AUT Journal of Modeling and Simulation

AUT J. Model. Simul., 54(2) (2022) 197-210
DOI: 10.22060/miscj.2023.21212.5275

ISPREC++: Learning Edge Type Importance in Network-Oriented Paper Recommendation

E. Jafari[1], B. Shams[2], S.Haratizadeh[1]*

[1] Faculty of new sciences and technologies, University of Tehran, Tehran, Iran
[2] Faculty of Mathematical sciences, Alzahra University, Tehran, Iran

ABSTRACT: With the spread of the Internet and the possibility of online access to articles, a wide range of scientific articles are available to researchers, while finding relevant articles among this substantial number of articles turns out to be a real dilemma. To solve this problem, several scientific paper recommendation algorithms have been proposed. Most of these algorithms suffer from some drawbacks that limit their usability. For example, many of these recommendation methods are designed to recommend papers only to users who had published articles before and can't support new researchers. Also, they usually do not utilize many important features of articles each of which can have a role in determining the relevance of the articles to users. To address these concerns, in this paper, we present the novel method of Integrated Scientific Paper Recommendation with an edge-weight learning approach, called ISPREC++, as an extended version of ISPREC that focuses on learning the weights of edge types in Heterogeneous Information Networks based on users' preferences. ISPREC++ sets the weights of edges in SPIN using a Bayesian Personalized Ranking (BPR) based method and utilizes Gradient Descent to optimize its objective function. Thereafter, it exploits a limited random-walk algorithm for a Top-N recommendation. Extensive experiments on a real-world dataset demonstrate the significant performance superiority of ISPREC++ compared to the state-of-the-art scientific paper recommendation algorithms.

Review History:
Received: Mar. 13, 2022
Revised: Jan. 16, 2023
Accepted: Feb. 27, 2023
Available Online: Feb. 28, 2023

Keywords:
Heterogeneous Information Network
Random-Walk with Restart
Bayesian Personalized Ranking
Paper Recommendation
Recommender Systems

1- Introduction

The expansion of the Internet induces researchers from all over the world to access online versions of all articles presented at conferences, published in scientific journals, and on open-publishing platforms. This significant number of articles makes researchers confused to distinguish the high-quality and relevant articles from others. They usually search keywords and filter out the results to reach a reliable paper. However, it is a time-consuming task, and sometimes they cannot find the exact content they are required. Accordingly, an efficient approach to recommending scientific articles that can create high-quality recommendations seems to be critical [1]. This problem leads to forming Scientific Paper Recommendation approaches, with the aim of recommending similar scientific articles in the scientific community. These methods can be grouped into three categories: Content-Based Filtering (CBF), Collaborative Filtering (CF), and Graph-Based methods(GB) [1].

In the CBF methods, the content of articles (e.g., titles, abstracts, and keywords) is mostly considered to find papers relevant to the ones that a user has read [2]. On the other hand, the CF class of methods concentrates on other users' actions or ratings for articles and refers to similar users as "neighboring users"[3]. Based on the interests of neighboring users, the relevant papers are found and recommended to the target user [4, 5]. These two models complement each other's shortcomings because each does not utilize some important aspects of the data.

In graph-based recommendation methods, deciding how to design a graphical representation that accurately and plenary reflects each piece of valuable information is challenging, and another important point is how to exploit the network to generate recommendations. This class of methods first uses graph structures to model various entities in the data and demonstrate their relations by edges. Second, the recommendation is made by analyzing the graphical model [6]. At the step of graph construction, most of the algorithms in the SPR task use homogeneous networks that only reflect one type of association among entities (e.g., citation network, author network, etc.) [7, 8]. The other group exploits heterogeneous networks with two types of edges: user-paper and paper-paper. A user-paper edge reflects an interaction between a user and a paper, while a paper-paper link represents the aggregated similarity between a pair of papers. Unfortunately, both of these approaches, homogeneous and heterogeneous, fail to model the complex relations among users and papers in a scientific paper recommendation system. The first approach, homogeneous graph structure, suffers

*Corresponding author's email: haratizadeh@ut.ac.ir

from the lack of information as it can only model one type of relation. The second approach, heterogeneous graph, loses valuable information by aggregating all types of relations among papers into a single similarity relation. Generally, current graph-based scientific article recommendation algorithms fail to model and exploit different kinds of entities and relationships in their integrated network representation of data.

The final step in graph-based methods is analyzing the graph to make recommendations. Some algorithms used graph structure to learn the representation of each user and paper based on their relationship in the network, then make recommendations based on their embedding [9]. Other popular methods in graph-based methods for generating recommendations are using random walks and meta-paths. In random walk-based methods, the importance of nodes (nodes with many inlinks) is considered in a recommendation. Also, meta-path-based methods try to determine the weight of different meta-paths. However, no proposed algorithm considers the importance of nodes and paths in an integrated framework. As each of these methods has some benefit in finding the users' preferences, we combined them in an integrated framework for the first time in the SPR task. To do that, first, the optimum weights of the edge types are learned, which determine various meta-paths importance. Then, the transition matrix is built similarly to random-walk methods for having the importance of nodes either.

In this paper, we present a novel graph-based algorithm, called Integrated Scientific Paper RECommendation system with an edge-weight learning approach (ISPREC++), which is an extended version of ISPREC [10]. We added a learning process to ISPREC to complete its shortcoming in not considering the weights of different features. Our proposed framework consists of different types of nodes and relations in a structure to get the most out of it. After the network construction, we need an algorithm to recognize and analyze the reliable paths and meta-paths in the network to make personalized recommendations. As each of these meta-paths has distinct importance in determining the value of articles for users, we used Bayesian Personalized Ranking(BPR) for reaching the best weights of their link types. To address this issue, we exploit a limited random walk with restart algorithm to rank papers and then recommend papers with the highest rank to each target user.

The main contributions of this paper are outlined as follows:

We develop a novel framework that considers both the importance of edge types and nodes. We learn the weights of different edge types using a BPR-based method and a novel negative sampling method for the first time in the SPR.

Our framework is evaluated on a real-world dataset, and the results show a significant improvement over the state-of-the-art algorithms. We also report a set of experiments to evaluate the performance of different modules of our algorithm separately.

The rest of the paper is organized as follows: In Section 2, the related work in the field of recommending scientific papers and recommendations in HINs are summarized. In section 3, we briefly explain the Preliminaries and Notations used in our algorithm, and the details of our proposed framework are explained in Section 4. Section 5 presents the experiments and results as well as descriptions of the dataset, metrics, and evaluation methods. Finally, the paper is summarized and concluded in Section 6.

2- Related Work

In this section, we intend to investigate related works in the scientific paper recommendation domain and a survey on related recommendation methods on HINs.

