



Scholarly recommendation systems: a literature survey

Zitong Zhang¹ · Braja Gopal Patra² · Ashraf Yaseen¹ · Jie Zhu¹ · Rachit Sabharwal¹ · Kirk Roberts³ · Tru Cao¹ · Hulin Wu¹

Received: 3 June 2022 / Revised: 28 April 2023 / Accepted: 9 May 2023 /
Published online: 4 June 2023
© The Author(s) 2023

Abstract

A scholarly recommendation system is an important tool for identifying prior and related resources such as literature, datasets, grants, and collaborators. A well-designed scholarly recommender significantly saves the time of researchers and can provide information that would not otherwise be considered. The usefulness of scholarly recommendations, especially literature recommendations, has been established by the widespread acceptance of web search engines such as CiteSeerX, Google Scholar, and Semantic Scholar. This article discusses different aspects and developments of scholarly recommendation systems. We searched the ACM Digital Library, DBLP, IEEE Explorer, and Scopus for publications in the domain of scholarly recommendations for literature, collaborators, reviewers, conferences and journals, datasets, and grant funding. In total, 225 publications were identified in these areas. We discuss methodologies used to develop scholarly recommender systems. Content-based filtering is the most commonly applied technique, whereas collaborative filtering is more popular among conference recommenders. The implementation of deep learning algorithms in scholarly recommendation systems is rare among the screened publications. We found fewer publications in the areas of the dataset and grant funding recommenders than in other areas. Furthermore, studies analyzing users' feedback to improve scholarly recommendation systems are rare for recommenders. This survey provides background knowledge regarding existing research on scholarly recommenders and aids in developing future recommendation systems in this domain.

Keywords Scholarly recommendation systems · Literature recommendation · Collaborator recommendation · Conference recommendation · Journal recommendation · Reviewer recommendation

Z. Zhang, B.G. Patra These authors contributed equally to this work.

✉ Ashraf Yaseen
Ashraf.Yaseen@uth.tmc.edu

Extended author information available on the last page of the article

1 Introduction

A recommendation or recommender system is a type of information filtering system that employs data mining and analytics of user behaviors, including preferences and activities, to filter required information from a large information source. In the era of big data, recommendation systems have become important applications in our daily lives by recommending music, videos, movies, books, news, etc. In academia, there has been a substantial increase in the extent of information (literature, collaborators, conferences, datasets, and many more) available online and it has become increasingly taxing for researchers to stay up to date with relevant information. Several recommendation tools and search engines in academia (Google Scholar, ResearchGate, Semantic Scholar, and others) are available for researchers to recommend relevant publications, collaborators, funding opportunities, etc. Recommendation systems are evolving rapidly. The initial scholarly recommender system was intended for literature by recommending publications using content-based similarity methods [1]. Currently, there are several recommendation systems available for researchers and these are widely used in different scholarly areas.

1.1 Motivation and research questions

In this article, we focus on different scholarly recommenders used to improve the quality of research. To the best of our knowledge, no article currently focusing on all scholarly recommendation systems together is available right now. Previous surveys on recommendation systems were conducted separately for each recommendation system. Most of these studies were based on literature or collaborator recommendation systems [2]. Currently, there is no comprehensive review that contains a description of different types of scholarly recommendation systems, particularly for academic use.

Therefore, it is necessary to provide a survey as a guide and reference to researchers interested in this area; a systematic review of scholarly recommendation system would serve this purpose. It helps to explore research achievements in scholarly recommendation, provide researchers with an overall presentation of systems for allocating academic resources, and identify improvement opportunities.

This article describes the different scholarly recommendation systems that researchers use in their daily activities. We are taking a closer look at the methodologies used for developing such systems. The research questions of our study are as follows:

- *RQ1* What different problems are addressed by scholarly recommendation systems?
- *RQ2* What datasets or repositories were used for developing these recommendation systems?
- *RQ3* What types of methodologies were implemented in these recommendation systems?
- *RQ4* What further research can be performed to overcome the drawbacks of the current research and develop new recommenders to enhance the field of scholarly recommendation?

To answer our first research question, we collected over 500 publications on scholarly recommenders from the ACM Digital Library, DBLP, IEEE Explorer, and Scopus. Literature and collaborator recommendation systems are the most studied recommenders in the literature, with many publications in each. Websites for searching publications host literature recommendations as a key function, almost all of which are free for researchers. However, a few collaborator recommendation systems have been implemented online; and are not free

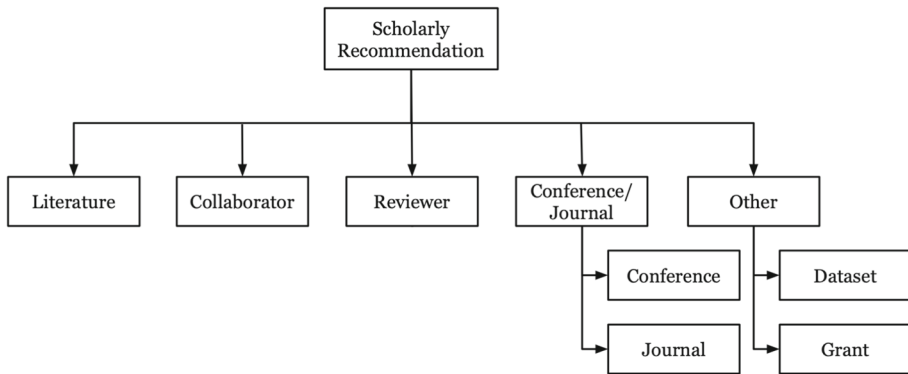


Fig. 1 Scholarly recommenders studied in this article

for all users. One of the reasons can be attributed to the large amount of personal information and preferences required by these recommenders.

Furthermore, we studied journal and conference recommendation systems for publishing papers and articles. Although many publishing houses have implemented their own online journal recommender systems, conference recommender systems are not available online. Next, we studied reviewer recommendation problems, in which reviewers are recommended for conferences, journals, and grants. Finally, we identified datasets and grant recommendation systems, which are the least studied scholarly recommendation systems. Figure 1 shows all currently available scholarly recommendations.

1.2 Materials and methods

An initial literature survey was conducted to identify keywords related to individual recommendation systems that can be used to search for relevant publications. A total of 26 keywords were identified to search for relevant publications (see Supplementary 17).

At the end of the full-text review process, 225 publications were included in this study. The number of publications on individual recommendation systems is shown in Fig. 2. To be eligible for the review, we focused on the description, evaluation, and use of natural language processing algorithms. During the full-text review process, we excluded studies that were not peer-reviewed, such as abstracts and commentary, perspective, or opinion pieces. Finally, we performed data extraction and analysis on 225 articles and summarized their data, methodology, evaluation metrics, and detailed categorization in the following sections. The PRISMA flowchart for our publication collection is shown in Fig. 3; with example search keywords.

The remainder of this paper is organized as follows. Section 2 describes different literature recommendation systems based on their methodologies and corresponding datasets. Section 3 describes different approaches for developing collaborator recommendation systems. Section 4 reviews the journal and conference venue recommendation systems. Section 5 describes the reviewer's recommendation system. In Sect. 6, we review all other scholarly recommendation systems available in the literature such as datasets and grant recommendation systems. Finally, Sect. 7 discusses future work and concludes the article.

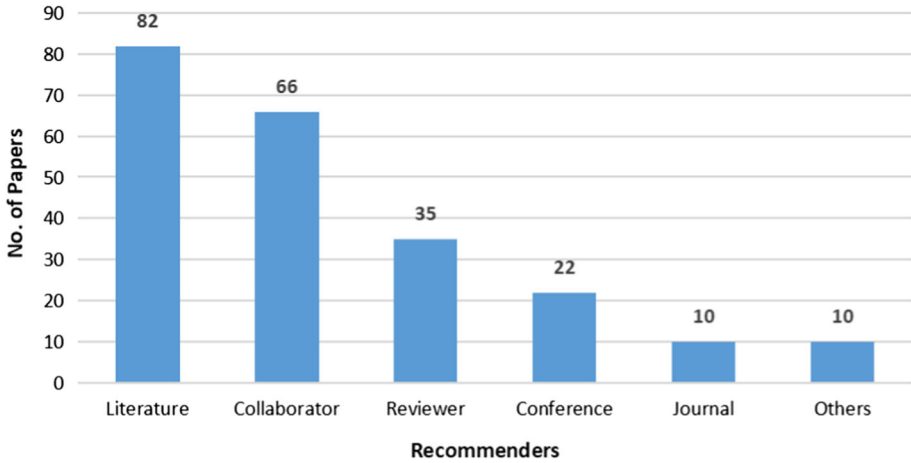


Fig. 2 Number of papers/articles collected for studying different recommenders

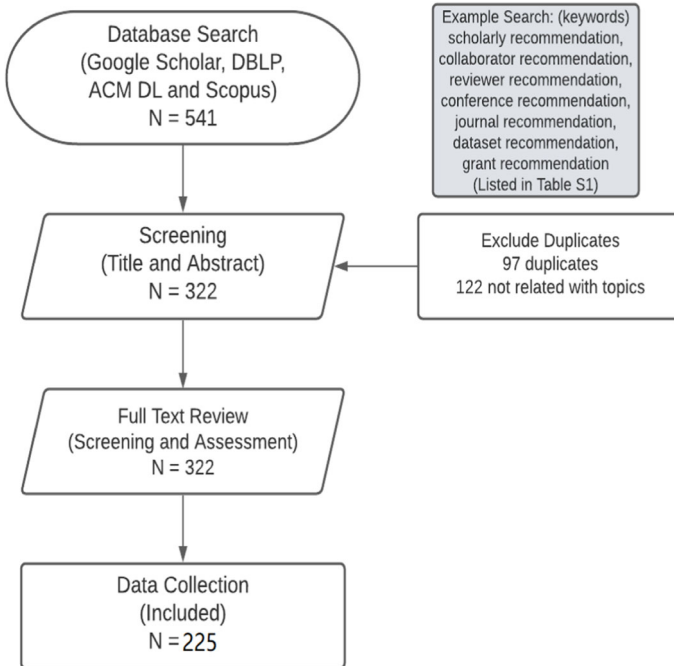


Fig. 3 PRISMA flowchart for including publications in scholarly recommendation

2 Literature recommendation

Literature recommendation is one of the most well-studied scholarly recommendation problems with several research articles published in the past decade. Recommender systems for scholarly literature have been widely used by researchers to locate papers, keep up with their

Table 1 Sources of datasets used for literature recommendation approaches

	CiteSeer	ACM	DBLP	CiteULike
Absolute	9	11	12	7
Relative	32%	40%	43%	25%

research fields, and find relevant citations for drafts. To summarize the literature recommendation systems, we collected 82 publications for scholarly papers and citations.

The first research paper recommendation system was introduced as a part of the CiteSeer project [1]. In total, 11 out of 82 publications (approximately 13%) used applications or methodologies based on a citation recommendation system. As one of the widest subsets of scholarly literature recommendation, citation recommendation aims to recommend citations to researchers while authoring a paper and finding work related to their ideas. It recommends citations based on the content of the researchers' work. Among the 11 citation recommender papers, content-based filtering (CBF) methodologies have been widely used on the fragments of the citations for the recommendation, and some of them applied collaborative filtering (CF) to develop a potential citation recommendation system based on users' research interests and citation networks [3].

2.1 Data

In this section, we describe the datasets used to develop literature recommendation systems. A total of 75 reviewed publications evaluated the methodologies using different datasets. The authors of 45 publications chose to construct their own datasets based on manually collected information or paid datasets that were rarely used. Several open-source published datasets are commonly used to develop literature recommendations.

Owing to the rapid development of modern websites for literature search, datasets for literature recommendation are readily available. There were 28 publications that used public databases for the testing and evaluation of the methods. The sources of these datasets are listed in Table 1. These websites collected publications from several scientific publishers and indexed them with their references and keywords. Using the information extracted from these public resources, researchers created datasets to perform recommendation methodologies and obtain the ground truth for offline evaluation.

DBLP was used in 12 reviewed publications and ACM was used in 11 reviewed publications to construct datasets for evaluation. DBLP hosts more than 5.2 million publications,¹ and obtains its database entries by using a limited number of volunteers who manually enter tables of contents of journals and conference proceedings. The CiteSeer dataset was used in 9 reviewed publications to conduct an offline evaluation. It currently contains over 6 million publications; and is continuously crawling the web to find new content using user submissions, conferences, and journals as data entries. Petricek et al. [4] proved that the application of autonomous acquisition through web crawling in CiteSeer introduces a significant bias against papers with a low number of authors. Among the reviewed papers, we can say that most of researchers constructed their own datasets for evaluation by combining the information from multiple databases. These self-constructed evaluation datasets based on different resources were used to avoid bias resulting from using information from only one source.

¹ <https://dblp.org>, accessed on October 16, 2020.

The CiteULike dataset was used in 7 reviewed publications. CiteULike is a web service that contains social tags added to research articles by users. The dataset was not originally intended for literature recommendation system research, but is still frequently used for this purpose.

2.2 Methods

Three main approaches were used to develop literature recommenders: CBF ($N = 37$ papers), CF ($N = 16$ papers), and hybrid ($N = 29$ papers). Next, we introduce the promising and popular approaches used in each recommendation class. We also provide an overview of the most important aspects and techniques used for literature recommendation.

2.2.1 Content-based filtering (CBF)

CBF is one of the most popular methods for recommending literature and is used in 37 of 82 publications. Based on the user-item model that treats textual contents as ‘items,’ CBF usually uses topic-based methods to measure the similarity of the publication’s topic that users are interested in and the topic of target publications. These methods performed well in terms of topic and content matching. A summary of CBF approaches used for literature recommendation can be found in Table 2.

CBF recommenders use keywords or topics as key features because they are used to describe a publication. The creation of a content-based profile of users usually concentrates on the user’s preference model, and the user’s interaction log with the recommendation system converted by a weighted vector of item features. For example, Hong et al. [9] constructed a paper recommendation methodology based on a user profile built with extracted keywords, and calculated the similarity between a given topic and collected papers by using cosine similarity to recommend initial publications for each topic.

Most of the reviewed publications used the term frequency and inverse document frequency (TF-IDF) representation to evaluate the similarities between text objects. TF-IDF negates the effect of high-frequency words while determining the importance of an item. Magara et al. [38] constructed methodologies for recommending serendipitous research papers from two large normally mismatched information spaces or domains using Bisociative Information Networks (BisoNets) and TF-IDF measures as weighting and filtering terms. Lofty et al. [11] combined TF-IDF with a cosine similarity measure to construct a methodology for paper recommendation using ontology. To address higher relevancy and serendipity, Sugiyama and Kan [25] also constructed feature vectors using the TF-IDF measure and user profiles utilizing the Co-Author Network (CAN), computed cosine similarity and recommended papers with higher similarity.

In summary, researchers claim that content-based recommender systems are independent for each user to build their own profiles so that the most suitable recommendation can be made for different users. Also, based on automatically generated user models, recommendation systems using CBF would spend less time and calculation on up-front classification.

