Received February 11, 2019, accepted March 1, 2019, date of publication March 19, 2019, date of current version April 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2906106

Academic Venue Recommendations Based on Similarity Learning of an Extended Nearby Citation Network

ABDULRHMAN M. ALSHAREEF [1,2], MOHAMMED F. ALHAMID [3], (Member, IEEE), AND ABDULMOTALEB EL SADDIK [2], (Fellow, IEEE)

[1] Information System Department, FCIT, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[2] Multimedia Communications Research Laboratory, University of Ottawa, Ottawa, ON K1N 6N5, Canada
[3] Software Engineering Department, CCIS, King Saud University, Riyadh 11451, Saudi Arabia

Corresponding author: Abdulrhman M. Alshareef (aalshareef@uottawa.ca)

This work was supported in part by the Deanship of Scientific Research, King Saud University, Riyadh, Saudi Arabia, under Grant RG-1437-042.

ABSTRACT The rapidly increasing number of potential academic venues for research publication and commentary has made sourcing the venue that would best contribute to promoting effective scientific cooperation more challenging. In this paper, we propose a similarity learning approach to determine the most appropriate venue to publish an article. We first analyze the article metadata and cited articles to build the citation network matrices of the given article and then apply these to learn and build similarity matrices between academic objects (i.e., articles, authors, and venues) at an extended nearby article citation network. Using the formed matrices, we estimate a collaborative anticipation confidence score of a relationship between the venues in the extended network. For our empirical studies, we used an actual academic dataset to validate the efficiency of our approach and recommend an appropriate academic venue. The experimental results highlight the effectiveness of our proposed approach to optimize overall recommendation quality, compared with other baseline approaches.

INDEX TERMS Citation network analysis, similarity analysis, similarity learning, academic venue recommendation, extended nearby citation network.

I. INTRODUCTION

With today's rapid growth of large volumes of scholarly data, the development of scientific publication indexing and recommendations is receiving significant interest from researchers who want to benefit from the advancements in data analysis techniques. Figure 1 illustrates the citation network surroundings and the relationships between articles, with each line representing a relationship between two articles and each node representing an article. The increase of publications has encouraged the research industry to move toward smarter environments, and the digital twin concept introduced in [1] is digital replication of a living or non-living physical entity. The aim is to generate a smarter environment for different entities by facilitating the means to monitor, understand and optimize the functions of the entity, and provide continuous feedback to improve quality. One component of the digital twin is to arranging a virtual collaboration between researchers in the same field. Hence, this work will form a virtual community that will help young researchers to find suitable venues through similarity learning of an extended nearby citation network [2].

Both the volume of scientific publications and their corresponding metadata databases, and the recommendation knowledge of academic objects, have grown substantially. Recently, academic recommender systems mainly focus on article, author and venue recommendation [3]. Now, the huge number and availability of venues regarding specific areas in the same disciplines has increased the challenge for younger researchers to reach the right community. Thus, it is increasingly difficult for them to determine the most appropriate target venues for their research publications [4], and this affects their research efficiency due to the time-consuming submission creation, out of scope rejections and lack of connections in a new community. The problem of finding the best academic venue for publication is not new.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan Bu.

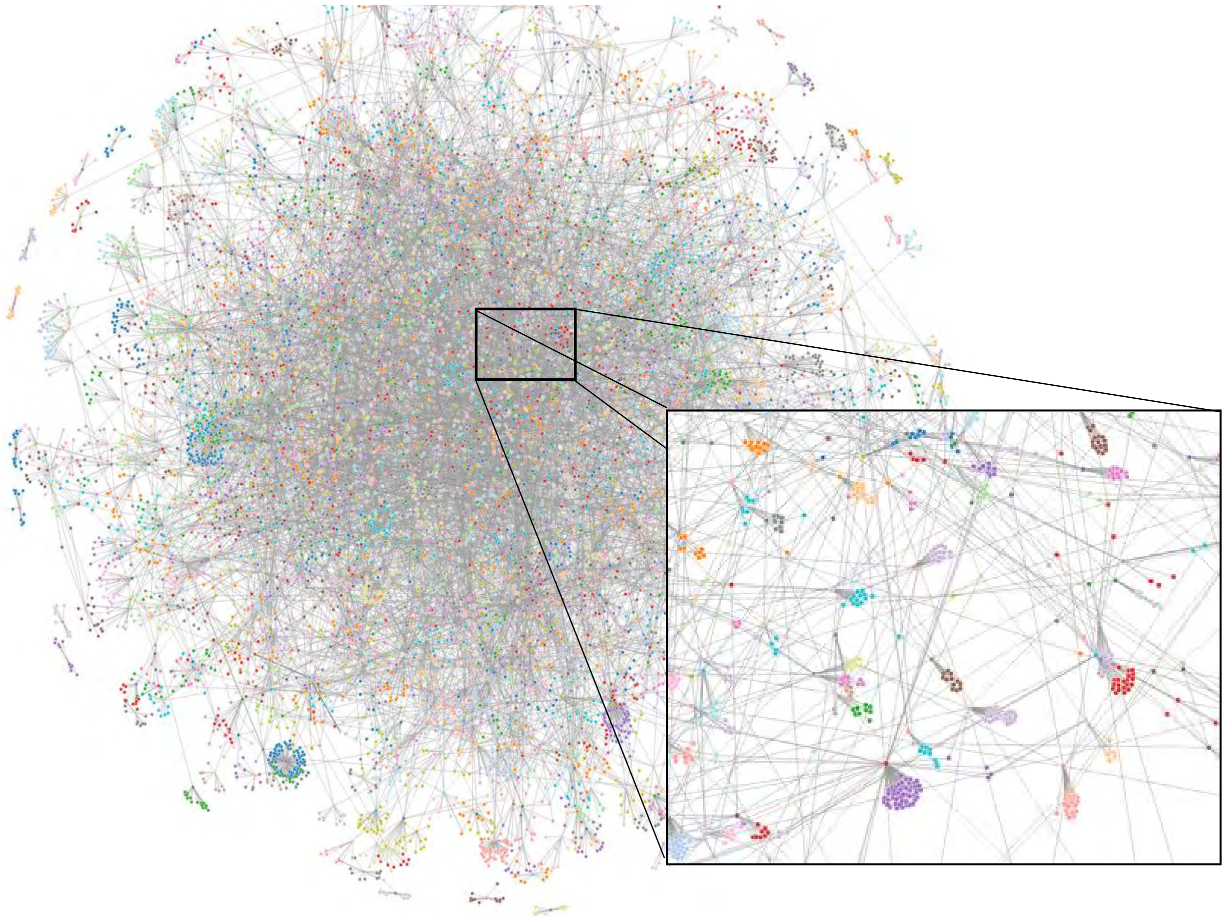

FIGURE 1. An overview of the citation network surroundings and relationships.

Many methodologies have been applied to create recommendation systems that help younger researchers and authors reach suitable venues [4]–[6].

In order to achieve the goal of recommending optimal academic venues, this paper has three objectives: 1) identify and build the citation network matrices of an extended nearby citation network; 2) learn and create the similarity matrices between academic objects; and 3) estimate the collaborative confidence score that best represents the anticipation of a potential relationship between venues in the extended network.

To address the first objective, we used the article metadata and the venue information to identify an extended nearby citation network for a given article. An extended nearby citation network includes the citing article, the cited articles and the related articles, as shown in Figure 2. We then apply the article, author and venue relation information to build the citation network matrices. The relation information is the article-author, article-venue and author-venue associations in the extended network.

To achieve the second objective of learning and creating the similarity matrices between academic objects, we employed the citation network matrices to determine the similarity between article-article, author-author and venue-venue. We propose an algorithm to estimate the similarity score between the academic objects based on the cosine between two academic objects using the information about academic associations in the citation network. We then use the similarity score to build the similarity matrices.

To realize the third objective we used the citation network and object similarity matrices to estimate the collaborative confidence score, considering the given article.

Though certain recommendation and citation analysis parameters have already been incorporated in the academic venue recommendation process of a number of existing works, there is a general lack of formulation on how to employ citation impacts as dimensions in the traditional academic venue recommendation model.

