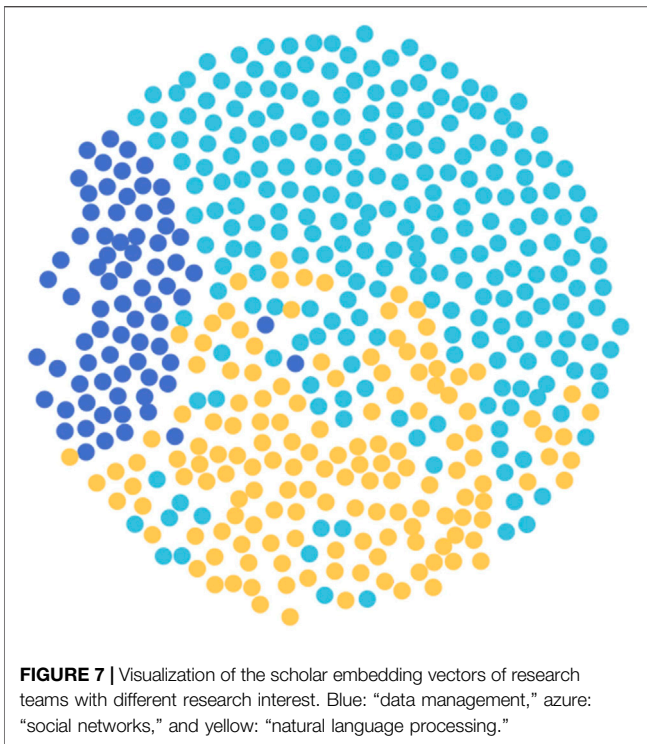


**FIGURE 6 |** Visualization of the scholar embedding vectors on SCHOLAT.

We take the real academic social networks SCHOLAT as an example. We select the 30 most frequent attributes features and textual features from SCHOLAT, respectively. We plot the heatmap of the similarity between these academic features as shown in **Figure 2**, where the darker the color is, the higher the similarities are. As we can observe from **Figure 2**, similarities between attributes features are higher than similarities between textual features, and the attributes features are more semantically related, indicating similar scholars share higher degrees of academic attributes features relevance. So attribute features-based user-user relationships could be better in building scholar’s academic friendships.

Impact of attribute features: the threshold parameters of academic features could affect the effectiveness of building academic scholar’s friendships, which in turn affects the results of scholar recommendations. Due to limited space, we

only select Precision and Recall at  $N = \{10, 20\}$ . **Figure 3** shows how the attributes features affect the recommendation performance on the SCHOLAT dataset. We see that with the increase of  $k_a$  (threshold of the top list of attributes features), the number of relationships shows nearly linear growth. However, the Precision and Recall values are continuously declined with  $k_a = \{15, 20, 25, 30, 35, 40, 45, 50\}$ . The only exception is when  $k_a$  is 10. This result suggests that the proper number of relationships is able to boost the recommendation performance. When a threshold is breached, a larger number of relationships will bring more noise data, which in turn harms the recommendation performance. We also observe that the number of relationships drops with the increase of  $\theta_a$  (threshold of attributes features similarity). The Recall values reach a turning point when  $\theta_a = 0.7$ . However, the Precision values slightly decline with the increase of parameter  $\theta_a$ . Similarly,



we find that only a proper number (100,000–200,000) of relationships brings good recommendation performance.

**Impact of textual features:** in **Figure 4**, similar observations can be obtained for the impact of textual features on SCHOLAT datasets. We notice that there is a steady growth in the number of relationships but a smooth decline in the Precision and Recall values with the increase of  $k_t$  (threshold of the top list of textual features). These results suggest that the more the relationships, the worse the Precision and Recall results. We also find a large increase in decline in the number of relationships when the textual features similarity  $\theta_t$  is over 0.7. On the contrary, the performance of Precision and Recall gets better with the increase of  $\theta_b$ , except for a minor drop at  $\theta_t = 0.7$ . From **Figures 3, 4**, we see that the impact of the attributes features threshold  $k_a$  is larger than the textual features threshold  $k_t$  due to the higher similarities between attributes features. This confirms previous findings in the heatmap of the similarities of academic features from SCHOLAT, where attribute features based on user-user relationships are more representative of scholars’ academic relationships.