The limitations of CBF can also be concluded. The improvements made in the papers we collected were mostly to overcome these limitations. CBF requires more calculation and resources to analyze each item for its features and build each user model individually. For example, to mark passages for citation recommendations, users are typically required to provide a representative bibliography. By examining the relevance between segments in a query manuscript and the representative segments extracted from a document corpus, He et al. [36]

Table 2 Overview of literature recommendation systems using CBF

Citation	Data	Method	Evaluation
[5]	-	-	-
[6]	-	TF-IDF, cosine similarity	-
[7]	-	CBF	-
[8]	-	SVM classifier	-
[9]	Self-conducted database	Cosine Similarity	Accuracy
[10]	Self-conducted database	CBF	Accuracy
[11]	ACM DL	TF-IDF, cosine similarity	Accuracy, P, Recall
[12]	DBLP, MAS(Microsoft Academic Search)	Markov Chain model	Compare Inference & Similarity
[13]	PubMed	Bisociative Information Networks	DTM
[14]	NLP	MFA (atrix-forest-based algorithm)	HITS ranking, co-citation graph
[15]	SfN (Society for Neuroscience)	TF, TF-IDF, log-entropy	Human curated topic distances
[16]	CiteSeerX	DBSCAN (density-based clustering)	Manual Evaluation
[17]	Docear's Digital Library	CBF	Manual Evaluation
[18]	NIIREO	TF-IDF, cosine similarity	Manual Evaluation
[19]	Self-conducted database	CBF, designed questionnaire	Manual Evaluation
[20]	Self-conducted database	CBF	Manual Evaluation, Accuracy
[21]	OpenEdition's search engine	CBF, centrality indicator	Manual rating
[22]	Rexa	Co-citation coupling Katz graph distance measure	Manual, P
[23]	ACM DL	PRF (pseudo relevance feedback) algorithms	MAP, NDCG
[24]	Google Scholar	Knowledge-based graph representation	MRR, MAP, NDCG
[25]	DBLP, ACM DL	TF-IDF, DU	MRR, NDCG

Table 2 continued

Citation	Data	Method	Evaluation
[26]	National Science and Technology Library	Filter user profile based on content	NDCG
[27]	ArnetMiner	Paper citation graph, TF-IDF, LDA	NDCG, MAP
[28]	Self-conducted database	CBF, keyphrase extraction	NDCG, recall
[29]	Information and Communication Engineers	HITS (analyzing the bipartite graph)	Precision
[30]	CiteULike	CombMNZ (computes the rank score)	P, MRR, NDCG
[31]	CiteSeer, CiteULike	IBM translation model, TF-ICF	P, Recall, MRR, Bpref
[32]	CiteULike	RankSVM	P, MRR, NDCG
[33]	Docear's PDF Spider crawled	CBF	offline evaluation
[34]	Computer Science Publication Dataset	VRRW (Vertex-Reinforced Random Walk)	Recall, MAP, CP
[35]	ACM DL, IEEE	Self-defined weighting schemes	Recall, NDCG
[36]	CiteSeerX	Contextual Similarity Dependency Feature Model	Recall, P, NDCG
[37]	ACM DL	CBF, self-designed criteria	Specificity, Coverage
[38]	Self-conducted database	BisoNets	—
[39]	Self-conducted database	CBF, NLP (natural language processing)	User tests
[40]	CiteSeerX	Context Graph (CG), TF-IDF	User tests, NDCG
[41]	—	CPA, COA	—

formulated a dependency feature model based on language model, contextual similarity, and topic relevance to produce a citation recommendation approach without author supervision. Neethukrishnan et al. [8] proposed a paper recommender methodology using an SVM classifier to find their users' personal ontology similarity to specify the conceptualization. Nasciment et al. [35] also proposed a novel source independent framework for research paper recommendation to reduce the resources required. They designed a framework that required only a single research paper as input, and generated several weighting candidate queries by using terms in that paper, and then applied a cosine similarity metric to rank the candidates to recommend the ones most related to the input paper.

In addition, the traditional CBF methods are not able to consider the popularity and rating of items, that is, it is difficult to differentiate between two research papers if they have similar terms in user model. To overcome this limitation, Ollagnier et al. [21] formulated a centrality indicator for their software, which was dedicated to the analysis of bibliographical references extracted from scientific collections of papers. This approach determines the impact and inner representativeness of each bibliographical reference according to their occurrences. Pera and Ng [30] adopted CombMNZ, a linear combination strategy that combines similarity degree and popularity score into a joint ranking, to build up their application, and a paper recommender system recommends papers considering both context similarity and popularity of the paper among users. Liu et al. [23] constructed a publication ranking approach with PRF (Pseudo Relevance Feedback) by leveraging a number of meta-paths on a heterogeneous bibliographic graph.

2.2.2 Collaborative filtering

We collected 16 studies that used the Collaborative Filtering (CF) method. CF methods find the users that are similar to the target user in their past ratings, and then recommend similar user options to the target user. These methods are suitable for extending the recommended range. A summary of literature recommendation papers using CF methods is presented in Table 3.

Common methodologies using a collaborative filtering algorithms can be categorized into two groups: model-based and memory-based. The main difference between the two approaches is that the model-based approach uses a matrix factorization-based algorithm, in which the preferences of users can be calculated by embedding factors. The memory-based approach calculates the preferences of users for items based on arithmetic operations (correlation coefficients or cosine similarity). Memory-based CF approaches are widely used in scholarly literature recommendation systems, which includes several different approaches, such as k-nearest neighbors (kNN), Latent Semantic Index (LSI), and Singular Value Decomposition (SVD). Pan and Li [48] used the LDA (Latent Dirichlet Allocation) model to construct a paper recommendation system using a thematic similarity measurement to transform a topic-based recommendation into a modified version of the item-based recommendation approach. Ha et al. [46] proposed a novel method using SVD for matrix factorization and rating prediction to recommend newly published papers that have not been cited by other papers by predicting the interests of the target researchers.

Compared to CBF methods and applications based on CF show the following advantages. First, given that CF approaches is independent of content, resource costs for error-prone item processing are reduced. In addition, popularity and quality assessments are often considered limitations of CBF, but CF can achieve them easily. Sugiyama and Kan [43] used the PageRank approach to rank the popularity factor and measure the importance of research papers, to enhance the user profile derived directly from the researchers' past works with

Table 3 Overview of literature recommendation system using collaborative filtering

Citation	Data	Method	Evaluation
[42]	DBLP	CF (machine learning)	MRR, NDCG, recall
[3]	DBLP, ACM DL	CF	MRR, NDCG
[43]	DBLP, ACL ARC	CF	NDCG, MRR
[5]	ACM DL	CF	recall, precision, and F1
[44]	DBLP, ACL ARC	CF (co-author network)	NDCG, MRR, recall, NITN
[45]	Self-conducted study	CF (multidimensional)	User rating
[46]	DBLP	CF (SVD-based method)	Recall, precision
[47]	DBLP	CF, data-intensive computing	Recall, accuracy, error rate
[48]	Self-conducted DB	CF LDA (latent Dirichlet allocation)	Manual Evaluation
[49]	CiteULike	Collaborative filtering	Recall, precision, and F1
[50]	BibSonomy	Collaborative Tagging, FolkRank	MAP(Mean Average Precision)
[51]	DBLP	FP (frequent pattern) growth algorithm	MRR, NDCG
[52]	Socrec Dataset	Information Filtering	Recall, precision
[53]	MovieLens dataset	Subspace clustering algorithm	Recall, precision, and F1
[54]	CiteULike, CiteSeer	Tag inference	Tag growth, tag reuse
[55]	Brazilian DL	CF, TF-IDF	Jaccard and Fuzzy Similarity Function

information coming from their referenced papers as well as papers that cite the work. CF approaches are also used for serendipitous recommendations; because they are usually based on user similarity and not item similarity. Tang and McCalla [44] constructed user profiles via a co-author network to build a serendipitous paper recommendation system based on a scholarly social network.

The limitations of CF are also shown in the reviewed papers. To make precise recommendations, a CF system requires a great volume of existing data to start the recommendation. This problem is called Cold Start. Loh et al. [55] used scientific papers written by users to compose user profiles, representing user interests or expertise in order to alleviate the cold start problem in the recommender system. Data sparsity is another problem, which represents active users only by observing a small subset of the dataset to rate the papers. Keshavarz and Honarvar [47] presented an approach for paper recommendation based on local sensitive hashing by converting the citations of papers to signatures and comparing these signatures to each other to detect similar papers according to their citations. Sugiyama and Kan [3] also applied CF to discover potential citation papers that help in representing target papers to recommend, in order to alleviate sparsity. The authors also attempted to improve the scalability of the approaches, to reduce the amount of calculation and resources required for recommendations.

2.2.3 Hybrid

Approaches to the previously introduced recommendation may be combined with hybrid approaches. We reviewed 29 studies that applied hybrid recommendation approaches. Table 4 summarizes the papers that we collected where literature recommendation was developed using hybrid approaches.

As a combination of CBF and CF, hybrid recommendation approaches can be categorized into four main groups. The first group implements CBF and CF methods separately and then combine their recommendation results. Liu et al. [70] constructed a citation recommendation method that employed an association mining technique to obtain the representation of each citing paper from the citation context. Then, these paper representations were compared pairwise to compute similarities between the cited papers for CF. Zarrinkalam and Kahani [62] used multiple linked data sources to create a rich background data layer and combine multiple-criteria CF and CBF to develop a citation recommender. Zhang et al. [65] constructed a paper recommendation method based on the semantic concept similarity computed from collaborative tags.

The second and third groups incorporate CBF characteristics into a CF method or incorporate some CF characteristics into a CBF method. West et al. [63] formulated a citation-based method for making scholarly recommendations. The method uses a hierarchical structure of scientific knowledge, making possible multiple scales of relevance for different users. Nart et al. [82] built a method that simplifies CF paper recommendations by extracting concepts from papers to generate and explain the recommendations. Zhou et al. [57] used the concepts and methods of community partitioning and introduced a model to recommend authoritative papers based on the specific community. Magalhaes et al. [67] constructed a user paper-based recommendation approach by considering the user's academic curriculum vitae.

The fourth group is to construct a general unifying model that incorporates both content-based and collaborative characteristics. Meng et al. [58] built a unified graph model with multiple types of information (e.g., content, authorship, citation, and collaboration networks) for efficient recommendation. Pohl et al. [64] treated access data as a bipartite graph of users and documents analogous to item-to-item recommendation systems to build a paper recommender method using digital access records (e.g., http-server logs) as indicators. Gipp et al. [41] developed a paper recommender system that used keyword-based search by combining it with citation analysis, author analysis, source analysis, implicit ratings, explicit ratings, and, in addition, innovative and yet unused methods like the 'Distance Similarity Index' (DSI) and the 'In-text Impact Factor' (ItIF).

2.3 Evaluation

The evaluation metrics for different recommendation methods vary, making it difficult to compare them. To objectively compare the performance of these approaches, 75 publications used two main evaluation metrics.

First, accuracy is the most widely used parameter for evaluating a recommendation system, and it is the capability to recommend the most relevant items based on the given information. Among the reviewed papers, many offline evaluation metrics were applied to measure the accuracy. The second factor is the recommendation system's ability to satisfy users. For example, considering serendipitous factors and user requirements instead of only considering the accuracy of the recommendation system. Some of the reviewed papers designed questionnaires for users to collect their feedback, or applied their methods to real-world systems

Table 4 Overview of literature recommendation system using hybrid method

Citation	Data	Method	Evaluation
[41]	–	ICFA, ICDA	–
[56]	–	RNN, MeSH-based similarities	–
[57]	cit-HepTh	GCE (Greedy Clique Expansion)	Accuracy
[58]	ACL Anthology Network	Unified Multilayer Graph Model LDA	MAP, recall
[59]	ACM DL, CiteSeerX	DNTC, TF-IDF	Accuracy, MAP
[60]	ACM DL, IEEE	CF, CBF	Manual rating
[61]	ACM Portal	Correlation Graph	Accuracy
[62]	ACM, DBLP, IEEE	multicriteria CF, CBF	Recall, NDCG
[63]	AMiner	Hierarchical clustering, CBF	Manual rating
[64]	arXIV record	Bipartite graph	MAP
[65]	CiteULike	concept graph	Manual rating
[66]	CiteULike	CF, CBF	Accuracy
[67]	CV-Lattes	TF-IDF, ontology-based user file	NDCG
[68]	DBLP, ACM IEEE, CiteSeer	IR, Ranking SVM	NDCG
[69]	Google Scholar	Chi-square test, ranking measure	Accuracy
[70]	HEP-PH, HEP-TH	CF, CBF	Precision, recall, F1
[71]	DBLP, CiteSeer	Hybrid recommender	Precision
[72]	Mendeley	CF, CBF	recall, precision, and F1
[73]	ML, HCI, DB	CBF, CF	Accuracy
[74]	Mr. DLib	–	–
[75]	Mr. DLib	RF, GBM, GLM	Precision, recall, F1
[76]	Papits	Graph Neural Networks	Precision
[77]	ResearchIndex (CiteSeer)	Hybrid (CF, CBF)	User acceptance
[78]	Self-conducted database	TF-IDF, LDA	Manual rating
[79]	Self-conducted database	Multiple-criteria Decision Aiding	Simply list the ranking score
[80]	Self-conducted database	CBF, CF	Accuracy
[81]	Self-conducted database	two-layer graph model	Precision, recall
[82]	SPC (Scientific Paper Collection)	CG (Context Graph), TF-IDF	RES Accuracy
[83]	World Wild Web	AdaBoost, multiclass classification	Feedback Website

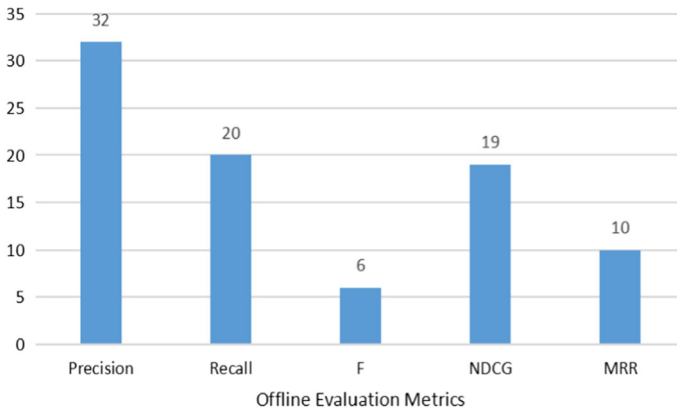


Fig. 4 Distribution for evaluation metrics used in literature recommendation

to evaluate user satisfaction. To quantify and compare the accuracy and user satisfaction of recommendation systems, evaluation methods can be divided into two groups: online and offline.

2.3.1 Online evaluation

A total of 17 publications evaluated their methods with a user study or a real-world system using an online evaluation. They created a rating scheme for users to rate the recommendation results. These manual rating results were then used to analyze and judge an method. In addition, 6 publications out of the 17 applied online evaluations, the methodology of recommendation methods in real-world systems and collected feedback from users for evaluation. Despite analyzing a method based on manually rated the results, online evaluation is typically based on users' acceptance results. Acceptance is commonly measured by the Click-Through Rate (CTR), that is, the ratio of recommendations clicked by users.

2.3.2 Offline evaluation

A total of 59 publications applied offline evaluations to analyze the recommendation algorithms based on the prepared offline datasets. Offline evaluations typically measure the accuracy of recommendation methods based on the ground truth, normally obtained from the information provided by the database, or obtained by manual tests.

To measure the accuracy, precision at position n ($P@n$) is often used to express how many items of the ground truth are recommended within the top n recommendations. Other decision support metrics including Recall and F-measure were also commonly used, often together with Precision as a reference. To evaluate the quality of recommendation, rank-aware evaluation metrics including mean reciprocal rank (MRR) and normalized discounted cumulative gain (nDCG) were also widely used to test highly relevant items that were ranked at the top of a recommendation list. The different evaluation metrics used are illustrated in Fig. 4.

Table 5 Sources of datasets used for collaborator recommendation approaches

	DBLP	ScholarMate (%)	ACMDL (%)	CiteSeerX (%)	Self-conducted (%)	Others (%)
Absolute	17	4	2	2	20	14
Relative	29	7	3	3	34	24

3 Collaborator recommendation

Currently, research in any area has expanded exponentially beyond its own fields to other research fields in the form of collaborative research. Collaboration is essential in academia to obtain good publications and grants. Identifying and determining a potential collaborator is challenging. Hence, a recommendation system for collaboration would be very helpful. Fortunately, many publications on recommending collaborators are available.