Our research goal is to identify different aspects of a citation network, and select the features that are best suited to choose and deliver different collaboration modules. The focus of the paper is to use similarity relationships to extract and identify user preferences, and employ similarity learning to analyze the data more intensively to determine which to use

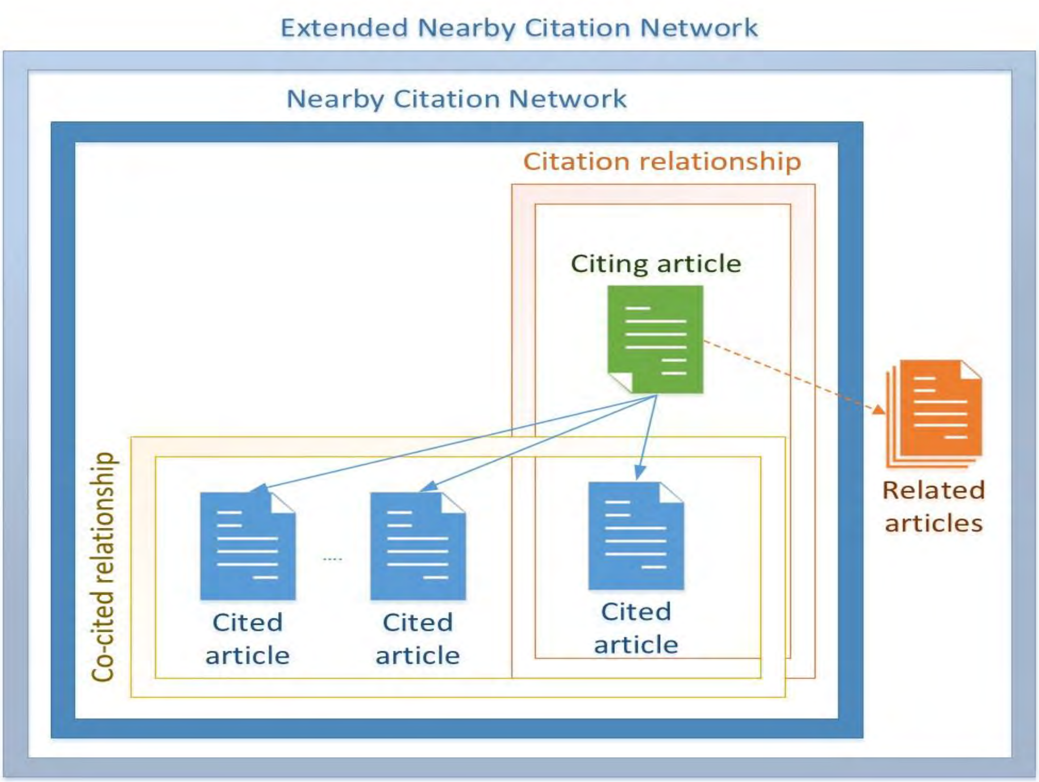

FIGURE 2. The structure of the extended nearby citation network.

in the recommendation process to enhance the analysis of the citation network for given article. This will be reflected in the selection quality of publishing venues, to contribute to a more effective and informed scientific community. Academic venue recommender systems are intended to increase user satisfaction, by analyzing the publication content or amount of information available about the citation network. However, user preferences and selections to cite different articles should not minimize the importance of citation impact. For example, user selections depend on various factors, including content similarity, bibliometric impact, author experience and article popularity. This information can increase the quality of the recommendation outcomes, and thereby enhance user experience.

The paper is organized as follows: Section II discusses related work, and Section III describes the proposed recommendation model. We evaluate the recommendation model in section IV with a detailed discussion of the achieved results, and conclude the paper in Section V with identification of current limitations, and suggestions for future work.

II. RELATED WORK

Analysis of the citation network facilitates the recognition of associations between academic objects in scientific society [7]. Various studies have used different foci on specific citation network levels to recognize relationships among academic objects such as research scope, society experience and institution locations [8], [9]. Some studies reflect the citation network association characteristics in their analysis, to recognize the scientific society and promote scientific cooperation among researchers.

De Vocht *et al.* [10] explained the relationship between scientific cooperation and social interaction among authors related to a given conference. They were able to envision author associations and social activities, and formed the network of scientific conferences associations. The growth in author communications on social networks with respect to a particular conference, raises the chance of detecting the academic scientific networks of the conference. However, they had only moderate recall outcomes due to the missing conference information on the targeted social network. Newman [7], [11] studied social network effects and researcher collaboration networks to determine scientific cooperation among the scientific society. He considered current co-author relationships in his analysis but did not include topical similarities in the citation network. He did, however, find that two researchers were more likely to collaborate if they shared a co-author.

Mu *et al.* [12] discussed a query-based citation network in order to personalized the outcome. They used the three layers of author, paper and word to analyze relations in the network, and employed a graph-based approach to identify the relations in the network. However, they disclosed reviewed the venue relation among the citation network.

Boyack and Klavans [13] discussed the clustering accuracy of the mapping approach in citation network analysis. They applied four typical approaches: co-citation, bibliographic coupling, direct citation and hybrid. The total accuracy was enhanced by the hybrid approach compared to the other individual approaches. They used bibliographic coupling on both references and words, and the efficiency of applying bibliographic features to form a collaboration network increased through the integration of different approaches. Cohen and Ebel [14] examined the collaborator relation in social networks, based on a given set of keywords to recommend cooperation. They also studied various analytical approaches to recommend an author candidate for scientific cooperation. The targeted approaches included co-occurrence of the number of collaborations, textual relevance and global importance. They concluded that the integration of co-occurrence and content-based approaches improves the results of the analysis.

Article submission to a non-relevant publishing venue is one of the major reasons for rejections in the research industry, and a publishing venue recommender system would help users find the appropriate venue for their submission. Yang and Davison [5] proposed a venue recommendation approach based on article writing styles, using both content and stylometric features to determine similarities between articles. However, they did not consider the bibliographic associations between authors and venues at the citation network. Chen *et al.* [4] proposed an approach centered on a random walk, with a restart to endorsing venues. They calculated the importance score of two nodes in a weighted graph by considering different academic factors based on the co-publishing frequency, author academic level and relation weight. However, they excluded the topical similarity and venue relation of the citation network in their analysis. Yu *et al.* [15] proposed journal recommendation systems that apply content-based and global relevance approaches. Their system used the term frequency model and term weighting scores to recommend related articles. They neglected the venue impact on different academic objects, however, as well as the relationship related to the given query.

An overall view of the relevant studies in the field of academic venue recommendation and citation analysis indicates that prior studies have typically focused on existing associations between scientific articles. To enhance analysis outcomes, this paper concentrates on learning the similarities of academic objects in an extended nearby citation network for a given query. We propose a recommendation model that measures the confidence score of venues at the extended network, using bibliographic features to propose an appropriate academic venue for article submission. We virtually arranged the venue profiles in a certain neighborhood to help us determine an appropriate venue. We also analyzed article metadata, due to its availability and accessibility in large data repositories, to represent scholarly publications and to reduce the computational overload.

III. DESIGN OF ACADEMIC VENUE RECOMMENDATION MODEL

In this section, we discuss the details of the recommendation model and explain how to build citation network matrices for a given article to determine the similarity between academic objects. We then explain the procedure to find the similarity between academic objects and build the similarity matrices.

TABLE 1. Summary of notations and their meaning.

Notations	Meaning
$\boldsymbol{H}$	Set of Authors.
$\boldsymbol{A}$	Set of Articles.
$\boldsymbol{V}$	Set of Venues.
$\boldsymbol{T}$	Set of Terms.
$\boldsymbol{E}$	Set of Elements.
$\mathbf{C}(\tilde{\mathbf{C}})$	Author-Article matrix (Normalized matrix of $\mathbf{C}$).
$\mathbf{P}(\tilde{\mathbf{P}})$	Venue-Article matrix (Normalized matrix of $\mathbf{P}$).
$\mathbf{L}(\tilde{\mathbf{L}})$	Author-Venue matrix (Normalized matrix of $\mathbf{L}$).
$\mathbf{R}(\mathbf{R}^k)$	Author-Author similarity matrix (Contains k most similar authors).
$\mathbf{G}(\mathbf{G}^k)$	Article-Article similarity matrix (Contains k most similar Articles).
$\overrightarrow{\mathbf{VA}}$	Latent preferences of venues toward articles matrix.
$\overrightarrow{\mathbf{AH}}$	Latent preferences of articles toward authors matrix.
$\overrightarrow{\mathbf{VH}}$	Latent preferences of venues toward authors matrix.
$\overrightarrow{\mathbf{HA}}$	Latent preferences of authors toward articles matrix.