**Impact of graph embedding module:** to investigate the effectiveness of graph embedding module in scholar recommendation in academic social networks task, we implement two different network representation learning models<sup>7</sup>, SDNE [44] and Node2Vec [45], as the replacement of our graph embedding module. **Figure 5** shows the impact of

different graph embedding-based models on SCHOLAT and APS datasets. Recommendation performance on the SCHOLAT dataset shows that the GESRec model outperforms SDNE and Node2Vec in all Precision, Recall, and F1 score metrics. Similar results can be observed on the APS dataset. However, the advantages of performance have been decreased. The possible reason might be that the dataset is based on co-author relationships instead of rich academic correlations between scholars on the SCHOLAT dataset. Overall, these improvements on both datasets verify the effectiveness of the graph embedding in scholar recommendation in ASNs tasks.

### 5.3 Visualizing Scholar Embedding Vectors

In this section, we visualize the scholar embedding vectors extracted from the SCHOLAT dataset in **Figure 6**. Besides, we randomly select three research teams (e.g., 56 scholars in “data management,” 149 scholars in “social networks,” and 110 scholars in “natural language processing”). **Figure 7** provides a visualization of the scholar embedding vectors in three research teams with different research interests. Nodes with the same color mean they are from the same research team. Specifically, we first get the scholar embedding vectors by the GESRec model and then map these vectors into a low-dimensional space. Finally, we further map the low-dimensional vectors to a 2D space using the t-SNE [46]. In **Figure 6**, we observe that there are many clusters in which the scholars belong to the same cluster, meaning they share similar academic features. These scholar embedding vectors are used on the following scholar recommendation task. In **Figure 7**, we find that except for a few blue nodes, other nodes from the same research team tend to close one another. This is because scholars in the same research team usually share similar research interests, making their scholar embedding vectors have a higher academic similarity. Hence, these visualizations of the scholar embedding vectors show that our GESRec model is able to recommend potential similar scholars to users with high academic similarities.

## 6 CONCLUSION AND FUTURE WORK

In this article, we tackle the problem of recommending potential similar scholars for users in academic social networks (ASNs). We propose a novel graph embedding-based scholar recommendation system by leveraging academic auxiliary information. The proposed model consists of three steps: First, we construct the enhanced academic social networks by combining two types of academic features (e.g., attributes features and textual features) from scholar’s academic information with user-user relationships, and the correlations between the scholar’s academic features are calculated by the similarity of the corresponding embedding vectors. Then, the refined feature representations of the scholars are obtained from the enhanced ASNs by the graph embedding framework. Finally, top- $n$  potential scholars are generated for final recommendation. In the future, we would like to consider more academic information to improve the performance of our model.

<sup>7</sup><https://github.com/palash1992/GEM>

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

All authors conceived the paper. CY drafted the initial manuscript. CY and RL developed the software and run the

analyses. YH participated in the discussion of theories. YT supervised the study and edited the manuscript. All authors reviewed the article.

## FUNDING

This work was supported by the National Natural Science Foundation of China (No. U1811263, 61772211), the Talent Research Start-Up Foundation of Guangdong Polytechnic Normal University (No. 2021SDKYA098).