2- 1- Scientific Paper Recommendation

Scientific paper recommendation methods from the end-user perspective can be divided into two categories: user-oriented and author-oriented. The author-oriented group of studies is only able to recommend articles to researchers who have at least one published paper. So, they limit users to only authors with some published papers and ignore the rest. Their recommendation is based on published papers of the target author, and they try to recommend papers that are kind of similar or relevant to them. Furthermore, they suffer from other weaknesses: first, the quality of recommendation declines when the number of articles published by the author is not enough to understand his/her research interests accurately. Second, they cannot recommend articles if the users start exploring other domains of research. Since our method does not place in this category, so we skip more details of these algorithms, although some of them have used graph-based approaches [9, 11].

The user-oriented group of algorithms can propose a recommendation as soon as a user registers and begins reading/searching articles. They assume that users have either published or read papers before and consider users and authors of papers as two separate entities. Our paper belongs to this category of algorithms and presents a novel approach for recommending papers to user-oriented scientific paper recommendation problems.

The early user-oriented scientific paper recommendation methods (USP-Recommendation) make recommendations based on the user's current Paper Of Interest (POI) [12-15]. These algorithms can be regarded as a general item-based recommendation in a session. They only exploit the relationship between papers to the POI and bypass the other available information concerning the interests of users. These algorithms are not practical in a real-world scenario in which users have not determined a certain specific paper as their POI.

The modern group of algorithms is more personalized: some algorithms follow a model-based algorithm to make a recommendation. For instance, it [16] proposes a two-level attentive neural network to capture the similarity between each word and each sentence to the title of an article. Two other algorithms, [17, 18], integrate social tag information and social friend information separately for Content-Based Filtering and Collaborative Filtering to enhance their performance, then make recommendations by combining

the result of both approaches. Another algorithm [19] learns the latent space of an article's content with two parallel autoencoders.

Some other algorithms of personalized algorithms in this class follow graph-based approaches: One algorithm exploits three different implicit social networks and two explicit social networks to calculate similarity among users and follows a collaborative filtering approach for recommending [7]. Our method differs from this algorithm as it only models users' relationships in a homogenous network structure, but we exploit a heterogeneous information network to simultaneously reflect different types of relationships among all entities of the system.

In another study, an algorithm called CARE [5] is proposed that constructs a graph of users and papers including user-paper interactions and paper-paper co-authorship links. Then, it exploits a random walk with a restart method to estimate the relevancy of articles to the target user. The differences between this method and ours are in the structure of the graphs, the fact that they only consider one type of use, and assuming that users prefer to choose articles based on their authors, while we consider different users, with different manners in choosing articles. For instance, a user follows illustrious venues, while, another one is interested in the papers with relevant keywords and tags.

UAGMT is another algorithm [20] that builds a bi-relational graph with users and papers as nodes and user paper interaction, citation, and content similarity as discrete edges. It uses content similarities to overcome the cold start problem and generates recommendations using random walk with the restart. In their algorithm, papers are connected together according to their content similarity (title, abstract, and tags) and citations while our approach considers a rich set of information (e.g. authors, venue, tags, and citations) about papers in the heterogeneous information network.

2- 2- Recommendation methods on HINs

There are several recommendation methods for HINs and most of them are based on random-walk. The random walk model assumes a random walker recursively moves to a neighbor in the network and it continues until the probability of locating the walker become converges. There are lots of similar methods to this model, such as the Local Random Walk [21], the Path-constrained random walk [22], HeteSim [23], and PathMining [24].

Random walk with restart is an important variant of random walk. In this method, a random walker moves to a random neighbor with probability A and returns to the starting node with probability 1-A recursively. This type of model was traditionally utilized in PageRank [25] and was familiar to recommender systems by Personalized PageRank [26]. Typical work includes ObjectRank [27], ItemRank [28], and the HIN-based method PathRank [29].

In SPR, these algorithms have a slight role, and researchers in this area have not paid much attention to these methods. Some of them only used simple random walk with restart for recommendations, but no SPR study utilize an algorithm for learning the weights of edge types. In addition, in all the mentioned methods above, they do not consider the importance of edge types, and edge weighting only depends on the number of outcoming edges from a specific node. Although in [30] they learned the probability of moving from different node types, they ignored the total number of out links and do not consider that the number of each edge type for different nodes could vary significantly. So, when the single probability that they learned spread between a various number of edges, limits their performance.

In ISPREC [10] a limited random-walk with restart is utilized to aggregate all meaningful meta-paths and avoid generalization, however, the importance of each meta-path is not regarded. So in this work, we extend ISPREC by learning the weights of links using Bayesian Personalized Ranking to contrast the importance of various relations in users' preferences for choosing articles. In the next sections, our novel framework is introduced in detail.

3- Preliminaries and Notations

Definition 3-1 Heterogeneous Information Network [31]. A heterogeneous information network is represented as a directed graph $G = (V, E)$ with two mapping functions: 1) a node type mapping function $\varphi : V \to A$, where each node $v \in V$ belongs to one particular node type $\varphi(v) \in A$, and 2) an edge type mapping function $\phi : E \to R$, and each edge $e \in E$ belongs to a particular relation $\phi(e) \in R$.

If the types of nodes $|A| > 1$ or the types of relations $|R| > 1$, the network is described as heterogeneous information network; otherwise, it is a homogeneous information network.

Definition 3-2 Network Schema [31]. The network schema is a meta template for a heterogeneous network $G = (V, E)$ with the object type mapping $\phi : V \to A$ and the link mapping $\psi : E \to R$, which is a directed graph defined over object types A, with edges as relations from R, denoted as $T_G = (A, R)$.

In HINs, two nodes may be connected by different paths, which may have various lengths and may contain different node types and link types. The concept of the meta path was proposed to describe the path types in a HIN [31].

Definition 3-3 Meta Path [31]. A meta path P is a path defined on the graph of network schema $T_G = (A, R)$ and takes the form $A_0 \xrightarrow{R_1} A_1 \xrightarrow{R_2} \cdots \xrightarrow{R_k} A_k$, which defines a composite relation $R_1 \circ R_2 \circ \cdots \circ R_k$ between node types A_0 and A_k, where $\circ$ denotes the composition operator on relations.

The length of a meta path P is the number of relations in it. In this paper, we use names of node types to denote the meta path: $P = A_1 A_2 \ldots A_k$ For instance, *UPAP* is used to denote the meta path *User* − *Paper* − *Author* − *Paper* in Fig. 1, whose length is 3.

Each meta path has its corresponding semantics [32]. For instance, the meta path *UPUP* explores the social relationships of users who read articles in common with the target user, while the meta path *UPAP* utilizes paper-author links to build relationships between users and papers.