A. OVERVIEW OF THE RECOMMENDATION MODEL

Our proposed model will determine a venue's latent preferred association to articles and authors, by learning from the similarity between academic objects. The model was motivated by researchers' behavior: they typically want to maintain contact with authors they encounter in venues, cite expert authors in a specific research area, look for high-quality and successful academic venues, join conferences that are closely related to their research and cite articles from high-quality venues and publishers. The model is an extension of our previous works [16], [17] that proposed academic recommendations based on topical, author and venue similarities, and achieved notable recommendation results. Here, we explain how to build a citation network and personalize the recommendation model by finding the latent preferences in a citation network for a given query. We also introduce a common set of symbols: Upper-case bold letters, such as **J**, represent matrices; while the corresponding lower-case italicized letters, such as *a,* represent entries in the matrices and capital italic letters, such as A, represent sets of entries. The lower case letter subscripts, for example A_y, represents an entry y from the set A. Table 1 summarizes the notations that has been used in the rest of this paper.

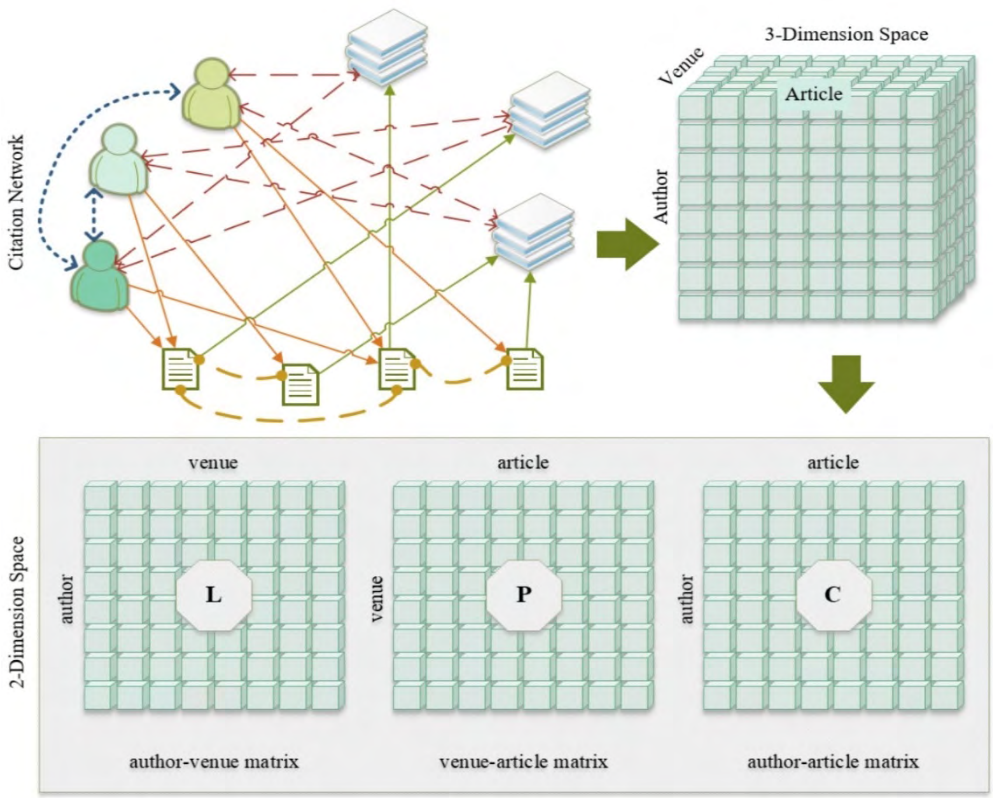

FIGURE 3. Illustrating the citation network in three-dimensional space and in two-dimensional space.

B. BUILDING THE ACADEMIC RECOMMENDATION MODEL

The citation network represents a group of authors, articles, and venues that have relationships with one another. The network relationships including citation, co-citation, and co-authorship and co-publishing information are available. For a set of authors, articles, and venues corresponding to our academic recommendation model, the set of authors $H = \{h_1, h_2, \ldots, h_{|H|}\}$, the set of articles $A = \{a_1, a_2, \ldots, a_{|A|}$, and a set of venues $V = \{v_1, v_2, \ldots, v_{|V|}$, the citation network formalized as a tuple $G := (H, A, V, Q)$ where $Q \subseteq H \times A \times V$ a trilateral relationship. Therefore, a citation network can be viewed as a three-dimensional space of authors, articles, and venues; consequently, this three-dimensional space can be projected onto three two-dimensional matrices, as shown in Figure 3.

We first define the three matrices obtained by aggregating over authors, articles, and venues, as follows:

- Author-Article matrix $\mathbf{C} = \left[c_{h,a}\right]_{|H|\times|A|}$, where $c_{h,a}$ represents the academic citation impact of article a that cited articles written by author h.
- Venue-Article matrix $\mathbf{P} = \left[p_{v,a}\right]_{|V|\times|A|}$, where $p_{v,a}$ represents the academic citation impact of article a that cited articles published in venue v.
- Author-Venue matrix $\mathbf{L} = \left[l_{h,v}\right]_{|H|\times|V|}$, where $l_{h,v}$ represents the academic citation impact of articles that author h have published in venue v.

C. CITATION NETWORK SIMILARITIES

In this paper, we determine two types of similarities in order to calculate the collaborative confidence score. The similarities types are the author-author similarity and article-article similarity. The similarity measure can be computed using different matrices. For instance, to compute the similarities between authors $\mathbf{D}_{|H|\times|H|}$, we use two matrices the author-article matrix $\mathbf{C}_{|H|\times|A|}$ and the author-venue matrix $\mathbf{L}_{|H|\times|V|}$. As mentioned earlier, If two authors frequently appear together in multiple venues, those authors have greater potential for collaboration based on the closely related semantics [15]. In this case, we could successfully measure the authors' similarities in terms of venues. The authors' similarities in terms of articles will be measured based on the co-occurrence of the event of co-authorship. In this paper, we examine the use of both approaches to determine the similarities.

As mentioned previously, Several similarity measures can be used to determine the similarity of any two vectors [18]. for example, the Euclidean Distance [18],

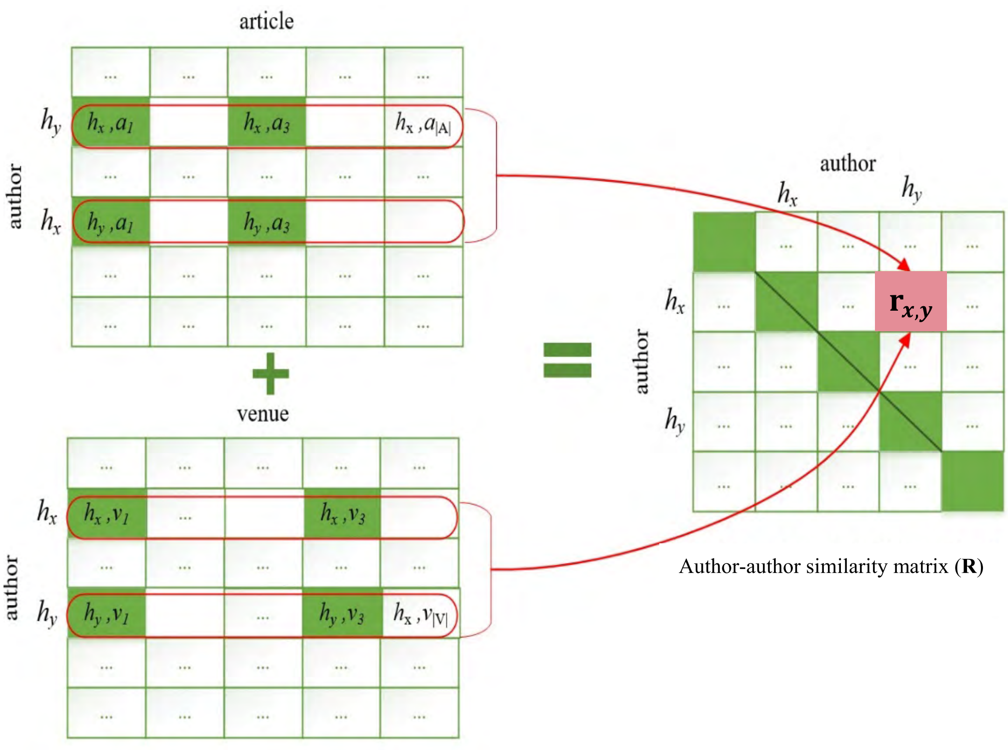

FIGURE 4. An illustration of the author-author similarity matrix ***R***.

cosine similarity [18], [19], and Pearson Correlation Coefficient [18], [20]. However, we chose to use the cosine similarity technique due to its high performance, which has been proven in previous studies [18], [20]. The Cosine Similarity represents the well-known method to compute the similarity between two sets of vectors.

