



A graph-based taxonomy of citation recommendation models

Zafar Ali¹ · Guilin Qi¹ · Pavlos Kefalas² · Waheed Ahmad Abro¹ · Bahadar Ali¹

Published online: 21 February 2020
© Springer Nature B.V. 2020

Abstract

Recommender systems have been used since the beginning of the Web to assist users with personalized suggestions related to past preferences for items or products including books, movies, images, research papers and web pages. The availability of millions research articles on various digital libraries makes it difficult for a researcher to find relevant articles to his/her research. During the last years, a lot of research have been conducted through models and algorithms that personalize papers recommendations. With this survey, we explore the state-of-the-art citation recommendation models which we categorize using the following seven criteria: platform used, data factors/features, data representation methods, methodologies and models, recommendation types, problems addressed, and personalization. In addition, we present a novel k-partite graph-based taxonomy that examines the relationships among surveyed algorithms and corresponding k-partite graphs used. Moreover, we present (a) domain's popular issues, (b) adopted metrics, and (c) commonly used datasets. Finally, we provide some research trends and future directions.

Keywords Recommender systems · Research paper recommender systems · Algorithms taxonomy · Information retrieval · Survey · Neural networks

✉ Zafar Ali
zafarali@seu.edu.cn

Guilin Qi
gqi@seu.edu.cn

Pavlos Kefalas
kefalasp@csd.auth.gr

Waheed Ahmad Abro
waheed@seu.edu.cn

Bahadar Ali
bahadardotali@gmail.com

¹ School of Computer Science and Engineering, Southeast University, Nanjing, China

² Department of Informatics, Aristotle University, Thessaloniki, Greece

1 Introduction

The Web contains billions of web pages with information on almost every aspect of daily activities. Moreover, the increasing rate of data on these pages, makes it difficult to support users with relevant information. To cope with these problems, recommender systems have been introduced to facilitate users with personalized suggestions. Additionally, with the rapid development of information technology, the number of research articles on various digital libraries is growing exponentially. Therefore, it is very difficult for a researcher to find research articles related to his/her research needs. To this direction, many models were presented in literature (Meng et al. 2013; Chakraborty et al. 2015; Caragea et al. 2013; Lee et al. 2013; Ebesu and Fang 2017; Alotaibi and Vassileva 2018; Sugiyama and Kan 2013; Bansal et al. 2016) providing personalized citation recommendations. In academic research, such models help researchers to find relevant articles that meet their research interests and information needs.

These models use one of the following techniques: (1) Collaborative filtering (CF) (Wang et al. 2014; Bansal et al. 2016; Wang and Li 2015), (2) Content-based (CB) (Amami et al. 2016; Bhagavatula et al. 2018), and (3) Hybrid (Cai et al. 2019; Chakraborty et al. 2015; Yang et al. 2018; Cai et al. 2018) approaches. CF-based models exploit users' past ratings along with other users' ratings to generate recommendations. CF maintains a user-item rating matrix to identify users or items with similar features and generate recommendations based on that. These models encounter the data sparsity issue, when there is not enough rating information about papers which can be used to generate relevant citation recommendation. In contrast, CB methods compute similarity between the descriptions and features of items (i.e., papers) and user profiles (Pazzani and Billsus 2007). In CB scenario, it is necessary that item descriptions along with user's profile information are available, otherwise CB system will face the cold start problem (Lops et al. 2011). Systems using hybrid models alleviate the sparsity using additional relationships among nodes in the network. Traditional graph-based models (Tian and Jing 2013; Chakraborty et al. 2015) consider recommendation as a link prediction problem. However, random walk based methods suffer of over-weighting old/outdated nodes in the network (Son and Kim 2017). To address such problems, recent studies (Gupta and Varma 2017; Cai et al. 2018, 2019) have employed network representation learning methods for graph-based citation recommendations. On the other hand, to capture the semantic representations of research papers and relevant contextual information, various models (Bansal et al. 2016; Yang et al. 2018; Ebesu and Fang 2017) have introduced deep learning methods to produce quality recommendations. Considering the indispensable role of paper recommendations in academia, a comprehensive review in this area is necessary for the researchers to better understand the strengths and weaknesses, application scenarios, issues and challenges, and evaluation methods/protocols of the state-of-the-art approaches.

Significance of this survey: To the best of our knowledge, only few studies (Bai et al. 2019; Beel et al. 2016) surveyed the domain till this day. Beel et al. (2016) was the first survey that classified models using three information filtering approaches i.e., content-based, collaborative filtering and hybrid recommender systems. This survey covered the literature until 2013, however, due to the advent of novel recommendation algorithms in last four years, a new inclusive framework and taxonomy is required. In the same direction (Bai et al. 2019) surveyed the domain, but both surveys lack of universal taxonomy that can be applied to any new model in the domain. Also, both surveys lack of presenting the trends and directions for future research. Due to the rise of new research methods

such as graph-based citation recommendations, deep learning, and representation learning methods, a comprehensive survey is necessary. Therefore, with this manuscript we aim to survey the domain by exploring new research areas, present a novel graph-based taxonomy, and discuss solutions for addressing the prominent issues in the domain.

Collection of the research papers: To identify relevant literature, we used the database of Web of Science along with Google Scholar. Additionally, we selected the most popular conferences such as KDD, WWW, ReCSys, AACL, JCDL, PAKDD, CIKM, and journals including AIRE, Knowledge-based systems, TKDE, Decision Support Systems, just to name a few, for finding relevant research works. Also, we searched for different keywords such as recommender systems, paper recommendation, citation recommendation, graph-based paper recommendation, deep learning, content-based paper recommendation, CF-based paper recommendation, article recommendation, cold-start, and sparsity, to download papers that meet the relevance criteria of research paper recommender systems.

Contributions of this survey: This paper surveys 47 state-of-the-art citation recommendation models published in the last eight years. First, we classify these models based on the following seven criteria: (a) platform used, (b) data factors/features used, (c) data representation method, (d) methodologies and models, (e) recommendation types, (f) problems addressed, and (g) the personalization. Then, we propose a hybrid k -partite graph taxonomy, acting as a hyper-taxonomy (Kefalas et al. 2015) which classifies algorithms and represents relationships among different kinds of entities. Then, we emphasize on trends, issues and solutions, while we analyze the most popular datasets and evaluation metrics used.

The remaining paper is organized as follows, in Sect. 2 we review the state-of-the-art methods used in the literature and we classify them based on multiple criteria. Then, in Sect. 3 we present a novel taxonomy based on hybrid k -partite graphs. Section 4 discusses popular issues, challenges and solutions, while Sect. 5 reports evaluation methods. Finally, Sect. 6 concludes this survey with new trends and future directions.

2 Algorithms classification

In this section, we examine and classify 47 state-of-the-art citation recommendation models. The classification is based on the following seven criteria: (a) platform used, (b) data factors/features, (c) data representation, (d) methodologies and models used, (e) provided recommendation types, (f) particular problems faced, and (g) the personalization type of the recommendations as shown in Table 1. Also, in Fig. 1 we depict all taxonomies used for that classification. The main purpose of this multilevel parallel classification is to point the strengths and the weaknesses of each model and at the same time to present the domain tendency. In the following subsections we will present the content of Table 1 in details.

2.1 Platform

In this section, we divide the models based on the fact of been integrated in an online system (i.e. website) or just working in off-line mode as shown in the third column of Table 1. This division is a good indicator of the systems that are supporting users in real-time. It is noticeable that most of the systems (Xia et al. 2016; Roy 2017; Manouselis and Verbert 2013; Tian and Jing 2013; Meng et al. 2013; Chakraborty et al. 2015; Caragea et al. 2013; Lee et al. 2013; Alotaibi and Vassileva 2018; Zhang et al. 2014; Son and Kim 2017;

Table 1 Classification of citation recommendation algorithms

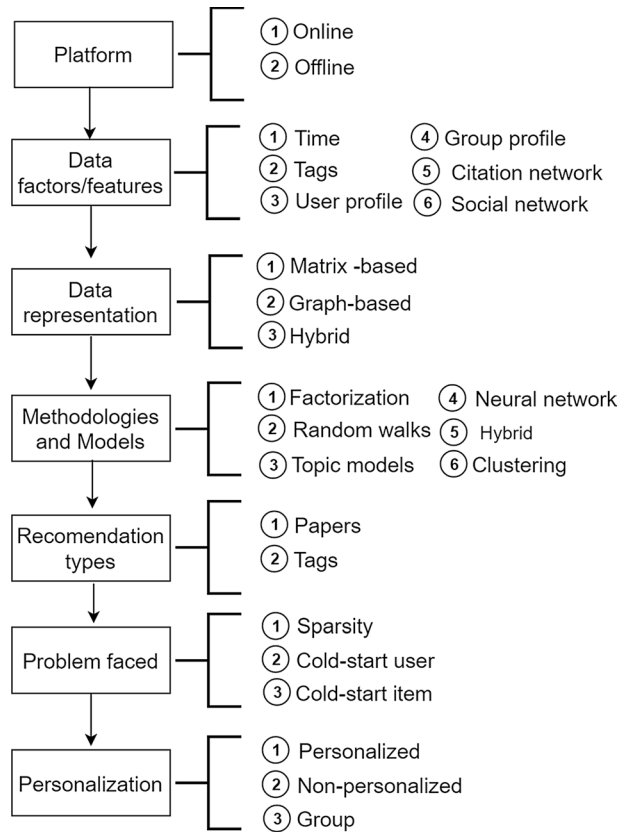
Model	Platform			Data factors/features						Data representation			Methodologies and models					Recommendation types		Problem faced		Personalization				
	Online	Offline	Time	Tags	User profile	Group profiles	Citation network	Social network	Matrix-based	Graph-based	Hybrid	Factorization	Random Walk	Probabilistic	Neural Network	Hybrid	Classification	Clustering	Papers	Tags	Sparsity	Cold-Start user	Cold-Start item	Non-Personalized	Personalized	Group
1 CARE (Xia et al. 2016)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
2 BR5 (Roy 2017)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
3 MAUT (Manouselis and Verbert 2013)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
4 BG-IterRWR (Tian and Jing 2013)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
5 APTW (Meng et al. 2013)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
6 DiSCern (Chakraborty et al. 2015)	-	✓	✓	-	-	-	-	-	✓	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
7 CRS (Caragea et al. 2013)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
8 APR (Lee et al. 2013)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
9 ISNR (Alotolabi and Vassileva 2018)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
10 CAT (Zhang et al. 2014)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
11 MSCN (Son and Kim 2017)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
12 AKR (Sesagiri Raamkumar et al. 2017)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
13 SPR (Sugiyama and Kan 2013)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
14 LDA-TM (Amami et al. 2016)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
15 tr-MF (Li et al. 2013)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
16 DIGTOBI (Kim et al. 2013)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
17 SciRecSys (Le Anh et al. 2014)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
18 BKS (El-Arini and Guestrin 2011)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
19 NPM (Huang et al. 2015)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
20 SARSF (Xia et al. 2014)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
21 CGRK (Blank et al. 2013)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
22 NCN (Ebesu and Fang 2017)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
23 ClusCite (Ren et al. 2014)	-	✓	✓	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
24 EFrec (West et al. 2016)	-	✓	✓	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
25 FeROSA (Chakraborty et al. 2016)	✓	-	-	-	-	-	-	-	✓	-	-	✓	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
26 SSAR (Sun et al. 2014)	✓	-	-	-	-	-	-	-	✓	-	-	✓	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
27 Science Concierge (Achakulvisut et al. 2016)	✓	-	-	-	-	-	-	-	✓	-	-	✓	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
28 CTR (Wang et al. 2014)	✓	-	-	-	-	-	-	-	✓	-	-	✓	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
29 Paper2vec (Ganguly and Pudi 2017)	-	✓	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
30 GRU-MI (Bansal et al. 2016)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
31 CPRec (Li et al. 2018)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
32 PCCR (Yang et al. 2018)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
33 VOPRec (Kong et al. 2019)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
34 BNR (Cai et al. 2019)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
35 TMR (Cai et al. 2018)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
36 CNRN (Waheed et al. 2019)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
37 TMLCCite (Dai et al. 2018)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
38 NREP (Yang et al. 2019)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
39 APR (Pan et al. 2015)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
40 DRDF-CR (Kobayashi et al. 2018)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
41 SAR (Gupta and Varma 2017)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
42 TopicCite (Dai et al. 2019)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
43 HRLHG (Jiang et al. 2018)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
44 RefSeer (Huang et al. 2014)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
45 MMRQ (Mu et al. 2018)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
46 PCTR (Wang and Li 2015)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
47 FCM (Guo et al. 2017)	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-

Sesagiri Raamkumar et al. 2017; Sugiyama and Kan 2013; Amami et al. 2016; Li et al. 2013; Kim et al. 2013; Le Anh et al. 2014; El-Arini and Guestrin 2011; Huang et al. 2015; Xia et al. 2014; Blank et al. 2013; Ebesu and Fang 2017; Ren et al. 2014; Ganguly and Pudi 2017; Bansal et al. 2016; Li et al. 2018; Yang et al. 2018; Kong et al. 2019; Cai et al. 2019, 2018; Waheed et al. 2019; Dai et al. 2018; Yang et al. 2019; Pan et al. 2015; Kobayashi et al. 2018; Gupta and Varma 2017; Dai et al. 2019; Jiang et al. 2018; Mu et al. 2018; Wang and Li 2015; Guo et al. 2017) are not integrated in an online systems against the minority which does support (West et al. 2016; Chakraborty et al. 2016; Sun et al. 2014; Achakulvisut et al. 2016; Wang et al. 2014; Huang et al. 2014) this feature.

2.2 Data factors/features

Recommender systems make use of various features and factors such as time, tags, user profile, group profiles, citation network, and social network while recommending suitable

Fig. 1 Overview of algorithms taxonomy



items to end users. In this section, we discuss about models with respect to such factors used for recommending research articles. These factors are discussed in more detail in the following subsections.

Time: In recent years, few time-aware approaches came up to meet users’ preference evolution (Kefalas and Manolopoulos 2017; Rafailidis et al. 2017; Kefalas et al. 2018). This evolution is related to users’ research interests, which diversify according to the current domain problems they face. For example, one researcher search for ‘tensor factorization’ approaches during this semester, while the same researcher alters her interest to ‘database coherence’ at the next semester. In literature, there are just three models (Ren et al. 2014; West et al. 2016; Chakraborty et al. 2015) that use time dimension to capture this evolution and further personalize their recommendations.

Tags: The most popular feature in citation recommendation models (Meng et al. 2013; Chakraborty et al. 2015; Caragea et al. 2013; Lee et al. 2013; Alotaibi and Vassileva 2018; Zhang et al. 2014; Son and Kim 2017; Sesagiri Raamkumar et al. 2017; Sugiyama and Kan 2013; Amami et al. 2016; Li et al. 2013; Kim et al. 2013; Le Anh et al. 2014; El-Arini and Guestrin 2011; Huang et al. 2015; Xia et al. 2014; Blank et al. 2013; Ebesu and Fang 2017; Ren et al. 2014; West et al. 2016; Chakraborty et al. 2016; Sun et al. 2014; Achakulvisut et al. 2016; Wang et al. 2014; Bansal et al. 2016; Dai et al. 2019; Huang et al. 2014; Mu et al. 2018; Wang and Li 2015) is tags since it correlates keywords with papers. Thus, each

paper is related with multiple words which give a short description of the paper's contents. This way, a content-based model can correlate a target user interests with similar papers based on these tags. It is noticeable that most of the systems that ignore tags (Xia et al. 2016; Roy 2017; Manouselis and Verbert 2013; Tian and Jing 2013; Ganguly and Pudi 2017; Li et al. 2018; Yang et al. 2018; Kong et al. 2019; Cai et al. 2019, 2018; Waheed et al. 2019; Dai et al. 2018; Yang et al. 2019; Pan et al. 2015; Kobayashi et al. 2018; Gupta and Varma 2017; Jiang et al. 2018) are hybrid approaches which use alternative features such as users profile and citation network.

User profile: Users' profile is another popular feature (Xia et al. 2016; Roy 2017; Manouselis and Verbert 2013; Tian and Jing 2013; Sugiyama and Kan 2013; Amami et al. 2016; Li et al. 2013; Kim et al. 2013; Le Anh et al. 2014; El-Arini and Guestrin 2011; Huang et al. 2015; Xia et al. 2014; Sun et al. 2014; Achakulvisut et al. 2016; Wang et al. 2014; Bansal et al. 2016; Li et al. 2018; Yang et al. 2018; Kong et al. 2019; Cai et al. 2019, 2018; Waheed et al. 2019; Dai et al. 2018; Yang et al. 2019; Pan et al. 2015; Kobayashi et al. 2018; Gupta and Varma 2017; Huang et al. 2014; Wang and Li 2015) exploited, since it contains rich information about them while searching for other users with similar taste to a target user. Also, users' past history such as purchase list, ratings, etc. may reveal their preferences and further personalize the final recommendations.