3.1 Data

A total of 59 publications were identified using databases to develop, test and evaluate recommender systems. In 20 publications, the authors constructed their own datasets based on manually collected information, unique social platforms, or paid databases that are rarely used. In 39 out of the 59 publications, the authors used open-source databases. Of these 39 publications, 17 used data from the DBLP library to evaluate the developed collaborator recommendation systems.

The datasets needed for developing collaborator recommendations usually include 2 major subjects: (1) contexts and keywords based on researchers' information; and (2) information networks based on academic relationships. Owing to the rapid development of online libraries and academic social networks, the extraction of information networks has become available. These datasets extracted relative information from different online sources and collected information to (i) construct profiles for researchers, (ii) retrieve keywords for constructing a structure, for specific domains and concepts, and (iii) extract weighted co-author graphs. In addition, data mining and social network analysis tools may also be used for clustering analysis and for identifying representatives of expert communities. The sources of datasets used in the 59 publications are listed in Table 5.

Among the reviewed studies, most researchers extracted information from these databases to construct training and evaluation datasets for their recommendations.

The DBLP dataset was used in 17 publications to evaluate the performance of the collaborator recommendation approaches. The DBLP computer science bibliography provides an open bibliographic list of information on major computer science fields and is widely used to construct co-authorship networks. In the co-authorship network graphs of DBLP bibliography, the nodes represent computer scientists and the edges represent a co-authorship incident.

ScholarMate, a social research management tool launched in 2007 was used in 4 publications. It has more than 70,000 research groups created by researchers for their own projects, collaboration, and communication. As a platform for presenting publication research outputs, ScholarMate automatically collects scholarly related information about researchers' output from multiple online resources. These resources include multiple online databases such as Scopus, one of the largest abstract and citation databases for peer-reviewed literature, including scientific journals, books, and conference proceedings. ScholarMate uses aggregated

data to provide researchers with recommendations on relevant opportunities based on their profiles.

3.2 Methods

Similar to other scholarly recommendation areas, research on methodologies to develop collaborator recommendations can be classified into the following categories: CBF, CF, and hybrid approaches. In this section, we introduce the approaches that are widely used in each recommendation class. In addition, we provide an overview of the most important aspects and techniques used in these fields.

3.2.1 Content-based filtering (CBF)

23 publications presented CBF methods for collaborator recommendation. CBF focuses on the semantic similarity between researchers' personal features, such as their personal profiles, professional fields, and research interests. Natural language processing techniques (NLP) were used to extract keywords from the associated documents to define researchers' professional fields and interests. A summary of publications on collaborator recommendation using CBF approaches is presented in Table 6.

The Vector Space Model (VSM) is widely used in content-based recommendation methodologies. By expressing queries and documents as vectors in a multidimensional space, these vectors can be used to calculate the relevance or similarity. Yukawa et al. [84] proposed an expert recommendation system employing an extended vector space model that calculates document vectors for every target document for authors or organizations. It provides a list in the order of relevance between academic topics and researchers.

Topic clustering models using VSM have been widely used to profile fields of researchers using a list of keywords with a weighting schema. Using a keyword weighting model, Afzal and Maurer [85] implemented an automated approach for measuring expertise profiles in academia that incorporates multiple metrics for measuring the overall expertise level. Gollapalli et al. [86] proposed a scholarly content-based recommendation system by computing the similarity between researchers based on their personal profiles extracted from their publications and academic homepages.

Topic-based models have also been widely applied for document processing. The topic-based model introduces a topic layer between the researchers and extracted documents. For example, in a popular topic modeling approach, based on the latent Dirichlet allocation (LDA) method, each document is considered as a mixture of topics and each word in a document is considered randomly drawn from the document's topics. Yang et al. [87] proposed a complementary collaborator recommendation approach to retrieve experts for research collaboration using an enhanced heuristic greedy algorithm with symmetric Kullback–Leibler divergence based on a probabilistic topic model. Kong et al. [88] applied a collaborator recommendation system by generating a recommendation list based on scholar vectors learned from researchers' research interests extracted from documents based on topic modeling.

As mentioned previously in the literature recommendation section, content-based methods usually suffer from a high calculation cost because of the large number of analyzed documents and vector space. To minimize this cost and maximize the preference, Kong et al. [100] presented a scholarly collaborator recommendation method based on matching theory, which adopts multiple indicators extracted from associated documents to integrate the preference matrix among researchers. Some researchers have also modified weighted fea-

Table 6 Overview of collaborator recommendation system using CBF

Citation	Data	Method	Evaluation
[89]	Self-conducted database	Knowledge Discovery in Databases Wu-Palmer's algorithm	
[90]	Self-conducted database	FLM (Fuzzy Linguistic Modeling)	Precision, Recall, F1
[91]	ScholarMate	Social-network analysis business intelligence bibliometric analysis	
[87]	ScholarMate	SKL(symmetric Kullback–Leibler) divergence	AR, NDCG
[84]	Self-conducted database	Concept-based Relevance Discerning Methods	
[92]	DBLP-AI, DBLP-DM	LDA, RWR (random walk and recommendation)	Precision, Recall
[93]	Recruitment site	MAHS (Main Activity Hierarchical System) TF-IDF, SVM (Machine Learning)	Precision, Recall, F measure, Accuracy
[94]	Resource Allocation and Tracking System	Science graph relationship network structure	
[95]	CiteSeerX, Wikipedia	Random forest	Consensus score
[96]	Wikipedia	DBPedia URLs, SPARQL queries (filter data)	
[85]	FacetedDBLP Journal of Universal Computer Science	Template-based Information Extraction using Rule-based Learning	
[97]	Self-conducted database	CKB (Context Knowledge Base)	Questionnaire
[98]	MEDLINE bibliographic database	SemRep (extract semantic) Neo4j Graph Database LBD (literature-based discovery)	
[99]	Self-conducted database	BoW(Bag-of-Words) representation Jaccard Similarity TF-IDF, LDA model	Precision, Recall
[100]	Database from MAG	PPG (Priority to Popular Groups) MCS (Minimum Cost in Step) Word2vec and Doc2vec	Precision
[101]	Self-conducted database	Modularity measure Newman algorithm	Precision
[102]	AI dataset (Artificial Intelligence journal) NN (Neural Networks journal) FGCS (Future Generation Computer Systems)	SLN (Semantic Link Network)	Precision, Recall, F1
[103]	DBLP co-authorship data	Leverage attributes of collaborations combined score function	TOPSCORE, Recall, MRR
[104]	MEDLINE bibliographic database	SemRep, LBD process	
[88]	APS1(American Physical Society)	LDA, CTPF, Network Embedding	Precision, Recall, F1, NDCG
[105]	–	CARS (Context-Aware Recommender System)	
[106]	ACM dataset	Keyword aggregation, TPC	Accuracy

Table 7 Overview of collaborator recommendation system using collaborative filtering

Citation	Data	Method	Evaluation
[107]	ACM DL, Scopus	HCI trust	
[108]	DBLP	SVM link prediction	Precision, Recall, F1
[109]	DBLP	RWR(derive 3 matrices)	Precision, Recall Coverage Rate
[110]	DBLP Frequent Itemset Mining Implementations	Weighted Association Rule based on Sociability	Lift Measure
[111]	Self-conducted database	Social Network	Precision, Recall, F1, Hit rate
[112]	–	RWR, PageRank	

tures and hybrid topic extraction methods with other factors to obtain higher accuracy. For example, Sun et al. [92] designed a career age-aware academic collaborator recommendation model consisting of authorship extraction from digital libraries, topic extraction based on published abstractions, and career age-aware random walk for measuring scholar similarity.

3.2.2 Collaborative filtering

Six publications presented a methodology based merely on collaborative filtering. Traditional CF-based recommendations aim to find the nearest neighbor in a social context similar to that of the targeted user. It selects the nearest neighbors based on the users' rating similarities. When the users rate a set of items in a manner similar to that of a target user, the recommendation systems would define these nearest neighbors as groups with similar interests and recommend items that are favored by these groups but not discovered by the target user. To apply this method to collaborator recommendation, the system would recommend persons who have worked with a target author's colleagues but not with the target author himself. Analogously, the system considers each author as an item to be rated and the scholarly activities such as writing a paper together as a rating activity, following the methodology of traditional CF-based recommendations. Researchers' publication activities are transformed into rating actions, and the frequency of co-authored papers is considered a rating value. Using this criterion, a graph based on a scholarly social network was built. A summary of the collaborator recommendation paper using CF approaches is presented in Table 7.

Based on this co-authorship network transformed from researchers' publication activities, several methods for link prediction and edge weighting have been utilized. Benchettara et al. [108] solved the problem of link prediction in co-authoring networks by using a topological dyadic supervised machine learning approach. Koh and Dobbie [110] proposed an academic collaborator recommendation approach that uses a co-authorship network with a weighted association rule approach using a weighting mechanism called sociability. Recommendation approaches based on this co-authorship network transformed from publication activities, where all nodes have the same functions, are called homogeneous network-based recommendation approaches.

The random walk model, which can define and measure the confidence of a recommendation, is popular in co-authorship network-based collaborator recommendations. Tong et al. [113] published Random Walk with Restart (RWR), a famous random walk model, which provides a good way to measure how closely related two nodes are in a graph. Applications and improvements based on RWR model are widely used for link prediction in co-authorship

networks. Li et al. [109] proposed a collaboration recommendation approach based on a random walk model using three academic metrics as the basics through co-authorship relationship in a scholarly social network. Yang et al. [112] combined the RWR model with the PageRank method to propose a nearest-neighbor-based random walk algorithm for recommending collaborators.

Compared with content-based recommendation approaches, which involve only the published profiles of researchers without considering scholarly social networks, homogeneous network-based approaches apply CF methods based on social network technology to recommend collaborators. Lee et al. [111] compared ASN-based collaborator recommendations with metadata-based and hybrid recommendation methodologies, and suggested it as the best method. However, homogeneous network-based collaboration recommendations do not consider the contextual features of researchers. As a combination of these two methods, a hybrid collaboration recommendation system based on a heterogeneous network is popular in current collaboration recommendation approaches and applications.

3.2.3 Hybrid

Approaches to previously introduced recommendation classes may be combined with hybrid approaches. 37 of the reviewed papers applied approaches with hybrid characteristics. As an improvement, heterogeneous network-based recommendations overcome these limitations. Table 8 summarizes all collaborator recommendation papers that we collected using hybrid approaches.

Heterogeneous networks are networks in which two or more node classes are categorized by their functions. Based on the co-authorship network used in most homogeneous network-based approaches, heterogeneous network-based approaches incorporate more information into the network, such as the profiles of researchers, the results of topic modeling or clustering, and the citation relationship between researchers and their published papers. Xia et al. [52] presented MVCWalker, an innovative method based on RWR for recommending collaborators to academic researchers. Based on academic social networks, other factors such as co-author order, latest collaboration time, and times of collaboration were used to define link importance. Kong et al. [114] proposed a collaboration recommendation model that combines the features extracted from researchers' publications using a topic clustering model and a scholar collaboration network using the RWR model to improve the recommendation quality. Kong et al. [115] proposed a collaboration recommendation model that considers scholars' dynamic research interests and collaborators' academic levels. By using the LDA model for topic clustering and fitting the dynamic transformation of interest, they combined the similarity and weighting factors in a co-authorship network to recommend collaborators with high prevalence. Xu et al. [116] designed a recommendation system to provide serendipitous scholarly collaborators that could learn the serendipity-biased vector representation of each node in the co-authorship network.

4 Venue recommendation

In this section, we describe recommendation systems that can help researchers identify scientific research publishing opportunities. Recently, there has been an exponential increase in the number of journals and conferences researchers can select to submit their research. Recommendation systems can alleviate some of the cognitive burden that arises when choosing

Table 8 Overview of collaborator recommendation system using hybrid methods

Citation	Data	Method	Evaluation
[117]	ScholarMate	HSN (heterogeneous social network) RAF (research analytical framework)	Precision, Recall, F
[116]	DBLP	RWR, serendipity-biased DeepWalk	Precision, Recall, Unexpectedness
[118]	16th ACM SIGKDD, DBLP	TF-IDF, k-means clustering algorithm	Questionnaire, Precision
[119]	Self-conducted database	Weight measure, Affin (Affiliation index)	Precision, Recall
[120]	DBLP	Weight measure, VSM (Vector Space Model)	Case study
[121]		Scent Score, Linkage generation	
[122]	Self-conducted database	TF-IDF, SVM (collaboration quality)	Precision, Recall, Average Precision
[123]	DBLP, hep-th database	heterogeneous bibliographic network RWR, edge weighting	RBP (Rank-Based Precision), RR (Reciprocal Rank)
[124]	CiteSeerX	Vertex Similarity Analysis Lexical Similarity Algorithm	User study, Kendall tau rank correlation coefficient
[125]	Self-conducted database	Community Detection Technique PSO (particle swarm optimization)	Louvain approach
[126]	SCHOLATE	Louvain method, friendship scores	
[127]	NIH dataset, NSF dataset	Prototypical Collaborations, taxonomies	Accuracy
[128]	Self-conducted database	Gradient Value Iteration algorithm competition function	Precision, MRR, NDCG
[129]	DBLP	CORALS (Collaboration Recommendation on Academic Social Networks) Affin (weight represents researchers relation)]	Recall, Diversity, Coverage
[130]	Self-conducted database	MTML (Multitheoretical Multilevel analytic)	Precision, ARC
[114]	DBLP	Word2vec, RWR	Precision, Recall
[131]	Self-conducted database	SimRank, Random Walk, PANTHER	Precision, Accuracy
[132]	ScholarMate	mutual reinforcement principle semantic keyword matching query-sensitive quality analysis edge weighting	AR (Average Rating score), NDCG
[133]	Self-conducted database	SVM, tensor-based group recommendation	
[134]	Self-conducted database	community detection algorithm PSO (Particle Swarm Optimization)	Modularity, Redundancy
[135]	Self-conducted database	Jaccard and Cosine Vertex Similarity	Accuracy

Table 8 continued

Citation	Data	Method	Evaluation
[52]	DBLP	MVCWalker (Innovative RWR)	Precision, Recall, Coverage Rate
[136]	USPTO	SVM-Rank	Precision, Recall, MAP (Mean Average Precision)
[115]	Self-conducted database	LDA (topic clustering), cosine similarity, RWR	Precision, Recall, F1
[137]	DBLP, ResearchGate	LDA, Euclidean distance cross-disciplinary recommendation	Precision
[138]	DBLP	TF-IDF, OTSU filter	Case study, interview
[139]	ISI database	TF-IDF, Adamic-Adar Similarity, SVM-Rank	Precision, Recall, F measure, MRR
[140]	Self-conducted database	Cross-Collaboration-Domain Data Preprocessing LDA, Hadoop Map-Reduce framework	Precision, Recall, F1
[141]	PubMed	LDA model	Manual evaluation
[142]	DBLP	RWR, link weight information	Precision, Recall, F1
[143]	Self-conducted database	Lucene data indexing system, T-Res	Precision, Recall (Manual evaluation)
[144]	DBLP	TBRSS (trend-based relation strength similarity)	Accuracy
[145]	Self-conducted database	ERGM (exponential random graph model), weighting	Precision, Recall, F1
[69]	Google Scholar	Chi-square test (estimate relevancy), scholar ranking	Chi-square test

the right conference or journal for publishing a work. In the following sections, we describe academic venue recommendation systems for conferences and journals.

4.1 Conference recommendation

The dramatic rise in the number of conferences/journals has made it nearly impossible for researchers to keep track of academic conferences. While there is an argument to be made that researchers are familiar with the top conferences in their field, publishing to those conferences is also becoming increasingly difficult due to the increasing number of submissions. A conference recommendation system will be helpful in reducing the time and complexity requirement to find a conference that meets the needs of a given researcher. Thus, conference recommendation is a well-studied problem in the domain of data analysis, with many studies being conducted using a variety of methods such as citation analysis, social networks, and contextual information.