## REFERENCES

- Xia F, Wang W, Bekele TM, Liu H Big Scholarly Data: A Survey. *IEEE Trans Big Data* (2017) 3:18–35. doi:10.1109/tbdata.2016.2641460
- Kong X, Shi Y, Yu S, Liu J, Xia F Academic Social Networks: Modeling, Analysis, Mining and Applications. *J Netw Comp Appl* (2019) 132:86–103. doi:10.1016/j.jnca.2019.01.029
- Jordan K From Social Networks to Publishing Platforms: A Review of the History and Scholarship of Academic Social Network Sites. *Front Digit Humanit* (2019) 6:5. doi:10.3389/fdigh.2019.00005
- Wan H, Zhang Y, Zhang J, Tang J Aminer: Search and Mining of Academic Social Networks. *Data Intellegence* (2019) 1:58–76. doi:10.1162/dint\_a\_00006
- Kyvik S, Reymert I Research Collaboration in Groups and Networks: Differences across Academic fields. *Scientometrics* (2017) 113:951–67. doi:10.1007/s11192-017-2497-5
- Wu D, Fan S, Yuan F Research on Pathways of Expert Finding on Academic Social Networking Sites. *Inf Process Manage* (2021) 58:102475. doi:10.1016/j.ipm.2020.102475
- Xu Y, Hao J, Lau RYK, Ma J, Xu W, Zhao D. A Personalized Researcher Recommendation Approach in Academic Contexts: Combining Social Networks and Semantic Concepts Analysis. In: Pacific Asia Conference on Information Systems; Taipei, Taiwan; July 2010. Taipei, Taiwan: PACISAISeL (2010). 9144–12.
- Yang C, Sun J, Ma J, Zhang S, Wang G, Hua Z. In: Scientific collaborator recommendation in heterogeneous bibliographic networks. 48th Hawaii International Conference on System Sciences, HICSS 2015; January 5–8, 2015; Kauai, Hawaii, USA. IEEE Computer Society (2015). 552–61.
- Wang Q, Ma J, Liao X, Du W A Context-Aware Researcher Recommendation System for university-industry Collaboration on R&D Projects. *Decis Support Syst* (2017) 103:46–57. doi:10.1016/j.dss.2017.09.001
- Xu Y, Zhou D, Ma J Scholar-friend Recommendation in Online Academic Communities: An Approach Based on Heterogeneous Network. *Decis Support Syst* (2019) 119:1–13. doi:10.1016/j.dss.2019.01.004
- Zhang C, Wu X, Yan W, Wang L, Zhang L Attribute-aware Graph Recurrent Networks for Scholarly Friend Recommendation Based on Internet of Scholars in Scholarly Big Data. *IEEE Trans Ind Inf* (2020) 16:2707–15. doi:10.1109/tii.2019.2947066
- Goyal P, Ferrara E Graph Embedding Techniques, Applications, and Performance: A Survey. *Knowledge-Based Syst* (2018) 151:78–94. doi:10.1016/j.knsys.2018.03.022
- Zhao Z, Zhang X, Zhou H, Li C, Gong M, Wang Y Hetnrec: Heterogeneous Network Embedding Based Recommendation. *Knowledge-Based Syst* (2020) 204:106218. doi:10.1016/j.knsys.2020.106218
- Forouzandeh S, Berahmand K, Rostami M Presentation of a Recommender System with Ensemble Learning and Graph Embedding: a Case on MovieLens. *Multimed Tools Appl* (2021) 80:7805–32. doi:10.1007/s11042-020-09949-5
- Hu X, Xu J, Wang W, Li Z, Liu A A Graph Embedding Based Model for fine-grained POI Recommendation. *Neurocomputing* (2021) 428:376–84. doi:10.1016/j.neucom.2020.01.118
- Khan HU, Daud A, Ishaq U, Amjad T, Aljohani N, Abbasi RA, et al. Modelling to Identify Influential Bloggers in the Blogosphere: A Survey. *Comput Hum Behav* (2017) 68:64–82. doi:10.1016/j.chb.2016.11.012
- Wang Y, Han X Attractive Community Detection in Academic Social Network. *J Comput Sci* (2021) 51:101331. doi:10.1016/j.jocs.2021.101331