Author-Author Similarity:

To compute the author-author similarity matrix $\mathbf{R}_{|H|\times|H|}$, we start by utilizing the author-article matrix $\mathbf{C}_{|H|\times|A|}$ and the author-venue matrix $\mathbf{L}_{|H|\times|V|}$. The idea behind relying on the detection of similar authors who attend the same venues is to use the list of venues attended by given authors to discover other interesting venues attended by similar authors. On the other hand, noticing the authors who are cited the same articles help to discover the degree of interest in cooperation between the authors, which helps to predict the possibility of future cooperation between them. We adopted the cosine-based similarity approach to determine the similarity between the two authors. The cosine-based similarity takes two vectors of shared venues and articles of authors h_x, and h_y, and computes their similarity according to their angle, as in Equation (1).

$$r_{(h_x,h_y)} = \cos\left(h_{x,v}, h_{y,v}\right) + \cos\left(h_{x,a}, h_{y,a}\right) = \frac{h_{x,v} \bullet h_{y,v}}{\|h_{x,v}\| \bullet \|h_{y,v}\|} + \frac{h_{x,a} \bullet h_{y,a}}{\|h_{x,a}\| \bullet \|h_{y,a}\|} \quad (1)$$

The author-author similarity matrix $\mathbf{R}_{|H|\times|H|}$ represents the similarity between authors where both rows and columns represent authors. We consider the top k nearest authors for each author in order to lower the computational time. Accordingly, if the corresponding similarity value between a pair of authors x and y is greater than the k highest similarity value, we set $r_{x,y}$ to the similarity value otherwise, we set $r_{x,y}$ value to zero. We use the non-zero entries to build the called the $\mathbf{R}^{\mathbf{k}}$ matrix that contains the k most similar authors. Figure 4 shows an example of building the similarity matrix **R**.

1) ARTICLE-ARTICLE SIMILARITY

To compute the article-article similarity matrix $\mathbf{G}_{|A|\times|A|}$, we consider the author-article matrix $\mathbf{C}_{|H|\times|A|}$ and the article-venue matrix $\mathbf{P}_{|A|\times|V|}$. To discover venues that most suitable to publish a given article, we detect the similar cited articles that have been published in the same venues to identify the list of venues that has the most similar article to the given one. Moreover, detecting the similar articles that have been cited by same authors help to forecast the possibility of future cooperation between authors that have a "common collaborator" (shared author) [7], [11].The cosine-based similarity approach has been selected to determine the similarity between two articles a_x, and a_y based on the shared venues and authors. The cosine-based computes their similarity according to the angle between two vectors of articles

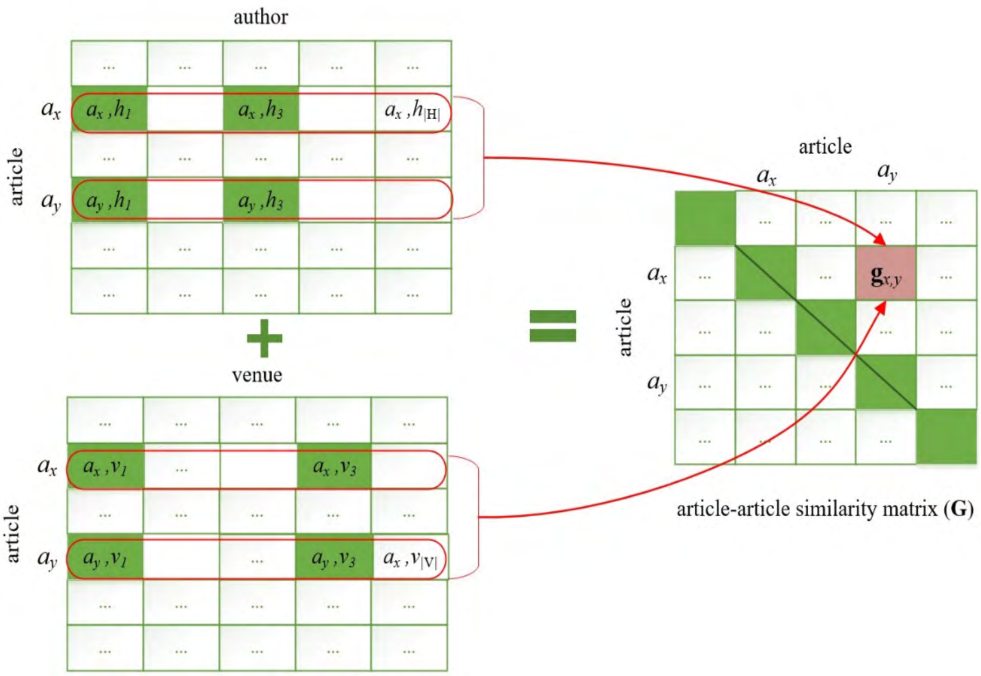

FIGURE 5. An illustration of the article-article similarity matrix G.

as in Equation (2).

$$\begin{aligned} g_{(a_x,a_y)} &= \cos\left(a_{x,v}, a_{y,v}\right) + cos\left(a_{x,h}, a_{y,h}\right) \\ &= \frac{a_{x,v} \bullet a_{y,v}}{\left\|a_{x,v}\right\| \bullet \left\|a_{y,v}\right\|} \\ &\quad + \frac{a_{x,h} \bullet a_{y,h}}{\left\|a_{x,h}\right\| \bullet \left\|a_{y,h}\right\|} \end{aligned} \tag{2}$$

Figure 5 shows an example of creating the article-article similarity matrix G. The $\mathbf{G}_{|A|\times|A|}$ matrix represents the similarity between articles where both rows and columns represent articles. We consider the top k nearest articles for each article to build the $\mathbf{G^k}$ matrix. Accordingly, if the corresponding similarity value between a pair of authors x and y is greater than the k highest similarity value, we set $g_{x,y}$ to the similarity value otherwise, we set $g_{x,y}$ value to zero. The non-zero entries will be consider to build the matrix that contains the k most similar articles.

2) IDENTIFY LATENT PREFERENCE FOR VENUE

Our proposed Latent Preference Identifier Model (LPIM) aims to reveal the concealed connection between the academic venues. For instance, authors who attended certain venues are more likely to collaborate in the production of joint scientific articles. This possibility increases if there is a common collaborator among them (joint author). In order to identify the hidden preferences for each venue, we analyze the bibliography information associated with the citation network of h, a, v in the dataset. In order to do that, we construct two latent models that represent the hidden preferences of the venues.

