Academic paper recommender system using multilevel simultaneous citation networks

Jieun Son, Seoung Bum Kim *

Department of Industrial Management Engineering, Korea University, 145 Anam-Ro, Seoungbuk-Gu, Seoul 136-713, South Korea

A R T I C L E   I N F O

A B S T R A C T

Researchers typically need to filter several academic papers to find those relevant to their research. This filtering is cumbersome and time-consuming because the number of published academic papers is growing exponentially. Some researchers have focused on developing better recommender systems for academic papers by using citation analysis and content analysis. Most traditional content analysis is implemented using a keyword matching process, and thus it cannot consider the semantic contexts of items. Further, citation analysis-based techniques rely on the number of links directly citing or being cited in a single-level network. Consequently, it may be difficult to recommend the appropriate papers when the paper of interest does not have enough citation information. To address these problems, we propose a recommendation system for academic papers that combines citation analysis and network analysis. The proposed method is based on multilevel citation networks that compare all the indirectly linked papers to the paper of interest to inspect the structural and semantic relationships among them. Thus, the proposed method tends to recommend informative and useful papers related to both the research topic and the academic theory. The comparison results based on real data showed that the proposed method outperformed the Google Scholar and SCOPUS algorithms.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Recommender systems have become popular and have begun to attract increasing attention from both academia and industry [1]. However, compared with other recommender applications, such as those for books, movies, and music, fewer studies have examined recommender systems for academic papers. Researchers typically need to filter a substantial number of academic papers to find those relevant to their research. This filtering is cumbersome and time consuming because the number of published academic papers is growing exponentially. Consequently, there is an urgent need for efficient academic paper recommender systems [2].

Previous studies have focused on finding better recommender systems for academic papers relevant to specific research topics. One of the most commonly used recommendation approaches is collaborative filtering. However, extensive studies indicated that this approach has inherent problems, such as data sparsity and ratings imbalance [3]. To address these problems, related recommendation techniques, such as content-based filtering, network analysis, and information retrieval, are being studied [4,5].

Content-based filtering approaches create relationships between items by analyzing their inherent characteristics. In most content-based filtering systems for textual applications (e.g., academic papers), item descriptions are keywords extracted from text. However, keyword-based systems create numerous complications that originate from natural language ambiguity. Further, keyword-based systems are unable to capture the semantics of user interests because they are primarily driven by string matching operations [6].

Citation networks based on citation-related connections within scientific literature have been used to compute relatedness among academic papers. Link-based techniques, such as co-citation and bibliographic coupling, measure relevance by focusing on neighbors. Co-citation is the similarity measure for two papers cited together by other papers, and bibliographic coupling is the similarity measure for two papers that refer to the same paper. However, recommender systems that use link-based techniques cannot consider complex relationships among papers because these techniques count the number of links directly cited in single-level networks [7].

Selecting relevant academic papers is similar to information retrieval, an activity in which information resources relevant to the desired information are selected from amongst a collection of such resources. The mainstream tool used for information retrieval research is ranking algorithms. The PageRank algorithm, in particular, has been used to produce a better global ranking of search results [8]. However, although

* Corresponding author.
E-mail addresses: jieunson@korea.ac.kr (J. Son), sbkim1@korea.ac.kr (S.B. Kim).

PageRank can estimate the authority of a paper, one of its drawbacks is that it ranks papers based on the citation count. Therefore, recent articles always score low and consequently are not recommended by the algorithm [9].

1.1. Contributions of the paper

The main contributions of our research can be summarized as follows. First, we propose a novel recommendation system for academic papers that combines citation analysis and network analysis. The proposed method is based on multilevel citation networks that compare all indirectly linked papers to the paper of interest to inspect the structural and semantic relationships among them. Our main research objective in this study is to consider the mutual relationships among the papers in a broad network beyond a single level, and to evaluate the significance of each paper through certain centrality measures. Second, the lack of a citation count notwithstanding, the proposed method can find influential papers using centrality measures that are derived from a citation network. Finally, we found that the proposed method outperformed the existing methods, Google Scholar and SCOPUS, based on user satisfaction data. We asked users to receive recommendations from these algorithms and rate the recommended item lists based on their satisfaction with the results.

1.2. Organization of the paper

The remainder of this paper is organized as follows. In the related work section, we review the academic paper recommendation approaches proposed in existing literature. In the proposed method section, we detail our approach to academic paper recommendation by using multilevel simultaneous citation networks. The experimental evaluation section presents and discusses our experimental results and evaluation. The last section concludes this paper.

2. Related work

2.1. Content-based filtering and collaborative filtering

Two types of algorithms are typically used in recommender systems: collaborative filtering and content-based filtering [10]. Collaborative filtering algorithms match users in a system based on the similarity of the past ratings provided by each user, and then recommend items that similar users have liked. Collaborative filtering requires a matrix consisting of user ratings for a particular item [11]. Therefore, collaborative filtering cannot generate accurate recommendations without sufficient initial ratings from users. This problem also occurs in the domain of academic papers because it is very difficult to gather user score information in digital libraries. To resolve this problem, researchers have focused their attention on creating a matrix of collaborative filtering ratings from the citation web between academic papers. In this ratings matrix, authors are represented as rows and papers as columns. Each entry is a rating for certain papers that the authors have cited [12]. However, in many cases the authors have published papers in various technical fields or changed their areas of interest during their careers. For example, a researcher who started his or her research activities in electrical engineering may later write many papers about statistics. Thus, it may be difficult to find a similar author group and recommend proper papers to target users when a database contains many such cases.

Content-based filtering algorithms recommend items to users based on their description [13]. Applications of content-based filtering in academic paper recommender systems rely on the ability to compare the similarities of complete text or keywords because text-based features are excellent for classifying papers [14]. The "related documents" function of SCOPUS is one example of a content-based filtering approach. The SCOPUS system defines keywords for a research paper, and the indexed keywords can be automatically imported as tags. Papers that contain one or more words in common with those in the paper of interest are returned as relevant. For the recommendation, the papers are selected in the order of the highest matching frequency of the keywords [15]. However, in some cases, text features are not as good at finding a related paper. Although a paper may be conceptually similar to the paper a user may be interested in, it may use a different vocabulary. In this case, the relevant paper may be overlooked. Conversely, the same word can be used in papers in many different fields; this can then result in the wrong papers being recommended to users. For example, "port" is an endpoint of communication in a computer operating system. However, "port" also has other meanings in unrelated contexts; it can mean "harbor" in the context of shipping, and it is also used to refer to a type of wine.

2.2. Information retrieval techniques

PageRank is one of the methods that Google uses to evaluate the importance of webpages to improve the quality of web search engines [7]. This algorithm has been widely applied not only to rank web search results, but also to recommend academic papers [16]. Google Scholar primarily uses PageRank techniques to identify papers related to the paper of interest. Quoting from https://scholar.google.com/intl/en/scholar/about.html, "Google Scholar aims to rank papers the way researchers do, weighing the full text of each paper, who it was written by, where it was published, as well as how often and how recently it has been cited in other papers" [17]. To provide recommendations, the "related articles" function of Google Scholar presents a list of closely related papers, ranked primarily by how similar these papers are to the paper of interest [18]. Although PageRank is a good method for determining the authority of a paper, it tends to rank papers based primarily on the number of citations. As a result, recent papers are always ranked low, even when the paper is known as eminent literature. This is an important limitation because recent papers may be important to researchers who wish to understand current issues and to set research directions. Bethard and Jurafsky [19] proposed integrating a keyword-based algorithm and citation information for learning literature search models. The main idea of their integrated approach is to look for similar terms and topics among the articles. Therefore, they include the classic term frequency-inverse document frequency (TF-IDF), which represents both the user query and the word counts in document and latent Dirichlet allocation (LDA). PageRank and citation count are only used to boost the article.

2.3. Citation networks

Citation analysis, used in large applications such as patent analysis and document analysis, refers references in one item to another item. While the two approaches presented above are based on similarity, the citation network is based on relational information. Therefore, it is useful for understanding the relationship between subjects, the flow of history, and publication trends [20]. The citation analysis of academic papers in particular is important because it can directly reveal papers closely related to the query paper. Several previous studies recommended papers for a manuscript containing a partial list of citations. Co-citation analysis, introduced by Small [21], is one of the first applications of co-occurrence. Small suggested that the more two papers are related to each other, the more often they are co-cited. Liang [22] presented graph networks that show how the papers are connected through citations. Connections are based on bibliographic coupling and co-citation strength [23,24]. Once a graph was built, graph metrics were used to find recommendation candidates. One or several input papers are given as the paper of interest and random walks were conducted to find the most popular items in the network graph [25,26]

Much of the literature on citation analysis considers just one level, directly linked to nodes [27]. However, in single-level analysis, the

relationship between indirectly connected papers cannot be comprehended.

3. Proposed method

3.1. Overview of proposed recommendation method

We propose a recommender system with a multilevel simultaneous citation network (MSCN) for paper references. MSCN considers the relationship among the papers in a broad area and evaluates the significance of each paper through certain centrality measures. Centrality identifies the most important papers within the network. The process for the proposed MSCN comprises three steps: (1) generation of multilevel citation networks, (2) selection of candidate papers, and (3) determination of ranking of each candidate paper for final recommendation. An overview of the procedure for the proposed recommendation method is shown in Fig. 1. First, we initially generate a directional multilevel citation network containing the papers that have a citation relation with the paper of interest. Once we generated the multilevel citation network, we computed a candidate score for each paper (i.e., each node of a multilevel citation network) to select for candidate papers. After the paper is determined to be relevant from the candidate scores, we calculate the average rank that can be used for final recommendation.

3.2. Generating multilevel citation networks

References are a list of cited papers appearing at the end of academic papers. The relations among papers are "cites" and conversely, "cited." Graphs describing these relationships are citation networks [28]. Fig. 2 shows an example of a multilevel citation network with six levels. The nodes represent papers and the links represent citations with direction. Beginning with paper of interest, I, we use its reference list to start a citation network. "Backward" is the name used to identify citation relationships for papers cited by I, and "forward" is the name used to identify citation relationships for papers citing I. Therefore, the level of a multilevel citation network is the sum of all levels in the backward and forward directions. The paper of interest cites three papers and is cited by four papers. Traditional citation analysis would only utilize a single-level network to consider the directly linked papers. In our proposed method, we extend the network to multiple levels. In the present study, we initially generate the multilevel citation network up to ten levels. We thought that ten is generally acceptable because using more than 10 levels would include the majority of papers, which are unrelated to the paper of interest.

3.3. Selection of candidate papers

Once we generated the multilevel citation network, we computed a candidate score for each paper (i.e., each node of a multilevel citation network) to select candidate papers. Candidate scores are calculated for all papers appearing in the multilevel citation network to quantify
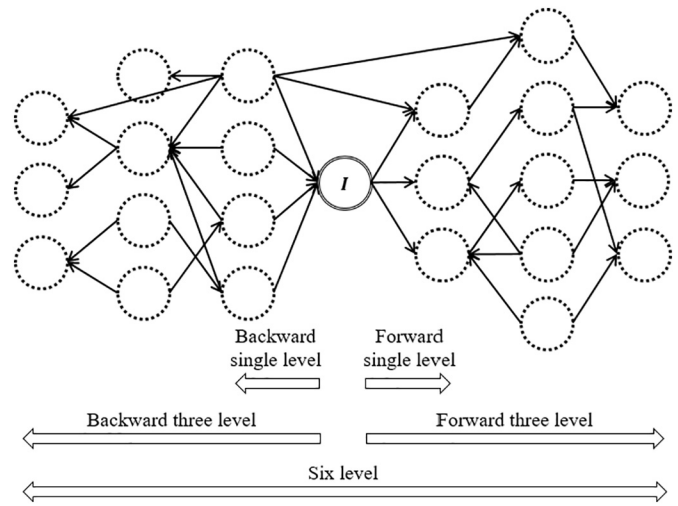
Fig. 2. Example of a multilevel citation network with six levels.

their relevance with the paper of interest. Two different citation relations between papers have been used to measure their similarity; namely, bibliographic coupling and co-citation. Bibliographic coupling is a measure of the similarity between two papers that refer to the same paper, whereas co-citation is the similarity measure for two papers cited together by other papers [29].

Fig. 3 illustrates bibliographic coupling and co-citation, showing that papers A and B both cite papers 1, 2, and 3.

Their bibliographic coupling strength (B.C strength) can be calculated as follows:

$$\text{B.C strength}(A,B) = \sum_{k=1}^{n} B.C\,(A_k, B_k),$$
$$B.C\,(A_k, B_k) = \begin{cases} 1, & \text{if paper } A \text{ and } B \text{ cite paper } k, \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

$B.\,C(A_k, B_k)$ is one if papers A and B cite paper k. In Fig. 3, the B.C strength for papers A and B is three. For papers C and D, which are both cited by papers 1, 2, and 3, and have a co-citation, the co-citation strength (C.C strength) can be calculated as follows:

$$\text{C.C strength}(C,D) = \sum_{k=1}^{n} C.C\,(C_k, D_k),$$
$$C.C\,(C_k, D_k) = \begin{cases} 1, & \text{if paper } C \text{ and } D \text{ are cited by paper } k, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

$C.\,C\,(C_k, D_k)$ is one if paper k cites papers C and D. In Fig. 3, the co-citation strength for papers C and D is three. The purpose of B.C strength and C.C strength is not to analyze the indirect relationship between the papers in a multilevel network, but to find the papers related to the paper of interest in a single-level network. Therefore, B.C strength and C.C strength of papers A and C are calculated independently.
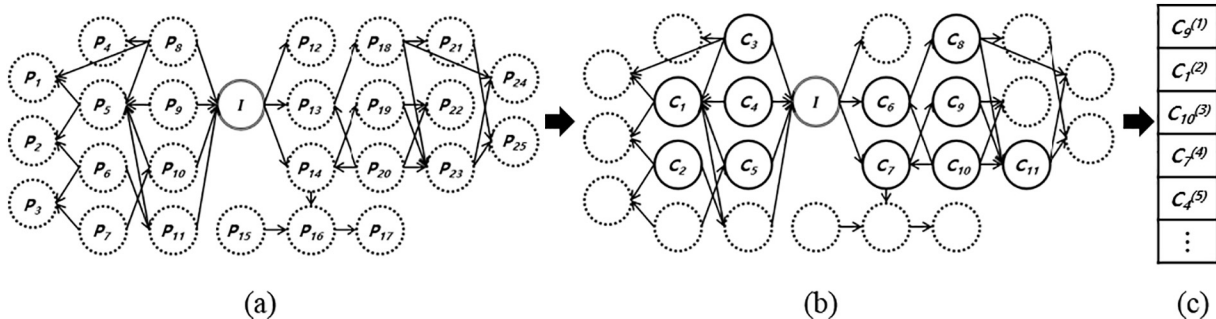
Fig. 1. Overview of the proposed MSCN recommendation method: (a) generation of multilevel citation networks, (b) selection of candidate papers, and (c) determination of ranking of each candidate papers for final recommendation.
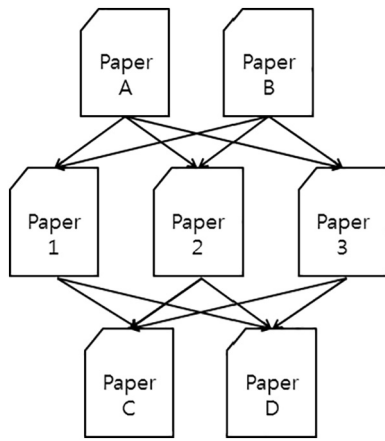
Fig. 3. Example of bibliographic coupling and co-citation analysis.

We propose merging these two measures to reflect both characteristics by defining the following candidate score (C-score):

$$\text{C-score }(P) = \frac{\sum_{J=1}^{n}(\text{B.C strength}(P, J) + \text{C.C strength}(P, J))}{d(I, P)}. \tag{3}$$

In the C-score, the numerator represents the similarity of two papers based on citation information, and the denominator is the distance between the paper and the paper of interest on the network. Therefore, C-score can be considered to be a combination of citation analysis and network analysis. The numerator is the sum of the bibliographic coupling strength and co-citation strength of paper P. J represents all papers except paper P, which is a target of the C-score. The C-score measures how strongly P is related with other papers, J, in both aspects. Thus, a high value of this numerator is an indication that P has a related subject matter with its neighbors. On the other hand, a low value indicates that P is not relevant to the contents of other papers. The C-scores consider the relevance of P with not only J, but also I. The denominator of the C-score determines the boundary of the research area, which has papers that are more relevant and closer to I. $d(I,P)$ is the distance, considered to be the number of links between I and P. The more distance there is between them, the more indicative that the topic or domain field of the two papers is different.

Table 1 shows an example of how the C-scores for papers P_5, P_{20}, and P_{24} are calculated in the seven-level citation network displayed in Fig. 4. For example, the B.C strength of P_5 is two ($[P_5, P_8 \rightarrow P_1]$, $[P_5, P_6 \rightarrow P_2]$) and the C.C strength of P_5 is six ($[P_5, P_1 \leftarrow P_8]$, $[P_5, P_4 \leftarrow P_8]$, $[P_5, I \leftarrow P_8]$, $[P_5, I \leftarrow P_9]$, $[P_5, I \leftarrow P_{10}]$, $[P_5, I \leftarrow P_{11}]$). The sum of the B.C strength and the C.C strength is eight, which is the numerator value of the C-score of P_5. The denominator value of the C-score of P_5 ($d(I,P)$) is two because the number of links between paper of interest I and P_5 is two. For example, P_5 and P_{20} have the same value for the total similarity. However, the C-score of P_{20} is lower than that of P_5 because P_{20} is farther than P_5 from paper of interest, I. This means that the degree of similarity of P_5 and P_{20} with their neighbor papers is the same, but P_5 is more dissimilar to the user topic than P_{20}. On the other hand, although P_5 and P_{24} are at the same distance, P_{24} has a lower C-score because the total similarity of P_{24} is lower than that of P_5. The papers with low C-scores tend to be isolated from the network community. In this case, P_{24} is likely to be

an irrelevant paper produced by self-citations and ceremonial citations. Ceremonial citations are citations that were used even though the authors did not read the cited publication [30]. In this way, the C-score is calculated for each of the papers to quantify the degree of the relationship between the paper of interest and all others. Having found the C-scores of all the papers, candidate papers are selected. The number of candidate papers is determined empirically. In this study, we selected 500 papers with large C-scores. Various experiments indicate that a proper network size for a given problem is between 500 and 800. In addition, our experiments have shown that there is no significant performance difference between networks of sizes 500 and 800. Networks with less than 500 papers may not accurately represent the field. In contrast, networks with more than 800 papers are too complicated to use in practice. The citation level for a network containing 500–800 papers (nodes) is usually six to eight. Note that C-scores will not be used in further steps because they are only used to determine whether the paper is relevant to paper of interest or not.

3.4. Determination of recommend papers

In the previous step, we selected 500 candidate papers ready for recommendation. A centrality analysis of the network is performed to examine the importance of each candidate paper (node) selected by the preceding step [31]. The centrality measure suggests the significance of individual papers due to their relationships with other papers [32]. We calculate four centrality measures (degree centrality, closeness centrality, betweenness centrality, eigenvector centrality) of 500 candidate papers to determine the most significant papers for recommendation. This approach combines the concepts of citation analysis and network analysis, because network analysis is performed only on the papers selected by citation analysis.

The degree centrality (C_D) is the most intuitive notion of centrality [33]. The more neighbors a given node has, the greater its influence is. We consider only the number of papers "cited," called the in-degree centrality:

$$C_D(P) = \frac{d(P)}{n-1}, \tag{4}$$

where $d(P)$ is the number of papers referring to paper P, and n is the total number of papers in the network. A high value of the in-degree centrality implies popularity.

The calculation of the degree centrality is limited by the number of nodes that are directly connected to the paper, and indirectly connected nodes are not included for the measurement. We use closeness centrality (C_c) to analyze the global hub and authority. Closeness centrality is based on the distance from a paper to all other nodes in the network, and is defined as the inverse total distance. The idea is that a paper is more central if it interacts with more of the other nodes, and it is considered relatively important [34].

$$C_C(P) = \frac{n-1}{\sum_{J \neq T} d(P, J)}, \tag{5}$$

where n is the total number of papers in the network. Therefore, $n-1$ is the minimum sum of distances for a paper that is adjacent to all other papers. P is a target paper of C_c, and J is all papers except paper P. $d(P,J)$ denotes the distance between paper P and other papers J. A node with high closeness centrality is located to the center.

Betweenness centrality (C_B) is based on the number of shortest paths passing through a vertex [35]. Vertices with high betweenness are potential deal makers. They are in a special position because most other nodes have to channel their communications through them. In other words, this measure is the extent to which a paper is positioned
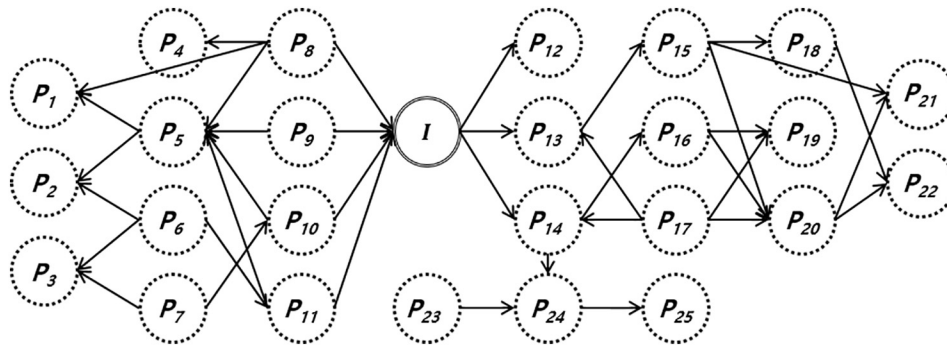
Table 1
The calculation of candidate score values for papers P_5, P_{20}, and P_{24}.

Paper	B.C strength	C.C strength	Total similarity	$d(I,P)$	C-score
P_5	2	6	8	2	4.0
P_{20}	2	6	8	3	2.6
P_{24}	0	1	1	2	0.3

Fig. 4. Seven-level citation network.

on the shortest path between other pairs of papers:

$$C_B(P) = \sum_{J \neq V \neq P} \frac{g_{JV}(P)}{g_{JV}}, \tag{6}$$

where g_{JV} is the number of links in the shortest route between paper J and V, and $g_{JV}(P)$ is the number of links in shortest route between J and V that pass through paper P. In the citation network, the papers with closely related citations constitute one community. A paper linking communities can control communication flow among communities, and thus is important. Typically, research papers that influence papers in various fields or that converge existing concepts tend to have a high score. In other cases, nodes play the role of bridging the flow and change in research trends.

Eigenvector centrality (C_E) depends on the number of neighbor nodes that are directly connected to a paper and the quality of the neighbor nodes [36]. Eigenvector centrality measures the influence of set B_T containing all papers linking to paper P.

$$C_E(P) = \frac{1}{\lambda} \sum_{J=B_P} A_{P,J} x_J, \tag{7}$$

where $A_{P,J}$ is the adjacency matrix in which its element is one if J is linked to P, and zero otherwise. x_J is the score of the eigenvector centrality of J, and λ is the eigenvalue of P. Eigenvector centrality measures not only how many papers are connected to a paper, but also how many important papers are connected to a paper.

The values of each of four centrality measures are converted to ranks to combine them. Note that the ranges of the four centrality measures described above are different. After the centrality values are converted to ranks, all four centrality measures have the same scale [1,25]. Now we use the following average ranks (AR) [37] that combines multiple rankings yielded by Eqs. (4)–(7):

$$AR(P) = \frac{\sum_{k=1}^{M} rank^k(P)}{M}, \tag{8}$$

where M is the number of centrality measures, and $rank^k(P)$ is ranking result with kth centrality measure on paper P.

Finally, candidate papers are sorted by the AR and the top-n papers in the list are recommended to users. The number of n can be determined by numerous factors, such as the characteristic of the domain field or the system environment.

4. Experimental evaluation

4.1. Experimental data and evaluation measure

We conducted experiments to evaluate the recommendation capabilities of the proposed MSCN and compared it with Google Scholar and SCOPUS. Google Scholar and SCOPUS are chosen as comparison

algorithms because they are the most recognized proprietary databases for journal content and provide recommendation services to customers [38].

Table 2 shows the eight academic papers that were selected as papers of interest in our experiments.

We chose papers 1, 5, and 6 because we wanted to compare the recommendation results when a recent paper was selected as a paper of interest. The main reason we chose papers 1, 4, 7, and 8 primarily is because we wanted to compare the recommendation results when an eminent paper in a specific field was selected as the paper of interest. We chose paper 3 mostly because we wanted to compare the recommendation results when a highly cited paper was selected as the paper of interest. Moreover, eight papers are selected from various fields, including business, psychology, biotechnology, medical, and social science.

We performed a blind test to assess the validity of the proposed method. For each paper of interest, we selected 500 papers as candidates and recommended 25 papers to the user. From the recommender system literature, we learn that performance evaluation is mostly conducted by using fewer than 25 items because recommending too many items can confuse users [39,40]. Twenty-four researchers with expertise in the subjects of the papers provided their recommendations. The titles, authors, year of publication, and journal names of the recommended papers are provided to the experts. The experts determine whether they are satisfied or not with recommended papers.

For evaluation, normalized discounted cumulative gain (NDCG) and mean reciprocal rank (MRR) are used [41,42]. NDCG measures the performance of a recommender system based on the graded relevance of the recommended items.

$$NDCG_p = \frac{DCG_p}{IDCG_p}, \tag{9}$$

$$DCG_p = \sum_{i=1}^{p} \frac{(2^r - 1)}{\log(1 + i)}, \tag{10}$$

$$IDCG_p = \sum_{i=1}^{p} \frac{1}{\log(1 + i)}. \tag{11}$$

$NDCG_p$ represents the total normalized gain accumulated at a particular rank p. DCG_p is the total gain accumulated at a particular rank p. The relevance value, r, of recommended item is binary; $r \in \{0, 1\}$. It is set to one if the user is satisfied with the recommended paper, or it is set to zero otherwise. $IDCG_p$ generates the maximum possible DCG until rank p for normalization. All NDCG calculations are then relative values on the interval 0.0 to 1.0. In a perfect recommendation, the NDCG value is one because DCG_p will be the same as $IDCG_p$.

Mean reciprocal rank (MRR) is widely used in the study of information retrieval and measures the ability of a recommender system to

Table 2
List of academic papers selected for experiments

	Title	Author	Year	Source	Cited by
1	Wafer classification using support vector machines	Baly, R., Hajj, H.	2012	IEEE Transactions on Semiconductor Manufacturing	33
2	A practical approach for interpreting multivariate T2 control chart signals	Mason, R. L. et al.	1997	Journal of Quality Technology	295
3	Nonlinear dimensionality reduction by locally linear embedding	Roweis, S. T., Saul, L. K.	2000	Science	11,052
4	Managing customer relationships through E-business decision support applications: a case of hospital–physician collaboration	Kohli, R., et al.	2001	Decision Support Systems	165
5	The power of the "like" button: The impact of social media on box office	Ding, C. et al.	2017	Decision Support Systems	0
6	A theory of social media dependence: Evidence from microblog users	Wang, C. et al.	2015	Decision Support Systems	19
7	Geographic information systems as a marketing information system technology	Hess, R. L. et al.	2004	Decision Support Systems	78
8	A literature network of human genes for high-throughput analysis of gene expression	Jenssen, T. K. et al.	2001	Nature genetics	974

return a relevant item at the top of the ranking.

$$MRR = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{rank_i}, \qquad (11)$$

where n is the number of users and $rank_i$ is the rank of the first correct item.

4.2. Experimental results

4.2.1. Correlation between centralities

In our experiment, four centrality measures are used to determine the most significant papers. Table 3 shows the correlation coefficients between four centrality measures, including degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality.

The result indicates that the betweenness centrality and the eigenvector centrality are highly correlated, regardless of paper type. Furthermore, the degree centrality is highly correlated with the closeness centrality, except in the case of paper 2.

4.2.2. Experimental results

In this section, we demonstrate the recommendation results from experiments that used eight papers. Figs. 5 and 6 compare the performances of Google Scholar, SCOPUS, and MSCN in terms of NDCG and MRR. The x-axes indicate the top ranked N papers among the 25 papers. The y-axes show the NDCG and MRR values calculated for the recommended papers.

Overall, the proposed MSCN outperformed Google Scholar and SCOPUS, except in the case of paper 7; In that case, Google Scholar performed slightly better than the proposed MSCN. Note that NDCG and MRR give more weight to the higher ranked papers, which are correctly recommended to the users. To minimize the impact of ranking, we also compared the average number of papers correctly recommended (regardless of ranks). Fig. 7 shows the comparative results of three recommendation methods for eight papers in terms of the average number of papers correctly recommended to 24 researchers.

Table 3
Correlation coefficients between centrality measures.

	C_D-C_C	C_D-C_B	C_D-C_E	C_C-C_B	C_C-C_E	C_B-C_E
Paper 1	0.55	0.21	0.19	0.19	0.31	0.76
Paper 2	0.42	0.35	0.34	0.20	0.28	0.75
Paper 3	0.59	0.27	0.44	0.15	0.27	0.69
Paper 4	0.50	0.20	0.26	0.21	0.22	0.58
Paper 5	0.58	0.57	0.57	0.05	0.21	0.79
Paper 6	0.68	0.41	0.41	0.13	0.40	0.51
Paper 7	0.67	0.24	0.18	0.42	0.42	0.74
Paper 8	0.77	0.21	0.45	0.07	0.41	0.73
Average	0.59	0.31	0.36	0.18	0.32	0.69

The results indicated that the proposed MSCN selected the most number of papers that were correctly recommended to the users. The keyword-based algorithm of SCOPUS suffers from some inherent problems [43,44]. For example, the keywords extracted from paper 1 such as "support vector machine (SVM)," "prediction methods," and "data models" can be used in various fields. Although semiconductors are the focus of paper 1, most of the recommended papers are in the pattern recognition or bioinformatics fields. In other words, paper 1 and the recommended papers have the SVM algorithm in common, but they have different applications. In general, there are two types of relevant papers: (1) papers related to algorithms and methods that are helpful for researchers to solve certain problems; and (2) papers focused on a similar research problem so that they are topically related. Superior recommender systems must cover both cases. However, keyword-based algorithms cannot distinguish between the two cases because they depend exclusively on the frequency of keywords. Therefore, results from keyword-based algorithms are often biased in favor of one type of relevancy. Further, the keyword-based algorithms of SCOPUS find only papers in which keywords frequently occur regardless of impact factor, status, or credibility. For example, 88% of all the recommended papers for paper 5 have never been cited, even if most of the recommended papers were published more than ten years ago. SCOPUS also performed poorly because it relies exclusively on occurrences of individual keywords. The extracted keywords form paper 3 are "locally," "linear," "embedding," "algorithm," "dimension," "reduction" and "data analysis." Users expect that the papers dealing with "locally linear embedding" algorithms for "dimension reduction" in the "data analysis'" field are recommended to them. However, SCOPUS ignores the semantic effects of word combinations.

The performance of Google Scholar is worse than the proposed MSCN. Papers 1, 5, and 6 are recent articles that were published after 2010. Google Scholar tends not to recommend suitable papers when a recent paper was selected as the paper of interest. Google Scholar tends to recommend papers that are published recently with a low citation count because the paper of interest does not have enough information for recommendation. Note that the average citation counts of the recommended papers from Google Scholar, papers 1, 5, and 6, is only 11.3. This result implies that when a recently published paper with few citations is selected as the paper of interest, it confuses the ranking algorithm used in Google Scholar that relies on citation information. Another limitation of Google Scholar arises when and older paper is selected as the paper of interest. In this case, most of the recommended papers by Google Scholar were out of date. For example, none of the papers in recent decades were recommended by Google Scholar for paper 2, paper 4, and paper 8. Furthermore, recommended papers are highly cited because citation counts significantly affect the ranking in Google Scholar. When the paper of interest is famous and highly cited by other papers, there is a strong possibility that a recommender system will recommend such famous papers to users because highly cited
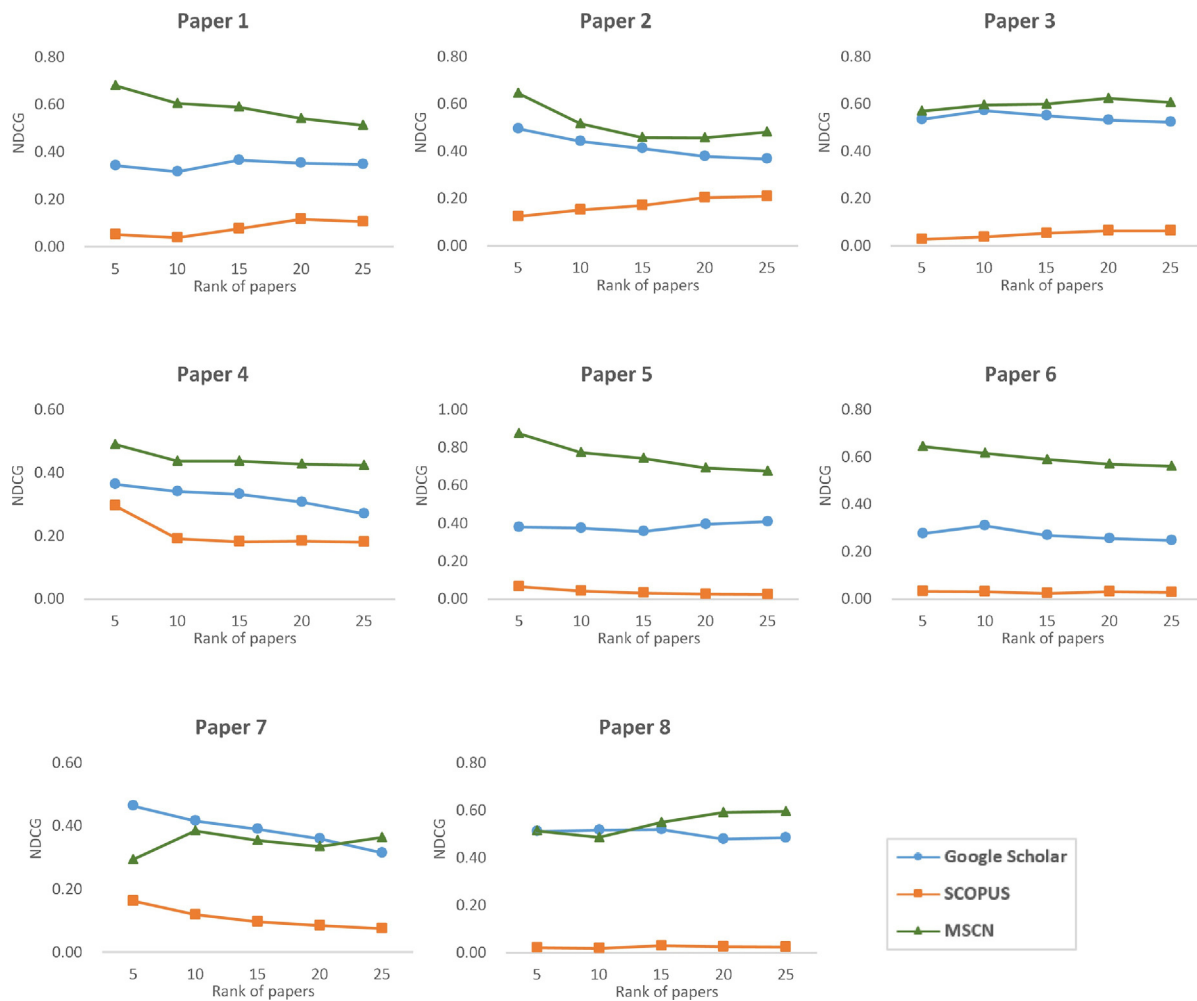
Fig. 5. Performance comparison of Google Scholar, SCOPUS, and MSCN with NDCG.

papers are provided to users repeatedly. This phenomenon is called an "over specialization" or "concentration bias" problem [45]. For paper 3, which has the highest number of citations in our experiments, 25 papers recommended by Google were published in 1991–2009. The average number of citations within these papers was 3685. Most of these 25 papers are papers that deal with comprehensive contents, or very famous technologies. On the other hand, the 25 papers recommended by the proposed method were published between 2000 and 2013, and include a variety of papers including old and famous papers, papers dealing with specific technologies, and recent papers with many citations. Furthermore, some studies also argue that Google Scholar can be suitable for searching for standard literature, but less suitable in searching for "gems" or papers whose authors are advancing views contrary to those of the mainstream [46–48]. This is neither good nor bad, but users should be aware of this bias because recently published but important papers may be omitted because such papers tend to have a low citation count. Therefore, Google Scholar may be inappropriate for users who want to read recently published papers with high impact.

The performance of MSCN is superior to the others. We can draw some inferences from our experiments. The citation analysis of MSCN is balanced in its coverage of both types of relevancy because its reference list involves all papers that are related to the topic or to its basic algorithms. Furthermore, when the paper of interest has high citation counts, such as paper 3, the recommended papers from MSCN are relatively new ones with high citation counts, or less famous papers but centrally important papers. We believe that it is suitable to use the proposed method when searching for the papers in advancing views contrary to those of the mainstream. In addition, the lack of citation

counts notwithstanding, it is possible to find influential papers through use of the four centrality measures derived from a citation network.

5. Conclusions

The vast amount of information currently available makes it difficult for researchers to find the academic papers that are most relevant to their current work or to the research field in which they are interested. Consequently, considerable research is focused on developing efficient academic paper recommender systems—defined as support systems that help users find papers. Content-based filtering, collaborative filtering, and information retrieval techniques have been widely used to construct better paper recommender systems. However, these approaches have limitations in that they cannot consider the mutual relationships among papers because they use only the similarity of the contents. In particular, some studies argue that content-based filtering and information retrieval techniques ignore the quality and popularity of items in academic paper recommender systems [4]. Citation analyses that use the reference list of the paper provide more accurate and reliable information because the authors can directly judge the relevance between papers [17,49]. However, it is difficult to understand the relationship between papers by citation analysis alone, such as in bibliographic coupling and co-citation measures, because the citation relation between papers is very complicated. To address this problem, we propose the multilevel citation network, which allows us to quantify direct, indirect, and integral relations among papers. Furthermore, multilevel citation networks can compare all the relevant papers more appropriately
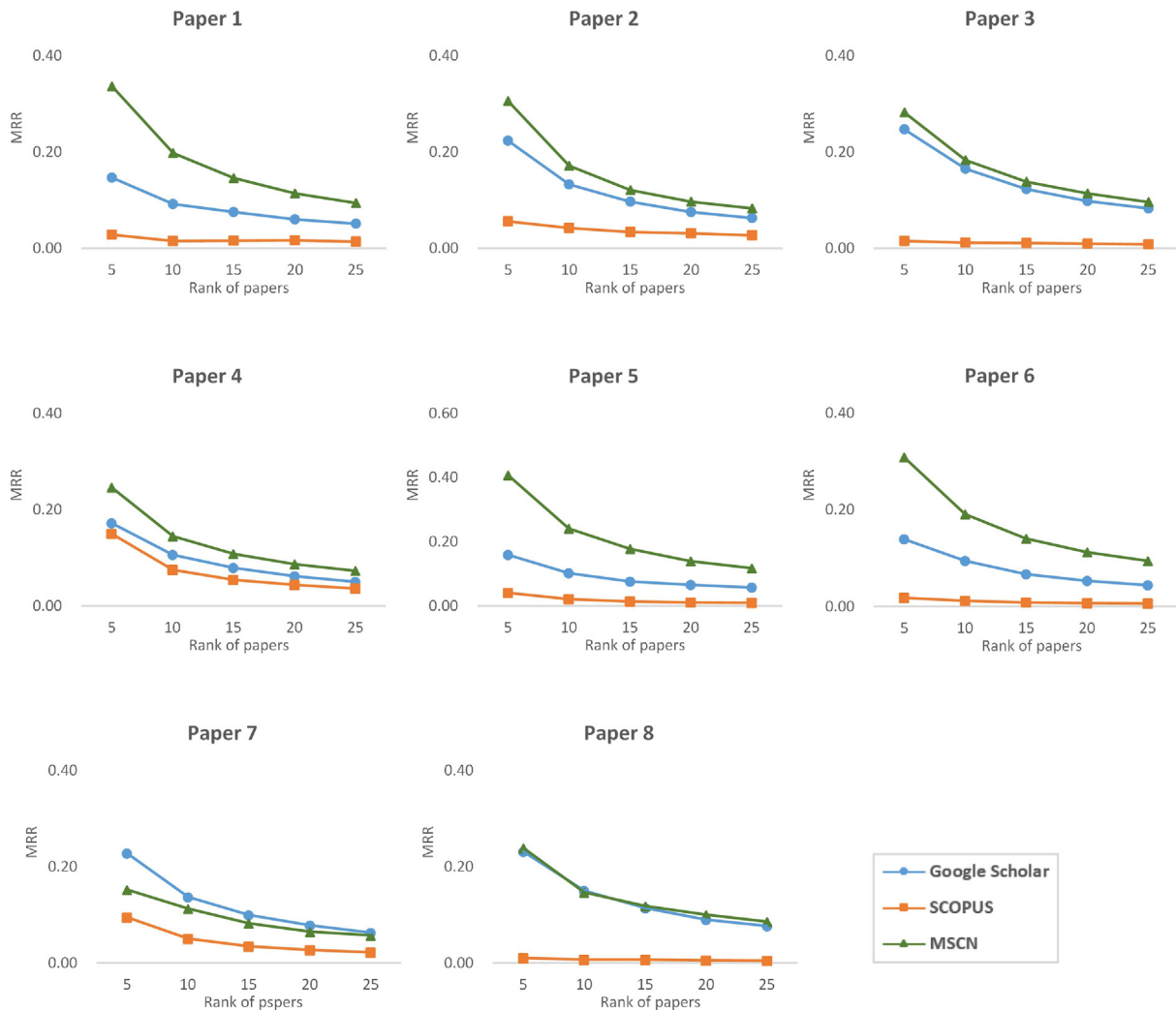
Fig. 6. Performance comparison of Google Scholar, SCOPUS, and MSCN with MRR.

than single level networks to inspect structural and complex relationships among papers.

From our experimental results and analysis, we highlight the following interesting findings and practical implications. Although PageRank in Google Scholar is a good method for determining the authority of a
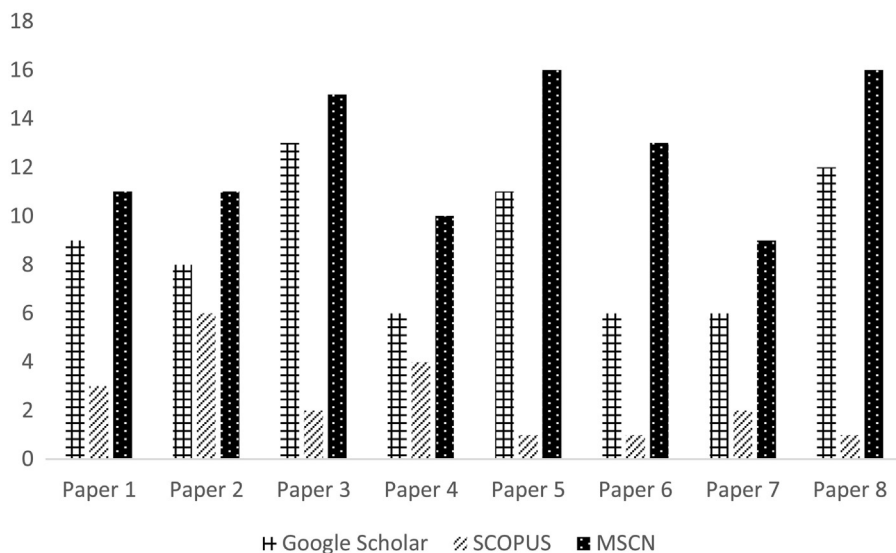

Fig. 7. Average number of correctly recommended papers.

paper, it tends to rank papers based largely on the number of citations. For example, when a recently published paper was selected as the paper of interest, Google Scholar tended to recommend papers published recently with a low citation count because a recent paper of interest does not have sufficient citation information. This is a deficiency that confused the ranking algorithm, which is based on citation information. Moreover, when an old paper is selected as the paper of interest, most of the recommended papers from Google Scholar are outdated. These results show that the PageRank technique originally designed for search engines can create limitations when used in paper recommendations. The proposed MSCN addresses this limitation by considering the relationship of "cite" and "cited by" equally. Consequently, the MSCN tended to find satisfactory papers regardless of the publication date of the paper of interest. The major limitation of SCOPUS is to consider only the co-occurrence of keywords without reflecting the context of the papers. SCOPUS recommended papers containing terminology or names of methodologies from the paper of interest, even when the context of the paper was not related to it. On the contrary, the proposed MSCN ensures quality and diversity of recommendation results because the proposed method uses citation relationships that provide reliable information beyond the feature similarity.

For future study, we would like to incorporate additional features other than citation information into our recommender system. Although citation information is important, it may be insufficient for recommending the most appropriate papers. Additional features such as authors and journals can be incorporated into each node to construct more informative networks. Further, we will consider the number of citations in the body of the paper to measure the relationship between the papers. If the paper of interest is cited multiple times in a particular paper, then that paper is more likely to be related to the paper of interest. However, our current method does not reflect the number of papers cited in the body of the paper because the method uses the reference section for citation analysis. We believe that if the number of citations in the body of the paper is included, our proposed recommendation method can be improved and more meaningful, assuming that the paper of interest is cited once in a specific paper. We will keep this issue as the priority for future study.

Jieun Son is currently a Ph.D. candidate in the Department of Industrial Management Engineering in Korea University at Seoul, Korea. Her research interests include recommender systems, association rule, social network analysis, and text mining. She won the Best Paper Award in SAS competition of Korea Business Intelligence Data Mining Society.

Seoung Bum Kim is a professor in the School of Industrial Management Engineering at Korea University. From 2005–2009, he was an Assistant Professor in the Department of Industrial & Manufacturing Systems Engineering at the University of Texas at Arlington. He received an M.S. in Industrial and Systems Engineering in 2001, an M.S. in Statistics in 2004, and a Ph.D. in Industrial and Systems Engineering in 2005 from the Georgia Institute of Technology. He has expertise in the machine learning algorithms and their applications including network analysis, recommender systems, manufacturing system, and healthcare system. He has published more than 100 internationally recognized journals and refereed conference proceedings. He was awarded the Jack Youden Prize as the best expository paper in Technometrics for the Year 2003. He is actively involved with the INFORMS, serving as president for the INFORMS Section on Data Mining.

Acknowledgments

References

[1] J. Lu, D. Wu, M. Mao, W. Wang, G. Zhang, Recommender system application developments: a survey, Decis. Support Syst. 74 (2015) 12–32.

[2] M.A. Domingues, A.M. Jorge, C. Soares, Dimensions as virtual items: improving the predictive ability of top-n recommender systems, Inf. Process. Manag. 49 (3) (2013) 698–720.

[3] O. Kassak, M. Kompan, M. Bielikova, User preference modeling by global and individual weights for personalized recommendation, Acta Polytech. Hung. 12 (8) (2015) 27–41.

[4] R. Dong, L. Tokarchuk, A. Ma, Digging friendship: paper recommendation in social network, Proceedings of Networking & Electronic Commerce Research Conference (NAEC 2009) 2009, pp. 21–28.

[5] T. Strohman, W.B. Croft, D. Jensen, Recommending citations for academic papers, Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM 2007, July, pp. 705–706.

[6] Lops, P., De Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: state of the art and trends. In Recommender Systems Handbook (pp. 73–105). Springer US.

[7] Y. Liang, Q. Li, T. Qian, Finding relevant papers based on citation relations, Web-age Information Management 2011, pp. 403–414.

[8] S. Brin, L. Page, Anatomy of a large-scale hypertextual web search engine, 7th Intl World Wide Web Conf, 1998.

[9] J. Beel, B. Gipp, Google Scholar's ranking algorithm: the impact of citation counts (an empirical study), 3rd International Conference on Research Challenges in Information Science, 2009. RCIS 2009, IEEE 2009, April, pp. 439–446.

[10] K.W. Cheung, J.T. Kwok, M.H. Law, K.C. Tsui, Mining customer product ratings for personalized marketing, Decis. Support Syst. 35 (2) (2003) 231–243.

[11] C. Cechinel, M.Á. Sicilia, S. Sánchez-Alonso, E. García-Barriocanal, Evaluating collaborative filtering recommendations inside large learning object repositories, Inf. Process. Manag. 49 (1) (2013) 34–50.

[12] H. Kautz, B. Selman, M. Shah, Referral Web: combining social networks and collaborative filtering, Commun. ACM 40 (3) (1997) 63–65.

[13] T.P. Liang, Y.F. Yang, D.N. Chen, Y.C. Ku, A semantic-expansion approach to personalized knowledge recommendation, Decis. Support Syst. 45 (3) (2008) 401–412.

[14] F.M. Belém, E.F. Martins, J.M. Almeida, M.A. Gonçalves, Personalized and object-centered tag recommendation methods for Web 2.0 applications, Inf. Process. Manag. 50 (4) (2014) 524–553.

[15] E. Negre, Information and Recommender Systems, John Wiley & Sons, 2015.

[16] B. Gipp, J. Beel, C. Hentschel, Scienstein: a research paper recommender system, Proceedings of the International Conference on Emerging Trends in Computing (ICETIC'09) 2009, January, pp. 309–315.

[17] Google Scholar, https://scholar.google.com/intl/en/scholar/about.html.

[18] N. Ford, The Essential Guide to Using the Web for Research, Sage, 2011.

[19] S. Bethard, D. Jurafsky, Who should I cite: learning literature search models from citation behavior, Proceedings of the 19th ACM International Conference on Information and Knowledge Management, ACM 2010, October, pp. 609–618.

[20] F. Galgani, P. Compton, A. Hoffmann, Summarization based on bi-directional citation analysis, Inf. Process. Manag. 51 (1) (2015) 1–24.

[21] H. Small, Co-citation in the scientific literature: a new measure of the relationship between two documents, J. Assoc. Inf. Sci. Technol. 24 (4) (1973) 265–269.

[22] Y. Liang, Q. Li, T. Qian, Finding relevant papers based on citation relations, Web-age Information Management 2011, pp. 403–414.

[23] Z. Huang, W. Chung, T.H. Ong, H. Chen, A graph-based recommender system for digital library, Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries, ACM 2002, July, pp. 65–73.

[24] D. Zhou, S. Zhu, K. Yu, X. Song, B.L. Tseng, H. Zha, C.L. Giles, Learning multiple graphs for document recommendations, Proceedings of the 17th International Conference on World Wide Web, ACM 2008, April, pp. 141–150.

[25] M. Gori, A. Pucci, Research paper recommender systems: a random-walk based approach, IEEE/WIC/ACM International Conference on Web Intelligence, 2006. WI 2006, IEEE 2006, December, pp. 778–781.

[26] N. Lao, W.W. Cohen, Relational retrieval using a combination of path-constrained random walks, Mach. Learn. 81 (1) (2010) 53–67.

[27] B. Gipp, J. Beel, C. Hentschel, Scienstein: a research paper recommender system, Proceedings of the International Conference on Emerging Trends in Computing (ICETIC'09) 2009, January, pp. 309–315.

[28] L. Egghe, R. Rousseau, Co-citation, bibliographic coupling and a characterization of lattice citation networks, Scientometrics 55 (3) (2002) 349–361.

[29] K.K. Lai, S.J. Wu, Using the patent co-citation approach to establish a new patent classification system, Inf. Process. Manag. 41 (2) (2005) 313–330.

[30] De Bellis, N. (2009). Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics. Scarecrow Press.

[31] B. Carrera, J. Lee, J.Y. Jung, Discovery of information diffusion process in online social networks, Int. J. Ind. Eng. Theory Appl. Pract. 23 (4) (2016).

[32] B. Zhu, S. Watts, H. Chen, Visualizing social network concepts, Decis. Support Syst. 49 (2) (2010) 151–161.

[33] T. Opsahl, F. Agneessens, J. Skvoretz, Node centrality in weighted networks: generalizing degree and shortest paths, Soc. Networks 32 (3) (2010) 245–251.

[34] K. Okamoto, W. Chen, X.Y. Li, Ranking of closeness centrality for large-scale social networks, Lect. Notes Comput. Sci. 5059 (2008) 186–195.

[35] L.C. Freeman, Centrality in social networks conceptual clarification, Soc. Networks 1 (3) (1978) 215–239.
[36] P. Bonacich, Some unique properties of eigenvector centrality, Soc. Networks 29 (4) (2007) 555–564.
[37] Y. Hong, S. Kwong, Y. Chang, Q. Ren, Consensus unsupervised feature ranking from multiple views, Pattern Recogn. Lett. 29 (5) (2008) 595–602.
[38] H.F. Moed, J. Bar-Ilan, G. Halevi, A new methodology for comparing Google Scholar and Scopus, J. Informet. 10 (2) (2016) 533–551.
[39] M. Deshpande, G. Karypis, Item-based top-n recommendation algorithms, ACM Trans. Inf. Syst. 22 (1) (2004) 143–177.
[40] P. Cremonesi, Y. Koren, R. Turrin, Performance of recommender algorithms on top-n recommendation tasks, Proceedings of the 4th ACM Conference on Recommender Systems, ACM 2010, September, pp. 39–46.
[41] R. Agrawal, S. Gollapudi, A. Halverson, S. Ieong, Diversifying search results, Proceedings of the 2nd ACM International Conference on Web Search and Data Mining, ACM 2009, February, pp. 5–14.
[42] H. Stuckenschmidt, Approximate information filtering on the semantic web, Annual Conference on Artificial Intelligence, Springer, Berlin, Heidelberg 2002, September, pp. 114–128.
[43] J.A. Torkestani, An adaptive learning to rank algorithm: learning automata approach, Decis. Support Syst. 54 (1) (2012) 574–583.
[44] P. Adamopoulos, On discovering non-obvious recommendations: using unexpectedness and neighborhood selection methods in collaborative filtering systems, Proceedings of the 7th ACM International Conference on Web Search and Data Mining, ACM 2014, February, pp. 655–660.
[45] T.P. Liang, Recommendation systems for decision support: an editorial introduction, Decis. Support Syst. 45 (3) (2008) 385–386.
[46] J. Beel, B. Gipp, Google Scholar's ranking algorithm: an introductory overview, Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09), 1, 2009, July, pp. 230–241.
[47] A. Pal, S. Ruj, CITEX: a new citation index to measure the relative importance of authors and papers in scientific publications, 2015 IEEE International Conference on Communications (ICC), IEEE 2015, June, pp. 1256–1261.
[48] D.F. Klosik, S. Bornholdt, The citation wake of publications detects Nobel laureates' papers, PloS One 9 (12) (2014), e113184. .
[49] M. Gori, A. Pucci, Research paper recommender systems: a random-walk based approach, IEEE/WIC/ACM International Conference on Web Intelligence, 2006. WI 2006, IEEE 2006, December, pp. 778–781.