Knowledge-Based Systems 210 (2020) 106438

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Paper recommendation based on heterogeneous network embedding

Zafar Ali [a,*], Guilin Qi [a], Khan Muhammad [b,*], Bahadar Ali [a], Waheed Ahmed Abro [a]

[a] School of Computer Science and Engineering, Southeast University, Nanjing, China
[b] Department of Software, Sejong University, Seoul 143-747, Republic of Korea

A R T I C L E   I N F O

Article history:
Received 7 March 2020
Received in revised form 6 September 2020
Accepted 8 September 2020
Available online xxxx

Keywords:
Recommender systems
Paper recommendation
Citation recommendation
Heterogeneous network embedding
Neural networks
Cold-start

A B S T R A C T

Researchers face millions of research papers on various digital libraries. Therefore, finding relevant research work that meets the preferences of a researcher is a challenging task. Hence, different paper recommendation models have been proposed to address this issue. However, these models lack in exploiting prominent information factors, namely: papers' citations proximity, authors' information, papers' topical relevance, venues' information, researchers' preference dynamics, and labels information to produce quality recommendations. Additionally, these models encounter problems such as cold start papers and data sparsity. To overcome these problems, this paper presents a weighted probabilistic paper recommendation model termed as PR-HNE, which jointly learns researchers' and papers' dynamics by encoding information from six information networks into a joint latent space. Specifically, it captures papers' citation proximity, authors' collaboration proximity, venues' information, labeled information, and topical relevance to generate personalized paper recommendations. Compared to state-of-the-art models, the results generated by PR-HNE over publicly available datasets prove 4% and 6% improvement in Mean Average Precision (MAP) and Mean Reciprocal Rank (MRR) metrics, respectively. Further, in the cold-start papers problem, the proposed model produced 8% better recall score than its counterparts.

1. Introduction

In recent years, recommendation systems [1] have got researchers' attention in various domains, such as movies [2], books [3,4], Point-of-Interests (POIs) [5,6], and cognitive models [7,8]. Such systems help users in recommending relevant items and information according to their needs and preferences. In the scientific literature domain, researchers need to find relevant publications that can help them in finding novel research ideas and conducting their research works [9]. Due to the availability of millions of research papers on the Web, it is difficult for researchers to find relevant articles. To assist researchers, paper recommendation models [10–14] generate useful recommendations that meet the researchers' information needs. Typically, paper recommendation models have been classified into three classes including: Content-based filtering (CBF) [12,15], Collaborative filtering (CF) [10,11,16], and graph-based [17–19]. CF-based paper recommendation models exploit the past ratings/feedback information of users and their friends to generate recommendations [20]. In this direction, a CF-based approach called Collaborative Topic Regression (CTR) [10] integrates the latent factor model along with the Latent Dirichlet Allocation (LDA) topic model to recommend articles. CF-based models can generate quality recommendations when user feedback/rating information is available. However, CF models suffer from sparsity problems [21] when the user–item/paper rating matrix is sparse with missing values. Thus, recommendations based on such limited information can lead to inadequate results [22]. Additionally, the efficiency of CF-based models become challenging due to the growing scale of users and items. To this end, hashing-based recommendation frameworks namely Discrete Collaborative Filtering (DCF) [23], Discrete Ranking-based Matrix Factorization (DRMF) [24], Deep Collaborative Hashing (DCH) [25], Binarized Collaborative Filtering with Distilling Graph Convolutional Networks (DGCN-BinCF) [26] and Quantization-based Hashing (QBH) [27] have been proposed to overcome the efficiency issues by representing users and items with binary codes. In contrast, CBF models [12,15] generate recommendations by utilizing the descriptions and features of items [28]. In this direction, Amami et al. [12] employed LDA topic modeling to the content of articles to generate recommendations. In CBF models, it is necessary that items/papers descriptions, as well as users' information, are available in enough detail, otherwise CB approach will face cold-start and limited content analysis issues [29].

Traditional graph-based models [30,31] consider recommendation as a link prediction problem by employing the properties

* Corresponding authors.
E-mail addresses: zafarali@seu.edu.cn (Z. Ali), gqi@seu.edu.cn (G. Qi), khan.muhammad@ieee.org (K. Muhammad), bahadar@seu.edu.cn (B. Ali), waheed@seu.edu.cn (W.A. Abro).

https://doi.org/10.1016/j.knosys.2020.106438
0950-7051/

of random walk methods. Nevertheless, models that use random walk methods suffer from over-weighting problem, i.e., old and outdated papers are over-weighted in the network [22]. Recently, more appealing network representation learning based methods such as VOPRec [32], Bibliographic Network Representation (BNR) [17], and Three-layered Mutually Reinforced model (TMR) [33] have been proposed, which exploit network structure and papers content to generate citation recommendations. However, these models are limited in terms of utilizing prominent information factors such as papers' citation proximity, venues' information, papers' topical relevance, and labels/tags, which can help exploiting researchers' preference dynamics and generating more personalized recommendations. Additionally, these models do not consider the significance of relations among the network objects, therefore they fail to capture meaningful semantics.

In paper recommender systems, researchers' preference dynamics correspond to multiple factors, such as authors information, research topics, venue information, and tags/labels of research articles. These factors play a vital role to assist researchers in obtaining relevant paper recommendations [34]. For instance, the author(s) of a paper can have a great influence on the readers, prevalence, and citations [35]. Often, a researcher follows a particular researcher or a research group/s with similar research preferences. Similarly, an author who does collaboration with another researcher is more important, compared to the one who has no collaboration with different research interests [17]. Meanwhile, researchers write papers that focus on a set of topics related to their scientific investigations. To this end, studies [10,12] point out that researchers are more interested in the papers that target the topic/s of their research interests. Also, a recent study [20] reveals that the most popular feature in citation recommendation models is tags since it correlates keywords with papers. Thus, each paper is marked with multiple tags that give a short description of its contents. In this manner, a paper recommendation model can correlate the interests of a target user with relevant papers based on these tags. In addition to that, a venue that publishes papers related to authors' published paper/s is of great importance to the researcher than a random venue. Also, recent studies [36,37] show that exploiting venue information has reasonable impact on the quality of citation recommendations.

In light of the above analysis, we argue that there is a need to exploit the aforementioned factors in a unified model which can generate personalized paper recommendations. In this paper, we present a paper recommendation model termed as Paper Recommendation based on Heterogeneous Network Embedding (PR-HNE), which captures researchers preference dynamics, authors information, topical relevance, labeled information, and venue information, encoding six information graphs among participating objects into a joint latent space. Additionally, the model captures researchers' preference dynamics and significance of relations among the objects to learn their vector representations, which is employed in generating personalized paper recommendations. As the model utilizes rich information sources, therefore, it can effectively learn researchers' preference dynamics and produces robust paper recommendations. The contributions of this paper are presented as follows:

- We present a weighted probabilistic approach for six information graphs to capture semantic relationships effectively among network objects and mitigate the cold-start papers and sparsity problems.
- We propose a novel network embedding approach, which jointly learns researchers' and papers' embedding by exploiting these weighted networks into a joint embedding space and generates personalized paper recommendations.
- We perform extensive experiments on two real-world datasets to evaluate the performance of the proposed model against state-of-the-art counterparts in terms of MAP, MRR, recall, and cold-start paper recommendations.

The remaining paper is organized as: Section 2 presents literature on state-of-the-art paper recommendation models. Section 3 introduces the problem statement and preliminaries. Section 4 demonstrates the methodology and detailed working process of the proposed model. Section 5 presents the experimental details and comparison of the proposed model against baseline methods. Section 6 provides a short case study. Section 7 concludes the paper, presents future research directions followed by references.

2. Related work

2.1. CBF and CF-based recommendation models

CF-based models employ the past ratings and feedback of users and their friends to generate recommendations [38]. For instance, Collaborative Topic Regression (CTR) [10] integrated the latent factor model along with the LDA topic model to exploit the best features of CF and CBF methods in generating paper recommendations. Similarly, Bansal et al. [11] generated paper predictions in the CF task by employing gated recurrent units (GRUs) to the textual information of research papers. Another CF-based model [39] applied the CF method to the paper-citation matrix and discovered potential citation papers, which are then used to generate feature vectors for the target papers. Also, researchers' profiles are developed using their published research papers. To produce final recommendations, the model employed the Cosine similarity between the vector representations of user profile and target papers. In contrast, CBF methods generate recommendations based on the descriptions and features of papers and users [28]. For instance, Amami et al. [12] employed LDA to the abstracts of articles to generate recommendations. To build a profile for a researcher, the model extracts topics from his/her authored papers using LDA topic modeling. On the other hand, the model utilized the language modeling approach to represent a candidate article. Finally, it computes similarities between these representations to generate final recommendations to a user. Similarly, Science Concierge [40] employed Latent Semantic Analysis (LSA) to the content of papers to produce recommendations. In contrast, NNRank [15] used a feed-forward neural model to the papers' content to encode them into a vector space. Then, it identifies k nearest neighbors for the query paper and employs another discriminative model between observed and unseen papers to re-rank the papers.

Though CBF and CF-based models can generate quality recommendations, however, such models are confronted with different problems. For instance, CF-based models face sparse ratings problem, therefore recommendations made on such limited information can cause inaccurate results [22]. On the other hand, CBF models can face limited content analysis and cold-start problems [20,29]. Furthermore, traditional CB and CF models do not consider auxiliary information sources, which can help in producing more robust recommendations.

2.2. Graph-based recommendation models

Graph-based paper recommendation systems employ k-partite graphs (i.e., uni-partite, bipartite, or hybrid) to exploit meaningful relations between the nodes [20]. Traditional graph-based paper recommendation models [30,31,41] employed random walk methods to generate citation recommendations. For instance, a Bi-Relational graph model [30] used an iterative random walk and restarts technique to recommend research articles. Similarly, Common Author Relation-based Recommendation (CARE) [41] utilized random walk with restart algorithm to rank research papers. Chakraborty et al. [31] adopted a time-aware random-walk method to produce papers recommendations. However, these models perceive recommendation like a link prediction task and therefore suffer from over-weighting old and outdated papers in the network [22]. Additionally, such models are limited in terms of utilizing textual and heterogeneous information in the network that can help in producing quality recommendations.

In recent years, there is a surge of different graph embedding [6,42,43] and network embedding [44–46] methods, which encode graphs or a community, and nodes into a low-dimensional space. In this direction, researchers have employed such representation learning methods for generating citation recommendations [32,33,47]. For instance, Gupta and Varma [47] learned the low-dimensional representations of network and papers' content using DeepWalk and Doc2Vec [48], respectively. Then, the model computes similarities between these representations to generate citation recommendations. Similarly, VOPRec [32] learned the vector representations of textual content as well as network structure to recommend research papers. That is, it integrates text-based nearest nodes and structured-based vectors learned using Paper2vec [49] and Struct2vec [50] embedding techniques, respectively. On the other hand, BNR [17] exploited the network structure as well as content of participating objects (authors, papers, and venues) to generate distributed representations of corresponding objects. That is, it employs Node2vec [46] algorithm on the corpus generated by biased random walk to obtain the vector representations of objects and produce citation recommendations. Existing NRL-based models such as BNR and VOPRec can generate promising recommendations compared to traditional graph-based models. However, such methods are limited in terms of utilizing auxiliary information sources such as authors' information, topical relevance, and labels/tags that can help in producing more justifiable recommendations. This research aim to overcome such problems by exploiting the aforementioned factors and producing adequate paper recommendations.

3. Problem formulation and preliminaries

In this section, we discuss different terminologies and concepts employed in the PR-HNE model. Fig. 1 presents different participating graphs that are used in the proposed model. Symbols used in this research are shown in Table 1. We discuss some important definitions as follows:

Table 1
List of symbols used in the paper.

Symbols	Description
A	set of authors $A = \{a_1, a_2, \dots a_n\}$
P	set of papers $P = \{p_1, p_2, \dots p_n\}$
L	set of labels $L = \{l_1, l_2, \dots l_n\}$
T	set of topics $T = \{t_1, t_2, \dots t_n\}$
V	set of venues $V = \{v_1, v_2, \dots v_n\}$
C	set of vertices $C = \{c_1, c_2, \dots c_n\}$
E	set of edges $e_{i,j}$ over each graph
S, N	number of samples and negative samples
W	set of weights $w_{i,j}$ over each graph
$w_{i,j}$	weight between two vertices
G	heterogeneous information network

Definition 1 (*Author–Author Graph*)**.** Author–author graph is an un-directed graph representing relationship between two authors. The graph is represented by $G_{aa'} = (A \cup A, E_{aa'}, W_{aa'})$ where A represents the set of authors and $E_{aa'}$ denotes the set of edges between authors. $W_{aa'}$ represents the set of weighted edges among authors, as defined in Eq. (1).

$$W_{aa'} = \frac{\sum_{\forall p_t \in (a,a')} |P_t|}{\sum_{\forall p_{t'} \in (a,a')} |P_{t'}|} \tag{1}$$

In Eq. (1), P_t denotes the total number of papers published by two authors in collaboration. $P_{t'}$ represents the total number of published works of author a and author a', which include papers written by both authors in collaboration as well as without collaboration.

Definition 2 (*Author–Paper Graph*)**.** The graph between authors and papers is a bi-partite graph indicating the relation between authors and papers. The graph is represented by $G_{ap} = (A \cup P, E_{ap}, W_{ap})$, where A and P denotes the set of authors and papers, respectively. E_{ap} denotes the set of edges among papers and authors. W_{ap} represents weighted edges between authors and papers, which is computed as using Eq. (2).

$$W_{ap} = \frac{\vec{a} \cdot \vec{p}}{Age_p + \xi} \tag{2}$$

where $\vec{a}$ and $\vec{p}$ represent the contextual vector representations of author a and paper p, respectively. Author's embedding is generated using the text of his/er query manuscript/s. We learn these embedding using pre-trained Sentence-BERT (SBERT) [51] model. Age_p represents the age of paper p, which is computed as the difference between its publishing year and current year. In addition, ξ denotes a positive constant that has value 1, which prevents the value of W_{ap} from being extremely large when the paper's p publishing year and current year are the same. As a result, the model assigns smaller weights to old papers and larger weights to recently published papers.

Definition 3 (*Paper–Topic Graph*)**.** The graph between papers and topics is a bi-partite graph indicating the relation between papers and topics. For instance, relation can be contain topic, i.e., a topic *Deep Learning* belongs to a paper p_i. The graph is represented by $G_{pt} = (P \cup T, E_{pt}, W_{pt})$, where P denotes the set of papers and T represents the set of latent topics. E_{pt} denotes the set of edges and W_{pt} is used for weighted edges between papers and topics computed using Eq. (3).

$$W_{pt} = pr(t|p) \tag{3}$$

where t, and p represent latent topics and papers, respectively. pr denotes topic distribution over papers discovered using a generative LDA [52] topic modeling. With this method, we characterize each topic as a distribution over all words. We identified *top-k* meaningful topics for papers with higher coherence score.

Definition 4 (*Paper–Paper Graph*)**.** The graph between papers is represented by $G_{pp'} = (P \cup P, E_{pp'}, W_{pp'})$, where P and $E_{pp'}$ denotes the set of papers and edges, respectively. Additionally, $W_{pp'}$ represents weighted edges among papers defined in the following Eq. (4).

$$W_{pp'} = \varrho \left(Sim \left(p, p' \right) \right) \tag{4}$$

where $Sim(p, p')$ represents similarity between two papers computed using Algorithm 1 (an improvement of CCIDF+) [53]. Aim is to maximize the similarity of two papers exploiting their textual information and network structure proximity.

where $Sim_{cont}(p_i, p_j)$ represents cosine between the contextual embedding vectors, i.e., $\vec{p}_i$ and $\vec{p}_j$ of two corresponding papers, which is computed employing SBERT. On the other hand, $weight_{in}(p_{co-ref[k]})$ and $weight_{out}(p_{co-cite[t]})$ represents similarity

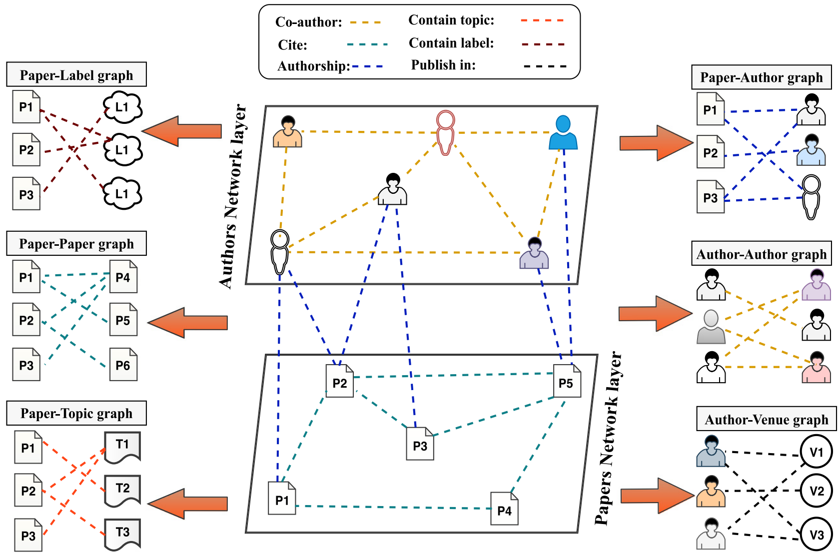


Fig. 1. An illustration of participating networks in the PR-HNE model, where objects such as authors, papers, venues, labels, and topics establish meaningful relations with each other. Graphs that are associated with the authors' relations contain Author–Author, Author–Paper, and Author–Venue, graphs which correspond to papers' connections including Paper–Topic, Paper–Paper, and Paper-Label.

Algorithm 1: Papers similarity computation based on network proximity and contextual information

Data: Directed Graph $G(C, E)$ and papers embedding vectors $(\vec{p}_i, \vec{p}_j)$
Result: Similarity score $Sim(p_i, p_j)$
forall $(p_i, p_j) \in C$ **do**
  Sim $(p_i, p_j) = 0$
end
foreach $p_i \in C$ **do**
  if $p_{ref} \neq \phi$ **then**
    // p_{ref} denotes list of paper p_i references.
    $weight_{out}(p_i) = \frac{1}{total(p_{ref})}$
  end
  if $p_{cit} \neq \phi$ **then**
    // p_{cit} denotes list of paper p_i citations.
    $weight_{in}(p_i) = \frac{1}{total(p_{cit})}$
  end
end
foreach $p_i \in C$ **do**
  Find co-references and co-citations of p_i, and p_j
  $Sim_{cc}(p_i, p_j) = \sum_{k=1}^{n} weight_{in}(p_{co-ref[k]}) + \sum_{t=1}^{m} weight_{out}(p_{co-cit[t]})$
end
foreach $(p_i, p_j) \in C$ **do**
  $Sim_{cont}(p_i, p_j) = cos(\vec{p}_i, \vec{p}_j)$
end
Sim $(p_i, p_j) = Sim_{cc}(p_i, p_j) + Sim_{cont}(p_i, p_j)$
Return Sim (p_i, p_j)

between two papers computed based on their co-citation relation and bibliographic coupling. In Eq. (4), ϱ denotes the normalization function to scale the values between 0 and 1.

Definition 5 (*Author–Venue graph*)**.** Author–venue graph is a bi-partite graph representing relationship between authors and venues. The graph is represented by $G_{av} = (A \cup V, E_{av}, W_{av})$, where A, V and E_{av} represent the sets of authors, venues, and edges, respectively. W_{av} represents the set of weighted edges between authors and venues computed as follows in Eq. (5).

$$W_{av} = \frac{1}{\sum_{i=1}^{n} |V_i|} \tag{5}$$

where V_i represents the total number of venues where author a has published papers.

Definition 6 (*Paper-Label Graph*)**.** By following the method adopted in [54], the graph between papers and labels/tags is a bi-partite graph indicating the relation between papers and corresponding labels. This graph encodes the supervised labeled information by capturing the category-level label co-occurrence relations. The network is represented by $G_{pl} = (P \cup L, E_{pl}, W_{pl})$ where P and L represent the sets of papers and labels, respectively, while E_{pl} and W_{pl} denotes the set of edges and their corresponding weights. W_{pl} represents the set of weighted edges between papers and labels, which is computed using TF–IDF (Term Frequency Inverse Document Frequency) scheme as follows:

$$W_{pl} = tf(l, p) * idf(l, D) \tag{6}$$

where tf represents term frequency of tag l in paper p, while idf denotes inverse document frequency. D represents the total number of papers.

Definition 7 (*Heterogeneous Information Network*)**.** Suppose we have a heterogeneous information network $G(C, E, W)$, where C represents the set of nodes. E and W denote the sets of edges and weights. An example of weighted heterogeneous network is depicted in Fig. 2.

We can notice that different objects (i.e., authors, papers, venues, topics, and labels) are linked with each other based on different relationships. For instance, paper P1 and paper P2 are connected based on citation relation. Additionally, the weights assigned to the edges represent the significance of the relation between the two objects. Similarly, paper P1 establishes relationships with venue *WWW*, tag *AI*, and topic *Ontology*. In contrast, the relationship between authors is maintained if they do collaboration research. In this research, the heterogeneous information

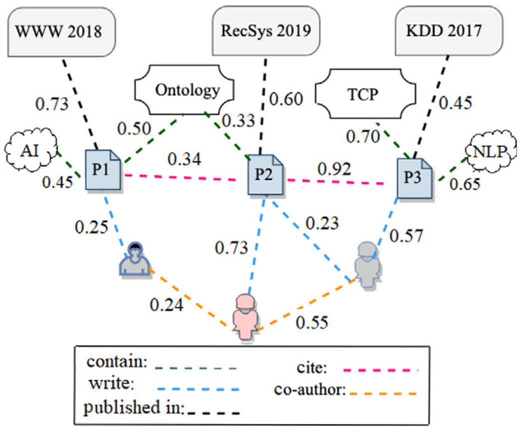


Fig. 2. A toy example of a weighted information network, where the weights between nodes represent the significance of the relationship between them.

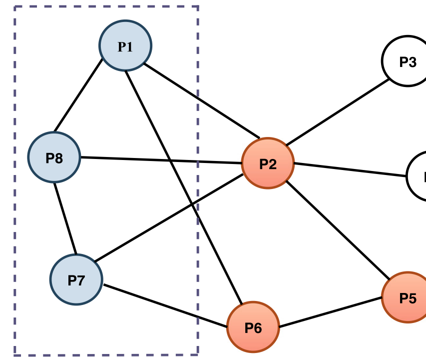


Fig. 3. A toy example of a citation network where vertices P2 and P5 are connected directly and hold first-order proximity, therefore they should be placed closely in the embedding space. Vertices P2 and P6 should also be kept close in the low-dimensional space as they share similar neighbors, such as P1 and P7.

network is comprised of six information graphs, as depicted in Fig. 1. Next, we define the problem statement as follows:

Problem Description: Given a user/researcher a and paper p, the proposed model aims to exploit the heterogeneous information network $G(C, E, W)$ and recommend *top-k* relevant research papers.

4. Proposed model

In this section, we present the main concepts and working of our proposed model. The model jointly learns the embedding of participating objects (i.e., authors, papers, topics, labels and venues) by exploiting the heterogeneous information network and presents a unified framework for generating personalized paper recommendations.

4.1. Learning bipartite network embeddings

In this section, we discuss the learning process of the proposed model for the bipartite network embedding. In literature, two proximity levels are generally considered in exploring the network, such as first-order proximity, and second-order proximity [42]. First-order proximity represents the local pairwise connectivity between vertices in the network. Naturally, each network embedding approach should maintain the first-order proximity, as it demonstrates a strong indication of similarity between nodes in the network. However, large networks are very diverse and such relations are very few. Therefore, to exploit the whole network, we employ the second-order proximity adopted in LINE [45]. That is, the proposed model uses the relationship between vertices which are not directly linked but possess similar neighbors. Second-order proximity assumes that vertices that have similar neighbors should be placed close to each other in the low-dimensional space. Fig. 3 illustrates the basic idea of the first-order and second-order proximity. Since vertices P5 and P6 are directly connected, therefore they possess first-order proximity and should be placed closely in the low-dimensional space. In contrast, vertices P2 and P6 are not directly linked but both have common neighbors, therefore they hold second-order proximity and should be kept close in the embedding space.