We analyze the hidden associations of venues' to identify the venues' latent preferences toward their articles and authors. In order to do that; we use the matrix **P** and the matrix **G** to catch the venues' latent preferences concerning articles. We constructed the new venue-article latent preference matrix $\vec{VA}$. The matrix represents the product of matrix **P** and **G**, as in Equation (3):

$$\overline{\mathbf{VA}} = \tilde{\mathbf{P}}\mathbf{G^k} \tag{3}$$

where the matrix $\tilde{\mathbf{P}}$ represents a normalized form of the matrix **P**, and the matrix $\mathbf{G^k}$ denotes the top k nearest articles. We take into consideration the top k nearest neighbors for each article to minimize the computational cost. The normalized venue-article matrix $\tilde{\mathbf{P}}$ can be defined as $\tilde{\mathbf{P}} = \left[\tilde{P}_{v,a}\right]_{|V|\times|A|}$, where $\tilde{p}_{v,a}$ is obtained as follows:

$$\tilde{p}_{v,a} = p_{v,a} \Big/ \sqrt{\sum_{i}^{|V|} / \left(p_{i,a}\right)^2} \tag{4}$$

The multiplication of the v-th row by the a-th column entails to identify the latent preferences of venues v, on articles a with regard to the articles k nearest neighbor. The details of the construction of the new matrix $\vec{\mathbf{VA}}$ indicated by Figure 6.

To illustrate a simple example of building the latent preference model for a venue, we consider six venues cited by seven articles. We utilize this article A_1 as reference examples throughout this section. When we aggregate the venue's association over the article, we can obtain the venue-article matrix shown in Table 2.

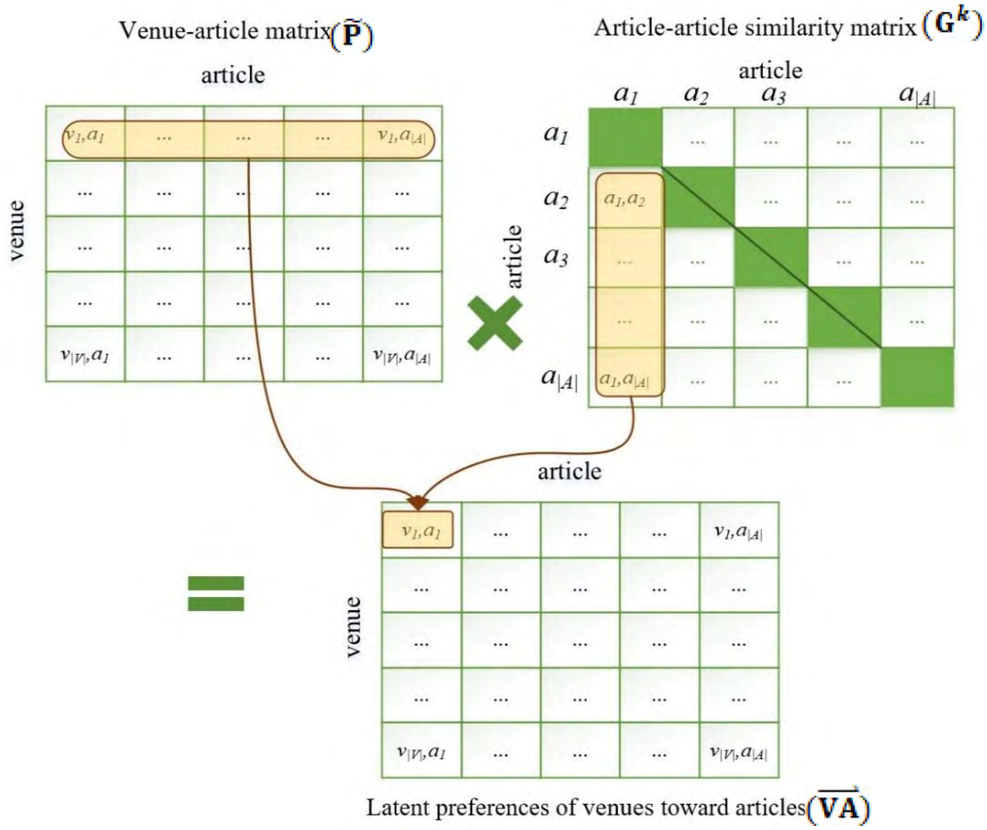

FIGURE 6. An illustration of the process of computing the venue-article matrix $\overrightarrow{\mathbf{VA}}$.

TABLE 2. An example of an author-article matrix, P.

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
V_1	3	0	2	0	3	1	0
V_2	2	1	0	0	1	2	2
V_3	3	1	2	3	0	1	3
V_4	1	2	3	2	3	1	2
V_5	1	0	2	3	1	2	3
V_6	1	2	3	2	0	1	2

The example matrix P allows us to easily derive the normalized matrix $\tilde{P}$. For example, the normalized value of A_1 for V_1 is calculated as $\tilde{p}_{V_1,A_1} = 3\Big/\sqrt{3^2+2^2+3^2+1^2+1^2+1^2} = 0.12$, whereas the value of article A_5 is $\tilde{p}_{V_1,A_5} = 3\Big/\sqrt{3^2+1^2+3^2+1^2} = 0.15$. Even though the number of citation by article A_1 for V_1 equals to the citation number for article A_5, the article A_5, retains more influence than the former article does with regard to his preferences.

To calculate V_1 latent preference for the article A_2, when we determine the article-article similarity between the article A_2 and every other article and we consider the five most similar articles, we determine the articles and similarities shown in Table 3. From the normalized and similarity values, V_1 latent preference value for the article A_2, $\overrightarrow{\mathbf{VA}}_{V_1,A_2}$, is calculated as follows: $\overrightarrow{\mathbf{VA}}_{V_1,A_2} = (0*1)+(0.07*0.88)+(0.15*0.67)+(0.08*0.65)+(0.12*0.45) = 0.27$. According V_1 cited by articles similar to article A_2, the latent value for the article A_2 can be expected as 0.27 even though V_1 has not previously cited by this article.

TABLE 3. Article similarities between the article and its five most.

	A_2	A_3	A_5	A_6	A_1
A_2	1	0.88	0.67	0.65	0.45

We use the matrix $\mathbf{L}$ and the matrix $\mathbf{R}$ to catch the venues' latent preferences concerning authors. We constructed the new venue-author latent preference matrix $\overrightarrow{\mathbf{VH}}$. The matrix represents the product of matrix $\mathbf{L}$ and $\mathbf{R}$, as in Equation (5). Where the matrix $\tilde{\mathbf{L}}$ represents a normalized form of the matrix $\mathbf{L}$, and the matrix $\mathbf{R}^k$ represents the top k nearest authors as clarified formerly. To reduce the computing cost, we consider the top k nearest neighbors for each author. The multiplication of the v-th row by the h-th column entails to identify the latent preferences of venues v, on authors h with regard to the authors k nearest neighbor. Figure 7 shows the

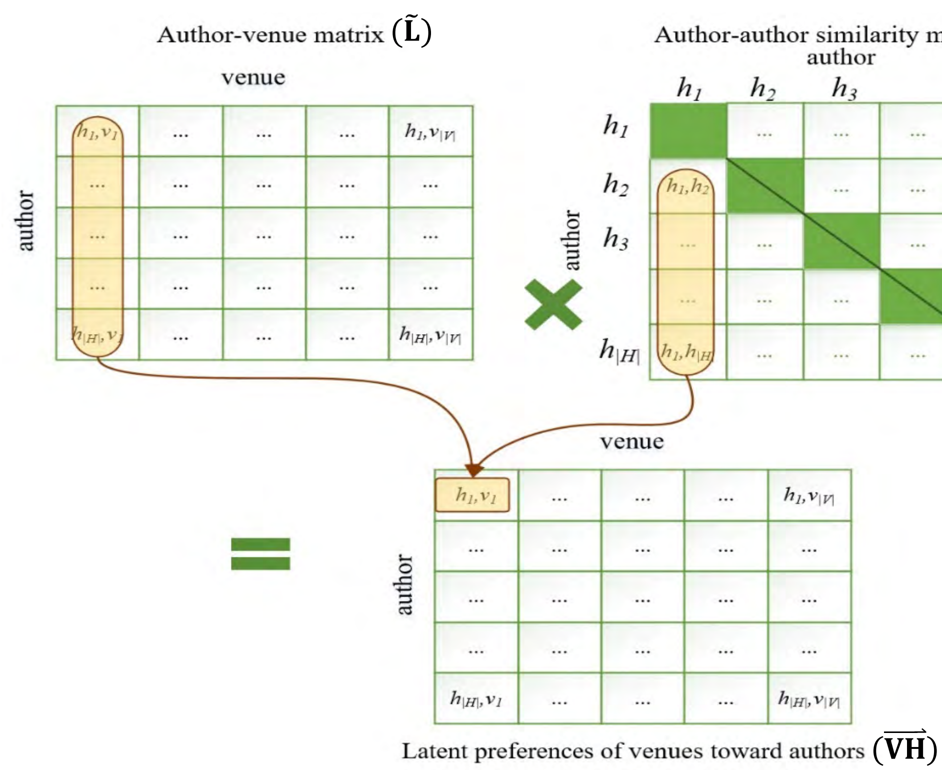

FIGURE 7. **An illustration of the process of computing the venue-author matrix $\overrightarrow{\mathbf{VH}}$.**

details of the construction of the new matrix $\overrightarrow{\mathbf{VH}}$.