4.1.1 Data

All reviewed publications used databases to test their methodology. Two publications chose to construct a custom dataset based on the manual collection of information and one publication used a rare paid dataset. The remaining 20 studies used published open-source databases to

Table 9 Sources of data used for Conference Recommendation Systems

	ACM	CORE	DBLP	WikiCFP	Custom	Other
Absolute	4	3	12	5	2	10
Relative	11.1%	8.3%	33.3%	13.9%	5.6%	27.8%

create the datasets used in their testing and evaluation environments. Table 9 provides a summary of the frequencies with which published open-source databases were used.

DBLP was the most used database with 12 occurrences, followed by ACM Digital Library and WikiCFP, both with 5 occurrences. The unique databases utilized in conference recommendation systems are Microsoft Academic Search, CORE Conference Portal, Epinion, IEEE Digital Library, and Scigraph.

Microsoft Academic Search hosts over 27 million publications from over 16 million authors and is primarily used to extract metadata on authors, their publications, and their co-authors. The CORE Conference portal provides rankings for conferences primarily in Computer Science and related disciplines. The CORE Conference provides metadata on conference publishers and rankings. The Epinion is a general review website founded in 1999 and utilized to create networks of ‘trusted’ users. The IEEE Digital Library is a database used to access journal articles, conference proceedings, and other publications in computer science, electrical engineering, and electronics. A scigraph is a knowledge graph aggregating metadata from publications in Springer Nature and other sources. WikiCFP is a website that collates and publishes calls for papers.

4.1.2 Methods

There are three main subtypes of conference recommendation systems: content-based, collaborative, and hybrid systems. The following section provides an overview of the most popular methods used by each sub types.

Content-based filtering (CBF)

Only 1 of the 23 publications in conference recommendations utilized pure CBF. Using data from Microsoft Academic Search, Medvet et al. [146] created three disparate CBF systems seeking to reduce the input data required for accurate recommendations: (a) utilizing Cavnar-Trenkle text classification, (b) utilizing two-step latent Dirichlet allocation (LDA), and (c) utilizing LDA alongside topic clustering.

Cavnar-Trenkle classification is an n -gram-based text classification method. Given a set of conferences $C = \{c_1, c_2, c_3, \dots\}$, it is necessary to define for each conference $c \in C$ a set of papers $P = \{p_1, p_2, p_3, \dots\}$ that were published in conference c . It creates an n -gram profile for each conference $c \in C$, using n -grams generated from each paper in the conference $p \in P$. Finally, it computes the distance between the n -gram profiles of each conference $c \in C$ and a publication of interest p_i and recommends an n number of conferences that optimize the minimum distance between c and p_i .

Collaborative filtering

Among 18 publications employed collaborative filtering strategies out of the 23 collected publications, the most popular filtering approach was based on around generating and analyzing a variety of networks on different types of metadata including citations, co-authorship, references, social proximity, etc.

Asabere and Acakpovi [147, 148] generated a user-based social context aware filter with breadth-first search (BFS) and depth-first search (DFS) on a knowledge graph created by computing the Social Ties between users, and added geographical, computing, social, and time contexts. Social Ties were generated by computing the network centrality based on the number of links between users and presenters at a given conference.

Other types of network-based collaborative filters include a co-author-based network that assigns weights with regard to venues where one's collaborators have published previously [149, 150], a broader metadata-based network that utilizes one or more distinct characteristics to assign weights to conferences (i.e., citations, co-authors, co-activity, co-interests, colleagues, interests, location, references, etc.) [146, 151–154], and RWR-based methods [155, 156].

Kucuktunc et al. [155] iterated the traditional RWR model by adding a directionality parameter (κ), which is used to chronologically calibrate the recommendations as either recent or traditional. The list of publications that used CF for conference recommendations is presented in Table 10.

Hybrid

A total of 6 publications used hybrid filtering strategies out of the total 23 publications. The most common hybrid strategy is to amalgamate standard topic-based content filtering with network-based collaborative filters. Table 11 summarizes publications that used hybrid filtering methods for conference recommendations.

4.2 Journal recommendation

As of April 14, 2020, the Master Journal List of the Web of Science Group contains 24,748 peer-reviewed journals for publishing articles from different publishing houses. The authors may face difficulties in finding suitable journals for their manuscripts. In many cases, a manuscript submitted to a journal is rejected because it is not within the scope of that journal. Finding suitable journals for a manuscript is the most important step in publishing articles. A journal recommendation system may reduce the burden of authors by selecting appropriate journals to publish as well as reducing the burden of editors from rejecting manuscripts that do not align with the scopes of the journals. Many publishing companies have their own journal finders that can help authors find suitable journals for their manuscripts.

In this section, we review all available journal recommendation systems by analyzing the methods used and their journal coverage. There are a total of ten journal recommendation systems, but we found only four papers describing details corresponding to their recommendation procedures. A detailed list of journal recommenders with their methods and datasets is provided in Table 12. Most journal recommenders were developed for different publishing houses. Most journal recommenders contain journals from multiple domains except eTBLAST, Jane, and SJFinder, where the journals are from the biomedical and life science domains.

TF-IDF, kNN, and BM25 were used to find similar journals using the keywords provided keywords. Kang et al. [172] used a classification model (using kNN and SVM) to identify the suitable journals. Errami et al. [169] used the similarity between provided keywords and journal keywords.

Rollins et al. [39] evaluated a journal recommender by using feedback from real users. Kang et al. [172] evaluated a system based on previously published articles. If the top three or top ten recommended journals contained the journal in which the input paper was published, then this would be counted as a correct recommendation; otherwise, it would be counted as

Table 10 Overview of conference recommendation systems using collaborative filtering

CitationData	Method	Evaluation
[147] Custom dataset from students at Accra Technical University (Ghana)	Breadth First Search (BFS), Social Context	Precision, Recall
[148] Custom dataset from students at Accra Technical University (Ghana)	Breadth First Search (BFS), Depth First Search (DFS), Social Context	Coverage, Recall, Precision, F1
[149] dblp: computer science bibliography	Co-Authorship Network Community Similarity Index, kNN	Rand Index, F1
[150] ACM Digital Library, Microsoft Academic Search	Author Network Analysis	Accuracy, Precision
[152] ACM Digital Library	Social Network, kNN	Serendipity, Accuracy, Precision, Recall
[152] dblp: computer science bibliography, wikicfp, CORE Conference Portal	Network-based ranking	Precision, Recall, F1
[152] CiteULike	Stochastic Gradient Descent (SGD), Singular Value Decomposition (SVD++)	Precision, Recall, DCG, nDCG, MAE, RMSE
[157] dblp: computer science bibliography, CiteSeerX	Direction Aware Random Walk with Restart (RWR)	N/A
[153] IEEE digital library	Citation Network	Precision, Recall, F1
[158] dblp: computer science bibliography, CORE Conference Portal	Current Personal Academic Venue Rating (CPAVR)	Precision, Recall, nDCG
[159] dblp: computer science bibliography	Network-based clustering, Modularity	Precision, Recall
[154] ACM Digital Library, Microsoft Academic Search	Social Network	Precision
[160] dblp: computer science bibliography	Random Walk with Restart (RWR)	Precision, Recall, F1, Average Venue Quality
[156] dblp: computer science bibliography	Random Walk with Restart (RWR)	Precision, Recall, F1
[161] International Conference on Web-Based Learning 2012	Social Context	Precision, Recall, F1
[162] dblp: computer science bibliography, Epinion	Social Network	Precision, Recall, MAE

a false recommendation. Similarly, eTBLAST [169] and Jane [170] were evaluated using previously published articles.

Deep learning-based recommenders perform better than traditional matching-based NLP or machine learning algorithms. However, none of the existing systems available for journal recommendations uses deep learning algorithms. One of the future goals may be the implementation of different deep learning algorithms. In addition to these publication houses, developing journal recommenders for different publication repositories (DBLP, arxiv, etc.) may be another future task.

Table 11 Overview of conference recommendation systems using hybrid filtering

Citation	Data	Method	Evaluation
[163]	dblp: computer science bibliography	kNN, Paper-Paper Peer Network (PPPN) Venue-Venue Peer Network (VVPN)	Accuracy, MRR, F1, Precision, nDCG, Average venue quality, Diversity, Stability
[164]	dblp: computer science bibliography, wikicfp, CORE Conference Portal	Citation Network	Precision
[165]	ACM Digital Library, CiteSeer Digital Library	Memory-based CF, Stylometric Feature LDA	Accuracy, MRR
[166]	Scigraph, wikicfp	TF-IDF, Singular Value Decomposition (SVD), Nonnegative Matrix Factorization (NMF), Doc2Vec Convolutional Neural Network (CNN)	Recall, MAP
[167]	dblp: computer science bibliography, wikicfp	Co-author Network, Word2Vec, Doc2Vec	Precision, Recall, F1
[168]	dblp: computer science bibliography, wikicfp	Interaction Strength, Word2Vec, Doc2Vec	Precision, Recall, F2

Table 12 Detailed overview of journal recommendation systems

Citation	Data sources	Method
[169]	MEDLINE Journals	Similarity score
[170]	MEDLINE Journals	TF-IDF K-nearest Neighbors (kNN)
[171]	MEDLINE Journals	NA
[172]	Elsevier Journals	Okapi BM25
[173]	IEEE Publications	NA
[174]	Springer Journals	NA
[39]	Journals from Clarivate Analytics Web of Science Journal Citation Reports	kNN, SVM
[175]	Wiley Journals	NA
[176]	Edanz Researcher Services	NA
[177]	Research Square	NA

5 Reviewer recommendation

In this section, we describe paper, journal, and grant reviewer recommendation systems that are available in literature. With the rapid increase in publishable research materials, pressure to find reviewers is overwhelming for conference organizers/journal editors. Similarly, it overwhelms program directors in finding appropriate reviewers for grants.

In the case of conferences, authors normally choose some research fields during the submission. The organizing committee of a conference typically has a set of researchers as reviewers who have been assigned from the same set of fields. Based on the matching of

Table 13 Sources of datasets used for reviewer recommendation

	NIPS	DBLP	IEEE	Others
Absolute	4	3	4	12
Relative	17%	14 %	17%	52%

the fields, the reviewers were assigned papers. However, the research fields are broad and may not exactly match those of the reviewer. In the case of journals, authors need to suggest that reviewers or editors need to find reviewers for manuscript reviewing. Whereas, for reviewing grant proposals, program directors are responsible for finding suitable reviewers for reviewing proposals.

The problem of finding reviewers can be solved by a reviewer recommendation system, which the system can recommend reviewers based on the similarity of contents or past experiences. The reviewer recommendation problem is known as the reviewer assignment problem. We searched for publications related to both reviewer recommendations and assignments.

5.1 Data

A total of 67 reviewed publications were retrieved using Google searches, and 36 publications were included in the final analysis after title, abstract, and full-text screening. Among these 36 publications, 23 conducted experiments to supplement the theoretical contents, and the sources of the datasets used are listed in Table 13.

5.2 Methods

Broadly, there are three major categories in terms of techniques used, one is based on information retrieval (IR), another one on optimization where the recommendation is viewed as an enhanced version of the generalized assignment problem (GAP), and the third includes techniques that fall between the first two categories.

5.2.1 Informational retrieval (IR)-based

IR-based studies generally focus on calculating matching degrees between reviewers and submissions.

Hettich and Pazzani [178] discussed a prototype application in the U.S. National Science Foundation (NSF) to assist program directors in identifying reviewers for proposals, named RevaId, which uses TF-IDF vectors for calculating proposal topics and reviewer expertise, and defined a measure called the Sum of Residual Term Weight (SRTW) for the assignment of reviewers. Yang et al. [179] constructed a knowledge base of expert domains extracted from the web and used a probability model for domain classification to compute the relatedness between experts and proposals for ranking expertise. Ferilli et al. [180] used Latent Semantic Indexing (LSI) to extract the paper topic and expertise of reviewers from publications available online, followed by Global Review Assignment Processing Engine (GRAPE), a rule-based expert system for the actual assignment of reviewers.

Serdyukov et al. [181] formulated a search for an expert to absorb a random walk in a document-candidate graph. A recommendation was made on reviewer candidate nodes with high probabilities after an infinite number of transitions in the graph, with the assumption

that expertise is proportional to probability. Yunhong et al. [182] used LDA for proposal and expertise topic extraction, and defined a weighted sum of varied index scores for ranking reviewers for each proposal. Peng et al. [183] built a time-aware reviewer's personal profile using LDA to represent the expertise of reviewers, then a weighted average of matching degree by topic vectors and TF-IDF of the reviewer and submitted papers were used for recommendation. Medakene et al. [184] used pedagogical expertise in addition to the research expertise of the reviewers with LDA in building reviewers' profiles and used a weighted sum of the topic similarity and the reference similarity for assigning reviewers to papers. Rosen-Zvi et al. [185] proposed an Author-Topic Model (ATM) that extends the LDA to include authorship information. Later, Jin et al. [186] proposed an Author-Subject-Topic (AST) model, with the addition of a 'subject' layer that supervises the generation of hierarchical topics and sharing of subjects among authors for reviewer recommendations. Alkazemi [187] developed PRATO (Proposals Reviewers Automated Taxonomy-based Organization) that first sorted proposals and reviewers into categorized tracks as defined by a tree of hierarchical research domains, and then assigned the reviewers based on the matching of tracks using Jaccard similarity scores. Cagliero et al. [188] proposed an association rule-based methodology (Weighted Association Rules, WAR) to recommend additional external reviewers.

Ishag et al. [189] modeled citation data of published papers as a heterogeneous academic network, integrating authors' h-index and papers' citation counts, proposed a quantification to account for author diversity, and formulated two types of target patterns, namely, researcher-general topic patterns (RSP) and researcher-specific topic patterns (RSP) for searching reviewers.

Recently deep learning techniques have been incorporated into feature representations. Zhao et al. [190] used word embeddings to represent the contents of both the papers and reviewers. Then, the Word Mover's distance (WMD) method was used to measure the minimum distances between paper and reviewer vectors. Finally, the Constructive Covering Algorithm (CCA) was used to classify reviewer labels for recommending reviewers. Anjum et al. [191] proposed a common topic model (PaRe) that jointly models topics to a submission and a reviewer profile based on word embedding. Zhang et al. [192] proposed a two-level bidirectional gated recurrent unit with an attention mechanism (Hiepar-MLC) to represent the semantic information of reviewers and papers and used a simple multilabel-based reviewer assignment strategy (MLBRA) to match the most similar multilabeled reviewer to a particular multilabeled paper.

Co-authorship and reviewer preferences were incorporated into collaborative filtering application. Li and Watanabe [193] designed a scale-free network combining preferences and a topic-based approach that considers both reviewer preferences and the relevance of reviewers and submitted papers to measure the final matching degrees between reviewers and submitted papers. Xu and Du [194] designed a three-layer network that combines a social network, semantic concept analysis and citation analysis, and proposed a particle swarm algorithm to recommend reviewers for submissions. Maleszka et al. [195] used a modular approach to determine a grouping of reviewers that consisted of a keyword-based module, a social graph module and a linguistic module. A summary of all IR-based reviewer recommendations can be found in Table 14.

5.2.2 Optimization-based

Optimization-based reviewer recommendations focus more on theory, modeling an algorithm of assignments under multiple constraints such as reviewer workload, authority, diversity, and conflict of interest (COI).