Group profiles: Group profiles contain information about a group of users with similar preferences. Though, this feature is not quite popular in the literature (Xia et al. 2016, 2014) since the personalization of the recommendations is diversified based on individuals preferences which constantly evolve. Also, some users may belong to multiple groups which makes this problem even harder.

Citation network: Citation network feature is related to the references among papers and consists an indicator of their strong bond. In such a network, the papers are represented as nodes that are connected with an edge if there is a reference from one to the other. Thus, this implicit information provided fruitful insights about (1) extensions, (2) correlations, and (3) references among the cited papers. It is noticeable that most of models i.e., 27 out of 47 models (Xia et al. 2016; Roy 2017; Son and Kim 2017; Sesagiri Raamkumar et al. 2017; Sugiyama and Kan 2013; Blank et al. 2013; Ebesu and Fang 2017; Ren et al. 2014; West et al. 2016; Ganguly and Pudi 2017; Li et al. 2018; Yang et al. 2018; Kong et al. 2019; Cai et al. 2019, 2018; Waheed et al. 2019; Dai et al. 2018; Yang et al. 2019; Pan et al. 2015; Kobayashi et al. 2018; Mu et al. 2018; Gupta and Varma 2017; Dai et al. 2019; Jiang et al. 2018; Huang et al. 2014; Wang and Li 2015; Guo et al. 2017) explored citation networks while generating citation recommendations.

Social network: Predicting a user's rating is one of the fundamental recommendation task. Traditional methods using the user-item rating matrix perform poorly when the matrix is sparse and they face sparsity and cold-start problems. In literature, it has been revealed that friends have a tendency to select the same items and give similar ratings in a social network environment. Approaches that make use of social networks (Xia et al. 2016; Tian and Jing 2013; Meng et al. 2013; Lee et al. 2013; Alotaibi and Vassileva 2018; Kim et al. 2013; Xia et al. 2014; Ebesu and Fang 2017; Ren et al. 2014; Sun et al. 2014; Wang et al. 2014; Wang and Li 2015) have improved rating prediction and addressed the sparsity and cold start problems encountered by traditional information filtering approaches.

2.3 Data representation

In this section, we discuss the data representation methods adopted as shown in the fifth column of Table 1. Please notice that there are the following three approaches: (1) Matrix-based, (2) Graph-based, and (3) Hybrid.

Matrix-based: This category belongs to the models that use matrix-based method as a data representation approach. Jianshan et al. presented Semantic-Social Aggregation Recommendation (SSAR) (Sun et al. 2014) model, which constructs a Keyword-Article (KA) matrix where the contents represent the weighted frequency scores of papers title, abstract and social tags. After computing similarities between keywords, the system recommends those articles matching researcher's profile. Moreover, it resolves sparsity which is a regular issue in matrices since the keywords are limited. On the other hand, a Scholarly Paper Recommendation model (SPR) (Sugiyama and Kan 2013) first constructs a paper-citation matrix using Pearson correlation between target and other papers to get the similarity score. Then, it updates the matrix to generate an intermediate imputed matrix where the missing values for the target paper with the neighborhood articles is computed. Citation Recommendation System (CRS) (Caragea et al. 2013) applied Singular Value Decomposition (SVD) (Koren et al. 2009) on the adjacency matrix, which is associated with the citation graph, hence brings related citing and cited papers close to each other. The SVD maps original high dimensional data into low-dimensional form that helps in identifying meaningful patterns in the data as defined below (Caragea et al. 2013):

$$\min \sum (c_{ui} - \mathbf{q}_i^T \mathbf{p}_u)^2 + \lambda (\|\mathbf{q}_i\|^2 + \|\mathbf{p}_u\|^2) \quad (1)$$

where (u, i) represent pairs of citing-cited papers, \mathbf{p}_u and \mathbf{q}_i are the vector representations of citing paper p_u and cited paper q_i . The correlation between them is computed using inner product $\mathbf{q}_i^T \mathbf{p}_u$. Similarly, Science Concierge (Achakulvisut et al. 2016) first creates a weight matrix X by computing term frequency (tf), tf-idf, and log-entropy scores to represent relationships between posters and relevant tags. To reduce the noise, the system employed SVD (Koren et al. 2009) to transform the weighting matrix X by taking the product of a left singular vectors (U), a diagonal singular value matrix (S) and a right singular vectors (V) as $X = USV^T$. Finally, by using the user's preference vector, indexed nearest approach is adopted to suggest posters to the conference participants. On the other hand, Collaborative Topic Regression (CTR) (Wang et al. 2014) model integrates latent factor models along with topic models such as Latent Dirichlet Allocation (LDA) to exploit the features of collaborative filtering and content analysis respectively. To compute latent representations, the model uses user-article rating matrix comprised of entries 0,1, where 1 represents that user has added an article to his/her library, and 0 otherwise.

Graph-based: Graph-based approach models information using k-partite graphs (i.e., uni-partite, bipartite, and higher order graphs etc.) to exploit relationships among entities. As an example, please notice Fig. 2a–c where uni-partite, bi-partite and higher order graphs are depicted respectively. In contrast to higher order graphs, uni-partite and bi-partite graphs are easy to deal with, however, heterogeneous graphs contain a vast number of relations to alleviate the sparsity problem.

Multiple techniques were presented in literature using of k-partite graphs. In this direction, Bi-relational Graph-based Iterative RWR (BG-IterRWR) (Tian and Jing 2013) uses an iterative random walk with restart model to find correlation between researcher–researcher and make predictions about researcher-article relevance. A bipartite graph among researchers and papers is used to explore readerships and calculate similarity between them. On the

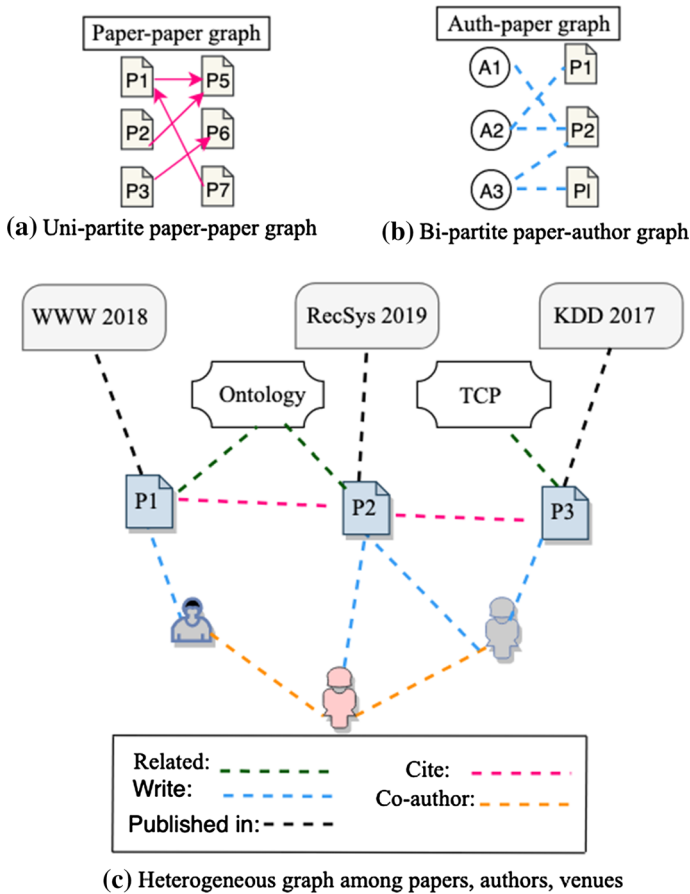


Fig. 2 An example of heterogeneous graph

other hand, (Meng et al. 2013) proposed a multi-layer unified graph-based personalized query oriented reference paper recommendation approach that models articles, users, articles' content and their relationship in the form of a multi-layer graph. Additionally, model integrates network connections i.e., citations and LDA generated topics using paper contents. Similarly, Bibliographic Network Representation (BNR) (Cai et al. 2019), exploits bibliographic network structure as well as content of different kinds of objects such as papers, authors and venues, to learn optimal representations of these objects. Then it generates recommendations by using similarity score between the representations of article, candidate articles as well as the representations of articles and authors.

Hybrid: Hybrid representation consists a combination of previous mentioned representations as shown in Fig. 3.

For example, matrix or graph representations can be used while exploring different kind of auxiliary information. Xia et al. presented a Common Author Relation-based Recommendation (CARE) (Xia et al. 2016) method that uses random walk with restart to exploit common author relationships along with researchers historical preferences. First, the pairwise relationship matrix W_{RA} between researchers and articles is constructed with:

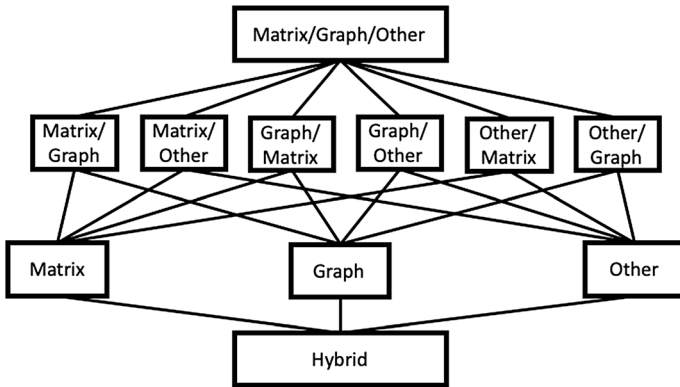


Fig. 3 Hybrid representation combinations

$$W_{RA}(i, j) = \begin{cases} 1 & \text{if } R_i \text{ expressed interest in } A_j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Likewise, the model creates another W_{AA} matrix to represent common author relations where 1 indicates the presence of common author(s) relations between two articles A_i and A_j . The model constructs a matrix of transition probabilities between vertices and utilize random walk with restarts method to rank articles using meta path-based approach:

$$T = \begin{bmatrix} T_{RR} & T_{RA} \\ T_{AR} & T_{AA} \end{bmatrix} = \begin{bmatrix} 0 & \frac{W_{RA}(i,j)}{\sum_k W_{RA}(i,k)} \\ \frac{W_{AR}(i,j)}{\sum_{k1} W_{AR}(i,k1) + \sum_{k2} W_{AA}(i,k2)} & \frac{W_{AA}(i,j)}{\sum_{k1} W_{AR}(i,k1) + \sum_{k2} W_{AA}(i,k2)} \end{bmatrix} \tag{3}$$

Notice, that each entry represents transition probability from one vertex to another. Equal values are given to each neighbor, no matter if the relation is between same type of vertices (i.e., article-article) or two different types of nodes (i.e., article-researcher). On the other hand, a Cluster-based Citation Recommendation (ClusCite) (Ren et al. 2014) model represents citation relation between papers (i.e., a directed sub-graph) using an adjacency matrix $Y \in \mathbb{R}^{n \times n}$, where $Y_{ij} = 1$ if there is citation relation between paper i and paper j and 0 otherwise. Similarly, the association between papers-authors and papers-venues are represented by bi-adjacency matrices where entries contain value 1 if there is relation between corresponding entities, otherwise 0.

Similarly, the model constructs a bi-partite graph between papers and terms, where weight matrix $R^{(\tau)} \in \mathbb{R}^{n \times |\tau|}$ is used to establish relationship between terms and papers, where entries in the matrix are filled using term frequency of terms in the corresponding paper. We can notice that only a few models (Xia et al. 2016; Roy 2017; Huang et al. 2015; Ebesu and Fang 2017; Ren et al. 2014; Yang et al. 2019) have adopted hybrid data representation approach.

2.4 Methodologies and models

In this section, we discuss various methodologies and models used to exploit information and generate recommendations as shown in sixth column of Table 1.

Random walks: Random walk is a dynamic technique when applied to a graph formalizes a path following a succession of random steps. It starts from root node and propagates forward or backward with probability p . That is, the probability of jumping to nearest entities is higher than that of distant nodes. Random walk with restart (Xia et al. 2016), a variant of random walk computes the transition probability as:

$$P(i, j) = \begin{cases} \frac{1}{deg_{out}(i)} & \text{if } (i, j) \in E \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where $deg_{out}(i)$ represents the number of outgoing edges from vertex i . Recommender systems use random walk to exploit relationships among entities by walking through graph structure. In this direction, BNR (Cai et al. 2019) makes use of biased random walk over the bibliographic network $G = (V, E)$ where V represents entities such as papers, authors, and venues, while E demonstrates relationships between entities. The model follows the concept of a neural language model, which considers each path generated using biased random walk as a sentence and each vertex as a word. To train the Skip-Gram model, Node2vec algorithm (Grover and Leskovec 2016) is employed on the corpus, obtaining vector representations of nodes in the graph. For a given vertex p_i the model aims to maximize the probability $\log P(N(p_i) | v_{p_i})$ where v_{p_i} represents output vector representation of current vertex p_i . The neighbor papers of vertex p_i i.e., $P(N(p_i))$ are generated employing a neighborhood sampling approach, where process gets start from source node $c_0 = u$, and the i^{th} vertex is generated taking fixed length walk as:

$$p(c_i = x | c_{i-1} = v) = \begin{cases} \frac{\pi_{xv}}{Z}, & \text{if } (x, v) \in E \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where Z is a normalizing constant, π_{xv} is used for unnormalized transition probability between nodes x and v computed as $\alpha_{pq}(t, x) \cdot w_{xv}$. Also, w_{xv} is a static edge weight if the graph is weighted otherwise 1, and $\alpha_{pq}(t, x)$ is computed as:

$$\alpha_{pq}(t, x) = \begin{cases} \frac{1}{p} & \text{if } d_{tx} = 0 \\ \frac{1}{q} & \text{if } d_{tx} = 1 \\ \frac{1}{q} & \text{if } d_{tx} = 2 \end{cases} \quad (6)$$

d_{tx} depicts the distance between t and x vertices, which is always among $\{0, 1, 2\}$. BG-IterRWR (Tian and Jing 2013) employed iterative random walk with restarts to a Bi-relational graph to identify correlation between researcher–researcher and make predictions about researcher–article relevance. The Bi-Relational graph is composed of sub-graphs including: researcher similarity, article similarity and a bi-partite graph to connect researchers and articles. The model is capable of recommending both new and old articles using the content of articles along with users correlation. Similarly, Faceted Recommendation System for Scientific Articles (FeRoSA) (Chakraborty et al. 2016) recommends articles also using random walk with restarts approach. However, it requires a matrix inversion and the available solutions are not scalable on large graphs. On the other hand, random walk with restart was also adopted by CARE (Xia et al. 2016) to rank research articles. The model exploits common author relationships along with researchers historical preferences. Meng et al. (2013) proposed a multi-layer unified graph-based recommendation model over articles, users, contents and their relationship in the form of multi-layer graph.

Additionally, it integrates network connections i.e., citations and collaborations along with words and topics generated through LDA.

Though, employing traditional random walk based methods such as Random Walk with Restart (RWR) (Meng et al. 2013) and PageRank (Brin and Page 1998) have generated promising results. However, these approaches have certain limitations including: lacks in utilizing mutual reinforcement rules among entities while ranking entities, mostly ignore to exploit semantic relatedness of the content associated with entities, and overweight old papers in the graph. To address the reinforcement problem, a Three-layered Mutually Reinforced model (TMR) (Cai et al. 2018) generated citation recommendations employing multi-layered mutual reinforcement rules, the model incorporates the personalized query information into a multi-layered graph to achieve a query-focused and mutually reinforced recommendation results. A Diversified Citation Recommendation System DiS-Cern (Chakraborty et al. 2015) used an enhanced time-variant of random walk to balance a trade-off between prestige and diversity to address the over-weight issue encountered by PageRank. Similarly, Eigenfactor Recommends (EFrec) (West et al. 2016) makes two modifications to the original PageRank algorithm (Brin and Page 1998) such as: (1) shortening the number of contiguous steps of the random walker before she walks to another part in the network and (2) it propagates to links instead of nodes. These modifications helped in resolving the over-weighting issue without causing any affect in exploring the network structure.