$$\overrightarrow{\mathbf{VH}} = \tilde{\mathbf{L}}\mathbf{R}^{k} \tag{5}$$

$$\overrightarrow{\mathbf{AH}} = \tilde{\mathbf{C}}\mathbf{R}^{k} \tag{6}$$

In order to facilitate the process of recommending an academic venue, we recognize the articles' latent preferences toward their selection of authors and vice versa. We analyze the hidden preferences of on the articles' choice toward authors and authors' toward articles. We utilize the matrix **C** and the matrix **R** to capture the articles latent preferences toward authors. We constructed the new author-article latent preference matrix $\overrightarrow{\mathbf{AH}}$. The matrix represents the product of matrix **C** and **R**, as in Equation (6). Where the matrix $\tilde{C}$ represents a normalized version of the matrix **C**, and the matrix $\mathbf{R}^{k}$ denotes the top k nearest authors as explained formerly. The multiplication of the a-th row by the h-th column entails to identify the latent preferences of articles a, on authors h with regard to the authors k nearest neighbor. We utilize the matrix **C** and the matrix **G** to capture the authors' latent preferences toward articles. We constructed the new author-article latent preference matrix $\overrightarrow{\mathbf{HA}}$. The matrix represents the product of matrix **C** and **G**, as in Equation (7):

$$\overrightarrow{\mathbf{HA}} = \tilde{\mathbf{C}}\mathbf{G}^{k} \tag{7}$$

where the matrix $\tilde{C}$ represents a normalized form of the matrix**C**, and the matrix $\mathbf{G^{k}}$ denotes the top k nearest articles.

3) VENUE COLLABORATIVE CONFIDENCE

We use the $\overrightarrow{\mathbf{VH}}$, $\overrightarrow{\mathbf{HA}}$, $\overrightarrow{VA}$ and $\overrightarrow{\mathbf{AH}}$ latent models to build a tailored recommendation model for rating venues according to a given query. The Collaborative Confidence Score (CCS) for each venue will be calculated to propose venues candidate list. The model will compute the CCS each venues based on its relation to the authors and venues using Equation (8).

$$CCS\,(v|q) = \frac{\sum_{h=1}^{|H|} vh_{v,h}.\mu_h}{|H|} + \frac{\sum\limits_{a=1}^{|A|} va_{v,a}.\mu_a}{|A|} \tag{8}$$

where $\mathbf{VH}_{v,h}$ denotes the latent preferences value of venue v toward a specific author h in the matrix $\overrightarrow{\mathbf{VH}}$ and $va_{v,a}$ denotes the latent preferences value of venue v toward a specific article a in the matrix $\overrightarrow{\mathbf{VA}}$, μ_h denotes the mean of latent preferences value of a specific author h toward the articles in the citation network where $\mu_h = \frac{\sum_{a=1}^{|A|} ha_a}{|A|}$.$\mu_a$ denotes the mean of latent preferences value of a specific article a toward the authors in the citation network where $\mu_a = \frac{\sum_{h=1}^{|H|} \mathbf{AH}_h}{|H|}$. The main idea here is to consider the value of the other dimension in the citation network and its effect to close the relation loop. $|V|$ represents the total number of venues. $|A|$ represents the total number of articles. $|H|$ represents the total number of authors. A latent preference value reflects the venues' effects toward authors and articles. Thus, it can increased the potential expansion of possibilities according to the venue's research scope. By utilizing latent models $\vec{\mathbf{VH}}$, $\vec{\mathbf{HA}}$, $\vec{\mathbf{VA}}$ and $\vec{\mathbf{AH}}$, venues that are suitable according to the user queries are ranked higher in the potential search list. We consider the other dimension impact to close the academic trinity that reflects the relation among the citation network of the given query upon user request.

We have presented an academic venue recommendation model in an extended nearby citation network that aims to search and rank academic venues relevant to users' inquiry to build an effective scientific community. The recommendation model computes the cosine similarity between the authors, and articles using the collaborative filtering approach to estimate the connection between the objects. After that, we build the latent preferences models for an academic venue to be used in the computation of the ranking score. The ranking model uses the latent preferences models to estimate the collaborative confidence score for each venue, depending on a user's query, to determine the most suitable venues that are personalized to the user's requests.

TABLE 4. Statistics of the IEEE dataset used to evaluate the recommendation system.

Dataset containing	Articles $\|A\|$	Authors $\|H\|$	Venues $\|V\|$
"database"	143,524	3,563	130
"multimedia"	95,156	11,086	204

IV. EXPERIMENTAL EVALUATION AND RESULTS

To evaluate the proposed system aiming article recommendation, we followed the evaluation procedure described in [4]. We used previously mentioned two sub-datasets of the IEEE digital library [21] for our experiments that defined in Table 4. Each dataset has been detached into two groups: 20% of the dataset randomly restrained as the test set and the rest of 80% used as the training set; then we performed five runs using various training-test partition datasets to avoid the possibility that the test set was biased.

Three types of evaluation metrics were adopted: precision at top-k using Equation (9), recall at top-k using Equation (10) and F1 measure using Equation (11) to assess the accuracy of the proposed system.

$$precision = \frac{|RC \cap RL|}{|RC|} \tag{9}$$

$$recall = \frac{|RC \cap RL|}{|RL|} \tag{10}$$

$$F_1 = \frac{2\,(precision.recall)}{precision + recall} \tag{11}$$

A. COMPUTATIONAL ANALYSIS

1) COMPUTING SIMILARITY

We analyze the computational complexity of the objects similarity matrices according to the number of articles $|A|$, the number of authors $|H|$, and the number of venues $|V|$. In order to decrease the time and cost of computing the offline similarities, we only consider the most similar neighbors for each similarity computation. Accordingly, we exclude the computed similarities of the academic objects that share few connections with others, and assign a zero similarity value if the similar objects is not among the top k nearest neighbors. In addition, we ignore computing the similarities of the same vector entries, For instance, we employ the matrix G, where $\mathbf{G} = \mathbf{G}^k$ to form the article-article similarity matrix. For any two column vectors $sim(g_x, g_y) = 1$, if $x = y$. Therefore, the outcome similarity matrix will be represented as a diagonal similarity matrix. Then, only the non-zero entries are used in the process of building the recommendation model. The worst-case computational cost of building the similarity matrices ($\mathbf{R}_{|H|\times|H|}$, and $\mathbf{G}_{|A|\times|A|}$) is $O(|H|^2\,|A|\,|V|)$, and $O(|A|^2\,|H|\,|V|)$ respectively.

2) IDENTIFYING THE LATENT PREFERENCES

We analyze the computational complexity of our LPIM model algorithm according to $|A|, |H|, |V|$, the number of similar articles k_a, and the number of similar authors k_h, and the number of similar venues k_v. Identifying the hidden preferences is completed offline in the recommender system. In the worst-case scenario, computing the six hidden preferences models $\vec{\mathbf{HA}}_{|H|\times|A|}$, $\vec{\mathbf{AH}}_{|A|\times|H|}$, $\vec{\mathbf{VH}}_{|V|\times|H|}$ and $\vec{\mathbf{VA}}_{|V|\times|A|}$ requires $O\,(k_a\,|H|\,|A|)$, $O\,(k_h\,|A|\,|H|)$, $O(k_h|V||H|)$ and $O(k_a|V||A|)$ respectively. However, this step depends on the creating the similarity matrix. Thus, the total cost of building each model contains the cost creating the similarity matrix in section A-1 and the cost of this step. Thus, we have considered the worst-case scenario for calculating the computational cost. Therefore, calculating each model costs $O\left(|H|^2\,|A|\,|V| + k_a\,|H|\,|A|\right)$, $O\left(|H|^2\,|A|\,|V| + k_v\,|H|\,|V|\right)$, $O\left(|A|^2\,|H|\,|V| + k_h\,|A|\,|H|\right)$, $O\left(|A|^2\,|H|\,|V| + k_v\,|A|\,|V|\right)$, $O(|V|^2\,|A|\,|H| + k_h|V||H|)$ and $O(|V|^2\,|A|\,|H| + k_a|V||A|)$ for models $\vec{\mathbf{HA}}$, $\vec{\mathbf{AH}}$, $\vec{\mathbf{VH}}$ and $\vec{\mathbf{VA}}$ respectively.