Table 14 Overview of reviewer recommendation systems, IR-based

Citation	Data	Method	Evaluation
[178]	U.S.NSF	TF-IDF	SRTW
[179]	National Science Council, Taiwan	Knowledge database, probability model	P@50, F-measure
[180]	IEA/AIE 2005	LSI, GRAPE	Accuracy
[181]	W3C2006, CSIRO2007	Absorbing random walk	MAP, MRR, P@5
[182]	NA	LDA, index score	NA
[183]	FCS journal	LDA, TF-IDF	Manual
[184]	ICA2IT19	LDA	MSE, RMSE, MAE, MAPE
[186]	WANFANG	AST	Perplexity, KL divergence, topic convergence
[187]	Umm Al-Qura University funding	Taxonomy	Manual
[188]	DBLP	WAR	Confidence
[189]	Clarivate Analytics' Web-of Science	Pattern mining	Coverage
[190]	Twitter, OHSUMED, 3-class MOV, 4-class MOV, Synthetic data	WMD-CC	Accuracy, P, R
[191]	NIPS	Word-embeddings, topic model	Relevance, P@5, P@10, manual
[192]	ACM Digital	Hiepar-MLC	R@k, NDCG@k, accuracy
[193]	CiteSeerX, DBLP	Scale-free network	Manual
[194]	NA	network analysis, particle swarm	NA
[195]	DBLP	hybrid	NA

Sun et al. [196] proposed a hybrid of knowledge and decision models to solve the proposal-reviewer assignment problem under constraints. Kolasa and Krol [197] compared artificial intelligence methods for reviewer-paper assignment problems, namely, genetic algorithms (GA), ant colony optimization (ACO), tabu search (TS), hybrid ACO-GA and GA-TS, in terms of time efficiency and accuracy. Chen et al. [198] employed a two-stage genetic algorithm to solve the project-reviewer assignment problem. In the first stage, reviewer were assigned by taking into consideration their respective preferences, and then, in the second stage, review venues were arranged in a way that allows the minimum times of change for reviewers.

Das and Gocken [199] used fuzzy linear programming to solve the reviewer assignment problem by maximizing the matching degree between expert sets and grouped proposals, under crisp constraints. Tayal et al. [200] used type-2 fuzzy sets to represent reviewers' expertise in different domains, and proposed using the fuzzy equality operator to calculate equality between the set representing the expertise levels of a reviewer and the set representing the keywords of a submitted proposal, and optimized the assignment under various constraints.

Wang et al. [201] formulated the problem into a multiobjective mixed integer programming model that considers Direct Matching Score (DMS) between manuscripts and reviewer, Manuscript Diversity (MD), and Reviewer Diversity (RD), and proposed a two-phased stochastic-biased greedy algorithm (TPGA) to solve the problem. Long et al. [202] studied the paper-reviewer assignment problem from the perspective of goodness and fairness, where they proposed maximizing topic coverage and avoiding the conflict of interest (COI) for the optimization objectives. They also designed an approximation method that provides 1/3 approximation.

Kou et al. [203] modeled reviewers' published papers as a set of topics and performed weighted-coverage group-based assignments of reviewers to papers. They also proposed a greedy algorithm that achieves a 1/2 approximation ratio compared with the exact solution. Kou et al. [204] developed a system that automatically extracts the profiles of reviewers and submissions in the form of topic vectors using the author-topic model (ATM) and assigns reviewers to papers based on the weighted coverage of paper topics.

Stelmakh et al. [205] designed an algorithm, PeerReview4All, which is based on an incremental max-flow procedure to maximize the review quality of the most disadvantaged papers (fairness objective) and to ensure the correct recovery of the papers that should be accepted (accuracy objective). Yesilcimen and Yildirim [206] proposed an alternative mixed integer programming formulation for the reviewer assignment problem whose size grows polynomially as a function of the input size. A summary of all the optimization-based reviewer recommendation papers is presented in Table 15.

5.2.3 Hybrid

Finally, we see hybrid of both methods in other studies. Conry et al. [207] modeled reviewer-paper preferences using CF of ratings, latent factors, paper-to-paper content similarity, and reviewer-to-reviewer content similarity and optimized the paper assignment under global conference constraints; therefore, the assignment was transformed into a linear programming problem. Tang et al. [208] formulated the problem of expertise matching to a convex cost flow problem which turned the recommendation into an optimization problem under constraints, and also used online matching algorithms to support user feedback to the system.

As one of the most popular systems for conference reviewer assignment, Charlin and Zemel [209] addressed the assignment by first using a language model and LDA for learning

Table 15 Overview of reviewer recommendation systems, optimization-based

Citation	Method
[196]	Knowledge and decision model
[197]	GA, ACO, TS, ACO-GA and GA-TS
[198]	Two-stage genetic algorithm
[199]	Fuzzy linear programming
[200]	Type-2 fuzzy sets
[201]	Mixed integer programming, TPGA
[202]	MaxTCPRA, greedy algorithm
[203]	WGRAP, greedy algorithm
[204]	WGRAP, greedy algorithm
[205]	Incremental max-flow procedure
[206]	Mixed integer programming

reviewer expertise and submission topics, followed by a linear regression for initial predictions of reviewers' preferences, combined with reviewers' elicitation scores (reviewers' disinterest or interests) in specific papers for the final recommendation, and optimized the objective functions under constraints. Liu et al. [210] constructed a graph network for reviewers and query papers using LDA to establish edge weights, and used the Random Walk with Restart (RWR) model on a graph network with sparsity constraints to recommend reviewers with the highest probabilities incorporating aspects of expertise, authority and diversity. Liu et al. [211] combined the heuristic knowledge of expert assignment and techniques of operations research, in which different aspects are involved, such as reviewer expertise, title and project experience. A multiobjective optimization problem was formulated to maximize the total expertise level of the recommended experts and avoid conflicts between reviewers and authors. Ogunleye et al. [212] used a mixture of TF-IDF, LSI, LDA and word2vec to represent the semantic similarity between submissions and reviewers' publications and then used integer linear programming to match submissions with the most appropriate reviewers. Jin et al. [213] extracted topic distributions of reviewers' publications and submissions using the Author-Topic Model (ATM) and Expectation Maximization (EM), then formulated the problem of reviewer assignment into an integer linear programming problem that takes into consideration the topic relevance, interest trend of a reviewer candidate, and authority of candidates. A summary of the reviewer recommendation papers is presented in Table 16.

6 Other scholarly recommendation

6.1 Dataset recommendation

In the Big Data era, extensive data have been generated for scientific discoveries. However, storing, accessing, analyzing, and sharing a vast amount of data is becoming a major challenge and bottleneck for scientific research. Furthermore, making a large amount of public scientific data findable, accessible, interoperable, and reusable (FAIR) is challenging. Many repositories and knowledge bases have been established to facilitate data-sharing. Most of these repositories are domain-specific, and none of them recommend datasets to researchers or users. Furthermore, over the past two decades, there has been an exponential increase in the number of datasets added to these dataset repositories. Researchers must visit each

Table 16 Detailed overview of reviewer recommendation systems, other

Citation	Data	Method	Evaluation
[207]	ICDM2007	Collaborative filtering, latent factor, linear programming	Manual
[208]	KDD08, KDD09, ICDM09, graduate CS courses@CMU, UIUC, Stanford and MIT	Topic model, online matching, convex cost flow	Matching score, expertise variance, coverage, confidence, P
[209]	NIPS10, ICML12	Language model, LDA, collaborative filtering, bipartite matching	NDCG@5, NDCG@10
[210]	UIUC, NIPS	LDA, graph network, RWR	P@1-10, matching score, authority, KL divergence
[211]	Simulation	Heuristic, decision model	Balance, rationality
[212]	NIPS	TF-IDF, LSI, LDA, word2vec, integer linear programming	NA
[213]	WANFANG, ArnetMiner	ATM, EM, integer linear programming	Distance@k, Interest@k, Authority@k

repository to find suitable datasets for their research. In this case, a dataset recommender would be helpful to researchers. This can save time and the visibility of the dataset.

A dataset recommender is not commonly used. However, dataset retrieval is a popular information retrieval task. Many dataset retrieval systems exist for general datasets as well as biomedical datasets. Google's Dataset Search² is a popular search engine for datasets from different domains. DataMed³ is another dataset search engine specific to biomedical domain datasets that combines biomedical repositories and enhances query searching using advanced natural language processing (NLP) techniques [214, 215]. DataMed indexes and provides the functionality to search diverse categories of biomedical datasets [215]. The research focus of DataMed is to retrieve datasets using a focused query. Search engines such as DataMed or Google Dataset Search are helpful when the user knows the type of dataset to search for, but determining the user intent of web searches is a difficult problem because of the sparse data available concerning the searcher [216].

A few experiments have been performed on data linking where similar datasets can be clustered together using different semantic features. Data linking or identifying/clustering similar datasets has received relatively less attention in research on recommendation systems. Specifically, only a few papers [217–219] have been published on this topic. Ellefi et al. [218] defined dataset recommendation as the problem of computing a rank score for each set of target datasets (D_T) such that the rank score indicates the relatedness of D_T to a given source dataset (D_S). The rank scores provide information on the likelihood of a D_T containing linking candidates for D_S . Similarly, Srivastava [219] proposed a dataset recommendation system by first creating similarity-based dataset networks, and then recommending connected datasets to users for each searched dataset. This recommendation approach is difficult to implement because of the cold start problem. Here, the cold start problem refers to the user's initial dataset selection, where the user has no idea what dataset to select/search for. If the user lands on an incorrect dataset, the system will always recommend the wrong dataset to the user.

² <https://datasetsearch.research.google.com>.

³ <https://datamed.org>.

Patra et al. [220, 221] and Zhu et al. [222] proposed a dataset recommendation system for the Gene Expression Omnibus (GEO) based on the publications of researchers. This system recommends GEO datasets using classification and similarity-based approaches. Initially, they identified the research areas from the publications of researchers using the Dirichlet Process Mixture Model (DPMM) and recommended datasets for each cluster. The classification-based approach uses several machine and deep learning algorithms, whereas the similarity-based approach uses cosine similarity between publications and datasets. This is the first study on dataset recommendations.

6.2 Grants/funding recommendation

Obtaining grants or funding for research is essential in academic settings. Grants help researchers in many ways during their careers. Finding appropriate funding opportunities is an important step in this process, and there are multiple grant opportunities available that a researcher may not be aware of. No universal repositories available for funding announcements worldwide. However, few repositories are available for funding announcements in the United States of America, such as, grants.gov, NIH, and SPIN. These websites host many funding opportunities in various areas. Furthermore, multiple new opportunities are available daily. Thus, it is difficult to find suitable opportunities for researchers. A recommendation system for funding announcements will help researchers find appropriate research funding opportunities. Recently, Zhu et al. [223] developed a grant recommendation system for NIH grants based on researchers' publications. They developed the recommendation as a classification using Bidirectional Encoder Representations from Transformers (BERT) to capture intrinsic, nonlinear relationships between researchers' publications and grant announcements. Internal and external evaluations were performed to assess the usefulness of the system. Two publications are available on developing a search engine to find Japanese research announcements [224, 225]. The titles of these papers suggest recommendation systems; however, the full text reveals that these publications describe the search for funding announcements in Japan. These publications describe a keyword-based search engine using TF-IDF and association rules.

7 Conclusion and future directions

Numerous recommendation systems have been developed since the beginning of the twenty-first century. In this comprehensive survey, we discussed all common types of scholarly recommendation systems outlining the data resources, applied methodologies and evaluation metrics.

Recommendation systems for the literature are still the most focused areas for scholarly recommendations. With the increasing need to collaborate with other researchers and publish research results, recommenders for collaborators and reviewers are becoming popular. Compared with these popular research targets, published recommendation systems for conferences/journals, datasets and grants are relatively less common.

To develop recommendation systems and evaluate their results, researchers commonly construct datasets using information extracted from multiple resources. Published open-source databases, such as DBLP, ACM and IEEE Digital Libraries, are the most commonly used sources for multiple types of recommendation systems. Some web services containing scholarly related information about its users, or social tags added by researchers, such as, ScholarMate and CiteULike, were also used to develop recommendation systems.

Content-based filtering (CBF) is the most commonly used approach for recommendation systems. Owing to the requirement of processing context information, measuring keywords and searching topics of academic resources, most recommendation systems were built based on CBF. It is difficult to consider the popularity and rating of objects in traditional CBF. To overcome these limitations, CF has been used to solve the problem, especially when recommending items based on researchers' interests and profiles. With the rapid development of recommendation systems and the need to overcome the high calculation costs, hybrid methods combining CBF and CF have been used by several recommenders to achieve better performance.

Based on the information gathered for the survey, we provide the following suggestions for better recommendation developments:

1. To Improve System Performance And Avoid The Limitations Of Existing Methodologies, A Combination Of Different Methods, Or Incorporating The Characteristics Of One Method Into Another May Be Helpful.
2. Evaluating The Efficiency Of The Recommendation System, Including Both Decision Support Metrics Such As Precision And Recall, And Rank-Aware Evaluation Metrics, Including Mrr And Ndcg, Will Make The Offline Evaluation More Applicable.
3. For Future Directions Of Scholarly Recommendation Research, We Suggest That Researchers Apply Recommendation Methodologies In Areas Less Studied, Such As Datasets And Grant Recommendations. We Believe That Researchers Would Benefit Significantly From These Areas From A Practical Perspective.

Based on extensive research, our literature review provides a comprehensive summary of scholarly recommendation systems from various perspectives. For researchers interested in developing future recommendation systems, this would be an efficient overview and guide.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A: Supplementary material

Table 17 List of keywords used to search publications for different recommendation systems

Types	Keywords
Literature recommendation	Scholarly recommendation, scholarly paper recommendation research paper recommendation, recommending academic paper recommender system, research paper recommender publications recommendation, scientific publications recommendation system scientific literature recommender
Collaborator recommendation	Collaborator recommendation, collaboration recommendation collaboration discovery, scholarly collaborator recommendation expert recommendation system, scientific collaborator recommender academic social networks recommendations
Reviewer recommendation	Reviewer recommendation, reviewer assignment reviewer matching, expert matching expert finding, expert recommendation
Conference recommendation	Conference venue recommendation journal venue recommendation
Others	Dataset recommendation grant recommendation

Table 18 Table of acronyms for scholarly recommendation researches

Acronyms	Definition
ACM	Association for Computing Machinery Digital Library
DBLP	The DBLP Computer Science Bibliography
CBF	Content-based filtering
TF-IDF	Term frequency and inverse document frequency
SVM	Support Vector Machine
P@n	Precision at position n
MAP	Mean Average Precision
MRR	Mean Reciprocal Rank
NDCG	Normalized Discounted Cumulative Gain
CF	Collaborative Filtering
kNN	k Nearest Neighbors
LDA	Latent Dirichlet Allocation
LSI	Latent Semantic Index
SVD	Singular Value Decomposition
RWR	Random Walk with Restart

References

1. Bollacker KD, Lawrence S, Giles CL (1998) Citeseer: an autonomous web agent for automatic retrieval and identification of interesting publications. Springer, Berlin, pp 116–123
2. Das D, Sahoo L, Datta S (2017) A survey on recommendation system. *Int J Comput Appl* 7:160