- Zhao P, Ma J, Hua Z, Fang S Academic Social Network-Based Recommendation Approach for Knowledge Sharing. *SIGMIS Database* (2018) 49:78–91. doi:10.1145/3290768.3290775
- Porcel C, Ching-López A, Lefranc G, Loia V, Herrera-Viedma E Sharing Notes: An Academic Social Network Based on a Personalized Fuzzy Linguistic Recommender System. *Eng Appl Artif Intelligence* (2018) 75:1–10. doi:10.1016/j.engappai.2018.07.007
- Brandão MA, Moro MM, Lopes GR, de Oliveira JPM. Using Link Semantics to Recommend Collaborations in Academic Social Networks. In: 22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13–17, 2013, Companion Volume. International World Wide Web Conferences Steering Committee/ACM (2013). 833–40. doi:10.1145/2487788.2488058
- Huang X, Tang Y, Qu R, Li C, Yuan C, Sun S, et al. Course Recommendation Model in Academic Social Networks Based on Association Rules and Multi-similarity. In: 22nd IEEE International Conference on Computer Supported Cooperative Work in Design; May 9–11, 2018; Nanjing, China. CSCWDIEEE (2018). 277–82. doi:10.1109/cscwd.2018.8465266
- Kong X, Mao M, Wang W, Liu J, Xu B Voprec: Vector Representation Learning of Papers with Text Information and Structural Identity for Recommendation. *IEEE Trans Emerg Top Comput.* (2021) 9:226–37. doi:10.1109/tetc.2018.2830698
- Nikzad-Khasmakhi N, Balafar MA, Feizi-Derakhshi M. The State-Of-The-Art in Expert Recommendation Systems. *Eng Appl Artif Intell* (2019) 82:126–47.
- Feng W, Zhu Q, Zhuang J, Yu S An Expert Recommendation Algorithm Based on pearson Correlation Coefficient and Fp-Growth. *Cluster Comput* (2019) 22:7401–12. doi:10.1007/s10586-017-1576-y
- Yang C, Ma J, Silva T, Liu X, Hua Z A Multilevel Information Mining Approach for Expert Recommendation in Online Scientific Communities. *Comp J* (2015) 58:1921–36. doi:10.1093/comjnl/bxu033
- Davoodi E, Afsharchi M, Kianmehr K A Social Network-Based Approach to Expert Recommendation System. In: Hybrid Artificial Intelligent Systems - 7th International Conference, HAIS 2012; Salamanca, Spain; March 28–30th, 2012, 7208. Salamanca, Spain: Lecture Notes in Computer Science (2012). 91–102. Proceedings, Part I (Springer). doi:10.1007/978-3-642-28942-2\_9
- Ding J, Chen Y, Li X, Liu G, Shen A, Meng X Unsupervised Expert Finding in Social Network for Personalized Recommendation. In: Web-Age Information Management - 17th International Conference, WAIM; Nanchang, China; June 3–5, 2016, 9658. Nanchang, China: Lecture Notes in Computer Science (2016). 257–71. Proceedings, Part I (Springer). doi:10.1007/978-3-319-39937-9\_20
- Ge H, Caverlee J, Lu H. TAPER: A Contextual Tensor-Based Approach for Personalized Expert Recommendation. In: Proceedings of the 10th ACM Conference on Recommender Systems; Boston, MA, USA; September 15–19, 2016. Boston, MA, USA: ACM (2016). 261–8.
- Chen J, Tang Y, Li J, Mao C, Xiao J. Community-based Scholar Recommendation Modeling in Academic Social Network Sites. In: Web Information Systems Engineering - WISE 2013 Workshops - WISE 2013 International Workshops BigWebData, MBC, PCS, STeH, QUAT, SCEH, and STSC 2013; Nanjing, China; October 13–15, 2013, 8182. Nanjing, China: Lecture Notes in Computer Science (2013). 325–34. Revised Selected Papers (Springer).

30. Mao D, Li C, Li J, Tang Y, Chen M, Yang X. Academic Social Network Scholars Recommendation Model Based on Community Division. In: 22nd IEEE International Conference on Computer Supported Cooperative Work in Design; May 9-11, 2018; Nanjing, China. CSCWDIEEE (2018). p. 265-70. doi:10.1109/cscwd.2018.8465149
  31. Shi C, Hu B, Zhao WX, Yu PS Heterogeneous Information Network Embedding for Recommendation. *IEEE Trans Knowl Data Eng* (2019) 31: 357-70. doi:10.1109/tkde.2018.2833443
  32. Huang X, Qian S, Fang Q, Sang J, Xu C. Meta-path Augmented Sequential Recommendation with Contextual Co-attention Network. *ACM Trans Multimed Comput Commun Appl* (2020) 16(1-52):5224. doi:10.1145/3382180
  33. Xie F, Zheng A, Chen L, Zheng Z Attentive Meta-Graph Embedding for Item Recommendation in Heterogeneous Information Networks. *Knowledge-Based Syst* (2021) 211:106524. doi:10.1016/j.knsys.2020.106524
  34. Li S, Zhao Z, Hu R, Li W, Liu T, Du X. Analogical Reasoning on Chinese Morphological and Semantic Relations. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018; Melbourne, Australia; July 15-20, 2018; editors, I Gurevych Y Miyao, 2. Melbourne, Australia: Short Papers Association for Computational Linguistics (2018). 138-43. doi:10.18653/v1/p18-2023
  35. Tang J, Qu M, Wang M, Zhang M, Yan J, Mei Q. LINE: Large-Scale Information Network Embedding. In: Proceedings of the 24th International Conference on World Wide Web; Florence, Italy; May 18-22, 2015. Florence, Italy: WWW 2015ACM (2015). 1067-77.
  36. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed Representations of Words and Phrases and Their Compositionality. In: Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems; Lake Tahoe, Nevada, United States, 5-8. Lake Tahoe, Nevada, United States: Proceedings of a meeting held December (2013). 3111-9.
  37. Xu Q, Qiu L, Lin R, Tang Y, He C, Yuan C. An Improved Community Detection Algorithm via Fusing Topology and Attribute Information. CSCWD 2021. In: 24th IEEE International Conference on Computer Supported Cooperative Work in Design; May 5-7, 2021; Dalian, China. IEEE (2021). p. 1069-74. doi:10.1109/cscwd49262.2021.9437681
  38. Tu C, Liu H, Liu Z, Sun M. CANE: Context-Aware Network Embedding for Relation Modeling. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017; Vancouver, Canada; July 30 - August 4, 1. Vancouver, Canada: Long Papers Association for Computational Linguistics (2017). 1722-31. doi:10.18653/v1/p17-1158
  39. Pennington J, Socher R, Manning CD. Glove: Global Vectors for Word Representation. In: A Moschitti, B Pang, W Daelemans, editors. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014; October 25-29, 2014; Doha, Qatar. Doha, Qatar: A meeting of SIGDAT, a Special Interest Group of the ACL (2014). p. 1532-43. doi:10.3115/v1/d14-1162
  40. Sun J, Zhang Y, Ma C, Coates M, Guo H, Tang R, et al. Multi-graph Convolution Collaborative Filtering. *CoRR abs/2020* 2001-00267.
  41. Zhou Y, Wilkinson DM, Schreiber R, Pan R. Large-scale Parallel Collaborative Filtering for the Netflix Prize. In: Algorithmic Aspects in Information and Management, 4th International Conference, AAIM 2008; Shanghai, China; June 23-25, 2008, 5034. Shanghai, China: Lecture Notes in Computer Science (2008). 337-48. Proceedings (Springer).
  42. Rendle S, Freudenthaler C, Gantner Z, Schmidt-Thieme L. BPR: Bayesian Personalized Ranking from Implicit Feedback. *CoRR abs/2012* 1205:2618.
  43. Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online Learning of Social Representations. In: The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; New York, NY, USA; August 24 - 27, 2014. New York, NY, USA: KDD '14ACM (2014). 701-10.
  44. Wang D, Cui P, Zhu W. Structural Deep Network Embedding. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Francisco, CA, USA; August 13-17, 2016. San Francisco, CA, USA: ACM (2016). p. 1225-34. doi:10.1145/2939672.2939753
  45. Grover A, Leskovec J. node2vec: Scalable Feature Learning for Networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Francisco, CA, USA; August 13-17, 2016. San Francisco, CA, USA: ACM (2016). 855-64. KDD. doi:10.1145/2939672.2939754
  46. van der Maaten L, Hinton G. Visualizing Data Using T-Sne. *J Machine Learn Res* (2008) 9:2579-605.
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2021 Yuan, He, Lin and Tang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.