3) COMPUTING THE ACADEMIC COLLABORATIVE RANKING

All the previous steps to build the citation network and find the latent preferences using the similarity matrices computation can be done offline whenever new information available. When building the latent model, objects scores for a given user query will calculated based on the extended nearby citation network. To find the authors, articles and venues candidate list for a given query, the required computational cost are $O\left(|h|^2\,|a|\,|v|\right)$, and $O\left(|a|^2\,|h|\,|v|\right)$, respectively, where $|h|$, $|a|$ and $|v|$ represent a collection of authors, articles and venues in the extended nearby citation network for given article.

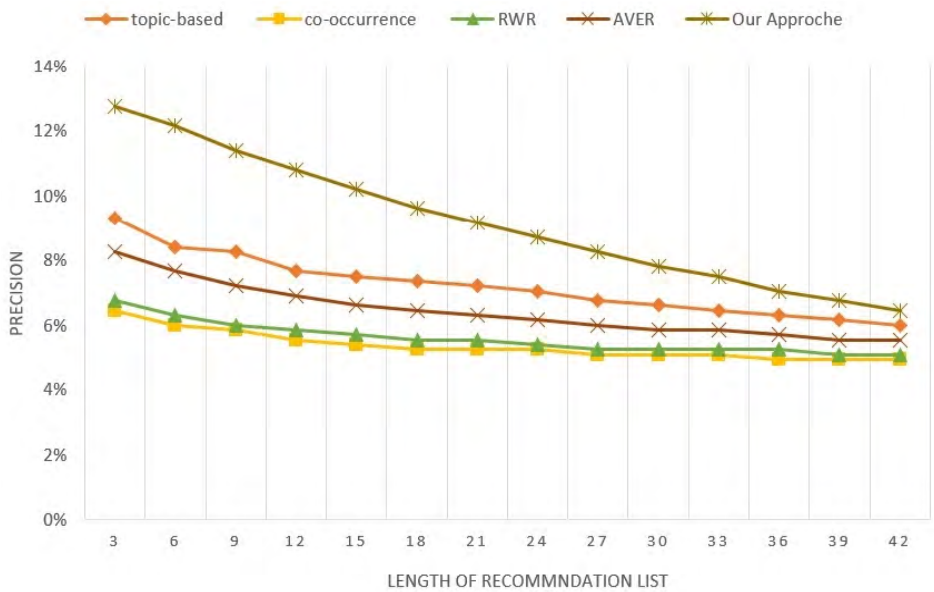

FIGURE 8. Precision comparison for different approaches to recommend venues.

B. RETRIEVAL ACCURACY AND COMPARISON

We compared our proposed approach with four recommendation models: Academic Venue Recommendation model (denoted AVER), Random Walks with Restarts model (RWR), topic-based model, and co-occurrence model all described in [4]. AVER is an extended version of random walks with restarts model where the edge's weight between two nodes based on co-publication frequency and researcher academic level. RWR method is a renowned method in the graph-based approach, has been widely used to determine the most related object based on the importance weight. It uses a weighted graph between nodes to endorse the candidate objects. It estimates the edge's weight between the nodes based on different academic features. we use the co-publishing enumeration to estimate the edge's weight between the nodes in this operation. In this implementation, we use the topic-based model to denote content-based approach. The cosine similarity has been used to compute the similarity between the venues. We calculate the similarity using the topic of the articles as a feature and we treated it as vectors. The co-occurrence model represents the collaborative filtering approach. This depends on a frequent occurrence of two items together. For instance, if author S and author F are attending the same venue X, and author S attended venue Y. Thus, venue Y can be recommended to author F to attend. By conducting this experiment, we aim to verify the accuracy of the proposed methodology and compare it with the different approaches to predict the most related venue where the article can be submitted.

We used the reference list for each article in the experimental dataset to obtain the bibliographic information about the authors and venues. We employed the place names for each tested article as the basis for assessing the quality of the recommendation. The goal was to compare the system recommendations to the actual user selections. With regard to the venue selections, the venue names from different articles were combined in a testing pool, and we then examined the relation between the venues in the testing pool and the given article. The system can then recognize the candidate list of venues that will be compared with the actual selection. For each given article we determined the extent to which the approach will predict the venues the article was published in, based on the given query.

Figure 8 and Figure 9 illustrate our experimental results compared to the other four approaches based on venue recommendation. We retrieve the top k venues ranked that should be recommended to the user. Our proposed approach outperformed the other approaches as shown in precision and recall curves. The precision and recall values of a specific top k retrieved venues are shown as data points of the graph curves, with the left points on the curves denoting top $k = 3$, and the points on the right denoting the top $k = 42$. Our approach showed 19.9%, 39.7%, 37.3%, and 28.5% precision improvement and 20.0%, 39.9%, 36.7%, and 29.1% recall improvement on the dataset average compared to topic-based, co-occurrence, RWR, and AVER models respectively. We continue to examine the F-measures of each approach, as shown in Figure 10, the proposed system obtained approximately 20.0%, 39.9%, 37.1%, and 28.7% improvement on the F-measures.

We found the accuracy of our approach had higher average accuracy and retrieval values than other methods when retrieving a different number of venues. When we compared the reported results over the long term we found that the topic-based model achieved better results than the AVER, RWR, and co-occurrence models. In addition, using an LPIM model

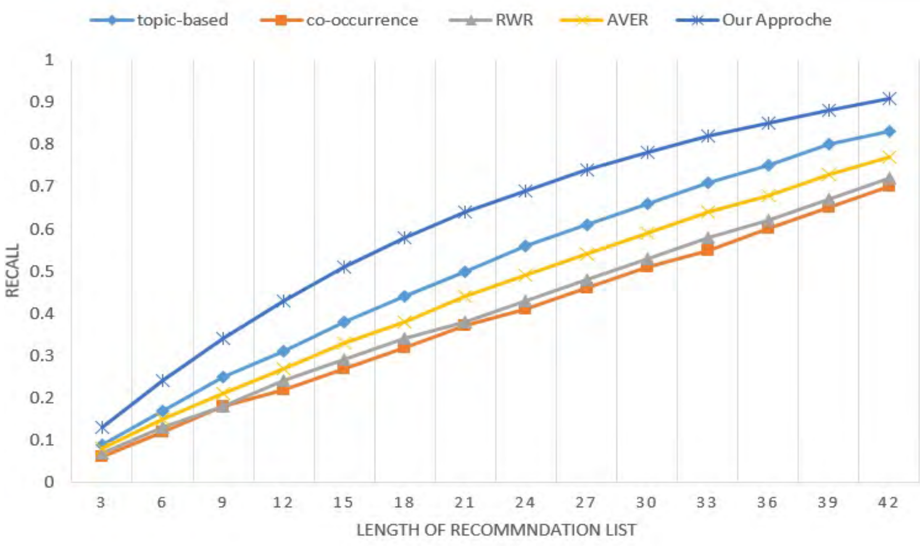

FIGURE 9. Recall comparison for different approaches to recommend venues.

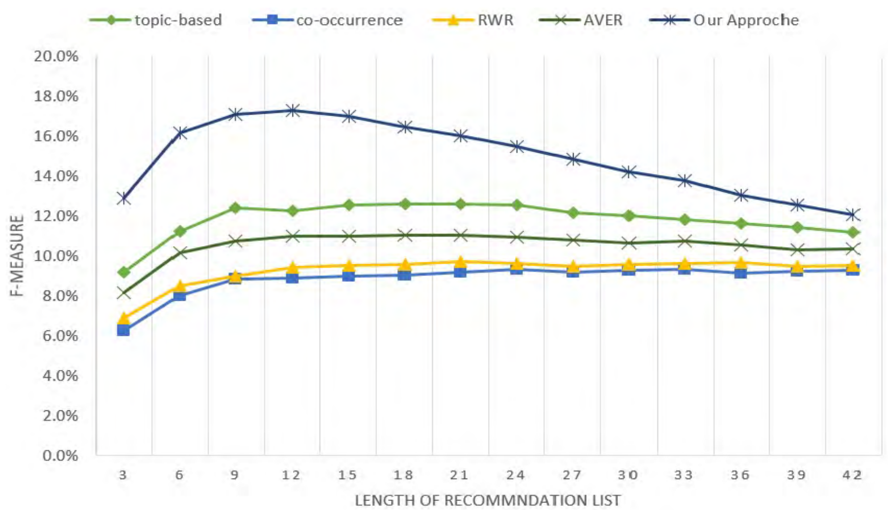

FIGURE 10. F-measurement comparison for different approaches to recommend venues.

based on multiple dimensions revealed more relevant venues than other filtering approaches, which indicates that it had positive impact on recommending more efficient candidate lists.

V. CONCLUSION AND FUTURE WORK

The focus of this paper is venue recommendation for researchers, based on similarity learning of extended nearby citation networks of a given article. We combined the citation network matrices that represent the relation between academic objects and the similarity matrices, to identify the latent preference in the extended network. We explain how we built and utilized the citation and similarity matrices to find the latent preference in the extended network. We conducted extensive experiments on an actual academic dataset to examine the accuracy of the proposed approach compared to other baseline approaches. The experiment results verify that our approach outperforms other baseline systems in terms of precision and recall. In addition, our method to identify latent preferences can enhance recommendation capabilities and retrieve more relevant venues, considering the similarity learning of the extended nearby citation network.

Potential improvements will be addressed in upgrades, including the influence of object impact index on a particular research scope or institution. We also intend to enrich the analysis by tuning the recommender system so it is capable of simultaneously suggesting different types of collaboration in the scientific community, including current and former affiliation information for authors. In addition, we aim to develop the analysis further by considering different factors affecting the venue impact, including the journal editorial board, the conference committee and publisher reputation.

REFERENCES

[1] A. E. Saddik, "Digital twins: The convergence of multimedia technologies," *IEEE Multimed.*, vol. 25, no. 2, pp. 87–92, Apr./Jun. 2018.
[2] A. M. Alshareef, M. F. Alhamid, and A. E. Saddik, "Article impact value for nearby citation network analysis," in *Proc. IEEE Int. Conf. Big Data Smart Comput.*, Jan. 2018, pp. 398–403.
[3] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Research-paper recommender systems: a literature survey," *Int. J. Digit. Libraries*, vol. 17, no. 4, pp. 305–338, Jun. 2015.
[4] Z. Chen, F. Xia, H. Jiang, H. Liu, and J. Zhang, "Aver: Random walk based academic venue recommendation," in *Proc. 24th Int. Conf. World Wide Web*, May 2015, pp. 579–584
[5] Z. Yang and B. D. Davison, "Venue recommendation: Submitting your paper with style," in *Proc. 11th Int. Conf. Mach. Learn. Appl.*, Dec. 2012, pp. 681–686.
[6] E. Medvet, A. Bartoli, and G. Piccinin, "Publication venue recommendation based on paper abstract," in *Proc. Int. Conf. Tools with Artif. Intell.*, Nov. 2014, pp. 1004–1010.
[7] M. E. J. Newman, "The structure of scientific collaboration networks," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 2, pp. 404–409, 2001.
[8] T. Hansen and R. F. Hansen, "Integration of the scientific community as exemplified by the biotech sector: An analysis based on bibliometric indicators in the Danish-Swedish border region," *GeoJournal*, vol. 67, no. 3, pp. 241–252, 2006.
[9] P. Tuire and E. Lehtinen, "Exploring invisible scientific communities: Studying networking relations within an educational research community. A finnish case," *High. Educ.*, vol. 42, no. 4, pp. 493–513, Dec. 2001.
[10] L. De Vocht *et al.*, "Visualizing collaborations and online social interactions at scientific conferences for scholarly networking," in *Proc. 24th Int. Conf. World Wide Web*, May 2015, pp. 1053–1054
[11] M. E. J. Newman, "Scientific collaboration networks: I. Network construction and fundamental results," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 1, pp. 1–8, Jun. 2001.
[12] D. Mu, L. Guo, X. Cai, and F. Hao, "Query-focused personalized citation recommendation with mutually reinforced ranking," *IEEE Access*, vol. 6, pp. 3107–3119, 2017.
[13] K. Boyack and R. Klavans, "Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?" *J. Amir. Soc. Inf. Sci. Technol.*, vol. 61, no. 12, pp. 2389–2404, May 2010.
[14] S. Cohen and L. Ebel, "Recommending collaborators using keywords," in *Proc. 22nd Int. Conf. World Wide Web*, Sep. 2013, pp. 959–962.
[15] S. Yu *et al.*, "PAVE: Personalized academic venue recommendation exploiting co-publication networks," *J. Netw. Comput. Appl.*, vol. 104, pp. 38–47, Dec. 2017.
[16] A. M. Alshareef, M. F. Alhamid, and A. E. Saddik, "Recommending scientific collaboration based on topical, authors and venues similarities," in *Proc. IEEE Int. Conf. Inf. Reuse Integr. (IRI)*, May 2018, pp. 55–61.
[17] A. M. Alshareef, M. F. Alhamid, and A. E. Saddik, "Toward citation recommender systems considering the article impact in the extended nearby citation network," *Peer-to-Peer Netw. Appl.*, vol. 1, pp. 1–10, Oct. 2018.
[18] A. Huang, "Similarity measures for text document clustering," *New Zeal.*, vol. 4, pp. 49–56, Apr. 2008.
[19] Y. K. Jeong, M. Song, and Y. Ding, "Content-based author co-citation analysis," *J. Informetr.*, vol. 8, no. 1, pp. 197–211, 2014.
[20] N. J. Van Eck and L. Waltman, "Appropriate similarity measures for author co-citation analysis," *J. Amir. Soc. Inf. Sci. Technol.*, vol. 59, no. 10, pp. 1653–1661, Jun. 2008.
[21] IEEE. *API Xplore Digital Library*. Accessed: Feb. 11, 2018. [Online]. Available: https://developer.ieee.org/

ABDULRHMAN M. ALSHAREEF received the M.Sc. degree in electronic business technologies from the University of Ottawa, where he is currently pursuing the Ph.D. degree with the Multimedia Communications Research Laboratory, with a focus on data mining, information retrieval, and recommendations systems. He is also a Lecturer with the Information System Department, King Abdulaziz University, Jeddah, Saudi Arabia. His current research interests include research impacts and research recommendations among scholar communities.

MOHAMMED F. ALHAMID (GS'08–M'10) received the Ph.D. degree in computer science from the University of Ottawa, Canada. He is currently an Assistant Professor with the Software Engineering Department, King Saud University, Riyadh, Saudi Arabia. His research interests include recommender systems, social media mining, big data, and ambient intelligent environment.

ABDULMOTALEB EL SADDIK (M'01–SM'04–F'09) is currently a Distinguished University Professor and the University Research Chair with the School of Electrical Engineering and Computer Science, University of Ottawa. He has authored or co-authored ten books and more than 550 publications and has chaired more than 50 conferences and workshops. His research interest includes the establishment of digital twins using AI, AR/VR, and the tactile Internet that allow people to interact in real time with one another as well as with their digital representation. He has received research grants and contracts totaling more than $18 million. He has supervised more than 120 researchers and has received several international awards. He is an ACM Distinguished Scientist and a Fellow of the Engineering Institute of Canada and the Canadian Academy of Engineers. He was a recipient of the IEEE I&M Technical Achievement Award, the IEEE Canada C.C. Gotlieb (Computer) Medal, and the A.G.L. McNaughton Gold Medal for important contributions to the field of computer engineering and science.