3. Sugiyama K, Kan M-Y (2013) Exploiting potential citation papers in scholarly paper recommendation. In: Proceedings of the 13th ACM/IEEE-CS joint conference on digital libraries, pp 153–162
4. Petricek V, Cox IJ, Han H, Councill IG, Giles CL (2005) Modeling the author bias between two on-line computer science citation databases. In: Special interest tracks and posters of the 14th international conference on World Wide Web, pp 1062–1063
5. Haruna K, Akmar Ismail M, Damiasih D, Sutopo J, Herawan T (2017) A collaborative approach for research paper recommender system. *PLoS ONE* 12(10):0184516
6. Philip S, Shola P, Oveye A (2014) Application of content-based approach in research paper recommendation system for a digital library. *Int J Adv Comput Sci Appl* 10:5
7. Peis E, del Castillo JM, Delgado-López JA (2008) Semantic recommender systems. Analysis of the state of the topic. *Hipertext Net* 6(2008):1–5
8. Neethukrishnan K, Swaraj K (2017) Ontology based research paper recommendation using personal ontology similarity method. In: 2017 second international conference on electrical, computer and communication technologies (ICECCT), pp 1–4. IEEE
9. Hong K, Jeon H, Jeon C (2012) Userprofile-based personalized research paper recommendation system. In: 2012 8th international conference on computing and networking technology (INC, ICCIS and ICMIC), pp 134–138. IEEE
10. Ghosal T, Chakraborty A, Sonam R, Ekbal A, Saha S, Bhattacharyya P (2019) Incorporating full text and bibliographic features to improve scholarly journal recommendation. In: 2019 ACM/IEEE joint conference on digital libraries (JCDL), pp 374–375. IEEE
11. Lofty M, Salama A, El-Ghareeb H, El-dosuky M (2014) Subject recommendation using ontology for computer science ACM curricula. *Int J Inf Sci Intell Syst* 1:3
12. Le Anh V, Hai VH, Tran HN, Jung JJ (2014) Scirecsys: a recommendation system for scientific publication by discovering keyword relationships. In: International conference on computational collective intelligence, pp 72–82. Springer
13. Maake BM, Ojo SO, Zuva T (2019) Information processing in research paper recommender system classes. In: Research data access and management in modern libraries, pp 90–118. IGI Global
14. Shimbo M, Ito T, Matsumoto Y (2007) Evaluation of kernel-based link analysis measures on research paper recommendation. In: Proceedings of the 7th ACM/IEEE-CS joint conference on digital libraries, pp 354–355
15. Achakulvisut T, Acuna DE, Ruangrong T, Kording K (2016) Science concierge: a fast content-based recommendation system for scientific publications. *PLoS ONE* 11(7):0158423
16. Habib R, Afzal MT (2017) Paper recommendation using citation proximity in bibliographic coupling. *Turkish J Electr Eng Comput Sci* 25(4):2708–2718
17. Beel J, Langer S, Genzmehr M, Nürnberger A (2013) Introducing docear's research paper recommender system. In: Proceedings of the 13th ACM/IEEE-CS joint conference on digital libraries, pp 459–460
18. Uchiyama K, Nanba H, Aizawa A, Sagara T (2011) Osusume: cross-lingual recommender system for research papers. In: Proceedings of the 2011 workshop on context-awareness in retrieval and recommendation, pp 39–42
19. Tang T (2006) Active, context-dependent, data-centered techniques for e-learning: a case study of a research paper recommender system. *Data Min E-Learn* 4:97–111
20. Hong K, Jeon H, Jeon C (2013) Personalized research paper recommendation system using keyword extraction based on userprofile. *J Converg Inf Technol* 8(16):106
21. Ollagnier A, Fournier S, Bellot P (2018) Biblme recsys: harnessing bibliometric measures for a scholarly paper recommender system. In: BIR 2018 Workshop on Bibliometric-enhanced Information Retrieval, pp 34–45
22. Strohman T, Croft WB, Jensen D (2007) Recommending citations for academic papers. In: Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval, pp 705–706
23. Liu X, Yu Y, Guo C, Sun Y, Gao L (2014) Full-text based context-rich heterogeneous network mining approach for citation recommendation. In: IEEE/ACM joint conference on digital libraries, pp 361–370. IEEE
24. Manrique R, Marino O (2018) Knowledge graph-based weighting strategies for a scholarly paper recommendation scenario. In: *KaRS@ RecSys*, pp 5–8
25. Sugiyama K, Kan M-Y (2015) A comprehensive evaluation of scholarly paper recommendation using potential citation papers. *Int J Digit Libr* 16(2):91–109
26. Zhang Z, Li L (2010) A research paper recommender system based on spreading activation model. In: The 2nd international conference on information science and engineering, pp 928–931. IEEE
27. Jiang Y, Jia A, Feng Y, Zhao D (2012) Recommending academic papers via users' reading purposes. In: Proceedings of the sixth ACM conference on recommender systems, pp 241–244

28. Hagen M, Beyer A, Gollub T, Komlossy K, Stein B (2016) Supporting scholarly search with keyqueries. In: European conference on information retrieval, pp 507–520. Springer
29. Ohta M, Hachiki T, Takasu A (2011) Related paper recommendation to support online-browsing of research papers. In: Fourth international conference on the applications of digital information and web technologies (ICADIWT 2011), pp 130–136. IEEE
30. Pera MS, Ng Y-K (2011) A personalized recommendation system on scholarly publications. In: Proceedings of the 20th ACM international conference on information and knowledge management, pp 2133–2136
31. Huang W, Kataria S, Caragea C, Mitra P, Giles CL, Rokach L (2012) Recommending citations: translating papers into references. In: Proceedings of the 21st ACM international conference on information and knowledge management, pp 1910–1914
32. Pera MS, Ng Y-K (2014) Exploiting the wisdom of social connections to make personalized recommendations on scholarly articles. *J Intell Inf Syst* 42(3):371–391
33. Beel J, Langer S, Gipp B, Nürnberger A (2014) The architecture and datasets of docear's research paper recommender system. *D-Lib Mag* 20(11/12)
34. Chakraborty T, Modani N, Narayanam R, Nagar S (2015) Discern: a diversified citation recommendation system for scientific queries. In: 2015 IEEE 31st international conference on data engineering, pp 555–566. IEEE
35. Nascimento C, Laender AH, da Silva AS, Gonçalves MA (2011) A source independent framework for research paper recommendation. In: Proceedings of the 11th annual international ACM/IEEE joint conference on digital libraries, pp 297–306
36. He Q, Kifer D, Pei J, Mitra P, Giles CL (2011) Citation recommendation without author supervision. In: Proceedings of the fourth ACM international conference on web search and data mining, pp 755–764
37. Sesagiri Raamkumar A, Foo S, Pang N (2015) Rec4lrw—scientific paper recommender system for literature review and writing. In: Proceedings of the 6th international conference on applications of digital information and web technologies, pp 106–120
38. Magara MB, Ojo SO, Zuva T (2018) Towards a serendipitous research paper recommender system using bisociative information networks (bisonets). In: 2018 international conference on advances in big data, computing and data communication systems (icABCD), pp 1–6. IEEE
39. Rollins J, McCusker M, Carlson J, Stroll J (2017) Manuscript matcher: a content and bibliometrics-based scholarly journal recommendation system. In: BIR@ ECIR, pp 18–29
40. De Nart D, Tasso C (2014) A personalized concept-driven recommender system for scientific libraries. *Procedia Comput Sci* 38:84–91
41. Gipp B, Beel J, Hentschel C (2009) Scienstein: a research paper recommender system. In: Proceedings of the international conference on emerging trends in computing (ICETIC'09), pp 309–315
42. Alzoghbi A, Ayala VAA, Fischer PM, Lausen G (2016) Learning-to-rank in research paper cbf recommendation: leveraging irrelevant papers. In: CBRecSys@ RecSys, pp 43–46
43. Sugiyama K, Kan M-Y (2010) Scholarly paper recommendation via user's recent research interests. In: Proceedings of the 10th annual joint conference on digital libraries, pp 29–38
44. Sugiyama K, Kan M-Y (2011) Serendipitous recommendation for scholarly papers considering relations among researchers. In: Proceedings of the 11th annual international ACM/IEEE joint conference on digital libraries, pp 307–310
45. Tang TY, McCalla G (2009) The pedagogical value of papers: a collaborative-filtering based paper recommender. *J Dig Inf* 10(2):458
46. Ha J, Kim S-W, Faloutsos C, Park S (2015) An analysis on information diffusion through blogcast in a blogosphere. *Inf Sci* 290:45–62
47. Keshavarz S, Honarvar AR (2015) A parallel paper recommender system in big data scholarly. In: International conference on electrical engineering and computer, pp 80–85
48. Pan C, Li W (2010) Research paper recommendation with topic analysis. In: 2010 International conference on computer design and applications, vol. 4, pp 4–264. IEEE
49. Choochaiwattana W (2010) Usage of tagging for research paper recommendation. In: 2010 3rd international conference on advanced computer theory and engineering (ICACTE), vol 2, pp 2–439. IEEE
50. Doerfel S, Jäschke R, Hotho A, Stumme G (2012) Leveraging publication metadata and social data into folkRank for scientific publication recommendation. In: Proceedings of the 4th ACM RecSys workshop on recommender systems and the social Web, pp 9–16
51. Igbe T, Ojokoh B et al (2016) Incorporating user's preferences into scholarly publications recommendation. *Intell Inf Manag* 8(02):27
52. Xia F, Chen Z, Wang W, Li J, Yang LT (2014) Mvwalker: random walk-based most valuable collaborators recommendation exploiting academic factors. *IEEE Trans Emerg Top Comput* 2(3):364–375

53. Agarwal N, Haque E, Liu H, Parsons L (2005) Research paper recommender systems: a subspace clustering approach. In: International conference on web-age information management, pp 475–491. Springer
54. Farooq U, Song Y, Carroll JM, Giles CL (2007) Social bookmarking for scholarly digital libraries. *IEEE Int Comput* 11(6):29–35
55. Loh S, Lorenzi F, Granada R, Lichtnow D, Wives LK, de Oliveira JPM (2009) Identifying similar users by their scientific publications to reduce cold start in recommender systems. In: Proceedings of the fifth international conference on web information systems and technologies (WEBIST 2009), vol 9, pp 593–600
56. Hassan HAM (2017) Personalized research paper recommendation using deep learning. In: Proceedings of the 25th conference on user modeling, adaptation and personalization, pp 327–330
57. Zhou Q, Chen X, Chen C (2014) Authoritative scholarly paper recommendation based on paper communities. In: 2014 IEEE 17th international conference on computational science and engineering, pp 1536–1540. IEEE
58. Meng F, Gao, D, Li, W, Sun X, Hou Y (2013) A unified graph model for personalized query-oriented reference paper recommendation. In: Proceedings of the 22nd ACM international conference on information and knowledge management, pp 1509–1512
59. Al Alshaiikh M, Uchyigit G, Evans R (2017) A research paper recommender system using a dynamic normalized tree of concepts model for user modelling. In: 2017 11th international conference on research challenges in information science (RCIS), pp 200–210. IEEE
60. Tang TY, McCalla G (2009) A multidimensional paper recommender: experiments and evaluations. *IEEE Int Comput* 13(4):34–41
61. Gori M, Pucci A (2006) Research paper recommender systems: a random-walk based approach. In: 2006 IEEE/WIC/ACM international conference on web intelligence (WI 2006 Main Conference Proceedings) (WI'06), pp 778–781. IEEE
62. Zarrinkalam F, Kahani M (2012) A multi-criteria hybrid citation recommendation system based on linked data. In: 2012 2nd international conference on computer and knowledge engineering (ICCKE), pp 283–288. IEEE
63. West JD, Wesley-Smith I, Bergstrom CT (2016) A recommendation system based on hierarchical clustering of an article-level citation network. *IEEE Trans Big Data* 2(2):113–123
64. Pohl S, Radlinski F, Joachims T (2007) Recommending related papers based on digital library access records. In: Proceedings of the 7th ACM/IEEE-CS joint conference on digital libraries, pp 417–418
65. Zhang M, Wang W, Li X (2008) A paper recommender for scientific literatures based on semantic concept similarity. In: International conference on asian digital libraries, pp 359–362. Springer
66. Jomsri P, Sanguansintukul S, Choochaiwattana W (2010) A framework for tag-based research paper recommender system: an ir approach. In: 2010 IEEE 24th international conference on advanced information networking and applications workshops, pp 103–108. IEEE
67. Magalhaes J, Souza C, Costa E, Fechine J (2015) Recommending scientific papers: Investigating the user curriculum. In: The twenty-eighth international flairs conference, pp 489–494
68. Xue H, Guo J, Lan Y, Cao L (2014) Personalized paper recommendation in online social scholar system. In: 2014 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM 2014), pp 612–619. IEEE
69. Wu C-J, Chung J-M, Lu C-Y, Lee H-M, Ho J-M (2011) Using web-mining for academic measurement and scholar recommendation in expert finding system. In: 2011 IEEE/WIC/ACM international conferences on web intelligence and intelligent agent technology, vol 1, pp 288–291. IEEE
70. Liu H, Kong X, Bai X, Wang W, Bekele TM, Xia F (2015) Context-based collaborative filtering for citation recommendation. *IEEE Access* 3:1695–1703
71. Liu X-Y, Chien B-C (2017) Applying citation network analysis on recommendation of research paper collection. In: Proceedings of the 4th multidisciplinary international social networks conference, pp 1–6
72. Hristakeva M, Kershaw D, Rossetti M, Knoth P, Pettit B, Vargas S, Jack K (2017) Building recommender systems for scholarly information. In: Proceedings of the 1st workshop on scholarly web mining, pp 25–32
73. Lee J, Lee K, Kim JG (2013) Personalized academic research paper recommendation system. arXiv preprint [arXiv:1304.5457](https://arxiv.org/abs/1304.5457)
74. Feyer S, Siebert S, Gipp B, Aizawa A, Beel J (2017) Integration of the scientific recommender system mr.dlib into the reference manager jabref. In: European conference on information retrieval, pp 770–774. Springer
75. Collins A, Beel J (2019) Meta-learned per-instance algorithm selection in scholarly recommender systems. arXiv preprint [arXiv:1912.08694](https://arxiv.org/abs/1912.08694)