Probabilistic Topic Models: Probabilistic models use probabilities throughout computation. Kim et al. presented a probabilistic model called DIGTOBI (Kim et al. 2013) using a generalized modified version of Probabilistic Latent Semantic Indexing (PLSI) (Hofmann 2017) which integrates another topic model for digging behaviors of users. Similarly, TopicCite (Dai et al. 2019) integrates feature regression with PLSA to mine correlations between topics and citation features. To cope with inherent issues of matrix factorization, Collaborative Topic Regression (CTR) (Wang et al. 2014) proposed another hybrid approach of LDA (Xu et al. 2018). The model integrates latent factor models along with LDA to exploit the features of collaborative filtering and content analysis respectively. Similarly, an extended version of previous system called Relational Collaborative Topic Regression (RCTR) (Wang and Li 2015) incorporates user-item feedback, papers content and network structure information into a principled hierarchical Bayesian model. PCTR has proposed a family of relation probability functions to model relationships between papers. PCTR outperforms CTR in terms of prediction accuracy, and computational cost. Yingming et al. (Li et al. 2013) introduced a novel matrix factorization approach called Topic Regression Matrix Factorization (tr-MF). The model makes use of Gaussian prior to regularize item factors along with LDA for regularizing user factors. In the same direction, to address the issues linked with (Wang et al. 2014) i.e., (a) lacks in making predictions for researchers with few ratings and, (b) inability to generate task-specific recommendations. LDA Topic Model (LDA-TM) (Amami et al. 2016) proposed a CB system employing LDA topic modeling. To build user profile, the system first extracts topics from the abstract of papers with LDA i.e., topics related to researcher's A_i corpus Q_i are represented as a probability distribution $p_i(w|k)$ over words. Then, each unseen paper $d_j \in D$ is formally represented through language model $p(w|d_j)$. The similarities between the two representations are computed as:

$$\text{Similarity}(i,j) = \frac{1}{SKL(k^*,j)} \quad (7)$$

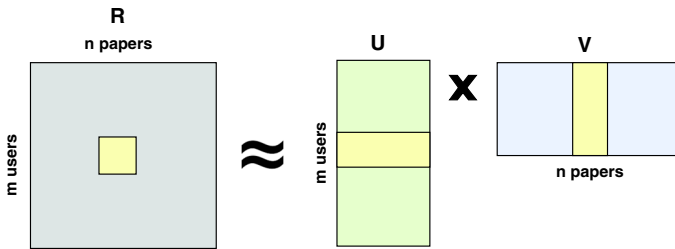


Fig. 4 Matrix factorization

where $SKL(k^*, j)$ is the Kullback Leibler divergence (Prendergast and Staudte 2014). From above discussion, it is evident that probabilistic topic models especially LDA in combination with other approaches has brought promising results.

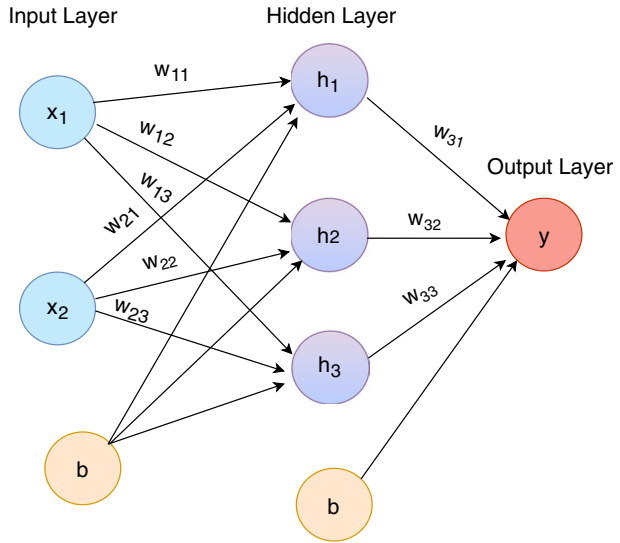
Factorization: To discover latent features and reduce computation cost, factorization is another method that decomposes matrices or tensors into smaller matrices using $\hat{Y} = V^T U$ (by minimizing $\min_{u,v} \|Y - UV^T\| + L(U, V)$) as shown in Figure 4. Matrices U and V , have (m, k) and (k, n) dimensions respectively, where m and n represents number of users and papers.

This way, we avoid over-fitting of factorized U and V matrices to the original matrix. In citation recommendation models, factorization is used to decompose the size of matrices or tensors into smaller matrices. For instance, a Citation Recommendation System (CRS) (Caragea et al. 2013) uses SVD (Koren et al. 2009) approach on the adjacency matrix associated with the CiteSeer citation graph to construct a latent ‘semantic’ space, where citing and cited papers are brought close to each other.

On the other hand, CTR (Wang et al. 2014) have utilized matrix factorization along with LDA topic model. The model generates a user latent vector $u_i \sim N(0, \lambda_u^{-1} I_k)$ for user i . Each paper j is represented as $v_j = \epsilon_j + \theta_j$, where θ_j represents topic distributions and ϵ_j denotes the paper (item) latent offset. Finally, for each (i, j) pair, the model predicts ratings $r_{ij} = (u_i^T v_j, c_{ij}^{-1})$ and generates the final recommendation list. Similarly, PCTR (Wang and Li 2015) extended the former model (Wang et al. 2014) by incorporating additional information sources. The process is almost similar to former model except the fact that it introduces an item relational vector τ_j used to identify binary link between pair of items. Like previous model, Content + Attributes Model (CAT) (Zhang et al. 2014) is a probabilistic latent factor model that exploits the content as well as attributes (i.e., venue, author, publishing year). User-item utility matrix is used to represent relationship between users and papers, where entities $r_{ij} = 1$ refer to the existence of relation a research article, otherwise 0. The model constructs a social graph $G = (V, E)$ where edges E represent relationship between users. For each user, the aim is to generate predictions r_{ij} for user p_i on item q_j not yet seen/rated by the user. The model uses topic modeling to generate low-dimensional latent representation for users $p_i \in \mathbb{R}^d$ and items $q_i \in \mathbb{R}^d$ using the content and descriptive attributes of papers. Finally, the model generates predictions by computing the inner product of users and item latent vectors as $r_{ij} = p_i^T q_j$.

Similarly, TMALCCite (Dai et al. 2018) presents a generative model to recommend citations in bibliographic network. The model exploits both the semantic topics and author communities combining LDA and matrix factorization. The generative process is such that, the model draws co-author links for each edge in the author-author graph G_{aa} . Then, for each paper j , it draws paper topic proportions θ_j and paper latent ϵ_j offset. Then, it

Fig. 5 A simple diagram of neural network



establishes a link P_{a_i,d_j} for each edge in author-paper graph G_{ad} . For each word $w_j:n \in d_j$, the model assigns topics $z_j:n \sim Multi(\theta)$ and $w_j:n \sim Multi(\beta_{z_j:n})$. Finally, a binary citation indicator is drawn for each edge in G_{dd} which is a conditional distribution $P(Y_{d_i,d_j}|\rho_{ij}) = \xi(\cdot|\rho_{ij})$, where $\rho_{ij} = \tau_1 \circ v_i^T v_j + \tau_3 \circ u_i^T u_j + \tau_3$ and \circ is Hadamard product.

Neural Networks: Recently, neural networks have produced promising results in the field of recommendation systems (Batmaz et al. 2019). Neural networks are a set of algorithms such as convolutional neural network (Kim 2014), recurrent neural network (Lipton et al. 2015; Abro et al. 2019), deep belief network (Mohamed et al. 2011), and tensor neural networks (Abedini et al. 2019c) which can learn a complex mapping between input and output space. Typically, a neural network model consists of an input layer, one or more hidden layers and an output layer. Each layer utilizes a set of neurons to produce real-valued activation, by regulating the weights, training the model (Liu et al. 2017). The input layer receives input data in the form of a vector representation i.e., $x = x_1, x_2, \dots, x_n$, which is passed through the hidden layer in the network, as shown in Figure 5. Hidden layers apply non-linearity to the input for learning/mapping between input and output. The output is described as $h = g(W * x + b)$.

Where W denotes weight metric between input layer neurons and hidden layer neurons and b is the bias vector, g indicates non-linear activation function such as sigmoid, tanh, relu etc. Finally, the weighted sums of one or more hidden layers is transmitted to the output layer, which predicts output of the neural network by applying activation function to the propagated value as defined:

$$\hat{y} = g(W * h + b) \tag{8}$$

where h represents the output of hidden layer and W, b are weight and bias of the output layer as depicted in Fig. 5. \hat{y} represents an output while g denotes the activation function that could be sigmoid or softmax depending on the task.

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (9)$$

$$\text{softmax}(x_m) = \frac{e^{x_m}}{\sum_k e^{x_k}} \quad (10)$$

Hidden layer apply non-linearity to the input for learning. To choose the number of hidden layers, researchers need to use cross-validation to test the accuracy on the test set. The optimal number of hidden units could easily be smaller than the number of inputs. If there is a lot of training examples, then multiple hidden units can produce good results, however sometimes just two hidden units works best with few training data.

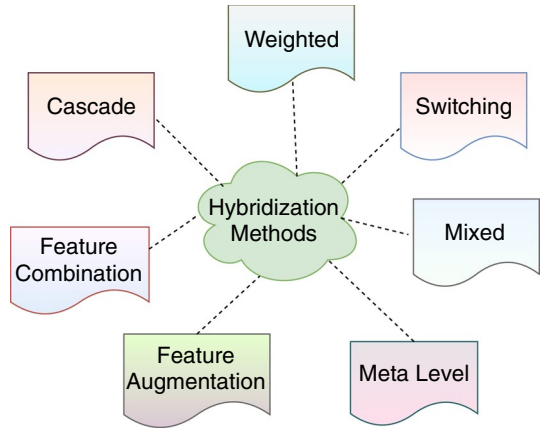
To this point, Huang et al. proposed a Neural Probabilistic Model (NPM) (Huang et al. 2015), a neural probabilistic model that exploits the semantic representation of cited papers and citation contexts. Multi-layer neural network is trained to learn the probability using Bayes rule of citing an article given citation context provided in the form of one or two sentences. Similarly, a Personalized Context-aware Citation Recommendation model called PCCR (Yang et al. 2018) presented a Long Short-term Memory (LSTM) (Abro et al. 2019) based approach which learns vector representations of citation context c as $v(c)$, and scientific paper d_i as $v(d_i)$ using context encoder and scientific paper encoders respectively. Then, the top relevant articles are selected by considering the relevance scores between c and d_i as:

$$r_c = V_D v(c)^T \quad (11)$$

where $V_D = [v_{d_1}, v_{d_2}, v_{d_3}, \dots, v_{d_n}]$. The model applies Softmax function on r_c to compute probability distribution for scientific papers over citation context c as defined in equation 10. Finally, the approach optimizes cross entropy loss function (Yang et al. 2018) to get refined predictions and recommend top N articles. Similarly, Gated Recurrent Unit Multi-task Learning (GRU-ML) (Bansal et al. 2016) model encodes the text sequence into a latent vector using gated recurrent units (GRUs) trained end-to-end on CF task. Also, the the embedding of users and tags are combined with papers representations for the final predictions. On the other hand, Neural Citation Network (NCN) (Ebesu and Fang 2017), uses the logic of neural machine translation approach. In particular, NCN is an encoder-decoder framework which learns semantic relationships between citation contexts and corresponding cited documents title by exploiting author relations. The encoder first encodes given context into vector space employing CNN max time delay neural network while decoder employs GRU for decoding encoder's representation by using attention mechanism and author networks. Finally, the system generates recommendations using the title of papers. In contrast, a CB neural model (Bhagavatula et al. 2018) embeds academic papers into a vector space using the textual content of documents. The model then finds K nearest neighbors of query document and re-rank the citations using another discriminative model which discriminate between seen and unobserved citations. The system outperforms other traditional CF-based approaches in the absence of metadata such as authors, venues, key phrases, and seed list of citations.

In recent years, citation recommendation models exploiting heterogeneous information sources has got much attention (Cai et al. 2019). Therefore, different graph embedding techniques (Goyal and Ferrara 2018) have been used to exploit the graph structure. In this direction, researchers have proposed different graph embedding techniques namely Paper2vec (Ganguly and Pudi 2017), Node2vec (Grover and Leskovec 2016), and LINE (Tang

Fig. 6 Hybridization methods



et al. 2015) capable of information modeling and exploiting rich information sources. For instance, Scientific Article Recommendation model SAR (Gupta and Varma 2017) exploits the distributed representation of bibliographic network using DeepWalk (Perozzi et al. 2014) and distributed representation of articles content using doc2vec (Le and Mikolov 2014). The model then transforms these two modalities into a common embedding space to maximize the correlation between these modalities. Finally, the linear combination of the projections produced using CCA such as:

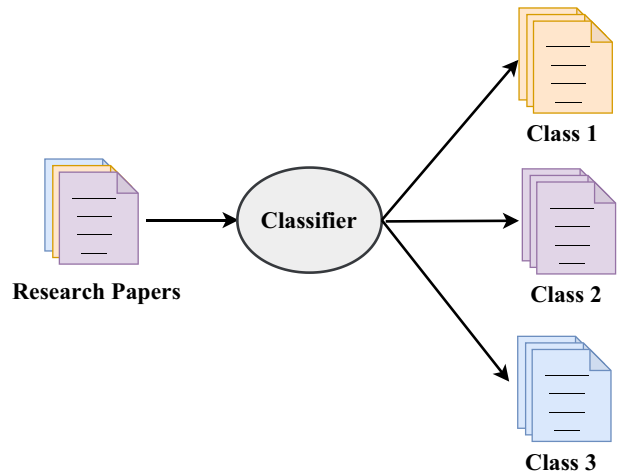
$$W = \alpha W_t + (1 - \alpha) W_g \quad (12)$$

here W_t and W_g are the projections of Doc2vec and DeepWalk (Perozzi et al. 2014) respectively while α is used to tune the contribution of each projection. Neural networks and network embedding techniques are emerging, because they can overcome the drawbacks of traditional filtering approaches by exploiting rich information sources and semantics.

Hybrid: Hybrid models combine two or more techniques to alleviate the inherent issues related to each approach separately (Sun et al. 2014; Ali et al. 2016), as shown in Fig. 6. Different hybridization techniques have been proposed by Burke (2002), where individual filtering approaches such as CB, CF, Knowledge-based, Utility-based and demographic approaches are combined to gain better performance and mitigate their inherent issues.

To this point, SSAR (Sun et al. 2014) combined heterogeneous connections including: social, behavioral and semantic connections with semantic content. This way they developed two types of profiles to represent researchers and then computes similarity scores between researcher-article and researcher-researcher. The model exploits the content of articles and users connections in social computing. First they analyze the semantic content of the articles by keyword similarity calculation method. In particular, it represents the profiles of researchers and candidate articles and match them. That is, the model constructs a Keyword-Article (*KA*) matrix in which matrix elements represent weighted term frequencies. Then similarities of keywords are calculated and articles are matched with the researcher's profile. A Social-Union approach (Symeonidis et al. 2011) retrieves the nearest neighbors of the local user with their corresponding similarity scores. Then, it assigns a voting score to the articles related to the closest neighbors based on nearest neighbors' interest. This way, the semantic content analysis and social connection analysis are exploited to recommend highly relevant and socially endorsed research papers.

Fig. 7 Example of research papers classification

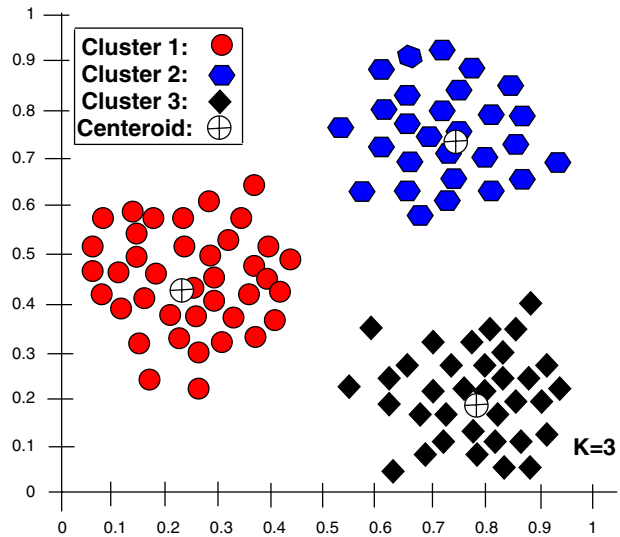


BG-IterRWR (Tian and Jing 2013) is a hybrid model employing article content and researchers readership information. The Bi-Relational Graph model uses iterative random walk and restarts techniques to find correlation between researcher–researcher and make predictions about researcher–article relevance. Meng et al. (2013) introduced another hybrid query-oriented personalized model which exploits diverse information sources such as papers’ content, authorship information and citation network. Additionally, LDA topic modeling is used during the procedure. In contrast, a hybrid Multilevel Simultaneous Citation Network (MSCN) (Son and Kim 2017) employs a multi-level citation network to compare indirectly linked papers to the interested paper, generating relevant papers matching both the research topic as well as academic theory.