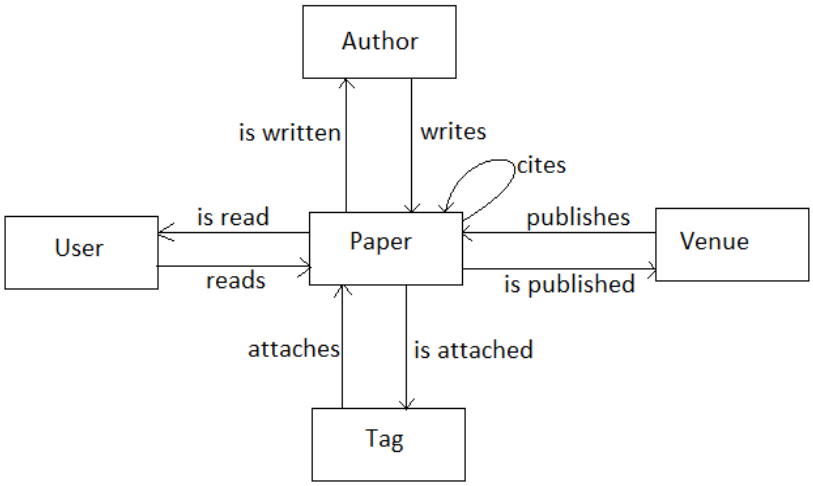

Fig. 1. SPIN Schema

4- ISPREC++ Framework

In this section, first, we describe the Network Construction and intuition behind the proposed structure. Then, our recommendation method and limited random walk with restart are explained. After that, we present our learning method for edge weights and the detailed algorithm.

4- 1- Network Construction

The most important thing about recommending articles to users is considering the various factors that users may have in mind when looking for an article to read. One clear factor is authorship, since researchers usually tend to follow the research of each other [5]. They want to be aware of the other articles by their favorite author. It occurs more if the author is prominent and reputable in his or her expertise.

In addition to authorship information, the content of the paper, especially the field of research that it belongs to is an important factor in selecting relevant papers. The sources of information for this decision factor include the title, abstract, or keywords of the paper, as well as the tags that users assign to articles. Tags reflect the domain of the article and the method that is used in it. Another way to find out about the scope of an article is to consider its venue of publication. Most researchers follow some important and reliable conferences or journals in their expertise to learn about the new studies in their field. Besides, when researchers find a relevant article, they analyze its citations to perceive what has been done before [33].

The challenging part is how to model such data to get the most out of it. One possible approach is to build a homogeneous network of papers and connect them based on their aggregated similarities. Another solution is to construct a bi-relational graph of users and papers and also connect papers based on their accumulative similarities. But as it is mentioned before, these approaches (similar to [5, 20]) fail to distinguish different entity relations. To keep as much information as possible in our graphical model of data, we develop our graph structure, SPIN, a heterogeneous information network that is shown in the Fig.1.

Using SPIN, users can access papers with different meta paths. For simplicity, we use U as the user, P as the paper, A as the author, V as the venue, and T as the tag. U-P-U-P denotes a meta-path that reflects the collaborative filtering approach, which means that the user is linked to a paper through a similar user who has a common read paper with the target user. The other meta path is U-P-A-P which connects the user to a paper with the same author. Similarly, U-P-T-P and U-P-V-P connect the user to a paper with the same tag and venue. U-P-P connects the user to a new paper cited by a paper that the user has read. The intuition of meaningful meta paths is summarized in Table 1.

4- 2- Recommendation Generation

To generate recommendations, we need to make sure that the suggested meta paths can reflect reliable relations among users and papers. The other point is that all the paths do not have equal importance, and the effect of paths that pass through important nodes should be more than others. Also, we investigate the appropriate length of paths to be used in our analysis.

A common technique for making recommendations with considering the importance of nodes is to use random walk with restart [5, 20]. In this method, for generating personalized recommendations, it is assumed a random walker starts from the node i that refers to the target user, and recursively moves to a random neighbor in the graph with probability α and returns to the node i with probability $1-\alpha$. The iterative equation is shown as (1). The algorithm calculates the probability for each node to be visited by the random walker at some random step t and represents those probabilities, for all the nodes, as a vector r. Then by sorting the probabilities of the vector r, the nodes with the highest

Table 1. Relations between the read paper by the user and the proposed paper

Corresponding Meta-path in SPIN	Intuition in scientific recommender systems
U-P-P	a citation of the paper
U-P-P-P	a citation of a cited paper
U-P-A-P	Two papers are written by the same author
U-P-V-P	Two papers are published by the same venue
U-P-T-P	The Same tags are assigned to both papers
U-P-U-P	Two papers are read by the same user

probability values will be recommended to users. This method is usually referred to as the Personalized PageRank approach [26]. The superiorities of this method that lead to its popularity in recommender systems are distinct. This method assists in finding similarities and closeness between nodes. It can indicate the distance between a target user and all the articles concerning the importance of nodes in the path to reaching them. In addition, this method considers all the existing meta paths between a user and a paper, so it is kind of an aggregation of all the meta paths that we mentioned in Table I. In addition, adding weights to the edges that distinguish meta-paths, could merge both differences in the importance of meta-paths and important nodes together.

Since the maximum length of our suggested meta-paths is 3, using random walks longer than 3 steps is not necessary and can even be misleading [31, 34, 35]. So, in this paper, a limited random walk with restart technique is used [30].

For applying limited random walk with restart, we first build the adjacency matrix of the SPIN graph, called P in which each row or column represents a node and $P[i][j]$s defined as follows:

$$P[i][j]=\begin{cases} p\left(\varphi(v_i)|\varphi(v_j)\right) & \text{if there is an edge between node } i \text{ and } j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The weight of outlinks of papers is determined by the learning algorithm and other probabilities are 1, then we normalized the matrix columns.

Assume a random walker starts from node i and at each step chooses a neighbor node randomly with probability α and moves to that node, or alternatively, decides to return to the starting node i with probability $1-\alpha$. We define $r^{(t)}$ as the distribution probability vectors that indicate how probable each node is to be seen in step t. The primary value of r refers to $r^{(0)}$ and captures $r_i^{(0)} \geq 0$ and $\sum_i r_i^{(0)} = 1$. Updating rule of each iteration follows (2), in which t refers to the iteration step.

$$r^{(t+1)} = \alpha P r^{(t)} + (1-\alpha) r^{(0)} \quad (2)$$

To make recommendations to a target user T, we set the initial $r_T^{(0)} = 1$ while the rest of the vector is set to 0, and $\alpha = 0.8$. Then we apply (2) three times. The resulting probabilities in r vector corresponding to paper nodes are then used to sort the papers. Finally, the papers that the user has read before are excluded from the sorted list and the Top-N highly ranked papers are recommended to the user.