76. Watanabe S, Ito T, Ozono T, Shintani T (2005) A paper recommendation mechanism for the research support system papits. In: International workshop on data engineering issues in E-commerce, pp 71–80. IEEE
77. Cosley D, Lawrence S, Pennock DM (2002) Referee: an open framework for practical testing of recommender systems using researchindex. In: VLDB'02: Proceedings of the 28th international conference on very large databases, pp 35–46. Elsevier
78. Zhao W, Wu R, Dai W, Dai Y (2015) Research paper recommendation based on the knowledge gap. In: 2015 IEEE international conference on data mining workshop (ICDMW), pp 373–380. IEEE
79. Matsatsinis NF, Lakiotaki K, Delias P (2007) A system based on multiple criteria analysis for scientific paper recommendation. In: Proceedings of the 11th panhellenic conference on informatics, pp 135–149. Citeseer
80. Vellino A (2010) A comparison between usage-based and citation-based methods for recommending scholarly research articles. *Proc Am Soc Inf Sci Technol* 47(1):1–2
81. Huang Z, Chung W, Ong T-H, Chen H (2002) A graph-based recommender system for digital library. In: Proceedings of the 2nd ACM/IEEE-CS joint conference on digital libraries, pp 65–73
82. De Nart D, Ferrara F, Tasso C (2013) Personalized access to scientific publications: from recommendation to explanation. In: International conference on user modeling, adaptation, and personalization, pp 296–301. Springer
83. Middleton SE, De Roue DC, Shadbolt NR (2001) Capturing knowledge of user preferences: ontologies in recommender systems. In: Proceedings of the 1st international conference on knowledge capture, pp 100–107
84. Yukawa T, Kasahara K, Kato T, Kita T (2001) An expert recommendation system using concept-based relevance discernment. In: Proceedings 13th IEEE international conference on tools with artificial intelligence. ICTAI 2001, pp 257–264. IEEE
85. Afzal MT, Maurer HA (2011) Expertise recommender system for scientific community. *J Univers Comput Sci* 17(11):1529–1549
86. Gollapalli SD, Mitra P, Giles CL (2012) Similar researcher search in academic environments. In: Proceedings of the 12th ACM/IEEE-CS joint conference on digital libraries, pp 167–170
87. Yang C, Ma J, Liu X, Sun J, Silva T, Hua Z (2014) A weighted topic model enhanced approach for complementary collaborator recommendation. In: 18th Pacific Asia conference on information systems, PACIS 2014. Pacific Asia Conference on Information Systems
88. Kong X, Mao M, Liu J, Xu B, Huang R, Jin Q (2018) Tnerec: topic-aware network embedding for scientific collaborator recommendation. In: 2018 IEEE smartworld, ubiquitous intelligence and computing, advanced and trusted computing, scalable computing and communications, cloud and big data computing, internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI), pp 1007–1014. IEEE
89. Guerrero-Sosas JD, Chicharro FPR, Serrano-Guerrero J, Menendez-Dominguez V, Castellanos-Bolaños ME (2019) A proposal for a recommender system of scientific relevance. *Procedia Comput Sci* 162:199–206
90. Porcel C, López-Herrera AG, Herrera-Viedma E (2009) A recommender system for research resources based on fuzzy linguistic modeling. *Expert Syst Appl* 36(3):5173–5183
91. Silva ATP (2014) A research analytics framework for expert recommendation in research social networks. Ph.D. thesis, City University of Hong Kong
92. Sun N, Lu Y, Cao Y (2019) Career age-aware scientific collaborator recommendation in scholarly big data. *IEEE Access* 7:136036–136045
93. Xu W, Lu Y, Zhao J, Qian M (2016) Complementarity: a novel collaborator recommendation method for smes. In: 2016 IEEE first international conference on data science in cyberspace (DSC), pp 520–525. IEEE
94. Vazhkudai SS, Harney J, Gunasekaran R, Stansberry D, Lim S-H, Barron T, Nash A, Ramanathan A (2016) Constellation: a science graph network for scalable data and knowledge discovery in extreme-scale scientific collaborations. In: 2016 IEEE international conference on big data (Big Data), pp 3052–3061. IEEE
95. Chen H-H, Treeratpituk P, Mitra P, Giles CL (2013) Csseer: an expert recommendation system based on citeseerx. In: Proceedings of the 13th ACM/IEEE-CS joint conference on digital libraries, pp 381–382
96. Chicaiza J, Piedra N, Lopez-Vargas J, Tovar-Caro E (2018) Discovery of potential collaboration networks from open knowledge sources. In: 2018 IEEE global engineering education conference (EDUCON), pp 1320–1325. IEEE
97. Petry H, Tedesco P, Vieira V, Salgado AC (2008) Icare. A context-sensitive expert recommendation system. In: ECAI'08, pp 53–58

98. Hristovski D, Kastrin A, Rindflesch TC (2016) Implementing semantics-based cross-domain collaboration recommendation in biomedicine with a graph database. *DBKDA 2016*:104
99. Araki M, Katsurai M, Ohmukai I, Takeda H (2017) Interdisciplinary collaborator recommendation based on research content similarity. *IEICE Trans Inf Syst* 100(4):785–792
100. Kong X, Shi Y, Yu S, Liu J, Xia F (2019) Academic social networks: modeling, analysis, mining and applications. *J Netw Comput Appl* 132:86–103
101. dos Santos CK, Evsukoff AG, de Lima BS, Ebecken NFF (2009) Potential collaboration discovery using document clustering and community structure detection. In: *Proceedings of the 1st ACM international workshop on complex networks meet information and knowledge management*, pp 39–46
102. Zhou J, Rafi MA (2019) Recommendation of research collaborator based on semantic link network. In: *2019 15th international conference on semantics, knowledge and grids (SKG)*, pp 16–20. IEEE
103. Cohen S, Ebel L (2013) Recommending collaborators using keywords. In: *Proceedings of the 22nd international conference on World Wide Web*, pp 959–962
104. Hristovski D, Kastrin A, Rindflesch TC (2015) Semantics-based cross-domain collaboration recommendation in the life sciences: preliminary results. In: *Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015*, pp 805–806
105. Li S, Abel M-H, Negre E (2019) Using user contextual profile for recommendation in collaborations. In: *The international research and innovation forum*, pp 199–209. Springer
106. Alinani K, Wang G, Alinani A, Narejo DH (2017) Who should be my co-author? recommender system to suggest a list of collaborators. In: *2017 IEEE international symposium on parallel and distributed processing with applications and 2017 IEEE international conference on ubiquitous computing and communications (ISPA/IUCC)*, pp 1427–1433. IEEE
107. Alinani K, Alinani A, Narejo DH, Wang G (2018) Aggregating author profiles from multiple publisher networks to build a list of potential collaborators. *IEEE Access* 6:20298–20308
108. Benchettara N, Kanawati R, Rouveiol C (2010) A supervised machine learning link prediction approach for academic collaboration recommendation. In: *Proceedings of the fourth ACM conference on recommender systems*, pp 253–256
109. Li J, Xia F, Wang W, Chen Z, Asabere NY, Jiang H (2014) Acrec: a co-authorship based random walk model for academic collaboration recommendation. In: *Proceedings of the 23rd international conference on World Wide Web*, pp 1209–1214
110. Koh YS, Dobbie G (2012) Indirect weighted association rules mining for academic network collaboration recommendations. In: *Proceedings of the tenth Australasian data mining conference*, vol 134, pp 167–173
111. Lee DH, Brusilovsky P, Schleyer T (2011) Recommending collaborators using social features and mesh terms. *Proc Am Soc Inf Sci Technol* 48(1):1–10
112. Yang C, Liu T, Liu L, Chen X (2018) A nearest neighbor based personal rank algorithm for collaborator recommendation. In: *2018 15th international conference on service systems and service management (ICSSSM)*, pp 1–5. IEEE
113. Tong H, Faloutsos C, Pan J-Y (2008) Random walk with restart: fast solutions and applications. *Knowl Inf Syst* 14(3):327–346
114. Kong X, Jiang H, Yang Z, Xu Z, Xia F, Tolba A (2016) Exploiting publication contents and collaboration networks for collaborator recommendation. *PLoS ONE* 11(2):0148492
115. Kong X, Jiang H, Bekele TM, Wang W, Xu Z (2017) Random walk-based beneficial collaborators recommendation exploiting dynamic research interests and academic influence. In: *Proceedings of the 26th international conference on World Wide Web companion*, pp 1371–1377
116. Xu Z, Yuan Y, Wei H, Wan L (2019) A serendipity-biased deepwalk for collaborators recommendation. *PeerJ Comput Sci* 5:178
117. Wang Q, Ma J, Liao X, Du W (2017) A context-aware researcher recommendation system for university-industry collaboration on r&d projects. *Decis Support Syst* 103:46–57
118. Davoodi E, Afsharchi M, Kianmehr K (2012) A social network-based approach to expert recommendation system. In: *International conference on hybrid artificial intelligence systems*, pp 91–102. Springer
119. Brandao MA, Moro MM (2012) Affiliation influence on recommendation in academic social networks. In: *AMW*, pp 230–234
120. Lopes GR, Moro MM, Wives LK, De Oliveira JPM (2010) Collaboration recommendation on academic social networks. In: *International conference on conceptual modeling*, pp 190–199. Springer
121. Payton DW (2004) Collaborator discovery method and system. Google Patents. US Patent 6,681,247
122. Huynh T, Takasu A, Masada T, Hoang K (2014) Collaborator recommendation for isolated researchers. In: *2014 28th international conference on advanced information networking and applications workshops*, pp 639–644. IEEE
123. Zhou X, Ding L, Li Z, Wan R (2017) Collaborator recommendation in heterogeneous bibliographic networks using random walks. *Inf Retr J* 20(4):317–337

124. Chen H-H, Gou L, Zhang X, Giles CL (2011) Collabseer: a search engine for collaboration discovery. In: Proceedings of the 11th annual international ACM/IEEE joint conference on digital libraries, pp 231–240
125. Ben Yahia N, Bellamine Ben Saoud N, Ben Ghezala H (2014) Community-based collaboration recommendation to support mixed decision-making support. *J Decis Syst* 23(3):350–371
126. Chen J, Tang Y, Li J, Mao C, Xiao J (2013) Community-based scholar recommendation modeling in academic social network sites. In: International conference on web information systems engineering, pp 325–334. Springer
127. Gunawardena CN, Hermans MB, Sanchez D, Richmond C, Bohley M, Tuttle R (2009) A theoretical framework for building online communities of practice with social networking tools. *Educ Media Int* 46(1):3–16
128. Zhang Y, Zhang C, Liu X (2017) Dynamic scholarly collaborator recommendation via competitive multi-agent reinforcement learning. In: Proceedings of the eleventh ACM conference on recommender systems, pp 331–335
129. Brandão MA, Moro MM, Almeida JM (2014) Experimental evaluation of academic collaboration recommendation using factorial design. *J Inf Data Manag* 5(1):52–52
130. Fazel-Zarandi M, Devlin HJ, Huang Y, Contractor N (2011) Expert recommendation based on social drivers, social network analysis, and semantic data representation. In: Proceedings of the 2nd international workshop on information heterogeneity and fusion in recommender systems, pp 41–48
131. Zhang J, Tang J, Ma C, Tong H, Jing Y, Li J, Luyten W, Moens M-F (2017) Fast and flexible top-k similarity search on large networks. *ACM Trans Inf Syst* 36(2):1–30
132. Sun J, Ma J, Cheng X, Liu Z, Cao X (2013) Finding an expert: a model recommendation system. In: Thirty fourth international conference on information systems, pp 1–10
133. Bukowski M, Valdez AC, Ziefle M, Schmitz-Rode T, Farkas R (2017) Hybrid collaboration recommendation from bibliometric data. In: Proceedings of 2nd international workshop on health recommender systems co-located with the 11th ACM conference recommender systems, pp 36–38
134. Rebhi W, Yahia NB, Saoud NBB (2016) Hybrid community detection approach in multilayer social network: scientific collaboration recommendation case study. In: 2016 IEEE/ACS 13th international conference of computer systems and applications (AICCSA), pp 1–8 D. IEEE
135. Huynh T, Hoang K (2012) Modeling collaborative knowledge of publishing activities for research recommendation. In: International conference on computational collective intelligence, pp 41–50. Springer
136. Wu S, Sun J, Tang J (2013) Patent partner recommendation in enterprise social networks. In: Proceedings of the sixth ACM international conference on web search and data mining, pp 43–52
137. Liang W, Zhou X, Huang S, Hu C, Jin Q (2017) Recommendation for cross-disciplinary collaboration based on potential research field discovery. In: 2017 fifth international conference on advanced cloud and big data (CBD), pp 349–354. IEEE
138. Olshannikova E, Olsson T, Huhtamäki J, Yao P (2019) Scholars' perceptions of relevance in bibliography-based people recommender system. *Comput Supp Coop Work* 28(3):357–389
139. Yang C, Sun J, Ma J, Zhang S, Wang G, Hua Z (2015) Scientific collaborator recommendation in heterogeneous bibliographic networks. In: 2015 48th Hawaii international conference on system sciences, pp 552–561. IEEE
140. Du G, Liu Y, Yu J (2018) Scientific users' interest detection and collaborators recommendation. In: 2018 IEEE fourth international conference on big data computing service and applications (BigDataService), pp 72–79. IEEE
141. Guerra J, Quan W, Li K, Ahumada L, Winston F, Desai B (2018) Scosy: a biomedical collaboration recommendation system. In: 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 3987–3990. IEEE
142. Wang W, Liu J, Yang Z, Kong X, Xia F (2019) Sustainable collaborator recommendation based on conference closure. *IEEE Trans Comput Soc Syst* 6(2):311–322
143. Datta A, Tan Teck Yong J, Ventresque A (2011) T-recs: team recommendation system through expertise and cohesiveness. In: Proceedings of the 20th international conference companion on World Wide Web, pp 201–204
144. Huynh T, Hoang K, Lam D (2013) Trend based vertex similarity for academic collaboration recommendation. In: International conference on computational collective intelligence, pp 11–20. Springer
145. Al-Ballaa H, Al-Dossari H, Chikh A (2019) Using an exponential random graph model to recommend academic collaborators. *Information* 10(6):220
146. Medvet E, Bartoli A, Piccinin G (2014) Publication venue recommendation based on paper abstract. In: 2014 IEEE 26th international conference on tools with artificial intelligence, pp 1004–1010. IEEE
147. Asabere N, Acakpovi A (2019) Rovets: search based socially-aware recommendation of smart conference sessions. *Int J Decis Supp Syst Technol* 11(3):30–46. <https://doi.org/10.4018/IJDSST.2019070103>

148. Asabere NY, Xu B, Acakpovi A, Deonauth N (2021) Sarve-2: exploiting social venue recommendation in the context of smart conferences. *IEEE Trans Emerg Top Comput* 9(1):342–353. <https://doi.org/10.1109/TETC.2018.2854718>
149. García GM, Nunes BP, Lopes GR, Casanova MA, Paes Leme LAP (2017) Techniques for comparing and recommending conferences. *J Braz Comput Soc* 23(1):1–14
150. Luong H, Huynh T, Gauch S, Do L, Hoang K (2012) Publication venue recommendation using author network's publication history. In: *Intelligent information and database systems*, pp 426–435
151. Zawali A, Boukhris I (2018) A group recommender system for academic venue personalization. In: *International conference on intelligent systems design and applications*, pp 597–606. Springer
152. Beierle F, Tan J, Grunert K (2016) Analyzing social relations for recommending academic conferences. In: *Proceedings of the 8th ACM international workshop on hot topics in planet-scale mObile computing and online social neTworking*, pp 37–42
153. Alshareef AM, Alhamid MF, Saddik AE (2019) Academic venue recommendations based on similarity learning of an extended nearby citation network. *IEEE Access* 7:38813–38825
154. Hiep L, Huynh T, Guach S, Hoang K (2012) Exploiting social networks for publication venue recommendations. In: *International conference on knowledge discovery and information retrieval*, pp 239–245. SciTePress, Spain
155. Küçükünç O, Saule E, Kaya K, Çatalyürek UV (2013) Theadvisor: A webservice for academic recommendation. In: *Proceedings of the 13th ACM/IEEE-CS joint conference on digital libraries. JCDL '13*, pp 433–434. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/2467696.2467752>
156. Chen Z, Xia F, Jiang H, Liu H, Zhang J (2015) Aver: Random walk based academic venue recommendation. In: *Proceedings of the 24th international conference on World Wide Web. WWW '15 companion*, pp 579–584. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/2740908.2741738>
157. Alhoori H, Furuta R (2017) Recommendation of scholarly venues based on dynamic user interests. *J Informet* 11(2):553–563. <https://doi.org/10.1016/j.joi.2017.03.006>
158. Mhirsi N, Boukhris I (2018) Exploring location and ranking for academic venue recommendation. In: *International conference on intelligent systems design and applications*, pp 83–91
159. Pham MC, Cao Y, Klamma R (2010) Clustering technique for collaborative filtering and the application to venue recommendation
160. Yu S, Liu J, Yang Z, Chen Z, Jiang H, Tolba A, Xia F (2018) Pave: personalized academic venue recommendation exploiting co-publication networks. *J Netw Comput Appl* 104:38–47
161. Asabere NY, Xia F, Wang W, Rodrigues JJPC, Basso F, Ma J (2014) Improving smart conference participation through socially aware recommendation. *IEEE Trans Hum-Mach Syst* 44(5):689–700. <https://doi.org/10.1109/THMS.2014.2325837>
162. Pham MC, Cao Y, Klamma R, Jarke M (2011) A clustering approach for collaborative filtering recommendation using social network analysis. *J Univ Comput Sci* 17(4):583–604
163. Pradhan T, Pal S (2020) Cnaver: a content and network-based academic venue recommender system. *Knowl-Based Syst* 189:105092
164. Boukhris I, Ayachi R (2014) A novel personalized academic venue hybrid recommender. In: *2014 IEEE 15th international symposium on computational intelligence and informatics (CINTI)*, pp 465–470. IEEE
165. Yang Z, Davison BD (2012) Venue recommendation: submitting your paper with style. In: *2012 11th international conference on machine learning and applications*, pp 681–686. IEEE
166. Iana A, Jung S, Naeser P, Birukou A, Hertling S, Paulheim H (2019) Building a conference recommender system based on scigraph and wikicfp. In: *Semantic Systems. The power of AI and knowledge graphs*, vol 11702, pp 117–123. Springer
167. Hoang DT, Hwang D, Tran VC, Nguyen VD, Nguyen NT (2016) Academic event recommendation based on research similarity and exploring interaction between authors. In: *2016 IEEE international conference on systems, man, and cybernetics (SMC)*, pp 004411–004416. IEEE
168. Hoang DT, Tran VC, Nguyen VD, Nguyen NT, Hwang D (2017) Improving academic event recommendation using research similarity and interaction strength between authors. *Cybern Syst* 48(3):210–230
169. Errami M, Wren JD, Hicks JM, Garner HR (2007) etblast: a web server to identify expert reviewers, appropriate journals and similar publications. *Nucleic Acids Res* 35(2):12–15
170. Schuemie MJ, Kors JA (2008) Jane: suggesting journals, finding experts. *Bioinformatics* 24(5):727–728
171. SJFinder: SJFinder Recommend Journals. <http://www.sjfinder.com/journals/recommend>
172. Kang N, Doornbal MA, Schijvenaars RJ (2015) Elsevier journal finder: recommending journals for your paper. In: *Proceedings of the 9th ACM conference on recommender systems*, pp 261–264
173. IEEE: IEEE Publication Recommender. <https://publication-recommender.ieee.org/home>
174. Springer: Springer Nature Journal Suggester. <https://journalsuggester.springer.com>