For instance, if we take a graph $G = (S_A \cup S_B, W)$, as a graph comprised of two disjoints sets, i.e., S_A and S_B, where the nodes in S_A that share many connections with S_B but are not directly linked with each other, would most likely possess the same distributions. Specifically, the conditional probability that node $v_j \in S_B$ is observed given node $v_i \in S_A$ is computed using Eq. (7).

$$p\left(v_j \mid v_i\right) = \frac{exp\left(\vec{v}_j^T \cdot \vec{v}_i\right)}{\sum_{c' \in S_B} exp\left(\vec{v}_j^T \cdot \vec{v}_i\right)} \tag{7}$$

where $\vec{v}_i$ and $\vec{v}_j$ are used to represent the embedding vectors of nodes v_i and v_j, respectively. For every $v_i \in S_A$, Eq. (7) is used to compute the conditional distribution for all corresponding nodes in set S_B. Also, a weight is linked with each edge, which determines the strength of the tie between any two vertices. Then, the model utilizes the conditional probability distribution to approximate the empirical distribution to retain the proximity of vertices (i.e., unlinked nodes) in the set S_A. The empirical distribution is computed by using the following Eq. (8):

$$\hat{p}\left(v_j \mid v_i\right) = \frac{w_{ij}}{\sum_{k \in N(i)} w_{ik}} \tag{8}$$

where $N(i)$ represents contexts set (i.e., out-neighbors of v_i) connected with v_i. In addition, w_{ij} denotes the associated weight of edge e_{ij}. The aim is to bring the probability score computed in Eq. (7) closer to the probability distribution shown in Eq. (8). To do so, the model minimizes the following objective function as defined:

$$O = \sum_{u_i \in S_A} \lambda_i D_{KL}(\hat{p}(\cdot \mid v_i), (p(\cdot \mid v_i))), \tag{9}$$

In Eq. (9), $\lambda_i = \sum_{k \in N_l(i)} w_{ik}$ denotes a regularization parameter employed for tuning the significance of v_i. $D_{KL}(. \mid .)$ represents the Kullback–Leibler divergence for empirical and conditional distributions. If we remove constants from Eq. (9), then it can be written as follows:

$$O = -\sum_{e_{i,j \in W}} w_{i,j} logp\left(v_j \mid v_i\right) \tag{10}$$

Similar to this loss function, we compute objective functions for different graphs discussed in Section 4.3.

4.2. Optimizing model using negative sampling

It is noteworthy that the optimization of $p(v_j \mid v_i)$ in Eq. (10) can be computationally expensive as we have to iterate through the entire set of nodes. To tackle this issue, we adopt negative sampling [55] over each edge to minimize computational complexity. More specifically, we employ the noisy distribution of each edge (i.e., $e_{i,j}$) individually to sample N negative edges defined in Eq. (11) as:

$$log\left(\vec{v}_j^T \cdot \vec{v}_i\right) + \sum_{k=1}^{N} W_{u_n} \sim p_n(u)\left[log\sigma\left(-\vec{v}_n^T \cdot \vec{v}_i\right)\right] \tag{11}$$

where σ represents the sigmoid function, its value for a variable y can be computed as $\sigma(y) = 1/(1+exp(-y))$, which gives value in range [0, 1]. Additionally, $P_n(v \propto d_u^{3/4})$ denotes the noise distribution, which selects negative samples based on unigram distribution in such a way that the occurrence of a node in the set does not depend on the occurrences of all other nodes. To further optimize Eq. (11), we employ Asynchronous Stochastic Gradient algorithm (ASGD) [56]. That is, during edge $e_{i,j}$ sampling process, the model adopts ASGD in such a way that the gradient of a node v_i over $\vec{v}_i$ embedding vector is calculated by using the following Equation:

$$\frac{\partial O}{\partial \vec{v}_i} = w_{ij} \cdot \frac{\partial logp\left(v_j \mid v_i\right)}{\partial \vec{v}_i} \tag{12}$$

We can notice that the weight w_{ij} representing the edge is multiplied with the gradient of $\vec{v}_i$. Problems may arise while tuning the learning rate when the weights of edges have a high variance. For instance, in a word co-occurrence network, some words co-occur many times (e.g., tens of thousands), while some words co-occur only a few times. In such networks, the scales of the gradients diverge and fining a good learning rate is harder. To address this problem, the proposed model samples a new edge by employing an alias table method adopted in [45], which brings time complexity to $O(1)$ results in effective embedding convergence. Finally, the edge sampling optimization approach [55] has the overall time complexity of $\mathcal{O}(N.E)$.

4.3. Joint learning of information graphs

The proposed heterogeneous information network consists of six graphs among authors, papers, venues, topics, and labels. The graphs which are related to the authors' relations include Author–Author, Author–Paper, and Author–Venue graphs. In contrast, the graphs that correspond to papers' relations contain Paper–Topic, Paper–Paper, and Paper-Label graphs. To exploit meaningful relations and semantics between users and papers, it would be intuitive to embed all these graphs jointly and try to minimize the following objective function:

$$O_{pr} = \underbrace{O_{aa'} + O_{ap} + O_{av}}_{\text{author graphs.}} + \underbrace{O_{pt} + O_{pp'} + O_{pl}}_{\text{paper graphs.}} \tag{13}$$

Next, we write each objective function presented in Eq. (13), where O_{pr} denotes the net loss of the proposed model.

$$Oaa' = -\sum_{e_{i,j} \in W_{aa'}} w_{ij} log(a_i \mid a_j') \tag{14}$$

$$Opt = -\sum_{e_{i,j} \in W_{pt}} w_{ij} log(p_i \mid t_j) \tag{15}$$

$$Opp' = -\sum_{e_{i,j} \in W_{pp'}} w_{ij} log(p_i \mid p_j') \tag{16}$$

$$Oap = -\sum_{e_{i,j} \in W_{ap}} w_{ij} log(a_i \mid p_j) \tag{17}$$

$$Oav = -\sum_{e_{i,j} \in w_{av}} w_{ij} log(a_i \mid v_j) \tag{18}$$

$$Opl = -\sum_{e_{i,j} \in w_{pl}} w_{ij} log(p_i \mid l_j) \tag{19}$$

To optimize the objective in Eq. (13), the model merges the edges of all graphs, and then at each step, the model is updated by sampling a new edge. The probability to sample an edge is based on the weight of the corresponding edge. Following this method, our model walks through the heterogeneous information network by exploring inner and outer vertices of different graphs along with weight influence. The overall joint training process is given in Algorithm 2.

Algorithm 2: Joint Training of the Model

Data: Graphs $G_{ap}, G_{pt}, G_{aa'}, G_{pp'}, G_{av}, G_{pl}$, number of samples S and number of negative samples N.
Result: $\vec{a}$: Author embedding vector, $\vec{p}$: Paper embedding vector, $\vec{l}$: Label embedding vector, $\vec{t}$: Topic embedding vector, $\vec{v}$: Venue embedding vector.

```
Initialize all embedding vectors a⃗, p⃗, v⃗, t⃗, l⃗.
while (iter ≤ S) do
    // Until we reach the required samples
    At every iteration on the corresponding graph, draw N
    negative edges
    S(w_aa') ∈ W_aa', update a.
    S(w_ap) ∈ W_ap, update a and p.
    S(w_av) ∈ W_av, update a and v.
    S(w_pt) ∈ W_pt, update p and t.
    S(w_pp') ∈ W_pp', update p.
    S(w_lp) ∈ W_lp, update l and p.
Return updated vectors for p, v, l, t, a.
```

4.4. Personalized paper recommendations

After learning the embeddings for candidate nodes in the heterogeneous information network, the next step is to make predictions for user a. That is, the aim is to recommend *top-k* research papers that meet researcher's a research preferences. The proposed method makes predictions by employing the score computation method defined as follows:

$$Q_{a,p} = \underbrace{\alpha \cdot (\vec{p}^T \cdot \vec{p_c})}_{\text{paper rel.}} + \underbrace{\beta \cdot (\vec{a}^T \cdot \vec{a_c})}_{\text{author inf.}} + \underbrace{\gamma \cdot (\vec{a}^T \cdot \vec{v_c})}_{\text{venue inf.}} + \underbrace{\theta \cdot (\vec{p}^T \cdot \vec{t_{cp}})}_{\text{topic rel.}} + \underbrace{\lambda \cdot (\vec{p}^T \cdot \vec{l_c})}_{\text{label inf.}} \tag{20}$$

where $\vec{p}$, and $\vec{a}$, denote the embedding vectors for query paper, and query paper's author, respectively. Similarly, $\vec{p_c}$, $\vec{a_c}$ $\vec{t_{cp}}$, $\vec{v_c}$, and $\vec{l_c}$ represent the embedding vectors of candidate paper, candidate paper's author, topic, venue, and label, respectively. Thus, the proposed model exploits multiple information networks by jointly learning embeddings to make predictions for a user. This way, the aim is to overcome the sparsity problem by employing auxiliary information sources. Also, the four parameters, such as $\alpha, \theta, \beta, \gamma$, and λ are employed for the regularization and tuning the importance of participating networks in the joint learning process.

5. Experiments and evaluation results

This section covers the evaluation of the recommendation results generated by our proposed model. That is, we assess the performance of our model against baseline approaches. For each target user, we partitioned the datasets into two sets namely: the training χ^t comprised of 80% of the papers and the test set χ^p containing 20% of the papers, where $\chi = \chi^t \cup \chi^p = \emptyset$. For a target researcher, we produced *top-k* recommendations using the papers available in χ^t. If the model recommends the ground

Table 2
Datasets specifications.