4- 3- Learning weights of different edges typed in SPIN

Far now, we explained how to construct a heterogeneous information network that can correctly model important factors and their relations in an integrated scientific recommendation framework. However, it is still under question how to determine the weight of each factor based on their importance in a network-oriented paper recommendation framework. In the following, we show how this problem can be through learning the importance of edge-type where maximizing BPR loss is our ultimate goal. It should be noted that BPR loss is one of the major important objective functions when our recommendation framework relies on implicit feedback.

Typically, model-based recommendation algorithms with implicit feedback seek to minimize the BPR Loss function. BPR is originally introduced by [36] and it approximates the number of irrelevant (unread papers) which ranked higher than relevant items (i.e. read papers) with a differentiable function.

Formally, $Loss_{BPR}$ is calculated through Eq.3 in

$$Loss_{BPR} = \sum_{(u,i,j)\epsilon D_s} \ln \sigma\left(r_{ui} - r_{uj}\right) \quad (3)$$

ere $D_s = \{(u,i,j) | i \epsilon I_u^+ \wedge j \epsilon I / I_u^+\} I_u^+$ are the items that the user u has observed and I is the set of all items. $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function and $\widehat{r_{ui}} = f(\theta, u, i)$ is the estimated relevance score of item i for the user u with regards to θ, a vector indicating system parameters.

In our recommendation framework, θ indicates the importance of each meta-path, and, or equivalently importance of each edge type. To clarify it more, let us consider the meaningful meta-paths represented in Table 1. As you can see, all meta-paths start from a user to a paper that the user has read before. As the users' node type is only connected to the papers' node type, the probability of moving from users to papers is 1. At the next step of the random walk, the random walker has to choose one node from the five node types that papers are connected to them. At this point, by learning the outlinks of papers, we can distinguish the probability of different meta-paths. After that, the random walker is located at one of the node types U, A, V, T, and P. From node types except P, it has only one choice to come back to node type P, and from node type P, the only meaningful link type is the citation type that brings the rand om walker to the P node type. Therefore, to learn the importance of each meta-path, it is enough to learn, P_τ denoted the weight of edge type $\tau \in R = \{PU, PA, PV, PT, PP\}$ and update the transition matrix based on P_τ.

4- 3- 1- Using gradient descent for BPR optimization

We utilize Mini-Batch Gradient Descent [37] to optimize the objective function and the updating rule is defined below where η is the learning rate.

$$\theta \leftarrow \theta + \eta . \sum_{(u,i,j)\epsilon Mini-Batch} \frac{1}{1 + e^{\widehat{r_{ui}} - \widehat{r_{uj}}}} . \left(\frac{\partial}{\partial P_\tau} \widehat{r_{ui}} - \frac{\partial}{\partial P_\tau} \widehat{r_{uj}} \right) \quad (4)$$

So, it is important to obtain how $\widehat{r_{ui}}$ elates to P_τ or each edge type $\tau \in R = \{PU, PA, PV, PT, PP$ and how its gradient $\frac{\partial}{\partial P_\tau} \widehat{r_{ui}}$ ould be calculated.

4- 3- 2- Gradient calculation

For estimating $\hat{r}$, we used the limited step random walk with restart and limited the steps to 3. We ignored meta-paths with one step, which is the step from a user to a paper that the user has already read, and it only increases the value of $\hat{r}$ for that paper. Our goal is to maximize the distance between items that users have seen before and the rest. Ignoring that meta-paths would increase our accuracy and avoid misleading. So, we assume:

$$r_{ui} = \alpha^3 \left(P^3\right)_{iu} + \alpha^2 (1-\alpha) \left(P^2\right)_{iu} \quad (4.1)$$

and

$$\left(P^3\right)_{iu} = \sum_{v_p, v_q \epsilon V} p(v_i | v_p) . p\left(v_p | v_q\right) . p\left(v_q | v_u\right) \quad (4.2)$$

$$\left(P^2\right)_{iu} = \sum_{v_p, v_q \epsilon V} p(v_i | v_p) . p\left(v_p | v_u\right) \quad (4.3)$$

the transition matrix P, we define:

$$P_{ij} = p\left(v_i | v_j\right) = \frac{1}{\sum_{v_k \epsilon D_j} w_{kj}} . p\left(\varphi(v_i) | \varphi(v_j)\right) \quad (4.4)$$

$$D_j \stackrel{\text{def}}{=} \{v_k \mid e(v_j, v_k) \in E\} \quad (4.5)$$

Regarding these considerations, the gradient for each edge type is:

$$\begin{aligned} \frac{\partial r_{ui}}{\partial p_\tau} = \alpha^3 . (& \sum_{\psi(ip)=\tau} w_{pq} w_{qu} . p_\tau + \\ & \sum_{\psi(pq)=\tau} w_{ip} w_{qu} . p_\tau + \sum_{\psi(qu)=\tau} w_{ip} w_{pq} . p_\tau) + \\ & \alpha^2 (1-\alpha) . \left(\sum_{\psi(ip)=\tau} w_{pu} . p_\tau + \sum_{\psi(pu)=\tau} w_{ip} . p_\tau \right) \end{aligned} \quad (5)$$

We used w_{ip} and $\psi(ip) = \tau$ as $p(v_i | v_p)$ and $\psi\left(e(v_p, v_i)\right) = \tau$ espectively, and p_τ as the current weight of edge type τ. As we require determining the probability of papers' outlinks, we can abbreviate the above equation to:

$$\begin{aligned} \frac{\partial r_{ui}}{\partial p_\tau} = & \alpha^3 . \left(\sum_{\psi(pq)=\tau} w_{ip} w_{qu} . p_\tau \right) + \\ & \alpha^2 (1-\alpha) . \left(\sum_{\psi(ip)=\tau} w_{pu} . p_\tau \right) \end{aligned} \quad (6)$$

In every batch, for each user u and his positive item $i \in I_u^+$, we sample a negative sample $j \in I / I_u^+$ and make a triplet we consider all item $i \in I_u^+$ one $j \in I / I_u^+$ as a negative sample in the training dataset to compute the gradients of $\widehat{r_{ui}}$ nd $\widehat{r_{uj}}$ We will discuss our negative sampling strategy in the next section. Then we update the learning parameters p_τ and reduce the learning rate after batch updating. And the learning scheme is summarized in Algorithm 1. At the end of each iteration, the parameters are normalized so that the sum of all the parameters is 1.