175. Wiley: Wiley Journal Finder. <https://journalfinder.wiley.com/>
176. edanz innovative scientific solutions: Edanz Journal Selector. <https://en-author-services.edanzgroup.com/journal-selector>
177. Guide J Journal Guide. <https://www.journalguide.com/bollacker1998citeeaser>
178. Hetlich S, Pazzani MJ (2006) Mining for proposal reviewers: lessons learned at the national science foundation. In: Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining, pp 862–871
179. Yang K-H, Kuo T-L, Lee H-M, Ho J-M (2009) A reviewer recommendation system based on collaborative intelligence. In: 2009 IEEE/WIC/ACM international joint conference on web intelligence and intelligent agent technology, vol 1, pp 564–567. IEEE
180. Ferilli S, Di Mauro N, Basile TMA, Esposito F, Biba M (2006) Automatic topics identification for reviewer assignment. In: International conference on industrial, engineering and other applications of applied intelligent systems, pp 721–730. Springer
181. Serdyukov P, Rode H, Hiemstra D (2008) Modeling expert finding as an absorbing random walk. In: Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval, pp 797–798
182. Yunhong X, Xianli Z (2016) A lda model based text-mining method to recommend reviewer for proposal of research project selection. In: 2016 13th international conference on service systems and service management (ICSSSM), pp 1–5. IEEE
183. Peng H, Hu H, Wang K, Wang X (2017) Time-aware and topic-based reviewer assignment. In: International conference on database systems for advanced applications, pp 145–157. Springer
184. Medakene AN, Bouanane K, Eddoud MA (2019) A new approach for computing the matching degree in the paper-to-reviewer assignment problem. In: 2019 international conference on theoretical and applicative aspects of computer science (ICTAACS), vol 1, pp 1–8. IEEE
185. Rosen-Zvi M, Griffiths T, Steyvers M, Smyth P (2012) The author-topic model for authors and documents. arXiv preprint [arXiv:1207.4169](https://arxiv.org/abs/1207.4169)
186. Jin J, Geng Q, Mou H, Chen C (2019) Author-subject-topic model for reviewer recommendation. *J Inf Sci* 45(4):554–570
187. Alkazemi BY (2018) Prato: an automated taxonomy-based reviewer-proposal assignment system. *Interdiscip J Inf Knowl Manag* 13:383–396
188. Cagliero L, Garza P, Pasini A, Baralis EM (2018) Additional reviewer assignment by means of weighted association rules. *IEEE Trans Emerg Top Comput* 2:558
189. Ishag MIM, Park KH, Lee JY, Ryu KH (2019) A pattern-based academic reviewer recommendation combining author-paper and diversity metrics. *IEEE Access* 7:16460–16475
190. Zhao S, Zhang D, Duan Z, Chen J, Zhang Y-P, Tang J (2018) A novel classification method for paper-reviewer recommendation. *Scientometrics* 115(3):1293–1313
191. Anjum O, Gong H, Bhat S, Hwu W-M, Xiong J (2019) Pare: A paper-reviewer matching approach using a common topic space. arXiv preprint [arXiv:1909.11258](https://arxiv.org/abs/1909.11258)
192. Zhang, D., Zhao, S., Duan, Z., Chen, J., Zhang, Y., Tang, J.: A multi-label classification method using a hierarchical and transparent representation for paper-reviewer recommendation. arXiv preprint [arXiv:1912.08976](https://arxiv.org/abs/1912.08976) (2019)
193. Li X, Watanabe T (2013) Automatic paper-to-reviewer assignment, based on the matching degree of the reviewers. *Procedia Comput Sci* 22:633–642
194. Xu Y, Du Y (2013) A three-layer network model for reviewer recommendation. In: 2013 sixth international conference on business intelligence and financial engineering, pp 552–556. IEEE
195. Maleszka M, Maleszka B, Król D, Hernes M, Martins DML, Homann L, Vossen G (2020) A modular diversity based reviewer recommendation system. In: Asian conference on intelligent information and database systems, pp 550–561. Springer
196. Sun Y-H, Ma J, Fan Z-P, Wang J (2007) A hybrid knowledge and model approach for reviewer assignment. In: 2007 40th annual Hawaii international conference on system sciences (HICSS'07), pp 47–47. IEEE
197. Kolasa T, Krol D (2011) A survey of algorithms for paper-reviewer assignment problem. *IETE Tech Rev* 28(2):123–134
198. Chen RC, Shang PH, Chen MC (2012) A two-stage approach for project reviewer assignment problem. In: *Advanced materials research*, vol 452, pp 369–373. Trans Tech Publ
199. Daş GS, Göçken T (2014) A fuzzy approach for the reviewer assignment problem. *Comput Ind Eng* 72:50–57
200. Tayal DK, Saxena P, Sharma A, Khanna G, Gupta S (2014) New method for solving reviewer assignment problem using type-2 fuzzy sets and fuzzy functions. *Appl Intell* 40(1):54–73
201. Wang F, Zhou S, Shi N (2013) Group-to-group reviewer assignment problem. *Comput Oper Res* 40(5):1351–1362

202. Long C, Wong RC-W, Peng Y, Ye L (2013) On good and fair paper-reviewer assignment. In: 2013 IEEE 13th international conference on data mining, pp 1145–1150. IEEE
203. Kou NM, U LH, Mamoulis N, Gong Z (2015) Weighted coverage based reviewer assignment. In: Proceedings of the 2015 ACM SIGMOD international conference on management of data, pp 2031–2046
204. Kou NM, U LH, Mamoulis N, Li Y, Li Y, Gong Z, (2015) A topic-based reviewer assignment system. Proc VLDB Endow 8(12):1852–1855
205. Stelmakh I, Shah NB, Singh A (2018) Peerreview4all: Fair and accurate reviewer assignment in peer review. arXiv preprint [arXiv:1806.06237](https://arxiv.org/abs/1806.06237)
206. Yeşilçimen A, Yıldırım EA (2019) An alternative polynomial-sized formulation and an optimization based heuristic for the reviewer assignment problem. Eur J Oper Res 276(2):436–450
207. Conry D, Koren Y, Ramakrishnan N (2009) Recommender systems for the conference paper assignment problem. In: Proceedings of the third ACM conference on recommender systems, pp 357–360
208. Tang W, Tang J, Lei T, Tan C, Gao B, Li T (2012) On optimization of expertise matching with various constraints. Neurocomputing 76(1):71–83
209. Charlin L, Zemel R (2013) The toronto paper matching system: an automated paper-reviewer assignment system
210. Liu X, Suel T, Memon N (2014) A robust model for paper reviewer assignment. In: Proceedings of the 8th ACM conference on recommender systems, pp 25–32
211. Liu O, Wang J, Ma J, Sun Y (2016) An intelligent decision support approach for reviewer assignment in r&d project selection. Comput Ind 76:1–10
212. Ogunleye O, Ifebanjo T, Abiodun T, Adebisi A (2017) Proposed framework for a paper-reviewer assignment system using word2vec. In: 4th Covenant University conference on E-Governance in Nigeria (CUCEN2016)
213. Jin J, Geng Q, Zhao Q, Zhang L (2017) Integrating the trend of research interest for reviewer assignment. In: Proceedings of the 26th international conference on World Wide Web Companion, pp 1233–1241
214. Roberts K, Gururaj AE, Chen X, Pournejati S, Hersh WR, Demner-Fushman D, Ohno-Machado L, Cohen T, Xu H (2017) Information retrieval for biomedical datasets: the 2016 biocaddie dataset retrieval challenge. Database 2017:1–9
215. Chen X, Gururaj AE, Ozyurt B, Liu R, Soysal E, Cohen T, Tiryaki F, Li Y, Zong N, Jiang M (2018) Datamed—an open source discovery index for finding biomedical datasets. J Am Med Inform Assoc 25(3):300–308
216. Jansen BJ, Booth DL, Spink A (2007) Determining the user intent of web search engine queries. In: Proceedings of the 16th international conference on World Wide Web, pp 1149–1150. ACM
217. Nunes BP, Dietze S, Casanova MA, Kawase R, Fetahu B, Nejdil W (2013) Combining a co-occurrence-based and a semantic measure for entity linking. In: Extended semantic web conference, pp 548–562. Springer
218. Ellefi MB, Bellahsene Z, Dietze S, Todorov K (2016) Dataset recommendation for data linking: an intensional approach. In: European semantic Web conference, pp 36–51. Springer
219. Srivastava KS (2018) Predicting and recommending relevant datasets in complex environments. Google Patents. US Patent App. 15/721,122
220. Patra BG, Roberts K, Wu H (2020) A content-based dataset recommendation system for researchers—a case study on gene expression omnibus (geo) repository. Database 2020:1–14
221. Patra BG, Soltanalizadeh B, Deng N, Wu L, Maroufy V, Wu C, Zheng WJ, Roberts K, Wu H, Yaseen A (2020) An informatics research platform to make public gene expression time-course datasets reusable for more scientific discoveries. Database 2020:1–15
222. Zhu J, Patra BG, Yaseen A (2021) Recommender system of scholarly papers using public datasets. In: AMIA summits on translational science proceedings, pp 672–679. American Medical Informatics Association
223. Zhu J, Patra BG, Wu H, Yaseen A (2023) A novel nih research grant recommender using bert. PLoS ONE 18(1):0278636
224. Kamada S, Ichimura T, Watanabe T (2015) Recommendation system of grants-in-aid for researchers by using jsps keyword. In: 2015 IEEE 8th international workshop on computational intelligence and applications (IWCIA), pp143–148. IEEE
225. Kamada S, Ichimura T, Watanabe T (2016) A recommendation system of grants to acquire external funds. In: 2016 IEEE 9th international workshop on computational intelligence and applications (IWCIA), pp 125–130. IEEE



Zitong Zhang is a PhD student in the Department of Biostatistics and Data Science at the University of Texas Health Science Center at Houston, working as a graduate research assistant under the supervision of Dr. Ashraf Yaseen. Zitong completed her master's degree in Computer Science at the University of Florida. Her thesis research focuses on applying natural language processing and machine learning for scholarly recommendation systems in biomedical field.



Braja Gopal Patra PhD, is a research associate in the Department of Population Health Sciences at Weill Cornell Medicine, located in New York, NY, USA. He obtained his PhD degree from Jadavpur University in Kolkata, WB, India. He earned his master's degree from the Department of Computer Science and Engineering at the National Institute of Technology in Agartala, India, and he completed his bachelor's degree in Computer Science and Engineering from West Bengal University of Technology in India. Prior to joining Weill Cornell Medicine, Dr. Patra served as a postdoctoral fellow at The University of Texas Health Science Center in Houston, TX, USA. He has contributed to many publications in reputable journals and conferences, showcasing his expertise in areas such as natural language processing, data science, recommendation systems, and information retrieval.



Ashraf Yaseen is an Assistant Professor of Data Science at the School of Public Health, University of Texas Health Science Center at Houston. He has extensive experience in computer programming, database design, implementation and management, and software development. His research interests include data management, deep learning in EHRs to address public health questions, machine learning, bioinformatics, and high-performance computing.



Gin Zhu received her PhD in Biostatistics and Data Science from the University of Texas Health Science Center with specialization in deep learning, machine learning, and natural language processing. She develops recommenders using state-of-the-art transformers and graph neural networks (GNN) and tackles recommender problems by the incorporation of statistical uncertainty measures. She is currently a Senior Data Scientist at Fidelity Artificial Intelligence Center of Excellence (AI COE).



Rachit Sabharwal is a PhD student in Biostatistics under the supervision of Dr. Ashraf Yaseen at the School of Public Health, University of Texas Health Science Center at Houston, where he also received his MS in Biostatistics. He has a BS in Environmental Health from the University of Rochester. His research interests include natural language processing, recommendation systems, machine learning with applications to public health informatics, and high-performance computing.



Kirk Roberts is an associate professor at the McWilliams School of Biomedical Informatics. His research focus is biomedical natural language processing and information retrieval. He earned his PhD in Computer Science from the University of Texas at Dallas and completed a postdoctoral fellowship at the US National Library of Medicine, National Institutes of Health.



Tru Cao is Associate Professor in the Department of Biostatistics and Data Science, School of Public Health, University of Texas Health Science Center at Houston. He obtained his bachelor degree from Ho Chi Minh City University of Technology, master degree from Asian Institute of Technology, and PhD degree from the University of Queensland. He then spent his postdoc at the University of Bristol and the University of California at Berkeley. Prior to his current position, he was Visiting Professor in the Department of Information Systems and Business Analytics, Eastern Washington University, and Full Professor in the School of Computer Science and Engineering, Ho Chi Minh City University of Technology. His current research interests include machine learning, natural language processing, text mining and retrieval, EHR data analytics and knowledge discovery, and intelligent health information systems.



Hulin Wu is the endowed Betty Wheless Trotter Professor and Chair, Department of Biostatistics and Data Science and Founding Director, Center for Big Data in Health Sciences, School of Public Health, University of Texas Health Science Center at Houston. Dr. Wu was trained in control engineering (BS and MS degrees) and statistics (PhD). His research interests include Big EHR data, complex high-dimensional data, differential equation models, computational systems biology and bioinformatics, clinical trials, and longitudinal data.

Authors and Affiliations

Zitong Zhang¹ · Braja Gopal Patra² · Ashraf Yaseen¹ · Jie Zhu¹ ·
Rachit Sabharwal¹ · Kirk Roberts³ · Tru Cao¹ · Hulin Wu¹

Zitong Zhang
Zitong.Zhang@uth.tmc.edu

Braja Gopal Patra
bgp4001@med.cornell.edu

Jie Zhu
ginnychu512@gmail.com

Rachit Sabharwal
Rachit.Sabharwal@uth.tmc.edu

Kirk Roberts
Kirk.Roberts@uth.tmc.edu

Tru Cao
Tru.Cao@uth.tmc.edu

Hulin Wu
Hulin.Wu@uth.tmc.edu

- ¹ Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA
- ² Department of Population Health Sciences, Weill Cornell Medicine, Cornell University, New York, NY 10065, USA
- ³ School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA