Citation Recommendation with a Content-Sensitive DeepWalk based Approach

Lantian Guo[1,2], Xiaoyan Cai[3], Haohua Qin[2], Yangming Guo[1], Fei Li[4], Gang Tian[5]
[1]School of Computer Science, Northwestern Polytechnical University, Xi'an, China 710072
[2]School of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao, China 266044
[3]School of Automation, Northwestern Polytechnical University, Xi'an, China 710072
[4]School of Information Engineering, Northwest A&F University, Xi'an, China 712100
[5]School of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China 266590

***Abstract*—Systems for recommending scientific papers mainly help researchers to find a list of references that related to the researcher's interest effectively and automatically. Many state-of-the-art technique have been used for recommendation system, however, the traditional approaches has the issues of data scarcities and cold start, and existing recommended approaches with network representation only focus on one aspect of node information and cannot leverage content information. In this paper, we proposed a Citation Recommendation method with a Content-Aware bibliographic network representation, called CR-CA, whose recommended process contains two levels: (1) At the node content level, the proposed approach calculates similarities between the target and candidate papers, selecting an initial seed set of papers; (2) At the citation network structure level, this approach exploits citation relationship between papers to study latent representation of the scientific papers based on a deep natural language method—-DeepWalk. The proposed approach was tested on the AAN dataset demonstrate that this approach outperforms baseline algorithms, in the true positive rate ($Recall$) and normalized discounted cumulative gain ($NDCG$).**

***Index Terms*—Citation recommendation, DeepWalk, Network structure, Content information**

I. Introduction

With the rapid development of information technology, the scientific publications have exponentially expansion, researchers are faced with the challenge is to search relevant scientific papers which satisfy their citation requirements in the massive data. In order to weaken the problems that this situation bring to researchers, the recommendation of scientific paper, which aims to suggest a small number of relevant publications that can be used as high-quality references to satisfy such citation requirements [15], have recently attracted increased attention [7].

The traditional work on scientific paper recommendation explored the use of Content-based filtering (CBF) and collaborative filtering (CF) techniques. Due to employ the content information, CBF has the problems of traditional information retrieval, such as semantic ambiguity [17]. The CF algorithm has been extensively developed in e-commerce but cannot be effectively applied for paper recommendation [2]. Although CF, as a typical recommendation technology, has its considerable application, it is usually limited by sparsity and scalability [22].

In a paper set, there are variety of different types of information. Neglecting the existence of some information will result in only one-sided correlations between the paper in the recommended set and the queries. Obviously, this is unreasonable. In recent years, the rise of heterogeneous information networks has led to the further development of graph-based recommendation [21]. The heterogeneous graph model can be constructed by utilizing multiple types of links from the paper set [24]. It can utilizes various relations among heterogeneous object, such as paper citation, author relationship and paper's content and so on [4, 9]. However, the traditional graph-based approaches have the problem of uncontrollable dimensions and sparsity. To address this problem, network representation [1, 16] encodes each node in a low-dimensional space while preserving the neighborhood relationship between node. It is latent features of the nodes that capture neighborhood similarity and community membership.

In this paper, we proposed a scientific paper recommendation approach in bibliographic network based on deep learning. In our approach, instead of having only citation information, we exploit network information from two parties: citation information and paper content. At the paper content level, we use the bag-of-words model generate paper vector according to paper content, Utilizing the similarity of between the papers to select the assumed citation of the target paper. At the citation information level, we integrated the assumed reference relationships and the candidate papers citations based on DeepWalk to explore the network topology, which can jointly learn good interpretable lower dimension spaces for paper nodes. Experimental results show that this approach significantly outperforms other baseline methods.

The main contributions of this paper are summarized as follows:

- When generating the paper network representation, not only the citation relationship between the papers is considered, but also the paper content information.
- Using text similarity to weaken some irrelevant papers because of the high number of citations, resulting in the

2375-9259/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDMW.2019.00082

representation of the target paper is close to the irrelevant papers learned by DeepWalk.

- A series of experiments conducted on AAN datasets are carried out to validate the effectiveness of the proposed approach.

The rest of this paper is arranged as follows. In Section II, we briefly review the related work on paper recommendation. Section II presents the details of our citation recommendation method. Section IV gives the experimental results and analysis. Section V gives the conclusion of this paper.

II. Related Work

The recommendation of scientific papers has been studied for decades in an effort to exploit various information in paper sets, such as citation relationships, author relationships, and paper contents. There are many methods to represent paper content, traditional such as bag-of-words (BOW) model [23], the term frequency-inverse document frequency (TF-IDF) [25] and Latent Dirichlet allocation (LDA) [19]. BOW model is a common way to obtain the representation of text content in the field of information retrieval, but this model does not consider the factor of word frequency. TF-IDF uses word frequency to extract keywords, and calculate the weight of the keywords in documents, represent the text as a sparse vector. However, the above two methods do not consider the semantic information. LDA can find the semantic content of papers for improving the chances of correct matches. Mikolov et al [10] proposed a skip-gram model architectures for computing continuous vector representations of words from large data sets. And this method used the target word's contextual information. Mikolov's paper provides a new idea for the representation of text content and graph node.

Node connections play an important role in graph-based approaches. Many graph-based approaches consider paper recommendation as a citation link predication task [20] and perform the recommendation process based on the random walk properties. Ni et al. [5] used a combination of path-constrained random walks for relational retrieval. They defined the proximity as a weighted combination of simple 'path experts'. Different papers have different contents, authors, and belong to different venues, so papers have different patterns of citation behaviour. Ren et al. [15] used citation, venue, author, and term of papers to construct a heterogeneous network and proposed a cluster-based citation framework for recommendation, called ClusCite. Pan et al. [11] built a heterogeneous graph to represent both citation and content information within papers. The form of network representation constructed by entity relationships is susceptible to problems of data sparsity [26].

The traditional graph-based representation uses a storage structure of a two-dimensional array (adjacency matrix) to indicate whether there is a connected edge between two nodes, existence is 1, otherwise 0. However, due to the long-tailed distribution, most of the nodes are not related, so the adjacency matrix is very sparse and not conducive to storage and calculation. With the success in many classification and link prediction, network representation has drawn a lot of attention to the researchers. Network Representation Learning (NRL, Graph Embedding Method (GEM)) [3] uses low-dimensional, dense, real-valued vectors to represent nodes in a network to facilitate computational storage without the need to manually extract features, you can project heterogeneous information into the same low-dimensional space to facilitate downstream calculations.

Perozzi et al. [13] proposed a NRL-DeepWalk, starting from a node in the graph uses random walk to generate sequence data similar to textual context, then using node as a 'word' training based on skip-gram to get latent representations of vertices in a network. Tang et al. [16]proposed two concepts: first-order proximity and second-order proximity, their model optimizes an objective which preserves both the local and global network structures through the two concepts above. This method suitable for large-scale date processing. Grover et al. [1] proposed a method similar to that of DeepWalk, the main innovation is to improve the strategy of random walk, define two parameters p and q, achieve a balance between Breadth-first Sampling(BFS) and Depth-first Sampling(DFS), take into account the local and macro information, and have high adaptability.

III. Proposed Method

In this paper, papers and keywords are used to build a two-layer graph model. Let $G(V,G)$ be a directed weighted graph. We set $V = V_p \cup V_w$ is the paper vertex set, $V_p = \{p_i\}\,(1 \leq i \leq n, \cdot n$ is the total number of papers. $V_w = \{w_j\}\,(1 \leq j \leq m, m$ is the number of keywords in the set of papers. $\mathrm{E} = \{E_{pp}, E_{pw}\}$ is the edge set, $E_{pp} = \{e_{ij}, p_i, p_j \in V_p\}$, $E_{pw} = \{e_{ij}, p_i \in V_p, \cdot p_j \in V_w\}$. Corresponding to the edges between papers(R_1), the edges between the papers and the keywords(R_2), respectively.

The model framework structure of the method proposed in this paper is shown in Fig. 1. The whole architecture consists of two main modules: regenerate the adjacency matrix module and network embedding module. Based on this module, the feature representation vector of the paper node can be learned. The vector includes vertex content information and network structure information so that it can be applied to citation recommendation work.

A. Regenerate the Adjacency Matrix

a) content similarity calculation

In this paper, the content of the paper includes two parts: title and abstract. All the text content associated with one paper vertex. According to aforesaid, the matrix $\mathbf{W}$ for the text content of the paper can be obtained. In the matrix $\mathbf{W}$, i-th row of matrix $\mathbf{W}$ is the content vector representation of p_i. The content similarity of the paper can be obtained as follows:

$$\text{content}\,(p_i, p_j) = \frac{\|p_i \cdot p_j\|}{\|p_i\| \cdot \|p_j\|} = \frac{\sum_{i=1}^{\text{size}} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{\text{size}} x_i^2} \cdot \sqrt{\sum_{i=1}^{\text{size}} y_i^2}} \tag{1}$$

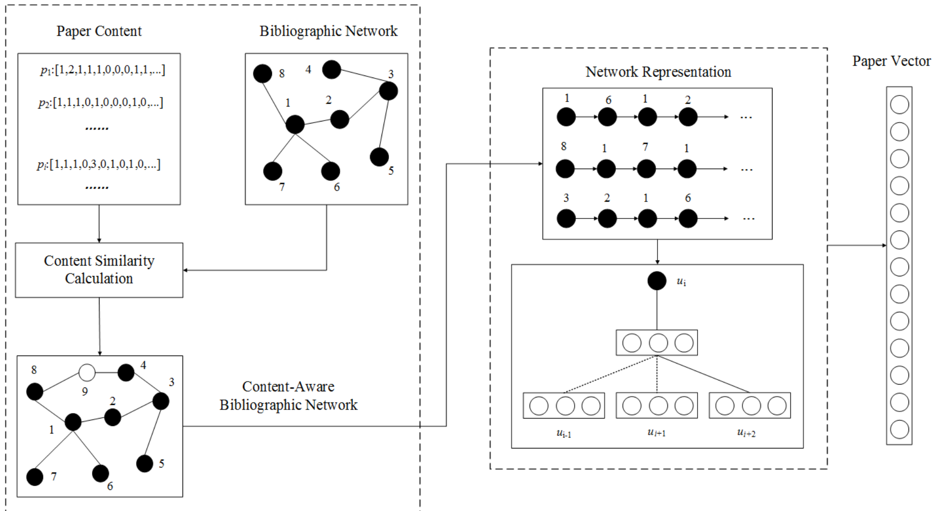

Fig. 1: Architecture of the Proposed Method

where x_i, y_i are the components of the paper vectors p_i, p_j.

b) Adjacency matrix regeneration

By calculating the similarity of the content of the paper, n papers with the largest similarity can be obtained. These n papers are called $TTSS$(Top Ten Similarity Set). However, due to the inaccuracy of similarity calculations, these n papers do not fully represent the feature of the paper. So $CTTS$(Cite Top Ten Set) and $TTCS$(Top Ten Cite Set) are defined. $CTTS$ is a set of papers that cite the paper in $TTSS$. $TTCS$ is also a set of papers that cited by the paper in $TTSS$. $TTSS$, $CTTS$ and $TTCS$ are obviously related.

Transitivity is an axiom of logic and mathematics, let R be the two elements relationship on a set of X, and x, y, $z \in X$, if the relationship $(x, y) \in R$ and $(y, z) \in R$, then get the relationship $(x, z) \in R$. $TTSS$, $CTTS$, and $TTCS$ are thus associated with the target paper. The definition of selecting the initial seed set CRS is represent as:

$$CRS = TTSS \cup CTTS \cup TTCS \tag{2}$$

According to Equation(2), the new adjacency matrix can be regained.

B. Bibliographic Network Representation

Bibliographic Network Embedding module employs the DeepWalk method [12]. DeepWalk is proposed for a node vectorization model based on Word2vec. The main idea is to use the random walk path of the construction node on the network to simulate the process of text generation, provide a sequence of nodes, and then use the Skip-gram and Hierarchical Softmax models to each of the local windows in the random walk sequence. The node pairs are probabilistically modeled to maximize the likelihood probability of the random walk sequence and use the final stochastic gradient to descend the learning parameters. Its objective function is:

$$L_s = \frac{1}{|S|} \sum_{i=1}^{|S|} \sum_{i-t \leq j \leq i+t, j \neq i} \log \mathcal{P}\left(v_j | v_i\right) \tag{3}$$

where

$$\mathcal{P}\left(v_j | v_i\right) = \frac{\exp(v_j^{'} \cdot v_i)}{\sum_{v_j^{'} \in V} \exp(v^{'} \cdot v_i)} \tag{4}$$

In our proposed method. the input for DeepWalk model is the . After the learning procedure with DeepWalk model, the vector representations for each paper node can be obtained. Then, these vector representations for paper are used to conduct the citation recommendation. For a given paper, we can choose its the candidate citaion papers by a Top-N ranking list, which is calculated by the similarities among the given paper and candidate papers.

IV. Experiments

A. Dataset

The experiment was conducted on the AAN 2013 Release dataset [14]. This dataset contains the complete collection of papers included in many ACL venues. The AAN 2013 Release[1] contains 21,236 papers published from 1965 to 2013. It provides information such as paper content, citation, and year of publication, author, journal, and title. The pretreatment of each paper in the AAN data set consisted of: (a) extracting the abstract and title, (b) removing the words with three characters or less, (c) removing these stop words, and (d) stemming the remaining words with a porter stemmer. We also removed papers that did not have a reference relationship. To reduce the noise, we also removed words appearing fewer than 10 times in the data set. Of the 21,236 papers, 12,504

[1] http://clair.eecs.umich.edu/aan/downloads/

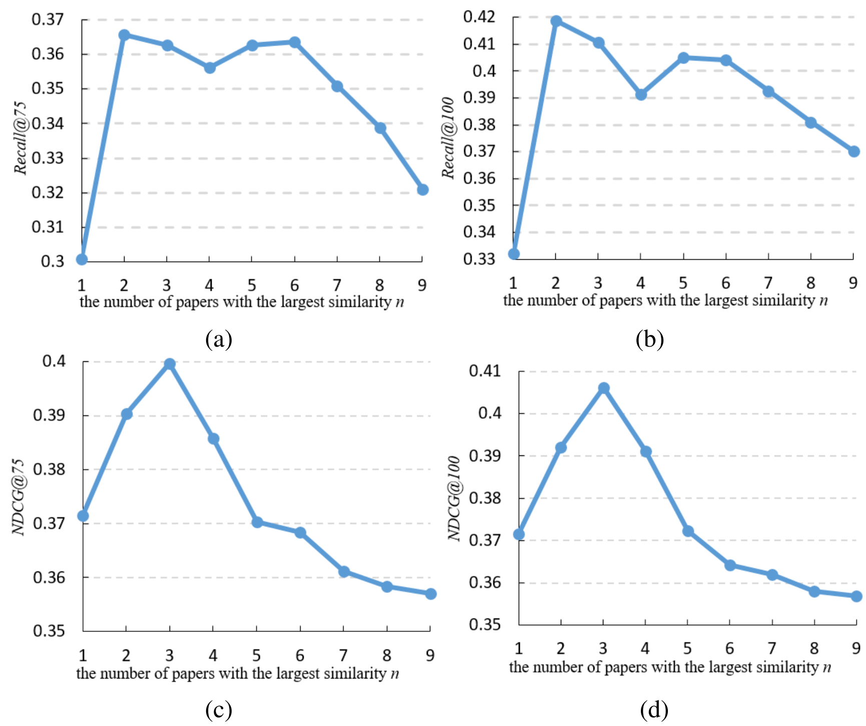

Fig. 2: $Recall$ and $NDCG$ for different n.

remained, containing 4918 distinct candidate words. Of these, 11,129 papers were published before 2013 and were used as a set of candidate papers (training set), and 1,375 papers were published in 2013 and were used as a test set. The abstract and title of a paper represented the content of the paper, providing the vector representation of the content of the paper using the BOW model. The words of each query consisted of the abstract and title. The reference list of each paper in the test setwas adopted as the actual reference of the paper.

B. Evaluation Metrics

In this section, two well-known measures, $Recall$ [8] and normalized discounted cumulative gain ($NDCG$) [18], were used for evaluating the accuracy and quality of the recommended results of the proposed approach. These two metrics have been widely used in the field of information retrieval and statistical classification. The formula for $Recall$ and $NDCG$ calculation is Equation 5 and Equation 6.

$$Recall@N = \frac{1}{C}\sum_{i=1}^{c}\frac{R(p)\cap T(p)}{T(p)} \tag{5}$$

$$NDCG@N = \frac{1}{C}\sum_{i=1}^{c}\left(\left(\sum_{j}^{N}\frac{2^{r_i}-1}{\log_2(j+1)}\right)/IDCG@N\right) \tag{6}$$

C. Experimental parameters analysis

A parameter of our approach: the number of papers with the largest similarity n that is the size of the set $TTSS$ are analysed and discussed.

Some relevant experiments have been conducted to determine the effect of the number of papers with the largest similarity n. Different values of n constructed different graphs and adjacency matrices, which inevitably led DeepWalk to generate different paper representations and then produced different effects on the recommendation results. There is no standard for setting the value of n, so we repeatedly ran experiments with different sizes of recommendation results to evaluate the impact of the number of papers with the largest similarity n. We all know that the more friends a man has, the more information about him we can get from his friends. We thus considered that the larger n was, the better the results generated. $Recall@75$ and $Recall@100$ had upward trends as n increased from 1 to 2 (Fig. 2a and b). $NDCG@75$ and $NDCG@100$ also had an upward trend as n increased from 1 to 3 (Fig. 2c and d).

We obtained $TTSS$ by calculating the similarity for incorporating the target paper into the citation network of the candidate paper to prepare for next step. $TTSS$ must be closer to the actual reference to be better. The calculation of paper-content similarity cannot be very accurate, however, so $TTSS$ is not exactly consistent with reality. A high n will contain some information that has nothing to do with the target paper, decreasing the accuracy of the recommended results.

TABLE I: Performance comparison of different methods in terms of $Recall$ and $NDCG$

Top-N	25		50		75		100	
Metric	$Recall$	$NDCG$	$Recall$	$NDCG$	$Recall$	$NDCG$	$Recall$	$NDCG$
PW	0.1972	0.3374	0.2769	0.3502	0.3182	0.3567	0.3787	0.3582
APW	0.2049	0.3227	0.2913	0.3471	0.317	0.3582	0.4031	0.3617
LINE	0.1212	0.2715	0.1628	0.2843	0.2173	0.3146	0.251	0.3419
Node2vec	0.1509	0.2587	0.227	0.2818	0.2767	0.3354	0.3152	0.3526
CRCA	**0.2285**	**0.3602**	**0.3113**	**0.3788**	**0.3658**	**0.3904**	**0.4188**	**0.392**

$Recall$@75 and $Recall$@100 had downward trends when n was 3 (Fig. 2a and b), and $NDCG$@75 and $NDCG$@100 had downward trends when n was 4 (Fig. 2c and d).

The best $Recall$ values were obtained for an n of 2, and the results were the worst when n was 1. The best $NDCG$ values were obtained for an n of 3, and the results were the worst when n was 9. We can therefore conclude that when n is 2, both $Recall$ and $NDCG$ will tend to increase at the beginning as n increases and will then tend to decrease. We ultimately decided to set n to 2 in our experiment. These relevant experimental analyses indicated that the effect of the recommendation results varied with n.

D. Comparison

We compared four baseline methods to validate the effectiveness of the proposed approach.

- **Baseline approach 1 (Paper Word, PW)**. PW is a two layer graph model that includes both citation and content information [11]. It applies a graph based similarity learning algorithm to for recommending papers. Its query is represented as $q = [0, q_w]$ using keywords.
- **Baseline Approach 2 (Author Paper Word, APW)**. AP-W represents a three-layer graph model [6] and contains author relationships , words in papers and paper citations for recommendation. A binary co-authorship graph is added to the PW model. It adopts an RWR framework to measure the query relevance for each publication. Its query is represented as$q = [q_a, 0, q_w]$, where q_a is the searcher information.
- **Baseline Approach 3 (LINE)**. For a large network $G(V, E)$, Line maps all nodes v in the network into a d-dimensional vector, and tries to keep the structure of the original network through first-order proximity and second-order proximity . The dimension for LINE is set to 75, and $rho = 0.025$.
- **Baseline Approach 3 (Node2vec)**. Node2vec defines a sequence of strategy generated bias random walk, still using skip-gram to train. Through the BFS and DFS of two different sampling methods, retain the information of different network structure. The dimension for Node2vec same to LINE, and $p = q = 0.25$.
- **Our Approach (CRCA)**. We proposed a Citation Recommendation method with Content-Aware bibliographic network representation (CRCA). We use the similarity between papers to obtain the adjacency matrix and use it to represent the relationship between the target and candidate papers. The vector dimension for CRCA same to LINE and Node2vec. Our model is also two layered, and the query representation is defined as $q = [q_w]$, which is different from PW and APW.

Table 2 shows the results of this approach in contrast to the PW, APW, LINE and Node2vec approaches. We set $n = 2$ and $d = 75$. The $Recall$ and $NDCG$ metrics increased for these three methods as the number of the recommendation list N gradual increased, because more papers are recommended as N increases. APW and CRCA outperformed PW for $Recall$ and $NDCG$. In other words, author-relationship information helped to obtain more accurate recommended results, and DeepWalk can produce a better recommendation than RWR if only paper contents and citations are used. In addition, CRCA always produced larger values than APW for $Recall$ and $NDCG$ as N increases. CRCA was an improvement over APW (on average 2.3% for $Recall$ and 3.29% for $NDCG$). These results indicated that the proposed approach could generate more accurate paper recommendations than APW, so DeepWalk can take advantage of citation information to obtain some features that RWR cannot and is more effective when information is missing. Furthermore, Our results show that under the same conditions, the recommended result of DeepWalk method is superior to LINE and Node2vec.

V. Conclusions

We introduced a new approach that used the content and citation relationship of papers to recommend relevant scientific papers for target papers based on network representation. We used content similarity to select an initial seed set of papers, the set CRS. We generated the matrix based on the paper vector, not the citation information. We used the DeepWalk to obtain the vector representation of papers for vectorising the candidate and target papers. We then used Euclidean distances to carry out the recommendation of the target paper. The experimental results indicated that this approach can cope with the problem caused by highly cited papers. The results also demonstrated that our proposed recommendation approach based on DeepWalk performed better than baseline methods. We concluded that the feature vectors generated by DeepWalk based on an adjacency matrix constructed by content similarity were able to improve the performance of citation recommendation.

Acknowledgments

This research was supported by National Natural Science Foundation of China [41771315, 61672433, 61872296, 61772429, 61702317, 61702305]; Natural Science Foundation of Shaanxi Province [2016JM6038] and the Fundamental Research Funds for the Central Universities [2452015060]. State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau [A314021402-1702].

References

[1] G. Aditya, "Leskovec jure node2vec: Scalable feature learning for networks," *855-864*, 2016.

[2] V. Andre, "A comparison between usage-based and citation-based methods for recommending scholarly research articles," *Proceedings of the American Society for Information Science and Technology*, vol. 47, no. 1, pp. 1–2, 2010.

[3] L. X. Chen Weizheng, Zhang Yan, "Network representation learning," *Big Data Research*, 2015.

[4] K. K. e. a. Ku?uktun Onur, Saule Erik, "Recommendation on academic networks using direction aware citation analysis," *Computer Science*, 2012.

[5] C. W. W. Lao Ni, "Relational retrieval using a combination of path-constrained random walks," *Machine Learning*, vol. 81, no. 1, pp. 53–67, 2010.

[6] W. W. e. a. Li Jing, Xia Feng, "Acrec: a co-authorship based random walk model for academic collaboration recommendation," in *In: Proceedings of the 23rd International Conference on World Wide Web:ACM*. Springer, 2014, pp. 1209–1214.

[7] B. X. e. a. C. Liu Haifeng, Kong Xiangjie, "Xiaomei, et al. context based collaborative filtering for citation recommendation," *IEEE Access*, vol. 3, pp. 1695–1703, 2015.

[8] X. H. e. a. Liu Qi, Chen Enhong, "Enhancing collaborative filtering by user interest expansion via personalized ranking," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 1, pp. 218–233, 2012.

[9] L. W. e. a. Meng Fanqi, Gao Dehong, "A unified graph model for personalized query-oriented reference paper recommendation," in *In: ACM International Conference on Conference on Information & Knowledge Management*, 2013, pp. 1509–1512.

[10] C. G. e. a. Mikolov Tomas, Chen Kai, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:13013781*, 2013.

[11] H. S. e. a. Pan Linlin, Dai Xinyu, "Academic paper recommendation based on heterogeneous graph," in *In: Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*. Springer, 2015, pp. 381–392.

[12] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014, pp. 701–710.

[13] S. S. Perozzi Bryan, Al-Rfou Rami, "Deepwalk: Online learning of social representations," in *In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining:ACM*, 2014, pp. 701–710.

[14] Q. V. e. a. Radev Dragomir R, Muthukrishnan Pradeep, "The acl anthology network corpus," *Language Resources and Evaluation*, vol. 47, no. 4, pp. 919–944, 2013.

[15] Y. X. e. a. Ren Xiang, Liu Jialu, "Cluscite: Effective citation recommendation by information network-based clustering," in *In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014, pp. 821–830.

[16] W. M. e. a. Tang Jian, Qu Meng, "Line:large-scale information network embedding," *1067-1077*, vol. 2, 2015.

[17] O. M. e. a. Totti Luam C., Mitra Prasenjit, "A query-oriented approach for relevance in citation networks," in *In: International Conference Companion on World Wide Web*, 2016, pp. 401–406.

[18] O. M. e. a. Totti Luam C, Mitra Prasenjit, "A query-oriented approach for relevance in citation networks," in *In: Proceedings of the 25th International Conference Companion on World Wide Web:International World Wide Web Conferences Steering Committee*, 2016, pp. 401–406.

[19] C. W. Wei Xing, "Bruce lda-based document models for ad-hoc retrieval," pp. 178–185, 2006.

[20] B. C. T. West Jevin D., Wesleysmith Ian, "A recommendation system based on hierarchical clustering of an article-level citation network," *IEEE Transactions on Big Data*, vol. 2, no. 2, pp. 113–123, 2016.

[21] B. T. M. e. a. Xia Feng, Wang Wei, "Big scholarly data: A survey," *IEEE Transactions on Big Data*, vol. pp, no. 99, pp. 1–1, 2017.

[22] Z. K. e. a. Yang Zhe, Wu Bing, "A survey of collaborative filtering-based recommender systems for mobile internet applications," *IEEE Access*, vol. 4, pp. 3273–3287, 2016.

[23] K. Youngjoong, "A study of term weighting schemes using class information for text classification," in *In: International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2012, pp. 1029–1030.

[24] G. C. e. a. Yu Yingying, Yu Yingying, "Meta-path-based ranking with pseudo relevance feedback on heterogeneous graph for citation recommendation," in *In: ACM International Conference on Conference on Information and Knowledge Management*, 2014, pp. 121–130.

[25] T. X. Zhang Wen, Yoshida Taketoshi, "A comparative study of tf*idf, lsi and multi-words for text classification," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2758–2765, 2011.

[26] Y. L. e. a. Zhiyuan Liu, Maosong Sun, "Knowledge representation learning: A review," *Journal of Computer Research and Development*, vol. 53, no. 2, pp. 247–261, 2016.