Datasets	Papers	Authors	Venues	Citation relations	Release time
DBLP	3,501,133	245,204	16,209	25,022,314	2019-05-05
AAN	21,455	17,342	312	113367	2016-12-01

truth in *top-k*, then we consider it an accurate recommendation, otherwise wrong. Further details on the datasets and metrics are given below.

5.1. Datasets

To conduct experiments, we used two commonly used datasets, namely the DBLP and the ACL anthology. Specifications and statistics of these datasets are given in Table 2.

DBLP: The DBLP dataset[1] consists of bibliography data in computer science and relevant domains. We used the latest version, i.e., DBLP-Citation-network V11. After removing papers with missing information, we have 3,501,133 research papers, 245,204 authors, 16,209 venues, and 25,022,314 citation relations. The dataset provides information such as papers' title, abstract, author, venue, keywords, and citation relations. To select labels/tags, one approach is to assign one label (i.e., machine learning, databases, NLP, etc.) to each paper considering their research field. However, this approach is more generic and it targets a very broad domain. To come up with a more fine-grained labeling, we used the keywords of research papers. As we know that authors assign keywords very carefully, and it is more appropriate to choose those keywords for the labeling task. Thus, we represented each paper by their corresponding keywords, which best represent the problem and domain of the paper. Moreover, to discover *top-k* meaningful topics with higher topic coherence, we employed a Gensim wrapper of LDA topic modeling from MALLET[2] following [12]. We extracted 11,135 meaningful topics using the abstracts of research papers.

ACL Anthology Network: ACL Anthology Network (AAN) dataset[3] contains papers related to computational linguistics, and NLP. The papers belong to different venues including: ACL, COLING, CoNLL, EACL, EMNLP, INLG, CL and TACL. After removing some papers with missing information, we have 21,455 research papers, 17,342 authors, 312 venues, and 113367 citation relations. For the labeling task, we adopted a similar approach as applied to the DBLP dataset. In contrast to the DBLP dataset, 1,757 topics are extracted using the full-text of research papers.

5.2. Evaluation metrics

To evaluate the performance of the proposed model against baseline methods, we employed three evaluation metrics such as recall, MAP, and MRR.

Recall: Recall is employed for measuring the percentage of original relevant papers which appear in the list of *top-k* recommendations. A system that has higher recall in lower *top-k*, produces more robust results. In our experiments, we choose $k = \{20, 40, 60, 80, 100\}$.

$$Recall = \frac{1}{Q}\sum_{j=1}^{Q}\frac{R_p \cap T_p}{T_p} \tag{21}$$

where Q denotes the total number of target articles and R_p represents the list of *top-k* recommendations produced against a target paper p.

Mean average precision: MAP measures the quality of recommendations by examining whether the relevant papers are listed in *top-k* results or not. To compute Average Precision (AP) for a query paper, we take the mean of the precision obtained after each Ground Truth Positives (GTP)) is retrieved, which is computed as follows.

$$AP@k = \frac{1}{GTP}\sum_{i=1}^{k}\frac{TPseen}{i} \tag{22}$$

where *TPseen* refers to the number of true positives seen till k.

MRR: Mean reciprocal rank analyzes the capability of a model to return a relevant paper in top k recommendations, defined as follows:

$$MRR = 1/Q_T \sum_{q\in Q_T} 1/rank_q \tag{23}$$

where Q_T denotes the testing set and rank(q) is the rank of its first ground truth paper.

5.3. Baseline approaches

In this section, we describe details about baseline paper recommendation models and variants of proposed model used for comparison. Further details of these models are given as follows:

- **LDA-TM** [12]. The model employs LDA topic modeling to the content of papers to generate paper recommendations. To generate topics, a Gensim wrapper of LDA topic modeling from MALLET is employed to the titles and abstracts of research papers. The model generates best results on the reported parameters settings. Therefore, we set the value of μ in the language model to 0.000001 and the optimal number of topics to 65.
- **ML-DTR Model** [11]. The ML-DTR Model encodes the articles text sequence into a vector representation employing GRUs. To be consistent with the ML-DTR, the content of articles are used to implement the model. The word embedding dimension is set to $K_w = 200$ and the first and second (output) layers of the RNNs has hidden states dimensions $K_{h1} = 200$, and $K_{h2} = 400$, respectively.
- **NNRank** [15]. It generates the vector representations of candidate papers and query manuscript employing a neural model, then cosine similarity between these vector representations is employed to recommend relevant papers. To conduct experiments, we adopted the parameters setting used in the model and code available at[4].
- **GAN-HBNR** [57]. Exploits network structure and relevant content using doc2vec and Denoising Autoencoder (DAE) to generate citation recommendations. In experiments, the generator used is of three layers, while the hidden state of DAE is set to 100. Additionally, the size of first two layers is set to 400, and the final layer has a size equal to the vocabulary. Additionally, we set the learning rate to 0.01 and the activation function used is tanh.
- **BNR** [17]. BNR is a bibliographic network representation learning method which employs both papers information along with authors information to generate citation recommendations for a query manuscript. To perform a fair comparison, we set the parameters as follows: dimensions=128, context size=10, and walks per vertex= 80.

[1] https://www.aminer.cn/citation.
[2] http://mallet.cs.umass.edu/.
[3] https://acl-arc.comp.nus.edu.sg/.
[4] http://labs.semanticscholar.org/ citeomatic/.

Table 3
Performance evaluation on the DBLP dataset.

Methods	MAP	MRR	Rec@20	Rec@40	Rec@60	Rec@80	Rec@100
LDA-TM	0.213	0.229	0.471	0.535	0.598	0.637	0.665
ML-DTR	0.263	0.288	0.508	0.575	0.628	0.668	0.705
GAN-HBNR	0.283	0.307	0.532	0.614	0.660	0.712	0.747
BNR	0.295	0.332	0.556	0.625	0.677	0.724	0.768
NNRank	0.526	0.571	0.520	0.579	0.636	0.672	0.715
PR-HNE	0.562	0.638	0.588	0.654	0.720	0.764	0.802

Table 4
Performance evaluation on the ACL Anthology dataset.

Methods	MAP	MRR	Rec@20	Rec@40	Rec@60	Rec@80	Rec@100
LDA-TM	0.238	0.243	0.551	0.617	0.652	0.693	0.712
ML-DTR	0.275	0.293	0.543	0.634	0.685	0.714	0.738
GAN-HBNR	0.291	0.317	0.565	0.663	0.704	0.748	0.763
BNR	0.304	0.353	0.575	0.673	0.714	0.755	0.772
NNRank	0.541	0.615	0.553	0.644	0.695	0.724	0.748
PR-HNE	0.593	0.656	0.604	0.692	0.753	0.794	0.820

- **Versions of our Model.** To evaluate the influence of papers relations, authors information, labels information, topical relevance, and venue information, we used five versions of our proposed model, i.e., PR-HNE$_{V_1}$, PR-HNE$_{V_2}$, PR-HNE$_{V_3}$, PR-HNE$_{V_4}$ and PR-HNE$_{V_5}$.
  - PR-HNE$_{V_1}$ is a relatively simplified version which employs the paper–paper network only.
  - PR-HNE$_{V_2}$ previous version is enriched by incorporating authors information.
  - PR-HNE$_{V_3}$ is the version that exploits author–venue networks, where we do not consider the paper-label, and paper–topic networks.
  - PR-HNE$_{V_4}$ is a modified version of the PR-HNE$_{V_3}$, where the previous version is enriched by incorporating paper–topic network.
  - PR-HNE$_{V_5}$ is the final proposed model, where all information graphs including paper-label are exploited in producing paper recommendations.

5.4. Comparison with other approaches

In this section, we present a comparison of baseline approaches against our method in terms of MAP, MRR, and Recall. Results in Table 3 reveal that the recommendations of the proposed approach yielded more precise results compared to all baseline approaches on the DBLP dataset. LDA-TM performs poorly in all baseline approaches because it exploits merely the content of papers and does not consider papers' citation proximity and other auxiliary information sources. In contrast, the NNRank model has produced second superior results in terms of MAP and MRR, its significance is attributed to its capability of exploiting meta data, such as abstracts, title and venue information to generate paper recommendations. However, the results generated by the proposed PR-HNE model outperform NNRank by achieving approximately 4% and 6% better results in terms of MAP and MRR, respectively. This is because the NNRank does not consider prominent factors such as papers' topics, and contextual information to produce more robust results. Additionally, it does not utilize the significance of relations among the network objects to produce quality recommendations.

The results on the ACL anthology are given in Table 4, which shows that all the models have produced better results compared to the results generated on the DBLP dataset. This is due to the fact that the ACL anthology provides rich content, which has a great impact on the results. It is apparent from the results that the BNR produces second-best results compared to other models in terms of Recall@100, since it utilizes the content of vertices along with network structure. In contrast, LDA-TM and ML-DTR generate insignificant results comparatively. This is because these models employ merely the textual content and therefore lack in terms of utilizing network structure and auxiliary side information to produce more robust results. However, the results generated by our model significantly outperformed all competitors, since our model employs the recently emerged HINs embedding approach by utilizing papers' citation proximity, authors' information, venue information, labels, and topical relevance. Finally, Fig. 4(b) shows that our model yielded almost 5% better recall score compared to the second best model BNR. It proves the fact that exploiting six information graphs can effectively generate diversified results.

Table 5
Impact of incorporating relation networks on the results.

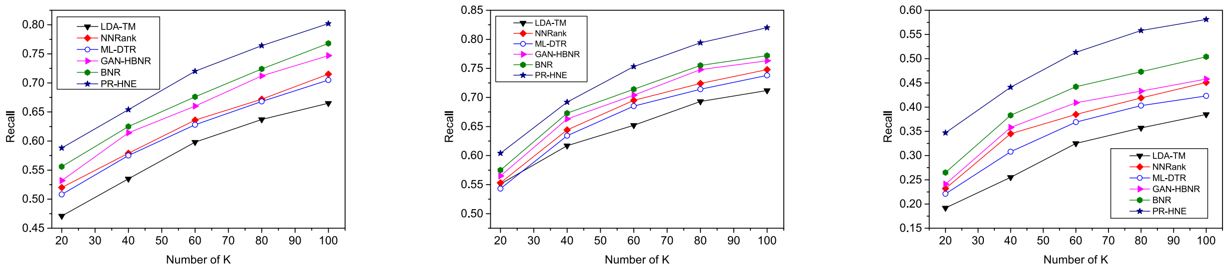
Dataset	Model	MAP	MRR	Rec@20
DBLP	PR-HNE$_{V_1}$	0.502	0.548	0.502
	PR-HNE$_{V_2}$	0.527	0.583	0.533
	PR-HNE$_{V_3}$	0.534	0.589	0.536
	PR-HNE$_{V_4}$	0.556	0.613	0.557
	PR-HNE$_{V_5}$	0.562	0.638	0.588
ACL Anthology	PR-HNE$_{V_1}$	0.532	0.568	0.522
	PR-HNE$_{V_2}$	0.551	0.593	0.543
	PR-HNE$_{V_3}$	0.557	0.601	0.548
	PR-HNE$_{V_4}$	0.579	0.638	0.573
	PR-HNE$_{V_5}$	0.593	0.656	0.604

5.5. Ablation study

To analyze the impact of different information networks, we conducted an ablation study on the DBLP and ACL Anthology datasets as depicted in Table 5. Specifically, we measured the impact of each information network on the significance of the PR-HNE. In this study, we examined five information networks such as paper–paper, author–author, paper–topic, paper-label, and author–venue. The results on the DBLP dataset demonstrate that the most important network is the authors relation network, as it has greater influence on the results. Additionally, the paper–topic and paper-label network show almost even performance on the DBLP dataset. In contrast, paper–topic and authors networks influence the results significantly on the ACL anthology dataset. Compared to the DBLP, the paper–topic network has a significant influence on the ACL anthology results. This is because ACL anthology provides rich content information, which helps the model to extract more meaningful topics and capture researchers' topical preferences. If we look at the results on the DBLP and ACL anthology datasets, it is evident that incorporating the author–venue network has less impact on the results comparatively. In a nutshell, exploiting authors' information and papers topics can better represent the preference dynamics of a researcher. Additionally, the results listed in Table 5 show that as we incorporate additional information networks, the performance of our model improves.

5.6. Cold-start paper recommendations

Cold-start paper problem arises when the system does not have enough information about the papers such as paper content (i.e., abstract, keywords, titles, etc.), citation relations, venue information, and author/s information, using which useful recommendations can be produced. We selected the DBLP dataset for choosing the cold-start papers since it is more sparse compared to the ACL anthology dataset. In the cold start papers' scenario, it is very challenging for the model to generate justifiable recommendations. In this connection, the proposed HNE model performs



(a) (b) (c)

Fig. 4. (a) Recall on the DBLP dataset (b) Recall on the ACL anthology dataset, (c) Recall on the cold-start papers.

Table 6
Analyzing the results by tuning parameters α, β, θ, γ and λ.

Dataset	α	β	θ	γ	λ	MAP	MRR	Rec@20
DBLP	0.1	0.2	0.2	0.4	0.1	0.532	0.605	0.533
	0.2	0.2	0.2	0.3	0.1	0.545	0.619	0.552
	0.2	0.3	0.2	0.2	0.1	0.558	0.622	0.558
	0.2	0.2	0.2	0.1	0.3	0.561	0.634	0.574
	0.2	**0.3**	**0.2**	**0.1**	**0.2**	**0.562**	**0.638**	**0.588**
	0.3	0.2	0.2	0.2	0.1	0.560	0.633	0.567
	0.2	0.2	0.3	0.2	0.1	0.561	0.635	0.579

well, as it utilizes auxiliary information sources such as citation proximity, authors' collaboration proximity, papers' topics, venues' information, and labels/tags to overcome the problem. To analyze the cold-start results, we used recall for different values of k as depicted in Fig. 4(c).

It is evident from the results that the recall score is degraded for all the models compared to the previous results on the DBLP and ACL anthology datasets that is natural, because there is limited information about the papers. The results demonstrate that LDA-TM has produced the least significant results comparatively. This is because it uses merely the content of research papers and ignore the network structure and other auxiliary information to overcome the cold-start problem. Though, BNR utilizes network structure and relevant content, its results are significantly lower compared to the PR-HNE, since it does not consider papers' labels, venue information, and topics relatedness. In contrast, our model has yielded the most significant results for the cold-start papers. It is attributed to the fact that the model captures the semantic representations of research articles and utilizes heterogeneous information which helped the model to tackle the cold-start problem.

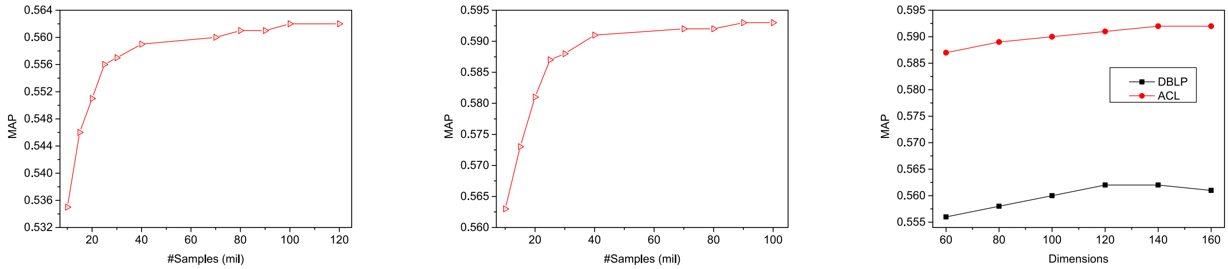
5.7. Parameters impact

This section examines the results of our model by tuning different parameters used in this paper. That is, we analyze the impact of (1) Samples and dimensions, and (2) regularization parameters, $\alpha, \beta, \theta, \gamma$, and λ. The behavior of the proposed model does not change greatly for different datasets to the number of dimensions. The MAP score increases with higher rate along with dimensionality until $d = 120$ for the DBLP and $d = 140$ for the AAN dataset as depicted in Fig. 5(c). After that, there is no significant change in the MAP score. In contrast, our model is more sensitive towards the sample size S and regularization parameters. Fig. 5(a) and Fig. 5(b) reveal the results on the DBLP and ACL Anthology datasets. It is evident that when sample size S is increased, the performance of our model converges. Therefore, it is better to keep samples size S large enough. We achieved better results by setting the value of S to 120 and 100 for the DBLP and ACL anthology, respectively.

On the other hand, the papers textual embeddings employed in paper–paper graph are generated using Sentence-BERT model, which fine-tuned the pre-trained BERT language model using a triplet loss function. Sentence-BERT encodes document sentences (i.e., title and abstract) and produces a final 768-dimensional paper embedding. Additionally, the model uses regularization parameters, i.e., α, β, θ, γ and λ for tuning the influence of different networks on the desired results. These parameters are used to check the impact of papers relations, authors' information, topics relations, venue information and labeled information, respectively. For simplicity, each parameter is assigned value in the range of [0,1], the overall sum of these values is equal to 1. We choose those values where the model achieves better results. The values of $\alpha, \beta, \theta, \gamma$ and λ on the DBLP are shown in Table 6. On the DBLP dataset, the quality of recommendation results are highly influenced by tuning the parameter of the authors relation network. It demonstrates that adding author information has great influence on the results. On the contrary, topic and author relation networks influence the results significantly on the ACL anthology dataset. The model gives best results when we set the values of $\alpha = 0.2, \beta = 0.3, \theta = 0.2$, $\gamma = 0.1$, and $\lambda = 0.2$ for the DBLP dataset as shown in Table 6. On the ACL anthology, we achieve best results by setting $\alpha = 0.2, \beta = 0.2, \theta = 0.3$, $\gamma = 0.1$, and $\lambda = 0.2$. The results reveal that tuning the parameter of author–venue network, i.e., γ, has less impact on the results comparatively. Additionally, the results given in Table 5 show that as long as we embed additional information networks, the performance of the model gets improved. Thus, our model mitigates the sparsity problem by exploiting multiple information sources regarding users or research papers.

6. Case study

In this section, we compare the recommendations generated by the proposed model against baseline approaches with the help of an example. That is, we present the recommendations generated by different models for a query manuscript, i.e., "A Novel Personalized Citation Recommendation Approach Basedon GAN", present in the DBLP dataset. This paper is generating citation recommendations by employing Generative Adversarial Network (GAN) to generate citation recommendations. To be specific, we present the top-5 results produced by the best performing models. As shown in Table 7, the results returned by our model have four records (i.e., marked as ✓) that match the ground truth of the query paper. On the other hand, both NNRank and BNR models have produced three accurate records. This is because both models cannot effectively exploit contextual information, topical relevance, and citation proximity while generating recommendations. In contrast, it is evident from the results that our model has produced more significant results in the top-5 results. The effectiveness of the PR-HNE is credited to its consideration of the most prominent factors and auxiliary



(a) (b) (c)

Fig. 5. (a) Sample size on the DBLP dataset (b) Sample size on the AAN (c) Dimensions size on the DBLP and AAN.

Table 7
Top-5 recommendations generated by three recommendation models for a query paper, where ✓shows an accurate recommendation generated by a model.

Title of the query paper	Models	Top-5 recommendation results
A Novel Personalized Citation Recommendation Approach Based on GAN	NNRank	1. Personalized Citation Recommendation via Convolutional Neural Networks. (✓) 2. Multi-step classification approaches to cumulative citation recommendation. 3. Multi-Modal Adversarial Autoencoders for Recommendations of Citations and Subject Labels. (✓) 4. Citation Worthiness of Sentences in Scientific Reports 5. CITEWERTs: A System Combining Cite-Worthiness with Citation Recommendation (✓)
	BNR	1. A LSTM based Model for Personalized Context-Aware Citation Recommendation (✓). 2. Improved Community Interaction Through Context Based Citation Analysis 3. CONTENT-BASED CITATION RECOMMENDATION (✓) 4. News Citation Recommendation with Implicit and Explicit Semantics. 5. A neural probabilistic model for context based citation recommendation (✓)
	PR-HNE	1. A Three-Layered Mutually Reinforced Model for Personalized Citation Recommendation (✓). 2. Multi-Modal Adversarial Autoencoders for Recommendations of Citations and Subject Labels. (✓) 3. Neural Citation Network for Context-Aware Citation Recommendation (✓) 4. RefSeer: a citation recommendation system 5. Generative Adversarial Network based Heterogeneous Bibliographic Network Representation for Personalized Citation Recommendation (✓)

information sources while generating recommendations. If we look at the results generated by PR-HNE in Table 7, it is worth pointing out that it contains papers published in recent years. The possible reason can be the consideration of the temporal dimension in paper–author network. In a nutshell, the results show that exploiting papers citation proximity, author information, papers topics relatedness, supervised labeled information, and venue information has produced more robust recommendation results comparatively.

7. Conclusions and future work

To find relevant research papers in accordance with researchers' information needs, most of the existing paper recommendation models ignore important factors such as papers citation proximity, authors collaboration proximity, papers' topics relatedness, and labeled information. Therefore, such models lack in generating quality recommendations, especially in cold-start papers and sparsity scenarios. To tackle these problems, we proposed a heterogeneous network embedding model called PR-HNE, which employed six information graphs and exploited the aforementioned factors while generating paper recommendations. That is, the model jointly learns embedding for papers and users employing the relationships between different weighted graphs and produce personalized recommendations. The experimental results of the proposed model against baseline approaches demonstrated that it outperformed state-of-the art citation recommendation models.

In the future, we will analyze the impact of introducing a deep neural network for a paper-label network to fine-tune the embeddings learned using unlabeled data. Meanwhile, to show the universal applicability of our model, we will conduct an online evaluation and analyze its significance on other datasets.

CRediT authorship contribution statement

Zafar Ali: Conceptualization, Methodology, Software, Data curation. **Guilin Qi:** Supervision, Conceptualization. **Khan Muhammad:** Conceptualization, Writing - review & editing. **Bahadar Ali:** Methodology, Software, Visualization. **Waheed Ahmed Abro:** Writing - original draft, Validation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Research presented in this paper was partially supported by the National Key Research and Development Program of China under grants (2018YFC0830200), the Natural Science Foundation of China grants (U1736204, 61602259), the Judicial Big Data Research Centre, School of Law at Southeast University, and the project no. 31511120201 and 31510040201.

References

[1] M. Kunaver, T. Požrl, Diversity in recommender systems–a survey, Knowl.-Based Syst. 123 (2017) 154–162.

[2] B. Walek, V. Fojtik, A hybrid recommender system for recommending relevant movies using an expert system, Expert Syst. Appl. 158 (2020) 113452.

[3] Z. Ali, S. Khusro, I. Ullah, A hybrid book recommender system based on table of contents (toc) and association rule mining. In: Proceedings of the 10th International Conference on Informatics and Systems, 2016, pp. 68–74.