Classification: Classification is a supervised learning technique used to classify items/papers into various labels or classes. These computational models are used to assign a class to an input which may either be feature vectors representing characteristics of elements being classified or data representing association between items. Classification process can be depicted in Fig. 7, where research papers have been classified into three classes.

Using classification, the model is able to classify one item based on its features. In our case, papers can be in relevant, non-relevant and undefined classes. Among the most popular algorithms are: Naïve Bayes, decision trees, Support Vector Machine (SVM), and neural networks. In surveyed models, classification approaches have been used by few models (Lee et al. 2013; Xia et al. 2014; Achakulvisut et al. 2016; Li et al. 2018; Yang et al. 2018; Kong et al. 2019; Cai et al. 2019). An Academic Paper Recommendation (APR) (Lee et al. 2013) model makes use of bag-of-words model to represent the data extracted from the titles, keywords and abstracts of research papers. In learning process, the model applies a lazy learning process similar to K-nearest neighbor. More specifically, it adopts a similar idea like KNN to retrieve most similar k papers to the target user’s previously published work. To compute similarity Pearson correlation or cosine similarity is used. In case the target user has more than one previously published works, then all candidate articles are clustered into different groups using K-means. After clustering, the score is computed calculating the distance between candidate article and its centroid. The computed score is used as a distance metric for K-nearest neighbor to produce top k recommendations for the target user. Science Concierge (Achakulvisut et al. 2016) employed Latent Semantic

Fig. 8 A simple illustration of clustering



Analysis (LSA) topic modeling approach to classify posters into topics by using human curated tree like classification approach. A socially-aware model is proposed by Xia et al. (2014) which finds similarities between active participants and other group members based on their research interests as well as social ties. After identifying unique tags, a tagged rating mechanism is used to classify preferences of each participant. Finally, the model recommends articles to those participants who have strong social ties with an Active Participant (AP).

Clustering: Clustering is an unsupervised learning technique used to identify groups of users/items (clusters) having similar features and properties. For instance, user clusters in recommendation systems can be used to make predictions about an individual's preferences consuming the opinion of users in the cluster. Quality clustering approach is the one which minimize the intra-cluster distances while keeping the inter-cluster similarity high. Clustering algorithms are broadly classified into partitional and hierarchical clustering (Rui and Wunsch 2005) considering the properties of clusters generated. Hierarchical approach clusters data objects with a sequence of partitions, either from singleton clusters to a cluster including all individuals or vice versa. On the other hand, partitional clustering divides data objects into predefined number of clusters without considering the hierarchical structure. Clustering process can be depicted in Fig. 8, where data objects have been clustered into three different clusters.

In this direction, APR (Lee et al. 2013) uses K-mean clustering approach to cluster candidate articles before going to select top N papers using K-Nearest Neighbor technique. More specifically, it clusters all candidate articles into different groups using K-means clustering algorithm and then applies a similar idea to that of K-nearest neighbors to generate top k recommendations to the target user.

West et al. (2016) proposed a novel approach called Eigenfactor Recommends(EFrec), employing hierarchical clustering approach to compute inter-cluster relevance scores of research papers. The model addresses the over-weighting issue faced by PageRank (Brin and Page 1998) using ALEF algorithm (Wesley-Smith et al. 2016) which (1) shortens the number of contiguous steps the random walker takes prior propagation to another part in

the graph and (2) propagates to links instead of nodes. Although EFrec is comparatively fast because of its insulation to textual data but at the same time it lacks in considering the semantic information to recommend articles. To retrieve diverse and relevant citations, DiSCern (Chakraborty et al. 2015) system first expand the input query and find communities/clusters of keywords in the network. Previous topic mixture models ignored an important factor of users' rating information and used conditional probability distributions to represent hidden topics. To address this issue, a generative topic modeling approach called DIGTOBI (Kim et al. 2013) generalizes PLSI model considering the digging as well as writing behaviors of users and clustering text collections. A Cluster-based Citation Recommendation model ClusCite (Ren et al. 2014) exploits information of heterogeneous bibliographic networks by considering different types of relationships, the model clusters citations into prominent interest groups. Furthermore, it exploits structural relevance features and deduct relative authority (i.e., importance) for each paper within every group.

2.5 Recommendation types

In this section, we divide the models based on the recommendation services they provide. There are three main types that are (1) article, (2) tags recommendations, or (3) both, as shown in seventh column of the Table 1. It is clear that the majority provide article recommendation services, while only one is not (Blank et al. 2013). Also, two of the models (Blank et al. 2013; Le Anh et al. 2014) support tag recommendations which can be interpreted that a little work has been done in this direction.

2.6 Problems faced

In this section, we discuss different problems researchers face in literature that are: (1) sparsity, (2) cold-start users, and (3) cold-start items discussed. It is evident from eighth column of Table 1 that only few research studies (Zhang et al. 2014; Sugiyama and Kan 2013; Amami et al. 2016; Li et al. 2013; Kim et al. 2013; Xia et al. 2014; Sun et al. 2014; Achakulvisut et al. 2016; Wang et al. 2014; Bansal et al. 2016; Wang and Li 2015) attempted to deal with these issues. Bellow, we discuss these problems briefly while detailed discussion is presented in Sect. 4.

Sparsity: Sparsity is one of the prominent problems studied in the field of recommender systems. The issue arises when the system has few information about the user or an item. In such cases, it is difficult to produce accurate predictions since we can not correlate target user with others. In surveyed algorithms, sparsity problem has been addressed by multiple models (Xia et al. 2014; Sun et al. 2014; Bansal et al. 2016; Sugiyama and Kan 2013).

Cold-start user: Cold start users and cold start papers/items are two notorious problems discussed in state-of-the-art literature. In cold start user case, system does not have enough information about the user. It occurs when new user enters the system and the system is not able to predict his/her tastes. To resolve cold-start user issue, research works (Li et al. 2013; Wang et al. 2014; Kim et al. 2013; Zhang et al. 2014; Wang and Li 2015) have presented different solutions.

Cold-start papers: Systems facing cold start paper problem do not have enough information about the papers, such as the case of a newly published paper when enters into the system. In that case, the system does not have the required information to correlate it with other similar papers or users preferences. In this direction, only five models (Amami et al.

2016; Kim et al. 2013; Bansal et al. 2016; Xia et al. 2014; Achakulvisut et al. 2016) have faced such issues.

2.7 Personalization

In this section, we categorize citation recommendation models by the recommendation services they provide i.e., personalized, non personalized recommendations, and group-based recommendations as shown in ninth column of Table 1. In personalized approach, users' previous history and behaviors with the system are taken into account. On the other hand, non-personalized recommender systems provides a list of papers which are more popular or top rated. A similar list of papers are generated for all the users and top N papers are recommended in a generalized manner without considering the preferences of an individual user. In explored literature, a great proportion of models have produced personalized recommendations (Xia et al. 2016; Roy 2017; Manouselis and Verbert 2013; Tian and Jing 2013; Meng et al. 2013; Chakraborty et al. 2015; Caragea et al. 2013; Lee et al. 2013; Alotaibi and Vassileva 2018; Zhang et al. 2014; Son and Kim 2017; Sesagiri Raamkumar et al. 2017; Sugiyama and Kan 2013; Amami et al. 2016; Kim et al. 2013; Le Anh et al. 2014; Huang et al. 2015; Blank et al. 2013; Ebesu and Fang 2017; Ren et al. 2014; Bansal et al. 2016; Li et al. 2018; Yang et al. 2018; Cai et al. 2019, 2018; Waheed et al. 2019; Yang et al. 2019; Pan et al. 2015; Kobayashi et al. 2018; Mu et al. 2018; Achakulvisut et al. 2016; Sun et al. 2014; West et al. 2016; Chakraborty et al. 2016; Gupta and Varma 2017; Dai et al. 2019; Jiang et al. 2018; Huang et al. 2014; Guo et al. 2017), while only few of them have targeted non-personalized (El-Arini and Guestrin 2011; Wang et al. 2014; Wang and Li 2015; Ganguly and Pudi 2017) or group-based recommendations (Li et al. 2013; Ren et al. 2014; Xia et al. 2014; Wang et al. 2014; Wang and Li 2015).

3 Graph-based taxonomy

In this section, we discuss a novel taxonomy based on the number of information networks used in a recommender system as shown in Fig. 9. In particular, we can see all information used that is related to the (1) research papers, (2) users, (3) venues, (4) tags, (5) behaviors/activities, and (6) relationships among these entities. In this figure, we have uni-partite (user–user, paper–paper, tag–tag networks), and bi-partite (user–paper, paper–tag etc.) networks that can be extended in k -partite.

The proposed approach can be extended into a hybrid k -partite graph with relations between more than two networks such as user–paper–tag. In light of this fact, we propose a novel taxonomy that classifies the explored models used in paper recommendation considering the number of integrated sub-networks. Table 2 denotes the classification of algorithms considering the proposed k -partite graph-based taxonomy. Where PTM represents Probabilistic TopicModels, MF denotes Matrix Factorization, TA represents Time-aware Algorithms, CB/CF refer to Content-based and Collaborative Filtering, NNs denotes Neural Networks, OA is for Other Approaches, HO refers to Higher Order graphs, and BP denotes Bi-partite graphs. Using the proposed taxonomy, we are going to classify all the models presented in Table 1 as follows:

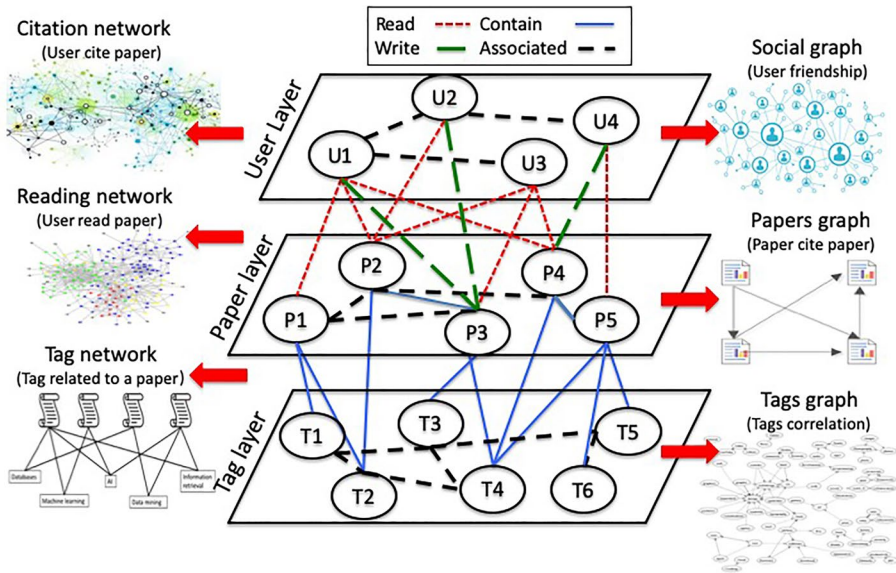


Fig. 9 A hybrid K-partite graph-based taxonomy

3.1 Unipartite graphs

This section covers algorithms using uni-partite graphs such as paper–paper, user–user, tag–tag etc. Unfortunately, none of the models of Table 2 use such network information.

- (1) **Social Graph**:: is an example of uni-partite graph representing social ties/relationships among users. Nodes in the graph represent a single user linked with another user.
- (2) **Paper Graph**:: is a uni-partite graph which demonstrates relationship between/among papers i.e., if paper A cite paper B then there is an edge between two articles.
- (3) **Tags/labels graph**:: represent items/research articles related to a tag or a label. Tags assigned to similar articles share a strong correlation.

3.2 Bipartite graphs

In most of the cases models use bipartite graphs such as paper–user, user–tag, tag–user etc. In particular, the ties between these two information layers denote a preference which can lead to better accuracy.

- (1) **Citation Graph**: denotes a tie between user and papers. Thus, there is an edge between two entities meaning that a user cites a research paper.
- (2) **Reading Graph**: represents that a user’s reading activity with an edge denoting that one user read one article.
- (3) **Tag Graph**: indicates association between papers and concerned tags. In this network, a paper can have multiple tags.

Table 2 A hybrid k-partite graph-based taxonomy of citation recommendation algorithms

Number	Algorithms	Methods/factors	Graph type	Number	Algorithms	Methods/factors	Graph type
1	BKS (El-Arini and Guestrin 2011)	OA	HO	25	CAT (Zhang et al. 2014)	MF+PTM	BP
2	SSAR (Sun et al. 2014)	OA	HO	26	Paper2vec (Ganguly and Pudi 2017)	NNs	BP
3	DiSCern (Chakraborty et al. 2015)	PTM+TA	HO	27	TMALCITE (Dai et al. 2018)	MF	BP
4	RefSeer (Huang et al. 2014)	PTM	HO	28	PCTR (Wang and Li 2015)	MF	BP
5	tr-MF (Li et al. 2013)	PTM	HO	29	SPR (Sugiyama and Kan 2013)	CB+CF	BP
6	TopicCite (Dai et al. 2019)	PTM	HO	30	MAUT (Manouselis and Verbert 2013)	CB+CF	BP
7	APTW (Meng et al. 2013)	RW+MTM	HO	31	LDA-TM (Amami et al. 2016)	CB+CF	BP
8	DIGTOBI (Kim et al. 2013)	MF	HO	32	BG-IterWR (Tian and Jing 2013)	RW	BP
9	Science Concierge (Achakulvisut et al. 2016)	MF	HO	33	FeKoSA (Chakraborty et al. 2016)	RW	BP
10	VOPRec (Kong et al. 2019)	MF+NNs	HO	34	MSCN (Son and Kim 2017)	RW	BP
11	DRDF-CR (Kobayashi et al. 2018)	NNs	HO	35	SAR (Gupta and Varma 2017)	NNs	BP
12	NREP (Yang et al. 2019)	NNs	HO	36	GRU-ML (Bansal et al. 2016)	NNs	BP
13	BNR (Cai et al. 2019)	NNs	HO	37	AKR (Sesagiri Raamkumar et al. 2017)	CB/CF	BP
14	HRLHG (Jiang et al. 2018)	NNs	HO	38	CRS (Caragea et al. 2013)	MF	BP
15	MMRQ (Mu et al. 2018)	NNs	HO	39	APR (Lee et al. 2013)	CB/CF	BP
16	APR (Pan et al. 2015)	NNs	HO	40	ISNR (Alotaibi and Vassileva 2018)	CB/CF	BP
17	ClusCite (Ren et al. 2014)	RW	HO	41	SciRecSys (Le Anh et al. 2014)	CB/CF	BP
18	FCM (Guo et al. 2017)	RW+NNs	HO	42	NPM (Huang et al. 2015)	NNs	BP
19	CARE (Xia et al. 2016)	RW+PTM	HO	43	SARSP (Xia et al. 2014)	CB/CF	BP
20	BRS (Roy 2017)	Hybrid	HO	44	CGRK (Blank et al. 2013)	CB/CF	BP
21	NCN (Ebesu and Fang 2017)	NNs	HO	45	CNRN (Waheed et al. 2019)	CB/CF	BP
22	CPRec (Li et al. 2018)	NNs	HO	46	EFrec (West et al. 2016)	TA	BP
23	PCCR (Yang et al. 2018)	NNs	HO	47	CTR (Wang et al. 2014)	MF	BP
24	TMR (Cai et al. 2018)	NNs	HO				

Next, we discuss models employing different recommendation techniques that exploit bipartite graphs.

3.2.1 Random walk based models

Many models use random walk methods over bi-partite graphs. For instance, a hybrid unified framework BG-IterRWR (Tian and Jing 2013) uses article content (abstract, title) and researcher readership information to recommend research articles. The model constructs a Bi-Relational Graph and use RWR learning process on a transition probability matrix. In contrast to traditional RWR method (Brin and Page 1998), BG-IterRWR iterates in such a way that if its walking on a node v_i of an article sub-graph G_A and v_i is linked with researchers sub-graph G_R . Then the probability to hop to another graph is $\beta \in [0, 1]$, otherwise it stays on the article sub-graph with probability $1 - \beta$. Similarly, a modified time-variant version of random walk i.e., Vertex Reinforced Random Walk (VRRW) is used by DiS-Cern (Chakraborty et al. 2015) to recommend citations. The model first constructs a citation graph $G(V, E)$ where vertices represent research papers and edges e_{ij} demonstrate citing relationship between papers v_i and v_j . Similarly, a weighted keyword-keyword graph is built $G_k(V_k, E_k)$ in which edges are established between two vertices i.e., keywords, if these keywords are contained in at least one article. A weight $w_{i,j}^k$ to the edge $e_{i,j}^k$ is assigned to the edges considering the number of papers in which both the keywords occur together. To do so, the model expands queries by clustering similar keywords in the topological structure. Then the system constructs an induced sub-graph of articles gathered corresponding to the expanded query, and reinforced random walk is applied on induced subgraph to generate relevant, recent as well as old citation recommendations.

On the other hand, FeRoSA (Chakraborty et al. 2016) adopted Random Walk with Restarts (RWR) that considers both the tags as well as citations links to generate recommendations. By using query paper as a starting node, RWR is performed on induced sub-graph having nodes related to query article. Moreover, the model arranges recommendation results into meaningful categories called facets/tags. Here, an induced sub-graph has been constructed where node importance is computed using random walk with restarts. MSCN (Son and Kim 2017) employs multi-level citation network which compare indirectly linked papers to the interested paper by exploiting the structural and semantic relations among articles. To compute similarity between articles, the model exploits two citation relationships namely bibliographic coupling and co-citation. Importance of a research article in the network is identified by determining its degree centrality i.e., papers with huge number of neighbors will have great influence.

3.2.2 CB and CF algorithms

Typically, two types of recommendation algorithms are used to generate recommendations namely: Collaborative Filtering (CF) and Content-based Filtering (CB) (Son and Kim 2017; Wang et al. 2018). CF-based models generate recommendations by using the past ratings information of users and their friends (Najafabadi and Mahrin 2016). Such systems maintain a user-item ratings matrix, using which predictions about items not yet observed by the users are made. In literature, two core CF-based approaches i.e., User-based CF (Breese et al. 1998) and item-based CF (Sarwar et al. 2001), are adopted to find nearest neighbors and nearest items respectively. These approaches are confronted with sparse ratings issue when the rating matrix is very sparse with few user ratings, therefore predictions

made on such limited information can lead to inaccurate predictions. More importantly, the problem gets more severe in research papers domain, as it is very difficult to find user ratings information on digital libraries (Son and Kim 2017). Additionally, CF algorithms can not guarantee justifiable recommendations if they follow traditional user-based CF or item-based CF (Zhang et al. 2014). Reason is that researchers not always prefer all the works of their co-authors, also due to the availability of huge number of papers on the Web, item-based approaches would generate too many recommendations (sometimes irrelevant). To overcome these issues, researchers make use of different CF variants such as exploiting the citation network, the content of articles, users activities and co-authorship information to generate author-papers ratings matrix and produce recommendations.

GRU-ML (Bansal et al. 2016) presented Gated Recurrent Units (GRUs) based CF model, which encodes the text sequence $X_j = (w_1, w_2, \dots, w_n)$ into K_w -dimensional embedding (i.e., e_1, e_2, \dots, e_n), for each word in the corpus. Then the model transforms these embedding into a single vector, $g(X_j)$ to get item content representation. To capture user behavior, the system maintains item-specific embedding \tilde{v}_j which is used to get the final representation of document as:

$$f(X_j) = g(X_j) + \tilde{v}_j \quad (13)$$

The system exploit explicit feedback matrix, where the ratings given by user i on paper j is set to 1 if user has viewed/liked the papers, and 0 otherwise. Finally, the model generates predictions $\tilde{r}_{i,j} = b_i + b_j + \tilde{u}_i^T f(X_j)$ for all unrated papers, where b_i and b_j are user-specific and item-specific biases.

In contrast to CF approaches, CB methods generate recommendations by utilizing the descriptions of items (Pazzani and Billsus 2007; Montaner et al. 2003). In paper recommendation domain, the importance of content can not be ignored, therefore most of the citation recommendations models exploit content of papers along with other auxiliary information sources. In this connection, LDA-TM (Amami et al. 2016), employs LDA topic modeling to the content of articles. First, the model represents a researcher's profile by extracting the topics using the abstract of papers authored by the corresponding researcher using LDA topic modeling approach. Then, each candidate paper is formally represented using a language model, and finally similarity between the two representations i.e., created for researcher's profile and a candidate paper is computed to suggest articles in descending order of the similarity score. On the other hand, a content-based citation recommendation model (Bhagavatula et al. 2018) embeds all academic papers employing a neural model into a vector space using the textual content of documents. The model then finds K nearest neighbors of query document and re-rank the citations using another discriminative model which discriminates between seen and unobserved citations.

In literature, researchers have used CB and CF for recommending articles, however these methods have certain limitations namely: cold start, sparsity, over-specialization and limited content, when applied alone. To overcome these problems, hybrid approaches have been proposed which try to exploit the merits of CB and CF approaches (Bobadilla et al. 2013). In this direction, SPR (Sugiyama and Kan 2013) exploits the content of research papers to generate feature vectors for both researchers and candidate papers. The model first constructs researcher's profile P_{user} using his/her past papers published in last three years using TF approach. To represent research papers, the model employs $TF - IDF$ scheme to the content of papers. Similarity between user's feature vector P_{user} and paper's feature vector F^{prec} is computed using cosine similarity as defined.

$$\frac{P_{user} \cdot F^{prec}}{|P_{user}| \cdot |F^{prec}|} \quad (14)$$

Furthermore, the system identifies potential citation papers (i.e., relevant papers not cited in the paper either intentionally or unintentionally) using CF, exploiting the paper-citation network using Pearson correlation to form a neighborhood. The discovered potential citations, papers that cite target paper p and papers that p refers to, are used to generate candidate paper's feature vector. Finally, cosine similarity is employed to generate final recommendations. On the other hand, SSAR (Sun et al. 2014) exploits different information sources including: user behavioral and social connections as well as articles semantic content (i.e., rich content in the title and abstract) to address the sparsity issue linked with CF-based models. Similarly, an Implicit Social Network based Recommendation model (ISNR) (Alotaibi and Vassileva 2018) exploits three implicit social networks including: readership implicit social network, co-readership implicit social network and tag-based implicit social network to generate recommendations and address cold-start and data sparsity issues encountered by CF-based models. The model constructs these social networks by using data collected from the users' publication list and social bookmarking websites.

3.2.3 Matrix factorization approaches

In surveyed models, multiple works have employed matrix factorization approaches in generating recommendations. CRS (Caragea et al. 2013) applied SVD (Koren et al. 2009) on the adjacency matrix associated with the citation graph, hence brings related citing and cited papers close to each other and address sparsity issue linked with CF algorithms. In contrast to standard CF-based adjacency matrix, here "users" from standard CF are replaced by citing papers in table rows while "items" are replaced with cited papers in matrix columns. The SVD maps original high dimensional data into low-dimensional form which helps in identifying meaningful patterns, correlation between papers is computed by taking the inner product $\mathbf{q}_i^T \mathbf{p}_u$. Where, \mathbf{p}_u and \mathbf{q}_i are the vector representations of citing paper p_u and cited paper q_i respectively. To cope with inherent issues of traditional matrix factorization (i.e., it cannot generate out-of-matrix predictions), Wang et al. (2014) have proposed a hybrid approach using the extension of LDA known as collaborative topic regression. The model integrates latent factor models along with LDA to recommend citations. The research work outperforms both matrix factorization and LDA on in-matrix and out-of-matrix predictions. Similarly, an extension of previous work (Wang et al. 2014) called Relational Collaborative Topic Regression (Wang and Li 2015) integrates user-item feedback, papers content and network structure information into a principled hierarchical Bayesian model. Both of these aforementioned models generate user latent vector u_i and paper latent vector v_j using LDA topic modeling, then for each (i, j) pair the model predicts ratings $r_{i,j}$ and generate final recommendation results. The difference between the models is that the latter model introduces an item relational vector τ_j which is used to identify binary link between pair of items along with other information employed by Wang et al. (2014).

CAT (Zhang et al. 2014) proposed a latent factor model, which recommends articles by applying probabilistic topic modeling and matrix factorization techniques to both the content such as title, abstract, plain text as well as attributes of research articles. Additionally, the model exploits authors social network information to find users with similar preferences and generate useful recommendations in cold start user scenarios. In contrast to CTR (Wang et al. 2014), which regularizes item factors employing a probabilistic topic

model, a novel matrix factorization approach tr-MF (Li et al. 2013) employs regression model and LDA to regularize item as well as user factors respectively. Additionally, the model exploits the network of similar taste users by employing their latent feature vectors generated using LDA, producing more adequate recommendations in cold start scenarios.

3.2.4 Time-aware approaches

The importance of time dimension can not be neglected, because researchers' interests change over time and therefore employing time can exploit patterns in researchers' behaviors. In this direction, different research studies have adopted time dimension in recommending citations. For instance, EFrec (West et al. 2016) tries to address the over-weighting problem faced by traditional random walk methods (i.e., over-weight old papers as they consider citation count in ranking articles) by using Article-Level Eigenfactor (ALEF) (Wesley-Smith et al. 2016) i.e., time-directed acyclic networks designed for paper-level citations, ALEF cope with the over-weight problem by (a) shortening the number of contiguous steps to move forward in the network and (b) teleporting to links instead of nodes. Similarly, DiSCern (Chakraborty et al. 2015) generates diversified recommendations using an enhanced time-variant of random walk (i.e., Vertex Reinforced Random Walk) to recommend recent as well as relevant citation recommendations. In contrast to general random walk methods such as PageRank algorithm (Page et al. 1999), VRRW considers both prestige and diversity in ranking nodes in the graph. That is, the method assumes that the transition probability of an edge does not remain constant, rather change over time. The direction awareness property adopted makes the model capable of recommending both recent as well as old citations.

On the other hand, Ayala-Gómez et al. (2018) exploit semantic features extracted from query manuscript using knowledge graphs along with temporal dimension (i.e., the age of papers in years since their publishing) and citation frequency features to recommend papers. The experimental results of the model on a criteria i.e., "evaluating on the next year", demonstrates that employing temporal features and entities relationships identified using DBpedia results in promising results. Recently, more sophisticated techniques have been proposed to enrich knowledge bases using Neural Tensor Networks (NTN) by integrating classic and meta-heuristic optimization methods (Abedini et al. 2019b).

3.3 Tripartite and higher order graphs

This section covers frameworks that integrate graphs consisting of three or more types of entities such as papers, authors, venues, tags etc. For instance, a graph consists of papers, authors and tags (i.e., Paper–Author–Tag Graph) is an example of tripartite graph. Similarly, a graph using information about papers, authors, venues, and tags (i.e., Paper–Author–Venue–Tag Graph) is an example of higher order graphs. We can see in Table 2 that 24 models have used tripartite and higher order graphs which are discussed in the following subsections.

3.3.1 Matrix factorization methods

Various citation recommendation models have adopted matrix factorization while recommending citations. For instance, matrix factorization approach has been used by Kim et al. (2013) in recommending digg articles by exploiting a modified version of PLSI along with user digging behaviors. Additionally, it utilizes diverse information gathered from multiple information networks including social network, tags network, and user profiles network.

Those models which consider merely the content of articles, ignore specific patterns represented by descriptive attributes. To overcome this issue, Zhang et al. (2014) have proposed a latent factor model using the content as well as attributes (i.e., venue, author, publishing year) while recommending citations. User-item utility matrix is used to represent relationship between users and papers, where entities $r_{ij} = 1$ refer to the existence of relation, otherwise 0. The model constructs a social graph $G = (V, E)$ where edges E represent relationship between users. For each user, the model generates predictions r_{ij} for user p_i on item q_j not yet seen/rated by the user. The system uses topic modeling to generate low-dimensional latent representations (i.e., $D = K + L$) for users $p_i \in \mathbb{R}^d$ and items $q_i \in \mathbb{R}^d$ using matrix factorization approach. Finally, predictions are computed by taking the inner product of users' and items' latent vectors as $r_{ij} = p_i^T q_j$.

In contrast, Science Concierge (Achakulvisut et al. 2016) uses LSA topic modeling to recommend articles considering the user's preference vector created from his/her provided votes. The model represents documents by using human curated tags assigned by the organizers of the conference. Moreover, the model constructs a weighting matrix X based on Term Frequency (TF), Term Frequency Inverse Document Frequency (TF-IDF) for representing relation between posters and human curated tags. To reduce the noise associated with such weighting techniques, the system decomposes weighting matrix using SVD (Koren et al. 2009) by computing the product of left singular vectors (U), a diagonal singular value matrix (S) and a right singular vectors (V) as $X = USV^T$. Finally, the model employs indexed nearest neighbor to generate personalized recommendations to the participants of conference.

VOPRec (Kong et al. 2019) recommends papers through learning vector representations using the textual as well as structural identity of research papers in multi-layer citation network. The model integrates text-based nearest nodes learned using Paper2vec (Ganguly and Pudi 2017) and structured-based vectors learned using struct2vec (Ribeiro et al. 2017) embedding techniques to build a citation network $G' = (V, E \cup E_1 \cup E_2, W)$, where E_1 and E_2 represents text-based and structure based edges. W and V are used for weights and vertices in the graph. The model derives network representation from matrix factorization, the idea is borrowed from (Yang et al. 2015) i.e., matrix factorization and graph-embedding techniques i.e., DeepWalk (Perozzi et al. 2014) are equivalent. That is, given a network $G = (V, E)$, DeepWalk decomposes a matrix M by taking product of two low-dimensional matrices i.e., $W \in \mathbb{R}^{k \times |V|}$ and $H \in \mathbb{R}^{k \times |V|}$ where $k \ll |V|$. After learning vector representations (i.e., exploiting text and network structure) employing network embedding technique, the system generates top Q citation recommendations.

3.3.2 Topic modeling

In this direction, Meng et al. (2013) have used smoothed LDA topic modeling along with symmetric dirichlet priors to recognize a paper as a probabilistic mixture of latent topics.

The unified multi-layer graph-based model uses diverse information including: content, authorship information and network connections i.e., citations and collaborations. Similarly, tr-MF (Li et al. 2013) has integrated LDA topic modeling with matrix factorization approach, the model uses Gaussian prior to regularize item factors along with LDA for regularizing user factors. In contrast, DiSCern (Chakraborty et al. 2015) has used Vertex Reinforced Random Walk (VRRW) in recommending citations, the transition probability used is reinforced and transition to a vertex is based on the previous history of visits to that vertex and hence the probability of jumping to other nodes does not remain constant always. The model employs a topic-sensitive transition probability to jump from one node to another in the network to generate more personalized and diversified recommendations. RefSeer (Huang et al. 2014), a probabilistic model makes use of an extended version of LDA called Cite-PLSA-LDA model, which in contrast to LDA, assume that the words in the citation context are related to the topics of both citing as well as cited documents. Thus the words in the citation contexts and citations are generated from topic-word $Pr(t | w)$ and topic-citation $Pr(d | t)$ multinomial distribution. In response to a query, the model uses the whole query to generate topics for local recommendations, while global citation recommendations are generated by processing every sentence as a separate query.

TopicCite (Dai et al. 2019) generates citation recommendations by investigating both the textual content and citation features employing feature regression and LSA topic model respectively. The model extracts 29 useful features including: cosine between title, abstract and keywords vectors, citations count, authors similarity using jaccard index, venue relevancy, and meta-path based similarity. In contrast to traditional regression problem, features are classified into topics while each feature possesses different weight corresponding different topics. To represent papers as topics, two types of distributions namely: paper-topic $p(z|d)$ and paper-word $p(w|z)$ distributions are generated where each paper is represented by K latent topics. Finally, feature regression and LSA mutually reinforce and jointly learns feature weights for topic learning and papers recommendations.

3.3.3 Neural networks

In surveyed models, multiple research works have employed neural networks to generate semantic-aware citation recommendations. For instance, a Discourse Facet-based Citation Recommendation model (DRDF-CR) (Kobayashi et al. 2018) learns the multi-vector distributed representations of text and citation graphs. The model learns distributed vector representations of papers, with each vector capturing a discourse facets (i.e., Method, Objective and Result) within an article, capable of generating context-dependent recommendations. On the other hand, a Network Representation-based Edge Prediction model (NREP) (Yang et al. 2019) incorporates node content information, observed structure information and hidden edge information, to learn node representation which are used for edge prediction and citation recommendation task. Similarly, BNR (Cai et al. 2019) model, incorporates bibliographic network structure and content of different kinds of objects (papers, authors and venues) to learn optimal representations of these objects and generate personalized recommendations. A novel multi-lingual network representation model called Hierarchical Representation Learning on Heterogeneous Graph (HRLHG) (Jiang et al. 2018) encapsulates both semantic and topological information into a low-dimensional joint embedding for Cross-language Citation Recommendations (CCR). Additionally, the model leverages both the global (i.e., task-dependent) and local (task-independent) semantic and utilizes a novel supervised random walk approach, hence results in optimized

representations of research papers by maximizing the probability of identifying the desired cross-language neighborhoods on the graph.

3.3.4 Other approaches

In this section, we discuss some novel algorithms which do not come under the umbrella of aforementioned techniques. For instance, Semantic-Social Aggregation Recommendation (SSAR Sun et al. 2014) has exploited heterogeneous information sources including: user behavioral and social connections as well as articles semantic content (i.e., rich content in the title and abstract) to address the inherited issues of CB and CF by ranking articles using social voting and pre-computed matching degree approaches. The framework introduces semantic expansion method in building researchers and article profiles by constructing a weighted term frequencies matrix for Keyword-Article (KA). On the other hand, a Beyond Keyword Search-based Recommendation model (BKS) (El-Arini and Guestrin 2011) utilizes a small set of papers as a query input and find more relevant and diverse set of literature matching the concept of the query set. The relevance weights and influence of various concepts as well as documents is computed by considering its existence in both query and candidate set papers. To define the concept c , a directed, acyclic graph is used in which nodes representing papers that contain c and citations as well as common authorship are demonstrated using edges. Also, the influence of one paper on another w.r.t concept c in the graph is computed by assigning concept-specific weights to the edges, which represents the probability of influence of paper x on paper y .

4 Research issues and solutions

In aforementioned sections, we have presented a detailed discussion on recommendation techniques used in literature. Furthermore, we present a novel taxonomy (i.e., a hybrid k -partite graph-based taxonomy) to classify the aforementioned models. Although existing paper recommendations try to assist researchers in suggesting relevant papers, they still possess inherent shortcomings and challenges which need to be addressed. Below, we present some core research issues in that field, namely: (1) cold-start, (2) complex network analysis, (3) limited user modeling, (4) network-specific solutions, (5) data sparsity, (6) lack of evaluation benchmarks, and (7) transformation of research into practice.

4.1 Cold-start

Cold start items (papers) (Khusro et al. 2016) and cold start users (Christoforidis et al. 2018) problem are two prominent issues in recommender systems. In such scenarios, the system does not have enough information to generate recommendations. It occurs when a new user or new paper enters the system and system does not have information such as user-paper ratings, paper's content and metadata, or user's historical records, to generate recommendations (Bobadilla et al. 2013). Moreover, CF-based models suffer from cold start problem as they solely rely on user-item rating matrix which are very sparse (Bansal et al. 2016). Furthermore, researchers are reluctant to provide ratings to research articles in digital libraries, and therefore most of papers does not have rating information (Son and Kim 2017). Employing user-based CF (Breese et al. 1998) or item-based CF (Sarwar et al. 2001) approaches on such sparse rating matrix leads to inadequate suggestions.

In contrast to CF, CB models can overcome the cold start problem by utilizing the content of research papers. Such systems exploit the content of papers and user profile to generate recommendations. However, CB algorithms rely on researchers profiles and their past history i.e., papers rating, reads, downloads, or bookmarks. Therefore, it is mandatory for the system to acquire auxiliary information otherwise it results in limited content analysis and cold start item problem (Adomavicius and Tuzhilin 2005).

To cope with these problems, multiple models have been proposed. In particular, Zhang et al. presented CAT (Zhang et al. 2014) that recommends citations by employing users social network along with paper's content. The proposed latent model exploits content (e.g., title and abstract in plain text), attributes (e.g., author, venue, publish year) along with user social relationships. In contrast, Wang et al. presented the Collaborative Topic Regression (CTR) (Wang et al. 2014) model that integrates CF with probabilistic topic modeling. This model uses articles content to compute similarity among researchers and research papers. Similarly, Yingming et al. proposed tr-MF (Li et al. 2013) model that uses probabilistic topic modeling with a novel matrix factorization approach. The system uses Gaussian prior to regularize item factors along with LDA topic modeling to regularize user factors. This way, the model can overcome cold start problem as long as the predictions of those users are computed using the content of the articles. Similarly, DIGTOBI model (Kim et al. 2013) also tackles cold start problem using a generic PLSI approach which exploits the writing as well as users searching behavior. In contrast to CTR, this model utilizes papers descriptions, votes, and papers content to generate recommendations in both cold-start and warm start scenarios.

4.2 Complex network analysis

Each network in consisted of different kind of entities i.e, papers, authors, venues, words and time as well as links such as cite, authorship, published in, representing relationships between entities in the network. Most research studies (Xia et al. 2016; Roy 2017; Son and Kim 2017) focused on the citation networks, using explicitly available relationships between papers such as citation and co-citation ignoring the rich auxiliary information sources and they fail to generate robust recommendations (Yang et al. 2018; Shi et al. 2019). The majority of the models using link-based techniques ignore complex relationships among papers since they exploit only single-level networks (Manouselis and Verbert 2013). The problem becomes harder if the papers if interest are not that popular. Traditional graph-based approaches use random walk methods such as PageRank (Brin and Page 1998), and therefore suffer from over-weighting old papers compared to new entries.

To overcome this issue, EFrec (West et al. 2016) used hierarchical clustering approach to compute relevance between nodes. Then, the articles with higher importance in these clusters were selected as candidate recommendations. Similarly, Discern (Chakraborty et al. 2015) introduced a time-variant random walk method which in contrast to PageRank algorithm assumes that the transition probabilities of edges is not constant. Therefore, transition probabilities between vertices are reinforced by the number of previous visits to these vertices. Similarly, (Waheed et al. 2019; Son and Kim 2017) exploiting multilevel citation network, to quantify direct, indirect, and integral relations among papers. These methods compare all the relevant papers in more than one hop and inspect the complex relationships among them using certain centrality measures namely degree centrality, betweenness centrality, closeness centrality, and Eigenvector centrality.

Though, the aforementioned approaches resolve the issues of single-level networks and random walks, however, researchers should not consider merely the links between papers but also exploit associated latent relations and heterogeneous information sources to recommend citations. Recently, the Heterogeneous Information Networks (HINs) (Shi et al. 2019) have generated promising results in generating recommendations. In contrast to traditional citation network analysis, the HINs-based methods have been demonstrated a more efficient way to model real world data for many applications (Dai et al. 2019; Huang and Mamoulis 2017), it represents an abstraction of the real world, focusing on the objects and the relationships among them. For instance, the diverse information sources in papers domain can include but not limited to papers-venues, papers-authors, paper-publication time, and research papers logical structure as well as textual content.

4.3 Limited user modeling

User modeling is the first step in recommender system which is used to identify user preferences and information needs. To generate satisfactory results, a recommendation system needs to infer more information about the user needs from user behaviors and interactions with the system (Sugiyama and Kan 2013). However, systems which ask users to explicitly provide their preferences by giving keywords and inputs behaves like typical search engines and hence lacks an important feature to recommend items to users even if they do not know about their needs. For instance, research works (El-Arini and Guestrin 2011; Ren et al. 2014; Dai et al. 2018) generate citation recommendations by asking researcher to provide a query article, hence their models behave like related document search. Of course, related-document search is not a trivial task, however they neglect the user modeling process and hence ignore a very crucial part of recommender systems. Among the reviewed models, a few authors have considered user modeling as an important part of their proposed work. However, most of the models built user's models by inferring information from user's authored and downloaded papers, which is not an adequate method. As research preferences change over the time and model based on such papers would lead to inappropriate recommendations. SPR (Sugiyama and Kan 2013), build user model by considering user's authored papers, weighting each paper based on the difference between a paper's publication year, and the year of the most recently authored paper. Moreover, their results demonstrate that models based on papers authored in last three years yielded satisfactory recommendations. More importantly, user modeling based citation recommendation methods are designed to generate personalized recommendation results. BNR (Cai et al. 2019) reveals that personalized citation recommendation approaches are superior than no-personalized in terms of generating relevant and accurate results.

4.4 Network-specific solution

Most of the existing models exploiting the citation network (Chakraborty et al. 2015, 2016), either ignore the content and logical structure of papers or use very limited meta data such as title, keywords, and citations. Such approaches confront with more critical problems when there is sparse rating information (i.e., sparsity issue in CF based systems) or the presence of false citations in systems merely using citation analysis. Bai et al. (2019) reported that most of the algorithms using graph-based approaches ignore the content of research articles and exploit graph structure to generate recommendations. Though, exploiting citation graph can help in diversifying recommendation results, however neglecting the

content can lead to inadequate recommendations (Ganguly and Pudi 2017). A research article is a rich source of content comprised of well structured logical sections, ignoring the content and structure of a paper during recommendations can lead to inadequate recommendation results (Waheed et al. 2019). Authors of research papers cite articles following certain procedural rules (Habib and Afzal 2019). That is, closely related articles are cited in the methodology and results parts while introduction contain papers related to the background knowledge. Additionally, authors cite baseline methods in experimental sections. Due to these facts, papers cited in methodology and results sections are considered to be more important than cited in introduction or related work sections (Sugiyama and Kan 2013). For instance, to find a baseline method or an alternative algorithm for a query manuscript employing citation network analysis approaches is not an easy task. Therefore, to generate such serendipitous recommendations (Kotkov et al. 2016), the logical structure and auxiliary textual information of papers need to be exploited.

4.5 Data Sparsity

Sparsity is a well-known notorious problem which occurs when the systems have enormous amount of data about items/papers but users are reluctant to provide ratings on items results in sparse rating matrix, therefore making predictions based on such matrices generate inaccurate results (Khusro et al. 2016; Bobadilla et al. 2013). Mostly, CF-based models suffer from sparse ratings issue where predictions are solely based on user's friends/nearest-neighbors ratings, and therefore makes it computationally difficult to find such neighbors. The issue can be more critical in citation recommender systems where the researchers would not afford to receive irrelevant recommendations beyond the scope of their research interests. In contrast to other domains like movies recommendations, the number of users are usually less than research papers in citation recommendations, which leads to data sparsity issue (Perozzi et al. 2014). To cope with sparsity issue, researchers have proposed different approaches in surveyed models. In this direction, RCTR (Wang and Li 2015) resolves sparsity issue encountered by traditional CF-based systems by integrating user-item feedback, papers content and network structure information into a principled hierarchical Bayesian model.

4.6 Lack of evaluation benchmarks

To evaluate the significance of a research work, researchers use publicly available datasets which helps researchers to assess their proposed methods against other baseline approaches. In explored literature, the most commonly used datasets include: CiteUlike, CiteSeer, DBLP, ACM, Aminer and others. Though, there exist some online datasets but most of the datasets reported by researchers are either not available publicly or possess erroneous and missing information. For instance, Bansal et al. (2016) in their proposed CF-based system has used Citeulike dataset reporting that the dataset they used is 99.99 % sparse, therefore generating recommendations on such datasets will lead to inadequate recommendation results and hence a big challenge for researchers. The availability of benchmark datasets encourage researchers to develop useful recommendation models. Unfortunately, most of the available datasets cannot be used for special kind of recommenders such as evaluating systems using neural networks, heterogeneous networks, and contextual information. For instance, these days neural networks are emerging in various fields such as NLP, text mining and recommendation. The significance of these models depend on the size and

quality of datasets, however, we do not have suitable datasets in paper recommendation domain which suits such models. Therefore, the development of a large annotated corpora for scientific papers is indispensable for the research community to come up with novel approaches. Furthermore, models exploiting heterogeneous information sources make use of datasets such as ACL anthology and DBLP, however these datasets do not have information about user feedback and ratings using which user preferences can be considered while generating recommendations. Most of the heterogeneous network-based models therefore ignore users preference dynamics.

4.7 Transformation of research into practice

After surveying the models, we can notice in Table 1 that out of 46 models only 6 models (Huang et al. 2014; Achakulvisut et al. 2016; Wang et al. 2014; Sun et al. 2014; Chakraborty et al. 2016; West et al. 2016) can be found as working applications. Of these 6 models, one of them is no longer online. Also, very few research studies i.e., 5 out of 47 have made their code publicly available. As most of the researchers ignore the reproducibility of experimental results. Therefore, evaluations are not handled consistently resulting in ambiguities in identifying effective recommendation and suitable baseline methods for conducting a novel research (Beel et al. 2016). Perhaps that is the reason why existing real world applications such as ResearchGate,¹ PubMed² and CiteULike³ did not inherit state-of-the-art research methods as most of them are using traditional information filtering mechanisms while generating citation recommendations.

5 Datasets and metrics for evaluation

To assess the performance of recommendation results, citation recommendation models make use of different evaluation metrics and datasets. Due to the emergence of novel information filtering methods and algorithms, it is indispensable to select appropriate datasets and metrics to evaluate the experimental results (Bobadilla et al. 2013). The experimental results evaluated by standard evaluation metrics on a dataset determines the effectiveness and significance of the proposed model against competitive baselines. In the following subsections, we present the most commonly used datasets and metrics employed in surveyed models.

5.1 Datasets

The most popular datasets used are given in Table 3. The second column of the table presents the names of 47 models we have already shown in Table 1. We discuss these datasets in the following subsections.

CiteULike: CiteULike was an online platform which allowed registered users to create personal libraries by saving papers which are of interest to them. The dataset⁴ consists of

¹ <https://www.researchgate.net/>

² <https://www.ncbi.nlm.nih.gov/pubmed/>

³ <http://www.citeulike.org/>

⁴ <http://www.citeulike.org/faq/data.adp>

Table 3 Evaluation datasets

Models	CiteUlike	CiteSeer	DBLP	ACL anthol- ogy	Aminer	Self-collected
1 CARE (Xia et al. 2016)	✓	–	–	–	–	–
2 BRS (Roy 2017)	–	✓	–	–	–	–
3 MAUT (Manouselis and Verbert 2013)	–	–	–	–	–	✓
4 BG–IteRWR (Tian and Jing 2013)	✓	–	–	–	–	–
5 APTW (Meng et al. 2013)	–	–	–	✓	–	–
6 DiSCern (Chakraborty et al. 2015)	–	✓	–	–	–	–
7 CRS (Caragea et al. 2013)	–	✓	–	–	–	–
8 APR (Lee et al. 2013)	–	–	–	–	–	✓
9 ISNR (Alotaibi and Vassileva 2018)	✓	–	–	–	–	–
10 CAT (Zhang et al. 2014)	–	–	–	–	✓	–
11 MSCN (Son and Kim 2017)	–	–	–	–	–	✓
12 AKR (Sesagiri Raamkumar et al. 2017)	–	–	–	–	–	✓
13 SPR (Sugiyama and Kan 2013)	–	–	–	–	–	✓
14 LDA–TM (Amami et al. 2016)	–	–	–	–	✓	–
15 tr–MF (Li et al. 2013)	✓	–	–	–	–	–
16 DIGTOBI (Kim et al. 2013)	–	–	–	–	–	✓
17 SciRecSys (Le Anh et al. 2014)	–	–	–	–	–	✓
18 BKS (El-Arini and Guestrin 2011)	–	–	–	–	–	✓
19 NPM (Huang et al. 2015)	–	✓	–	–	–	–
20 SARSP (Xia et al. 2014)	–	–	–	–	–	✓
21 CGRK (Blank et al. 2013)	–	✓	–	–	–	–
22 NCN (Ebesu and Fang 2017)	–	✓	–	–	–	–
23 ClusCite (Ren et al. 2014)	–	–	–	–	✓	✓
24 EFrec (West et al. 2016)	–	–	✓	–	✓	–
25 FeRoSA (Chakraborty et al. 2016)	–	–	–	–	–	✓
26 SSAR (Sun et al. 2014)	✓	–	–	–	–	–
27 Science Concierge (Achakulvisut et al. 2016)	–	–	–	–	–	✓
28 CTR (Wang et al. 2014)	✓	–	–	–	–	–
29 Paper2vec (Ganguly and Pudi 2017)	–	–	✓	–	–	–
30 GRU–ML (Bansal et al. 2016)	✓	–	–	–	–	–
31 CPRec (Li et al. 2018)	–	–	–	–	–	✓
32 PCCR (Yang et al. 2018)	–	–	✓	✓	–	–
33 VOPRec (Kong et al. 2019)	–	–	–	–	–	✓
34 BNR (Cai et al. 2019)	–	–	✓	✓	–	–
35 TMR (Cai et al. 2018)	–	✓	✓	✓	–	–
36 CNRN (Waheed et al. 2019)	–	–	–	–	✓	–
37 TMALCCite (Dai et al. 2018)	–	–	–	✓	✓	–
38 NREP (Yang et al. 2019)	–	–	✓	✓	–	–
39 APR (Pan et al. 2015)	–	–	–	✓	–	–
40 DRDF–CR (Kobayashi et al. 2018)	–	–	–	✓	–	–
41 SAR (Gupta and Varma 2017)	–	–	–	–	✓	–

Table 3 (continued)

	Models	CiteUlike	CiteSeer	DBLP	ACL anthol- ogy	Aminer	Self-collected
42	TopicCite (Dai et al. 2019)	–	–	✓	✓	–	–
43	HRLHG (Jiang et al. 2018)	–	–	–	–	–	✓
44	RefSeer (Huang et al. 2014)	✓	✓	–	–	–	–
45	MMRQ (Mu et al. 2018)	–	–	–	✓	–	–
46	PCTR (Wang and Li 2015)	✓	–	–	–	–	–
47	FCM (Guo et al. 2017)	–	–	–	✓	–	–

research papers in the users' libraries which are treated as 'likes', user provided tags on the papers, the authors, the title and abstract of the papers. There are two variants of the dataset namely CiteUlike-a and CiteUlike-t, made available by CTR (Wang et al. 2014). CiteUlike-a consists of 5551 users, 16980 papers and 3629 tags with a total of 204,987 user-item pairs, and CiteUlike-t having 5219 users, 25975 papers and 4222 tags with a total of 134,860 user-item likes. Also, it is noteworthy that Citeulike-t is much more sparse (99.90%) than Citeulike-a (99.78%). If we look at column 4 of table then we can observe that 8 out of 47 surveyed models have used CiteUlike datasets.

ACL Anthology Network: The ACL Anthology Network (AAN)⁵ dataset contains full text information of conference and journal papers in computational linguistic and natural language processing. The dataset contains a set of tuple fields such as research papers, authors, venues, terms, and citation. The dataset has two versions, old version was launched in February 2007 containing 10,921 papers, while the other version has been launched in march 2016 consisting of 22,878. The latest version covers the PDFs, OCR'd text, logical structure of papers along with text, venues, authors and citations. In explored studies, we can observe that most of existing graph-based approaches have adopted ACL Anthology dataset to evaluate their results because it provides information about diverse information sources and entities.

DBLP: The DBLP dataset⁶ consists of bibliography data in computer science and related fields. The original dataset is extracted from DBLP,⁷ comprised of eleven versions. The latest version i.e., DBLP-Citation-network V11 contains 4,107,340 articles and 36,624,464 citations. The dataset covers title, abstract, author, venue, year of publication and citation relations.

Aminer: Aminer dataset⁸ covers information related to research papers, authors and citation relations. The citation data is extracted from DBLP, ACM, and other sources. The first version contains 629,814 papers and 632,752 citations. Second version of the dataset contains 1,397,240 papers and 3,021,489 citations. Each paper is associated with abstract, authors, year, venue, and title.

⁵ <https://acl-arc.comp.nus.edu.sg/>

⁶ <https://www.aminer.cn/citation>

⁷ <https://dblp.uni-trier.de/>

⁸ <https://www.aminer.cn/data>

Self-Collected: Various recommendation models have created their own datasets by crawling data from different digital libraries such as ACM, Springer, IEEE and PubMed etc.

The results presented in Table 3 reveal that, most researchers prefer to use their own datasets (i.e., 15 models exploited self-created datasets) from different online repositories and digital libraries such as ACM, IEEE, and PubMed. Anthology is the most commonly used dataset which is reported 10 times (Kobayashi et al. 2018; Meng et al. 2013; Pan et al. 2015; Cai et al. 2019; Yang et al. 2019, 2018; Cai et al. 2018; Dai et al. 2019; Mu et al. 2018; Dai et al. 2018), as the dataset provides access to more rich information about different objects including papers, authors, venues, citation relations and textual content. CiteSeer dataset which provides full access to the content of articles and citation relations holds second position as 8 models have chosen it for conducting their experiments. The DBLP and Aminer datasets collected from DBLP stand at third position. Moreover, we noticed that most of the recent models (Cai et al. 2019; Yang et al. 2019; Ganguly and Pudi 2017; Yang et al. 2018; Cai et al. 2018; Dai et al. 2019; Mu et al. 2018; Yang et al. 2019; Dai et al. 2018) exploiting network embedding techniques employed DBLP and ACL Anthology as these datasets provide rich information including: papers content, citation relations, information about authors and venues which best suit such models (Table 4).

5.2 Evaluation metrics

In this section, we discuss different evaluation metrics used in surveyed models.

Recall: Recall is used to measure the percentage of original relevant citations which appear in top- N recommendation list. Citation recommendation models with higher recall in lower top- N , results in better recommendation results. This metric is defined as:

$$recall = \frac{\sum relevant\ paper \cap \sum retrieved\ paper}{\sum relevant\ paper} \quad (15)$$

Precision-MAP: Precision on the other hand measure the quality of citation recommendation system by checking whether the most relevant papers ranked higher in top N list. This metric is defined bellow:

$$Precision = \frac{\sum relevant\ paper \cap \sum retrieved\ paper}{\sum retrieved\ paper} \quad (16)$$

On the other hand, MAP is derived from average precision. The average precision for a query manuscript can be computed by taking mean of the precisions obtained after each relevant papers (here relevant means the ground truth positives (GTP) papers of the submitted manuscript) paper is retrieved as:

$$AP@k = \frac{1}{GTP} \sum_{i=1}^k \frac{TP_{seen}}{i} \quad (17)$$

After computing AP, the corresponding quantity is averaged over all queries to compute MAP as follows:

Table 4 Evaluation methods

Models	Precision- MAP	F-measure	MAE-RMSE	Coverage	nDCG	MRR	Recall	User study	Other
1 CARE (Xia et al. 2016)	✓	✓	-	-	-	-	✓	-	-
2 BRS (Roy 2017)	✓	-	-	-	-	-	-	-	-
3 MAUT (Manouselis and Verbert 2013)	-	-	✓	✓	-	-	-	-	-
4 BG-IterWR (Tian and Jing 2013)	-	-	-	-	-	-	✓	-	-
5 APTW (Meng et al. 2013)	✓	-	-	-	-	-	✓	-	-
6 DiSCem (Chakraborty et al. 2015)	✓	-	-	-	-	-	✓	-	-
7 CRS (Caragea et al. 2013)	✓	-	-	-	-	✓	✓	-	-
8 APR (Lee et al. 2013)	-	-	-	-	-	-	-	-	-
9 ISNR (Alotaibi and Vassileva 2018)	-	-	-	-	-	-	-	-	✓
10 CAT (Zhang et al. 2014)	-	-	✓	✓	-	-	✓	-	-
11 MSCN (Son and Kim 2017)	-	-	-	-	✓	✓	-	-	-
12 AKR (Sesagiri Raamkumar et al. 2017)	-	-	-	-	-	-	-	✓	✓
13 SPR (Sugiyama and Kan 2013)	-	-	-	-	✓	✓	-	-	-
14 LDA-TM (Amami et al. 2016)	-	-	-	-	-	-	✓	-	-
15 tr-MF (Li et al. 2013)	-	-	-	-	-	-	✓	-	-
16 DIGTOBI (Kim et al. 2013)	✓	-	-	-	-	-	✓	-	✓
17 SciRecSys (Le Anh et al. 2014)	-	-	-	-	-	-	-	-	✓
18 BKS (El-Arini and Guestrin 2011)	-	-	-	-	-	-	-	✓	-
19 NPM (Huang et al. 2015)	✓	-	-	-	✓	✓	-	-	-
20 SARSP (Xia et al. 2014)	✓	-	-	-	-	-	✓	-	-
21 CGRRK (Blank et al. 2013)	✓	-	-	-	-	-	✓	-	-
22 NCN (Ebesu and Fang 2017)	✓	-	-	-	✓	✓	✓	-	-
23 ClusCite (Ren et al. 2014)	✓	-	-	-	-	✓	✓	-	-
24 EFrec (West et al. 2016)	-	-	-	✓	-	-	-	-	-
25 FeRoSA (Chakraborty et al. 2016)	✓	-	-	-	-	-	-	-	✓
26 SSAR (Sun et al. 2014)	✓	-	-	-	-	✓	-	-	-

Table 4 (continued)

Models	Precision- MAP	F-measure	MAE-RMSE	Coverage	nDCG	MRR	Recall	User study	Other
27 Science Concierge (Achakulvisut et al. 2016)	-	-	-	-	-	-	-	-	✓
28 CTR (Wang et al. 2014)	-	-	-	-	-	-	✓	-	-
29 Paper2vec (Ganguly and Pudi 2017)	-	✓	-	-	-	-	-	-	-
30 GRU-ML (Bansal et al. 2016)	-	-	-	-	-	-	✓	-	-
31 CPRec (Li et al. 2018)	✓	-	✓	-	-	-	✓	-	-
32 PCCR (Yang et al. 2018)	✓	-	-	-	-	✓	✓	✓	-
33 VOPRec (Kong et al. 2019)	✓	-	✓	-	✓	-	✓	-	-
34 BNR (Cai et al. 2019)	✓	-	-	-	-	✓	-	-	-
35 TMR (Cai et al. 2018)	✓	-	-	-	-	✓	✓	-	-
36 CNRN (Waheed et al. 2019)	✓	-	-	-	✓	✓	-	-	-
37 TMALCCite (Dai et al. 2018)	✓	-	-	-	-	✓	✓	-	-
38 NREP (Yang et al. 2019)	✓	-	-	-	✓	-	✓	-	-
39 APR (Pan et al. 2015)	-	✓	-	-	✓	✓	-	-	-
40 DRDF-CR (Kobayashi et al. 2018)	-	-	-	-	✓	-	-	-	-
41 SAR (Gupta and Varma 2017)	✓	-	-	-	✓	-	-	-	-
42 TopicCite (Dai et al. 2019)	✓	-	-	-	-	✓	✓	-	-
43 HRLHG (Jiang et al. 2018)	✓	-	-	-	✓	✓	-	-	-
44 RefSeer (Huang et al. 2014)	-	-	-	-	-	✓	-	-	-
45 MMRQ (Mu et al. 2018)	-	-	-	-	✓	-	✓	-	-
46 PCTR (Wang and Li 2015)	-	-	-	-	-	-	✓	-	-
47 FCM (Guo et al. 2017)	-	-	-	-	✓	-	✓	-	-

$$MAP = \frac{1}{Q} \sum_{k=1}^Q AP(k) \quad (18)$$

where Q represents the whole number of queries.

F-measure: The f-measure metric is used to balance the trade-off between precision and recall by assigning them equal weights. F-measure can be calculated by taking harmonic mean of precision and recall as defined below:

$$F - measure = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (19)$$

Mean average error (MAE): measures the the average absolute divergence from the predicted rating \hat{y}_j and the true rating y_j as given in the following equation.

$$MAE = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j) \quad (20)$$

Root mean square error (RMSE): measure the square root of the average squared divergence from the predicted rating \hat{r}_{ij} and the true rating r_{ij} . In contrast to MAE, RMSE penalizes larger errors, moreover it is more appropriate for prediction accuracy as it considers both positive and negative ratings in measuring accuracy (Schafer et al. 2007).

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (21)$$

where j represents the number of the document, y_j is the true rating value, $y_j - \hat{y}_j$ is the predicted rating value and n denotes the number of ratings in the testing set.

nDCG: nDCG (Kobayashi et al. 2018) stands for Normalized Discounted Cumulative Gain, it analyzes the rank in the top N of the true relevant paper in the list of recommended citations by the model. It evaluates the significance of a recommender system based on the graded relevance of the recommended papers as defined:

$$\begin{aligned} nDCG_g &= \frac{DCG_g}{IDCG_g} \\ DCG_g &= \sum_{j=1}^G \frac{2^{g(j)} - 1}{\log(1 + j)} \\ IDCG_g &= \sum_{j=1}^G \frac{1}{\log(1 + j)} \end{aligned} \quad (22)$$

where $nDCG_g$ represents the normalized gain accumulated at a particular position g , while G denotes the list of relevant papers (ordered by their relevance) in the corpus up to position g . Discounted Cumulative Gain (DCG) is the weighted sum of the degree of relevancy of the ranked citations, it emphasizes highly relevant papers to appear early in the result list. Whereas, $IDCG_g$ represents the DCG of Ideal ordering which normalizes DCG values.

Finally, it is critical to mention that researchers do not have an integrated protocol using which the performance of all the models can be evaluated globally. Therefore, we

can not clearly declare that a specific evaluation metric can be used as standard to compare and evaluate the experimental results of all the models.

6 Conclusion and implications

These days, there are millions of research articles on the Web. To find relevant articles in accordance with researchers preferences and research needs is quite a daunting task. To address this problem, researchers have proposed different citation recommendation algorithms, using which researchers are assisted by generating more accurate and useful recommendations. In this regard, very few surveys exist exploring the area of citation recommendations, limited in deriving implications and in depth classification of these algorithms. In this survey, we explored 47 citation recommendation models by considering seven prominent aspects including: platform, data factors, data representation methods, methodologies used, recommendation types, problems addressed, and personalization. Moreover, we have proposed a hybrid k-partite graph-based taxonomy exploring the relationships among entities and corresponding citation recommendation algorithms. Additionally, the survey presented a comparison of the datasets and evaluation methods used for evaluating the performance of experimental results in each algorithm. Finally, the paper covered issues related to citation recommendation models and presented solutions in light of reviewed algorithms. After surveying the area of citation recommendation, we have discovered some useful implications and future directions given as follows:

- The survey highlights that most of the citation recommendation models i.e., 42 out of 47 are offline systems, they are not part of real life systems. It reveals the fact that few research studies have translated their research works into practice.
- The data factors/features column in Table 1 reveals that 26 out of 47 models have used tags, while 22 have maintained user models. On the other hand, models exploiting social network analysis and time dimension are still limited. The importance of time dimension is indispensable, as researchers interests changes over time and therefore exploring time may exploit patterns in researchers' behaviors. Additionally, social network analysis can help in resolving cold start as well as sparsity issues. Henceforth, we hope that researchers will enrich their recommendation results by incorporating time dimension and social network analysis in their works.
- The literature reveal that, recommendation systems employing neural networks and graph-embedding techniques are emerging, as 15 systems have adopted these methods in last three years. Significance of such approaches lies in exploiting the semantic representation of research papers and associated auxiliary information sources namely authors, venues, tags, and papers. However, existing models lack in considering user interests and require a reasonable training time to generate useful results.
- Most of the graph-based recommendation models i.e., 16 systems have employed random walk and its variants i.e., random walk with restarts, and reinforced random walk to generate recommendations. Though graph-based approaches have been used by a good number of models, however most of these approaches consider recommendation as a link prediction task and therefore face issues such as lack of personalization and over-weighting old papers in the graph. To overcome these problems, recent studies have exploited heterogeneous information networks to generate more adequate recommendations. Additionally, it is noteworthy that no research work has exploited knowl-

edge graph embedding techniques to generate citation recommendations. In future, researchers can employ different KBs (Knowledge bases) embedding methods such as TransD, TransE, and ComplEx, using which KB triples can be transformed into a vector space (Abedini et al. 2019a; Wang et al. 2017) and more semantic-aware recommendations can be generated.

- The survey reveals that the algorithms employing auxiliary information sources such as papers textual information, papers citation network, users social network, and other contextual information can generate more justifiable recommendations, as such techniques exploits rich information and fertilize their knowledge concerning researchers preferences.
- Finally, we can observe that a great proportion of surveyed models i.e., 42 out of 47 have generated personalized recommendation results, because researchers tend to prefer those articles which meet their research preferences. It is noticeable that only 4 systems have support for producing group recommendations while 3 models do not consider knowledge about users and hence generate non-personalized recommendations.

In conclusion, this survey will help researchers in exploring the area of citation recommendation and providing guidelines for future research.

References

- Abedini F, Keyvanpour MR, Menhaj MB (2019a) Epci: an embedding method for post-correction of inconsistency in the RDF knowledge bases. *IETE J Res* 66:1–13
- Abedini F, Keyvanpour MR, Menhaj MB (2019b) Neural tensor network training using meta-heuristic algorithms for RDF knowledge bases completion. *Appl Artif Intell* 33:656–667
- Abedini F, Menhaj MB, Keyvanpour MR (2019c) An mlp-based representation of neural tensor networks for the rdf data models. *Neural Comput Appl* 31:1135–1144
- Abro WA, Qi G, Gao H, Khan MA, Ali Z (2019) Multi-turn intent determination for goal-oriented dialogue systems. In: 2019 international joint conference on neural networks (IJCNN). IEEE, pp 1–8
- Achakulvisut T, Acuna DE, Ruangrong T, Kording K (2016) Science concierge: a fast content-based recommendation system for scientific publications. *PLoS ONE* 11:1–11
- Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans Knowl Data Eng* 17:734–749
- Ali Z, Khusro S, Ullah I (2016) A hybrid book recommender system based on table of contents (toc) and association rule mining. In: Proceedings of the 10th international conference on informatics and systems. ACM, New York, NY, USA, pp 68–74
- Alotaibi S, Vassileva J (2018) Implicit social networks for social recommendation of scholarly papers. In: Moshirpour M, Far B, Alhajj R (eds) Highlighting the importance of big data management and analysis for various applications. Springer, Cham, pp 79–92
- Amami M, Pasi G, Stella F, Faiz R (2016) An LDA-based approach to scientific paper recommendation. In: Natural language processing and information systems: 21st international conference on applications of natural language to information systems, NLDB 2016, Salford, UK, 22–24 June 2016, proceedings. Springer, pp 200–210
- Ayala-Gómez F, Daróczy B, Benczúr A, Mathioudakis M, Gionis A (2018) Global citation recommendation using knowledge graphs. *J Intell Fuzzy Syst* 34:3089–3100
- Bai X, Wang M, Lee I, Yang Z, Kong X, Xia F (2019) Scientific paper recommendation: a survey. *IEEE Access* 7:9324–9339
- Bansal T, Belanger D, McCallum A (2016) Ask the gru: multi-task learning for deep text recommendations. In: Proceedings of the 10th ACM conference on recommender systems, New York, NY, USA, pp 107–114
- Batmaz Z, Yurekli A, Bilge A, Kaleli C (2019) A review on deep learning for recommender systems: challenges and remedies. *Artif Intell Rev* 52:1–37
- Beel J, Gipp B, Langer S, Breitinger C (2016) Research-paper recommender systems: a literature survey. *Int J Digit Libr* 17:305–338

- Bhagavatula C, Feldman S, Power R, Ammar W (2018) Content-based citation recommendation. In: Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: human language technologies, Volume 1 (long papers), New Orleans, Louisiana, pp 238–251
- Blank I, Rokach L, Shani G (2013) Leveraging the citation graph to recommend keywords. In: Proceedings of the 7th ACM conference on recommender systems (RecSys), pp 359–362
- Bobadilla J, Ortega F, Hernando A, Gutiérrez A (2013) Recommender systems survey. *Knowl Based Syst* 46:109–132
- Breese JS, Heckerman D, Kadie C (1998) Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the fourteenth conference on uncertainty in artificial intelligence, pp 43–52
- Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. *Comput Netw ISDN Syst* 30:107–117
- Burke R (2002) Hybrid recommender systems: survey and experiments. *User Model User Adapt Interact* 12:331–370
- Cai X, Han J, Li W, Zhang R, Pan S, Yang L (2018) A three-layered mutually reinforced model for personalized citation recommendation. *IEEE Trans Neural Net Learn Syst* 29:6026–6037
- Cai X, Zheng Y, Yang L, Dai T, Guo L (2019) Bibliographic network representation based personalized citation recommendation. *IEEE Access* 7:457–467
- Caragea C, Silvescu A, Mitra P, Giles CL (2013) Can't see the forest for the trees? A citation recommendation system. In: Proceedings of the 13th ACM/IEEE-CS joint conference on digital libraries (JCDL), pp 111–114
- Chakraborty T, Modani N, Narayanam R, Nagar S (2015) Discern: a diversified citation recommendation system for scientific queries. In: 31st IEEE international conference on data engineering, (ICDE) 2015, Seoul, South Korea, 13–17 April 2015. Seoul, South Korea, pp 555–566
- Chakraborty T, Krishna A, Singh M, Ganguly N, Goyal P, Mukherjee A (2016) Ferosa: a faceted recommendation system for scientific articles. In: Proceedings, part II, of the 20th Pacific-Asia conference on advances in knowledge discovery and data mining—Volume 9652 (PAKDD), pp 528–541
- Christoforidis G, Kefalas P, Papadopoulos A, Manolopoulos Y (2018) Recommendation of points-of-interest using graph embeddings. In: 5th IEEE international conference on data science and advanced analytics, DSAA 2018, Turin, Italy, 1–3 Oct 2018, pp 31–40
- Dai T, Zhu L, Cai X, Pan S, Yuan S (2018) Explore semantic topics and author communities for citation recommendation in bipartite bibliographic network. *J Ambient Intell Humaniz Comput* 9:957–975
- Dai T, Zhu L, Wang Y, Zhang H, Cai X, Zheng Y (2019) Joint model feature regression and topic learning for global citation recommendation. *IEEE Access* 7:1706–1720
- Ebesu T, Fang Y (2017) Neural citation network for context-aware citation recommendation. In: Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval, pp 1093–1096
- El-Arini K, Guestrin C (2011) Beyond keyword search: discovering relevant scientific literature. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining (KDD), pp 439–447
- Ganguly S, Pudi V (2017) Paper2vec: combining graph and text information for scientific paper representation. In: Jose JM, Altingovde IS, Song D, Albakour D, Watt S, Tait J (eds) *Advances in information retrieval*. Springer, Aberdeen, pp 383–395
- Goyal P, Ferrara E (2018) Graph embedding techniques, applications, and performance: a survey. *Knowl Based Syst* 151:78–94
- Grover A, Leskovec J (2016) Node2vec: scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 855–864
- Guo L, Cai X, Hao F, Mu D, Fang C, Yang L (2017) Exploiting fine-grained co-authorship for personalized citation recommendation. *IEEE Access* 5:12714–12725
- Gupta S, Varma V (2017) Scientific article recommendation by using distributed representations of text and graph. In: Proceedings of the 26th international conference on world wide web companion, Republic and Canton of Geneva, Switzerland, pp 1267–1268
- Habib R, Afzal MT (2019) Sections-based bibliographic coupling for research paper recommendation. *Scientometrics* 119:643–656
- Hofmann T (2017) Probabilistic latent semantic indexing. In: *ACM SIGIR forum*, pp 211–218
- Huang Z, Mamoulis N (2017) Heterogeneous information network embedding for meta path based proximity. *arXiv preprint arXiv: 1701.05291*

- Huang W, Wu Z, Mitra P, Giles CL (2014) Refseer: a citation recommendation system. In: IEEE/ACM joint conference on digital libraries, pp 371–374
- Huang W, Wu Z, Liang C, Mitra P, Giles CL (2015) A neural probabilistic model for context based citation recommendation. In: Proceedings of the twenty-ninth AAAI conference on artificial intelligence, pp 2404–2410
- Jiang Z, Yin Y, Gao L, Lu Y, Liu X (2018) Cross-language citation recommendation via hierarchical representation learning on heterogeneous graph. In: The 41st international ACM SIGIR conference on research and development in information retrieval, New York, NY, USA, pp 635–644
- Kefalas P, Manolopoulos Y (2017) A time-aware spatio-textual recommender system. *Expert Syst Appl* 78:396–406
- Kefalas P, Symeonidis P, Manolopoulos Y (2015) A graph-based taxonomy of recommendation algorithms and systems in lbsns. *IEEE Trans Knowl Data Eng* 28:604–622
- Kefalas P, Symeonidis P, Manolopoulos Y (2018) Recommendations based on a heterogeneous spatio-temporal social network. *World Wide Web* 21:345–371
- Khusro S, Ali Z, Ullah I (2016) Recommender systems: issues, challenges, and research opportunities. In: Information science and applications (ICISA) 2016. Springer, pp 1179–1189
- Kim Y (2014) Convolutional neural networks for sentence classification. arXiv preprint [arXiv:1408.5882](https://arxiv.org/abs/1408.5882)
- Kim Y, Park Y, Shim K (2013) Digtobi: a recommendation system for digg articles using probabilistic modeling. In: Proceedings of the 22nd international conference on World Wide Web (WWW), pp 691–702
- Kobayashi Y, Shimbo M, Matsumoto Y (2018) Citation recommendation using distributed representation of discourse facets in scientific articles. In: Proceedings of the 18th ACM/IEEE on joint conference on digital libraries, New York, NY, USA, pp 243–251
- Kong X, Mao M, Wang W, Liu J, Xu B (2019) Voprec: vector representation learning of papers with text information and structural identity for recommendation. *IEEE Trans Emerg Top Comput* 99:1–1
- Koren Y, Bell R, Volinsky C (2009) Matrix factorization techniques for recommender systems. *Computer* 42:30–37
- Kotkov D, Wang S, Veijalainen J (2016) A survey of serendipity in recommender systems. *Knowl Based Syst* 111:180–192
- Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: Xing EP, Jebara T (eds) Proceedings of the 31st international conference on machine learning, Beijing, China, pp 1188–1196
- Le Anh V, Hoang Hai V, Tran HN, Jung JJ (2014) Scirecsys: a recommendation system for scientific publication by discovering keyword relationships. In: Computational collective intelligence. Technologies and applications: 6th international conference, ICCCI 2014, Seoul, Korea, 24–26 Sept 2014. Proceedings, Springer, pp 72–82
- Lee J, Lee K, Kim JG (2013) Personalized academic research paper recommendation system. CoRR [arXiv:1304.5457](https://arxiv.org/abs/1304.5457)
- Li Y, Yang M, Zhang ZM (2013) Scientific articles recommendation. In: Proceedings of the 22nd ACM international conference on information and knowledge management (CIKM), pp 1147–1156
- Li S, Brusilovsky P, Su S, Cheng X (2018) Conference paper recommendation for academic conferences. *IEEE Access* 6:17153–17164
- Lipton ZC, Berkowitz J, Elkan C (2015) A critical review of recurrent neural networks for sequence learning. arXiv preprint [arXiv:1506.00019](https://arxiv.org/abs/1506.00019)
- Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE (2017) A survey of deep neural network architectures and their applications. *Neurocomputing* 234:11–26
- Lops P, de Gemmis M, Semeraro G (2011) Content-based recommender systems: state of the art and trends. , Boston MA, pp 73–105
- Manouselis N, Verbert K (2013) Layered evaluation of multi-criteria collaborative filtering for scientific paper recommendation. In: International conference on computational science, pp 1189 – 1197
- 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 knowledge management (CIKM), pp 1509–1512
- Mohamed AR, Dahl GE, Hinton G (2011) Acoustic modeling using deep belief networks. *IEEE Trans Audio Speech Lang Process* 20:14–22
- Montaner M, López B, de la Rosa JL (2003) A taxonomy of recommender agents on the internet. *Artif Intell Rev* 19:285–330
- Mu D, Guo L, Cai X, Hao F (2018) Query-focused personalized citation recommendation with mutually reinforced ranking. *IEEE Access* 6:3107–3119
- Najafabadi MK, Mahrin MN (2016) A systematic literature review on the state of research and practice of collaborative filtering technique and implicit feedback. *Artif Intell Rev* 45:167–201

- Page L, Brin S, Motwani R, Winograd T (1999) The PageRank citation ranking: bringing order to the web. Technical report, Stanford InfoLab
- Pan L, Dai X, Huang S, Chen J (2015) Academic paper recommendation based on heterogeneous graph. In: Sun M, Liu Z, Zhang M, Liu Y (eds) Chinese computational linguistics and natural language processing based on naturally annotated big data, pp 381–392
- Pazzani MJ, Billsus D (2007) Content-based recommendation systems. In: Brusilovsky P, Kobsa A, Nejdl W (eds) The adaptive web. Springer, Berlin, pp 325–341
- Perozzi B, Al-Rfou R, Skiena S (2014) Deepwalk: online learning of social representations. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining, New York, NY, USA, pp 701–710
- Prendergast LA, Staudte RG (2014) Better than you think: interval estimators of the difference of binomial proportions. *J Stat Plan Inference* 148:38–48
- Rafailidis D, Kefalas P, Manolopoulos Y (2017) Preference dynamics with multimodal user-item interactions in social media recommendation. *Expert Syst Appl* 74:11–18
- Ren X, Liu J, Yu X, Khandelwal U, Gu Q, Wang L, Han J (2014) Cluscite: effective citation recommendation by information network-based clustering. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining, pp 821–830
- Ribeiro LF, Saverese PH, Figueiredo DR (2017) struc2vec: learning node representations from structural identity. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp 385–394
- Roy D (2017) An improved test collection and baselines for bibliographic citation recommendation. In: Proceedings of the 2017 ACM on conference on information and knowledge management (CIKM), pp 2271–2274
- Rui X, Wunsch D (2005) Survey of clustering algorithms. *IEEE Trans Neural Netw* 16:645–678
- Sarwar BM, Karypis G, Konstan JA, Riedl J et al (2001) Item-based collaborative filtering recommendation algorithms. *WWW* 1:285–295
- Schafer JB, Frankowski D, Herlocker J, Sen S (2007) Collaborative filtering recommender systems. Springer, Berlin, pp 291–324
- Sesagiri Raamkumar A, Foo S, Pang N (2017) Using author-specified keywords in building an initial reading list of research papers in scientific paper retrieval and recommender systems. *Inf Process Manag Int J* 53:577–594
- Shi C, Hu B, Zhao WX, Yu PS (2019) Heterogeneous information network embedding for recommendation. *IEEE Trans Knowl Data Eng* 31:357–370
- Son J, Kim SB (2017) Academic paper recommender system using multilevel simultaneous citation networks. *Decis Support Syst* 105:24–33
- Sugiyama K, Kan MY (2013) Exploiting potential citation papers in scholarly paper recommendation. In: Proceedings of the 13th ACM/IEEE-CS joint conference on digital libraries (JCDL), pp 153–162
- Sun J, Ma J, Liu Z, Miao Y (2014) Leveraging content and connections for scientific article recommendation in social computing contexts. *Comput J* 57:1331–1342
- Symeonidis P, Tiakas E, Manolopoulos Y (2011) Product recommendation and rating prediction based on multi-modal social networks. In: Proceedings of the fifth ACM conference on recommender systems, pp 61–68
- Tang J, Qu M, Wang M, Zhang M, Yan J, Mei Q (2015) Line: large-scale information network embedding. In: Proceedings of the 24th international conference on World Wide Web, Republic and Canton of Geneva, Switzerland, pp 1067–1077
- Tian G, Jing L (2013) Recommending scientific articles using bi-relational graph-based iterative rwr. In: Proceedings of the 7th ACM conference on recommender systems (RecSys), pp 399–402
- Waheed W, Imran M, Raza B, Malik AK, Khattak HA (2019) A hybrid approach toward research paper recommendation using centrality measures and author ranking. *IEEE Access* 7:33145–33158
- Wang C, Blei DM (2011) Collaborative topic modeling for recommending scientific articles. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining (KDD), pp 448–456
- Wang H, Li W (2015) Relational collaborative topic regression for recommender systems. *IEEE Trans Knowl Data Eng* 27:1343–1355
- Wang Q, Mao Z, Wang B, Guo L (2017) Knowledge graph embedding: a survey of approaches and applications. *IEEE Trans Knowl Data Eng* 29:2724–2743
- Wang D, Liang Y, Xu D, Feng X, Guan R (2018) A content-based recommender system for computer science publications. *Knowl Based Syst* 157:1–9
- Wesley-Smith I, Bergstrom CT, West JD (2016) Static ranking of scholarly papers using article-level eigenfactor (alef). arXiv preprint [arXiv:1606.08534](https://arxiv.org/abs/1606.08534)

- 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* 2:113–123
- Xia F, Asabere NY, Liu H, Deonauth N, Li F (2014) Folksonomy based socially-aware recommendation of scholarly papers for conference participants. In: *Proceedings of the 23rd international conference on World Wide Web (WWW)*, pp 781–786
- Xia F, Liu H, Lee I, Cao L (2016) Scientific article recommendation: exploiting common author relations and historical preferences. *IEEE Trans Big Data* 2:101–112
- Xu K, Qi G, Huang J, Wu T, Fu X (2018) Detecting bursts in sentiment-aware topics from social media. *Knowl Based Syst* 141:44–54
- Yang C, Liu Z, Zhao D, Sun M, Chang E (2015) Network representation learning with rich text information. In: *Twenty-fourth international joint conference on artificial intelligence*
- Yang L, Zheng Y, Cai X, Dai H, Mu D, Guo L, Dai T (2018) A lstm based model for personalized context-aware citation recommendation. *IEEE Access* 6:59618–59627
- Yang L, Zhang Z, Cai X, Guo L (2019) Citation recommendation as edge prediction in heterogeneous bibliographic network: a network representation approach. *IEEE Access* 7:23232–23239
- Zhang C, Zhao X, Wang K, Sun J (2014) Content + attributes: a latent factor model for recommending scientific papers in heterogeneous academic networks. In: *Proceedings of the advances in information retrieval: 36th European conference on IR research (ECIR)*. Springer, Amsterdam, pp 39–50

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.