Algorithm 1 Weight Calculation in ISPREC++

Input: SPIN = (V, E), the training set DS, the parameters α ,β and k, the initial learning rate η

Output: the learned parameters $\Theta = \left(p_{\tau_1}, p_{\tau_2}, \ldots, p_{\tau_{|R|}}\right)$: initialize $\Theta = (\frac{1}{|R|}, \frac{1}{|R|}, \ldots, \frac{1}{|R|})$: repeat

3: grad = 0
4: for each u in users
5: for each i in trainingset(u)
6: j← random negative sample
7: grad += compute gradients
8: grad /= number of users
9: update parameters
10: normalize parameters
11: $\eta \leftarrow \eta / \beta$ 2: until convergence
13: return $p_{\tau_1}, p_{\tau_2}, \ldots, p_{\tau_{|R|}}$ . .

4- 3- 3- Negative Sampling

According to the BPR Loss function, we should minimize $Loss_{BPR} = \sum_{(u,i,j)\in D} \ln \sigma\left(\widehat{r_{ui}} - \widehat{r_{uj}}\right)$ or all triples (u,i,j) n which i is a read paper by the user u and j is an unread paper by him. It is obvious that the number of read papers is far lower than the number of unread papers, and it is unpractical to consider all pairs of read papers and unread papers which at computational complexity of $O(m_u n)$ here n is the number of items and m_u is the number of read papers by user u. Therefore, several recommendation systems randomly sample a small set of negative items. Recently, it has been shown that this approach, known as random negative sampling, is not a well-suited choice for implicit recommender systems.

To be more clear, let us categorize negative items into three groups: a) unseen items that might be read by users in the future. b) The group of unseen items that obviously are not in the interest of the user. C) Negative items that are not in the interest of the user but are hard to distinguish among others.

The random negative sample approach typically selects a second group of items, called soft negative samples. It can be simplified through a toy example. Assume that you are a researcher in the area of recommender systems, and you are using Google Scholar to find your relevant articles. Google manipulates some articles in different disciplines and sub-disciplines, but it simply learns that you are not interested in geography, and it will not recommend those papers to you. If you consider the number of journals and the various topics they include, you will clearly understand that papers in the area of recommender systems constitute only a small percentage of the whole scientific papers. So, if the Google recommender system generates its samples randomly, the negative sample would probably lie in another discipline and Google only concentrates on an apparent example which adds no value to the system's intelligence.

On the one hand, if the sampling approach selects a negative sample k that belongs to the first category, it mistakenly pushes the recommendation algorithm to rank k less than positive items and among negative ones. Therefore, the recommender system cannot distinguish k as a positive item, and it will not be recommended to the user. Consequently, it is important to use a negative sampling method that focuses on the third group of negative items, which are called hard negative samples. To do that, we used a 3-step limited random walk with restart for choosing negative samples. For each user, we acquire a Top-K list and randomly choose items between ranks [100 ,300] as they potentially are informative to the systems. An item that is ranked among the top-100 items might be in the interest of the target user, and choosing them might confuse the recommender system. On the other hand, items that are not among the top-300 items, and are soft and useless to the system. Our idea is based on the proposed approach presented by [38] with the difference that it uses personalized PageRank, but we exploit limited random-walk scoring.

5- The results and discussion

We evaluate our proposed method on a real-world dataset for recommending papers to users with various evaluation metrics and compare the performance of our model with the state-of-the-art and baseline models. More clearly, we conduct a comprehensive experiment to answer the following research questions:

RQ1- What improvements have our algorithm made in comparison with other baseline and state-of-the-art algorithms?

RQ2- Was our graph structure standard and effective?

RQ3- How reliable is our recommendation method?

RQ4- How was the effect of setting weight for edge types?

These questions are discussed and answered in the rest of this section.

Table 2. Statistics of CiteUlike-SPIN

Users	Papers	Authors	Venues	Tags
5551	16980	5037	1230	7386

5- 1- Dataset

The dataset we used, is initially collected by [33] and after a while [39] added the tag information from CiteULike and citations from Google Scholar. Recently, in [10] they used a crawler to collect other data from Google Scholar, such as authors' names and IDs, years, and venue of publications referring to CiteULike-SPIN, which we used in this work.

We chose authors whose IDs were repeated in at least two papers. For the venue of publications, we did the same. For noise reduction, we do not consider the tags which are used less than five times. We used 5-fold cross-validation, so in each iteration, we used 80% of the data as a training set and the remaining 20% as a testing set. The statistics of the dataset are described in Table 2.

5- 2- Evaluation metrics

We used several popular evaluation metrics to evaluate the recommendation performance of the introduced algorithm [1]. First, we calculate these evaluation metrics for each user, then report the average amount among all users.

5- 2- 1- Precision: indicates how relevant the recommended papers have been. Precision is calculated as *(7)*:

$$Precision\ @u = \frac{Relevant\ recommended\ papers\ (u)}{Total\ recommended\ papers\ (u)} \quad (7)$$

5- 2- 2- call: quantifies the fraction of relevant papers that have been identified and recommended.

$$Recall\ @u = \frac{Relevant\ recommended\ papers\ (u)}{Total\ relevant\ papers\ (u)} \quad (8)$$

5- 2- 3- score: is defined as the weighted harmonic average of Precision and Recall.

$$Fscore\ @u = \frac{2(\text{Precision}@\text{u} \times \text{Recall}@\text{u})}{(\text{Precision}@\text{u} + \text{Recall}@\text{u})} \quad (9)$$

5- 2- 4- ccess: is the probability of finding a relevant paper among the Top-N recommended papers for each user. It is defined as (10).

$$success\ @u = \begin{cases} 1 & if\ a\ true\ paper\ in\ the\ TopN \\ 0 & otherwise \end{cases} \quad (10)$$

5- 2- 5- CG (Normalized Discounted Cumulative Gain): measures the quality of a recommended list of items [25] and is computed as the following in (11).

$$nDCG\ @K = \frac{1}{|U|}\sum_{u=1}^{|U|}\frac{DCG\ @K}{IDCG\ @K} \quad (11)$$

Such that

$$DCG\ @K = \sum_{i=1}^{K}\frac{\alpha(i)}{\log_2(i+1)} \quad (12)$$

$$IDCG\ @K = \sum_{i=1}^{\min(R,K)}\frac{1}{\log_2(i+1)} \quad (13)$$

such that $|U|$ refers to the number of users, i is the article rank, K length of the recommendation list, R is the number of relevant articles, and $\alpha(i)$ is a variable that takes value 1 if the article is relevant, and 0 if otherwise.

5- 2- 6- MRR (Mean Reciprocal Rank): like NDCG, this metric is applied to determine the quality of the recommended list by considering the rank of the first relevant paper in the recommended list, and averaging overall users. The definition is:

$$MRR\ @u = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{rank_i} \quad (14)$$

Where N represents the number of users and $rank_i$ is the rank of the first relevant paper in the Top-N list. This method evaluates the quality of the total recommendation list, but the other metrics evaluate the specific length of the Top-N list.