[4] I. Ullah, S. Khusro, Social book search: the impact of the social web on book retrieval and recommendation, Multimedia Tools Appl. 79 (11–12) (2020) 8011–8060.
[5] M. Aliannejadi, D. Rafailidis, F. Crestani, A joint two-phase time-sensitive regularized collaborative ranking model for point of interest recommendation, IEEE Trans. Knowl. Data Eng. 32 (6) (2020) 1050–1063.
[6] G. Christoforidis, P. Kefalas, A. Papadopoulos, Y. Manolopoulos, Recommendation of points-of-interest using graph embeddings, in: 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), IEEE, 2018, pp. 31–40.
[7] Y. Yang, J. Hu, Y. Liu, X. Chen, Doctor recommendation based on an intuitionistic normal cloud model considering patient preferences, Cogn. Comput. (2018) 460–478.
[8] C. Angulo, Z. Falomir, D. Anguita, N. Agell, E. Cambria, Bridging cognitive models and recommender systems, Cogn. Comput. 12 (2020) 426–427.
[9] J. Beel, B. Gipp, S. Langer, C. Breitinger, Paper recommender systems: a literature survey, Int. J. Digit. Libr. 17 (4) (2016) 305–338.
[10] C. Wang, D.M. Blei, Collaborative topic modeling for recommending scientific articles, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, in: KDD '11, Association for Computing Machinery, New York, NY, USA, 2011, pp. 448–456.
[11] T. Bansal, D. Belanger, A. McCallum, Ask the GRU: Multi-task learning for deep text recommendations, in: Proceedings of the 10th ACM Conference on Recommender Systems, New York, NY, USA, 2016, pp. 107–114.
[12] M. Amami, G. Pasi, F. Stella, R. Faiz, An LDA-based approach to scientific paper recommendation, in: Natural Language Processing and Information Systems: 21st International Conference on Applications of Natural Language To Information Systems, NLDB 2016, Salford, UK, June 22-24, 2016, Proceedings, Springer International Publishing, 2016, pp. 200–210.
[13] F. Ayala-Gómez, B. Daróczy, A. Benczúr, M. Mathioudakis, A. Gionis, Global citation recommendation using knowledge graphs, J. Intell. Fuzzy Systems 34 (5) (2018) 3089–3100.
[14] L. Galke, F. Mai, I. Vagliano, A. Scherp, Multi-modal adversarial autoencoders for recommendations of citations and subject labels, in: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, 2018, pp. 197–205.
[15] C. Bhagavatula, S. Feldman, R. Power, W. Ammar, Content-based citation recommendation, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), Association for Computational Linguistics, New Orleans, Louisiana, 2018, pp. 238–251.
[16] H. Wang, W. Li, Relational collaborative topic regression for recommender systems, IEEE Trans. Knowl. Data Eng. 27 (5) (2015) 1343–1355.
[17] X. Cai, Y. Zheng, L. Yang, T. Dai, L. Guo, Bibliographic network representation based personalized citation recommendation, IEEE Access 7 (2019) 457–467.
[18] D. Roy, An improved test collection and baselines for bibliographic citation recommendation, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM), 2017, pp. 2271–2274.
[19] L. Yang, Z. Zhang, X. Cai, L. Guo, Citation recommendation as edge prediction in heterogeneous bibliographic network: A network representation approach, IEEE Access 7 (2019) 23232–23239.
[20] Z. Ali, G. Qi, P. Kefalas, W.A. Abro, B. Ali, A graph-based taxonomy of citation recommendation models, Artif. Intell. Rev. (2020) 1–44.
[21] S. Khusro, Z. Ali, I. Ullah, Recommender systems: issues, challenges, and research opportunities, in: Information Science and Applications (ICISA) 2016, vol. 376, Springer, 2016, pp. 1179–1189.
[22] J. Son, S.B. Kim, Academic paper recommender system using multilevel simultaneous citation networks, Decis. Support Syst. 105 (2018) 24–33.
[23] H. Zhang, F. Shen, W. Liu, X. He, H. Luan, T.-S. Chua, Discrete collaborative filtering, in: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, in: SIGIR '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 325–334.
[24] Y. Zhang, H. Wang, D. Lian, I.W. Tsang, H. Yin, G. Yang, Discrete ranking-based matrix factorization with self-paced learning, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, in: KDD '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 2758–2767.
[25] Y. Li, S. Wang, Q. Pan, H. Peng, T. Yang, E. Cambria, Learning binary codes with neural collaborative filtering for efficient recommendation systems, Knowl.-Based Syst. 172 (2019) 64–75.
[26] H. Wang, D. Lian, Y. Ge, Binarized collaborative filtering with distilling graph convolutional networks, in: Proceedings of the 28th International Joint Conference on Artificial Intelligence, AAAI Press, 2019, pp. 4802–4808.
[27] Y. Zhang, D. Liu, G. Yang, L. Hu, Quantization-based hashing with optimal bits for efficient recommendation, Multimedia Tools Appl. (2020) 1–18.
[28] M.J. Pazzani, D. Billsus, Content-based recommendation systems, in: The Adaptive Web, vol. 4321, Springer, 2007, pp. 325–341.
[29] P. Lops, M. de Gemmis, G. Semeraro, in: F. Ricci, L. Rokach, B. Shapira, P.B. Kantor (Eds.), Content-based Recommender Systems: State of the Art and Trends, Springer US, Boston, MA, 2011, pp. 73–105.
[30] G. Tian, L. Jing, Recommending scientific articles using bi-relational graph-based iterative RWR, in: Proceedings of the 7th ACM Conference on Recommender Systems (RecSys), 2013, pp. 399–402.
[31] T. Chakraborty, N. Modani, R. Narayanam, S. Nagar, DiSCern: A diversified citation recommendation system for scientific queries, in: 31st IEEE International Conference on Data Engineering, (ICDE) 2015, Seoul, South Korea, April 13–17, 2015, Seoul, South Korea, 2015, pp. 555–566.
[32] X. Kong, M. Mao, W. Wang, J. Liu, B. Xu, VOPRec: VEctor representation learning of papers with text information and structural identity for recommendation, IEEE Trans. Emerg. Top. Comput. (2019) 1.
[33] X. Cai, J. Han, W. Li, R. Zhang, S. Pan, L. Yang, A three-layered mutually reinforced model for personalized citation recommendation, IEEE Trans. Neural Netw. Learn. Syst. 29 (12) (2018) 6026–6037.
[34] Z. Ali, P. Kefalas, K. Muhammad, B. Ali, M. Imran, Deep learning in citation recommendation models survey, Expert Syst. Appl. 162 (2020) 113790.
[35] T. Ebesu, Y. Fang, Neural citation network for context-aware citation recommendation, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, pp. 1093–1096.
[36] J. Chen, Y. Liu, S. Zhao, Y. Zhang, Citation recommendation based on weighted heterogeneous information network containing semantic linking, in: 2019 IEEE International Conference on Multimedia and Expo (ICME), 2019, pp. 31–36.
[37] L. Yang, Z. Zhang, X. Cai, T. Dai, Attention-based personalized encoder-decoder model for local citation recommendation, Comput. Intell. Neurosci. 2019 (2019).
[38] A. Gazdar, L. Hidri, A new similarity measure for collaborative filtering based recommender systems, Knowl.-Based Syst. 188 (2020) 443–452.
[39] K. Sugiyama, M.-Y. Kan, Exploiting potential citation papers in scholarly paper recommendation, in: Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL), 2013, pp. 153–162.
[40] T. Achakulvisut, D.E. Acuna, T. Ruangrong, K. Kording, Science concierge: A fast content-based recommendation system for scientific publications, PLOS ONE 11 (7) (2016) 1–11.
[41] F. Xia, H. Liu, I. Lee, L. Cao, Scientific article recommendation: Exploiting common author relations and historical preferences, IEEE Trans. Big Data 2 (2) (2016) 101–112.
[42] P. Goyal, E. Ferrara, Graph embedding techniques, applications, and performance: A survey, Knowl.-Based Syst. 151 (2018) 78–94.
[43] S. Cavallari, V.W. Zheng, H. Cai, K.C.-C. Chang, E. Cambria, Learning community embedding with community detection and node embedding on graphs, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017, pp. 377–386.
[44] P. Cui, X. Wang, J. Pei, W. Zhu, A survey on network embedding, IEEE Trans. Knowl. Data Eng. 31 (5) (2019) 833–852.
[45] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, Q. Mei, LINE: Large-scale information network embedding, in: Proceedings of the 24th International Conference on World Wide Web, in: WWW '15, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2015, pp. 1067–1077.
[46] A. Grover, J. Leskovec, Node2vec: Scalable feature learning for networks, in: Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, USA, 2016, pp. 855–864.
[47] S. Gupta, V. Varma, Scientific article recommendation by using distributed representations of text and graph, in: Proceedings of the 26th International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2017, pp. 1267–1268.
[48] Q. Le, T. Mikolov, Distributed representations of sentences and documents, in: E.P. Xing, T. Jebara (Eds.), Proceedings of the 31st International Conference on Machine Learning, PMLR, Bejing, China, 2014, pp. 1188–1196.
[49] S. Ganguly, V. Pudi, Paper2vec: Combining graph and text information for scientific paper representation, in: J.M. Jose, C. Hauff, I.S. Altıngovde, D. Song, D. Albakour, S. Watt, J. Tait (Eds.), Advances in Information Retrieval, Springer International Publishing, Cham, 2017, pp. 383–395.
[50] L.F. Ribeiro, P.H. Saverese, D.R. Figueiredo, Struc2vec: Learning node representations from structural identity, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2017, pp. 385–394.
[51] N. Reimers, I. Gurevych, Sentence-BERT: Sentence embeddings using siamese BERT-networks, in: EMNLP/IJCNLP, 2019, pp. 3982–3992.
[52] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent Dirichlet allocation, J. Mach. Learn. Res. 3 (null) (2003) 993–1022.
[53] T. Huynh, K. Hoang, L. Do, H. Tran, H. Luong, S. Gauch, Scientific publication recommendations based on collaborative citation networks, in: 2012 International Conference on Collaboration Technologies and Systems (CTS), 2012, pp. 316–321.

[54] J. Tang, M. Qu, Q. Mei, PTE: Predictive text embedding through large-scale heterogeneous text networks, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, in: KDD '15, ACM, New York, NY, USA, 2015, pp. 1165–1174.

[55] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Proceedings of the 26th International Conference on Neural Information Processing Systems, 2013, pp. 3111–3119.

[56] B. Recht, C. Re, S. Wright, F. Niu, Hogwild: A lock-free approach to parallelizing stochastic gradient descent, in: Advances in Neural Information Processing Systems, 2011, pp. 693–701.

[57] X. Cai, J. Han, L. Yang, Generative adversarial network based heterogeneous bibliographic network representation for personalized citation recommendation, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018, pp. 265–272.