5- 3- Baseline Methods

To evaluate the ISPREC++ framework, we compare its performance with the following base-line and state-of-the-art recommendation methods:

5- 3- 1- Graph-Based Collaborative Filtering(GB-CF) [40]

This method is the main baseline of graph-based recommendation algorithms that construct a bipartite graph of users and papers as nodes, and connect to each user node, the nodes corresponding to the papers in the user's library (papers that the user read it previously). It generates recommendations using a random walk with restart method, in which we set α equal to 0.8.

5- 3- 2- CARE [5]

Constructs a graph with users and articles as nodes and considers users' historical preferences and common author relations as edges. Then recommends articles to users using random walk with restart algorithm. We used $\alpha = 0.8$ s it is mentioned in their paper that this value induces the best result.

5- 3- 3- UAGMT [20]

Constructs a graph with users and articles as nodes. The readership is considered as edges between users and papers, and content similarities and citations are used as edges between papers. Then generate recommendations using random walk with restart. According to the paper, we set $\alpha = 0.75$ for the best performance.

5- 3- 4- CATA++ [19]

This is a model-based approach that uses the content and tags of articles to learn latent factors for users and articles using two parallel autoencoders. Then, the model's prediction scores are computed as the dot product of the latent factors of users and articles. By sorting the scores in descending order, Top-K articles are recommended. We used the same setting for parameters as they set for this dataset.

5- 3- 5- ISPREC [10]

Construct a novel Scientific Paper Information Network (SPIN) consisting of users, papers, authors, venues, and tags. They also include citation relations between papers, then generate recommendations using a limited step random walk with restart that the number of steps is limited to 3 and $\alpha = 0.8$ Also, to evaluate the effectiveness of our SPIN heterogeneous information network and limited random-walk recommendation algorithm, in ISPREC, we implement Random Walk with Restart, without limiting random-walk steps, and $\alpha = 0.8$ on SPIN and report the results as **SPIN-RWR.**

5- 4- Parameter Setting

In our experiments on ISPREC++, our main parameter is α which takes values between 0 and 1 and defines the probability that the random walker moves to a random neighbor at each step. If $\alpha = 1$ there is no restarting and the random walker only follows the links that means we consider the paths with the length of 3. However, if we set $\alpha = 0$, only one-step paths from the target node to one of its neighbors are considered, so, using larger values for α increases the effect of longer paths in the final result. To set the value for α, we evaluated the performance of the algorithm with three different α values (that is α=0.2, α=0.5, and α =0.8) on the validation data and we chose $\alpha = 0.8$ which led to the best performance in terms of all evaluation metrics that we used for every recommendation list size [5, 10]. The performance of the algorithms when using these three settings is presented in Fig. 2.

Following the common approach, the initial value of the learning rate is set to 0.5 and it is divided by 2 after each batch updating [30, 34, 41].

5- 5- Experiment Results

We evaluate our results in various aspects by answering the research questions:

5- 5- 1- Performance comparison (RQ1)

The evaluation results are reported for the suggested algorithm and the baseline models, on the CiteULike-SPIN dataset. The detailed results are displayed in Table 3. It is apparent from it that in all sizes of recommendation lists (N= 3, 30, 60), the proposed ISPREC++ algorithm outperforms the baseline methods on all the evaluation metrics with a significant improvement. These results show the superiority and high performance of the structure of our SPIN and ISPREC++.

5- 5- 2- Effectiveness of SPIN Network (RQ2)

To ensure that each of these meta-paths has a positive impact on recommendations, in the first step we added each of them individually to the user-paper network. We observed positive changes in the validation results in all evaluation metrics, which confirms the reliability of all suggested meta-paths We show the results of these different networks with a limited random walk with restart by recall as an evaluation metric in Fig. 3. In addition, the success of ISPREC and SPIN-RWR, both using SPIN, show the usefulness of the suggested graphical modeling approach.

5- 5- 3- Reliability of ISPREC(RQ3)

To show the reliability and effectiveness of ISPREC, we compared our recommendation method with the commonly used recommendation algorithm random walk with restart by applying it on our network as SPIN-RWR. Better results of ISPREC compared to SPIN-RWR prove that limiting the steps of the random walk process, not only raises the speed of the model but also increases the accuracy and quality of the recommendation list, probably due to the more personalized nature of the recommendations.

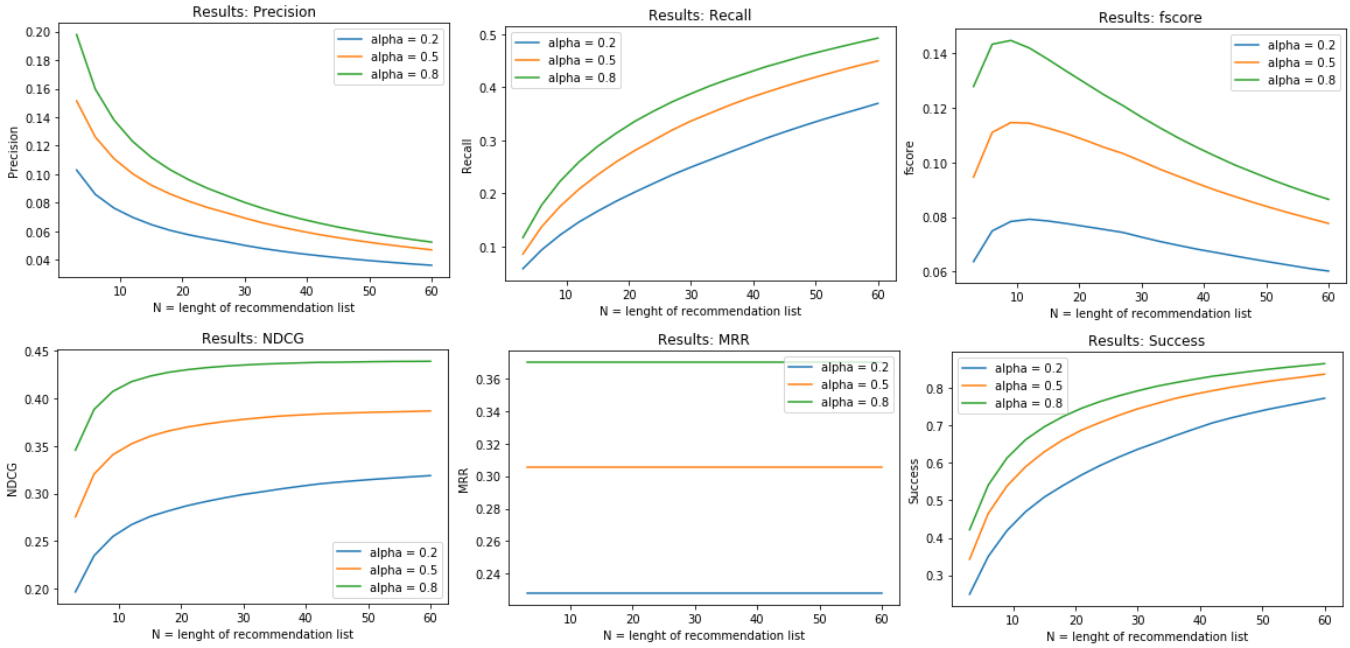

Fig. 2. Comparison of different values α in ISPREC++ algorithm

Table 3. Results of ISPREC++ in comparison with other state-of-the-art methods.

		GB-CF	CARE	UAGMT	CATA++	SPIN-RWR	ISPREC	ISPREC++
Precision	@3	0.136612	0.104546	0.117108	0.0508	0.135243	0.18554	0.197934
	@30	0.060286	0.049499	0.055648	0.0303	0.062215	0.078005	0.080225
	@60	0.040322	0.034836	0.03855	0.02424	0.042687	0.051448	0.052393
Recall	@3	0.080413	0.057676	0.065611	0.0261	0.077724	0.106929	0.116362
	@30	0.295534	0.23763	0.274005	0.1743	0.309169	0.379078	0.388132
	@60	0.388118	0.340395	0.379722	0.2698	0.420764	0.486875	0.492546
F-score	@3	0.0884	0.063985	0.07285	0.0296	0.085781	0.118049	0.127941
	@30	0.087954	0.07121	0.080969	0.0462	0.090775	0.113421	0.116646
	@60	0.0669	0.057656	0.064066	0.0412	0.070941	0.084998	0.086481
Success	@3	0.303008	0.245613	0.277139	0.136	0.313565	0.402918	0.421221
	@30	0.685102	0.615673	0.674149	0.5126	0.722645	0.787786	0.793082
	@60	0.781877	0.741272	0.783751	0.6333	0.822086	0.863736	0.86543
NDCG	@3	0.244071	0.19711	0.219264	0.0542	0.250417	0.328392	0.345315
	@30	0.344708	0.293969	0.324019	0.096	0.356205	0.422685	0.434535
	@60	0.35626	0.311207	0.337869	0.1241	0.367137	0.427434	0.43853
MRR		0.273139	0.227752	0.249855	0.1349	0.281711	0.355044	0.369612

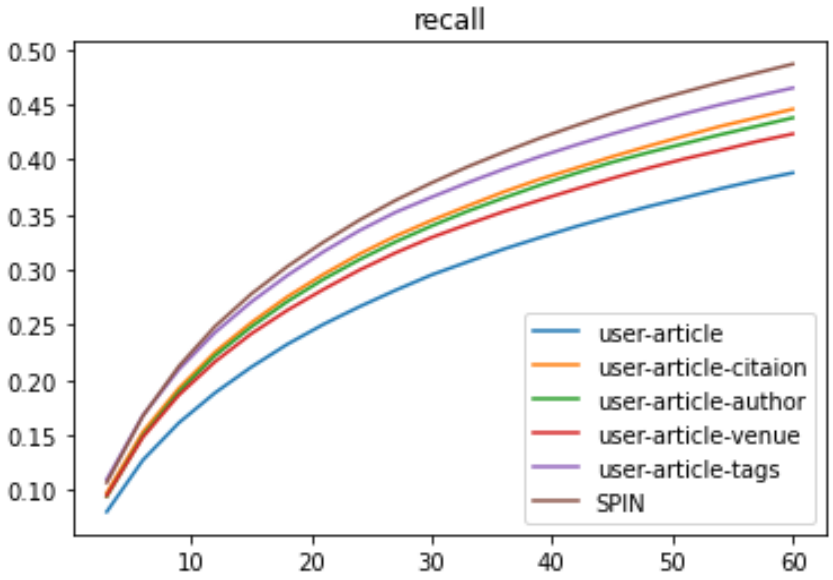

Fig. 3. Results of different networks

Effect of weighting the edges(RQ4): To show the importance of learning the weights of edges in our network, we compared ISPREC and ISPREC++. As it is clearly seen in Table 2, ISPREC++ has a better performance in all evaluation metrics. The most significant improvements were in two evaluation metrics that demonstrate the quality of the recommendation list, which are NDCG and MRR. This means by weighting the edges, in addition to more accuracy, we have a better order of items in the recommendation list.

6- Conclusions

In this paper, an extended version of the integrated scientific paper recommendation approach, ISPREC, was proposed for recommending relevant papers based on users' interests, referring to ISPREC++. For generating recommendations, a novel scientific paper information network (SPIN) is constructed. Then, a Bayesian Personalized Ranking based method is utilized to learn the weights of edges in the network, according to users' preferences. Next, it is analyzed using a limited random-walk–based approach. Our evaluation of a real scientific dataset shows the superiority of the suggested framework over state-of-the-art baseline models.

For future research, we suggest investigating the importance of each meta path for recognizing each user's interests separately. People can have unique factors in mind for choosing articles to read, so it is reasonable to assign appropriate weights to meta paths for each target user before analyzing the graph for extracting relevant papers for him. As another direction for future studies, the SPIN graph structure can be used and evaluated for recommending papers to authors as well.

References

[1] X. Bai, M. Wang, I. Lee, Z. Yang, X. Kong, F. Xia, Scientific paper recommendation: A survey, Ieee Access, 7 (2019) 9324-9339.

[2] J. Sun, J. Ma, Z. Liu, Y. Miao, Leveraging content and connections for scientific article recommendation in social computing contexts, The Computer Journal, 57(9) (2014) 1331-1342.

[3] M.S. Pera, Y.-K. Ng, Exploiting the wisdom of social connections to make personalized recommendations on scholarly articles, Journal of Intelligent Information Systems, 42(3) (2014) 371-391.

[4] K. Sugiyama, M.-Y. Kan, Serendipitous recommendation for scholarly papers considering relations among researchers, in: Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries, 2011, pp. 307-310.

[5] F. Xia, H. Liu, I. Lee, L. Cao, Scientific article recommendation: Exploiting common author relations and historical preferences, IEEE Transactions on Big Data, 2(2) (2016) 101-112.

[6] W. Zhao, R. Wu, H. Liu, Paper recommendation based on the knowledge gap between a researcher's background knowledge and research target, Information processing & management, 52(5) (2016) 976-988.

[7] S. Alotaibi, J. Vassileva, Personalized Recommendation of Research Papers by Fusing Recommendations from Explicit and Implicit Social Network, in: UMAP (Extended Proceedings), 2016.

[8] Q. Wang, W. Li, X. Zhang, S. Lu, Academic paper recommendation based on community detection in citation-collaboration networks, in: Asia-Pacific web conference, Springer, 2016, pp. 124-136.

[9] X. Ma, R. Wang, Personalized scientific paper recommendation based on heterogeneous graph representation, IEEE Access, 7 (2019) 79887-79894.

[10] E. Jafari, B. Shams, S. Haratizadeh, ISPREC: Integrated Scientific Paper Recommendation using heterogeneous information network, in: 2021 12th International Conference on Information and Knowledge Technology (IKT), IEEE, 2021, pp. 112-118.

[11] X. Ma, Y. Zhang, J. Zeng, Newly published scientific papers recommendation in heterogeneous information networks, Mobile Networks and Applications, 24(1) (2019) 69-79.

[12] K. Haruna, M.A. Ismail, A.B. Bichi, V. Chang, S. Wibawa, T. Herawan, A citation-based recommender system for scholarly paper recommendation, in: International Conference on Computational Science and Its Applications, Springer, 2018, pp. 514-525.

[13] N. Sakib, R.B. Ahmad, K. Haruna, A collaborative approach toward scientific paper recommendation using citation context, IEEE Access, 8 (2020) 51246-51255.

[14] J. Son, S.B. Kim, Academic paper recommender system using multilevel simultaneous citation networks, Decision Support Systems, 105 (2018) 24-33.

[15] W. Waheed, M. Imran, B. Raza, A.K. Malik, H.A. Khattak, A hybrid approach toward research paper recommendation using centrality measures and author ranking, IEEE Access, 7 (2019) 33145-33158.

[16] G. Guo, B. Chen, X. Zhang, Z. Liu, Z. Dong, X. He, Leveraging title-abstract attentive semantics for paper recommendation, in: Proceedings of the AAAI conference on artificial intelligence, 2020, pp. 67-74.

[17] L. Berkani, R. Hanifi, H. Dahmani, Hybrid recommendation of articles in scientific social networks using optimization and multiview clustering, in: International Conference on Smart Applications and Data Analysis, Springer, 2020, pp. 117-132.

[18] G. Wang, X. He, C.I. Ishuga, HAR-SI: A novel hybrid article recommendation approach integrating with social information in scientific social network, Knowledge-Based Systems, 148 (2018) 85-99.

[19] M. Alfarhood, J. Cheng, CATA++: A collaborative dual attentive autoencoder method for recommending scientific articles, IEEE Access, 8 (2020) 183633-183648.

[20] T. Cai, H. Cheng, J. Luo, S. Zhou, An efficient and simple graph model for scientific article cold start recommendation, in: International Conference on Conceptual Modeling, Springer, 2016, pp. 248-259.

[21] W. Liu, L. Lü, Link prediction based on local random walk, EPL (Europhysics Letters), 89(5) (2010) 58007.

[22] N. Lao, W.W. Cohen, Fast query execution for retrieval models based on path-constrained random walks, in: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, 2010, pp. 881-888.

[23] C. Shi, X. Kong, Y. Huang, S.Y. Philip, B. Wu, Hetesim: A general framework for relevance measure in heterogeneous networks, IEEE Transactions on Knowledge and Data Engineering, 26(10) (2014) 2479-2492.

[24] S. Lee, S. Lee, B.-H. Park, Pathmining: A path-based user profiling algorithm for heterogeneous graph-based recommender systems, in: The Twenty-Eighth International Flairs Conference, 2015.

[25] L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank citation ranking: Bringing order to the web, Stanford InfoLab, 1999.

[26] T.H. Haveliwala, Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search, IEEE transactions on knowledge and data engineering, 15(4) (2003) 784-796.

[27] A. Balmin, V. Hristidis, Y. Papakonstantinou, Objectrank: Authority-based keyword search in databases, in: VLDB, 2004, pp. 564-575.

[28] M. Gori, A. Pucci, V. Roma, I. Siena, Itemrank: A random-walk based scoring algorithm for recommender engines, in: IJCAI, 2007, pp. 2766-2771.

[29] S. Lee, S. Park, M. Kahng, S.-g. Lee, PathRank: Ranking nodes on a heterogeneous graph for flexible hybrid recommender systems, Expert Systems with Applications, 40(2) (2013) 684-697.

[30] Z. Jiang, H. Liu, B. Fu, Z. Wu, T. Zhang, Recommendation in heterogeneous information networks based on generalized random walk model and bayesian personalized ranking, in: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018, pp. 288-296.

[31] Y. Sun, J. Han, X. Yan, P.S. Yu, T. Wu, Pathsim: Meta path-based top-k similarity search in heterogeneous information networks, Proceedings of the VLDB Endowment, 4(11) (2011) 992-1003.

[32] N. Li, Y. Yu, Z.-H. Zhou, Diversity regularized ensemble pruning, in: Joint European conference on machine learning and knowledge discovery in databases, Springer, 2012, pp. 330-345.

[33] 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, 2011, pp. 448-456.

[34] H. Liu, Z. Jiang, Y. Song, T. Zhang, Z. Wu, User preference modeling based on meta paths and diversity regularization in heterogeneous information networks, Knowledge-Based Systems, 181 (2019) 104784.

[35] R. Pan, Y. Zhou, B. Cao, N.N. Liu, R. Lukose, M. Scholz, Q. Yang, One-class collaborative filtering, in: 2008 Eighth IEEE International Conference on Data Mining, IEEE, 2008, pp. 502-511.

[36] S. Rendle, C. Freudenthaler, Z. Gantner, L. Schmidt-Thieme, BPR: Bayesian personalized ranking from implicit feedback, arXiv preprint arXiv:1205.2618, (2012).

[37] S. Ruder, An overview of gradient descent optimization algorithms, arXiv preprint arXiv:1609.04747, (2016).

[38] R. Ying, R. He, K. Chen, P. Eksombatchai, W.L. Hamilton, J. Leskovec, Graph convolutional neural networks for web-scale recommender systems, in:

Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, 2018, pp. 974-983.

[39] H. Wang, B. Chen, W.-J. Li, Collaborative topic regression with social regularization for tag recommendation, in: Twenty-Third International Joint Conference on Artificial Intelligence, 2013.

[40] N. Chiluka, N. Andrade, J. Pouwelse, A link prediction approach to recommendations in large-scale user-generated content systems, in: European Conference on Information Retrieval, Springer, 2011, pp. 189-200.

[41] Y. Shi, M. Larson, A. Hanjalic, Unifying rating-oriented and ranking-oriented collaborative filtering for improved recommendation, Information Sciences, 229 (2013) 29-39.

HOW TO CITE THIS ARTICLE

E. Jafari, B. Shams, S.Haratizadeh, ISPREC++: Learning Edge Type Importance in Network-Oriented Paper Recommendation. AUT J. Model. Simul., 54(2) (2022) 197-210.

DOI: 10.22060/miscj.2023.21212.5275