

A Novel Personalized Academic Knowledge Sharing System in Online Social Network

Pengfei Zhao
University of Science and Technology of China
City University of Hong Kong
pengfei.zhao@outlook.com

Jian Ma
Department of Information Systems
City University of Hong Kong
isjian@cityu.edu.hk

Zhongsheng Hua
School of Management
University of Science and Technology of China
zshua@ustc.edu.cn

Shijian Fang
School of Management
University of Science and Technology of China
fangsj@ustc.edu.cn

Abstract

Information overload is a major problem for both readers and authors due to the rapid increase in scientific papers in recent years. Methods are proposed to help readers find right papers, but few research focuses on knowledge sharing and dissemination from authors' perspectives. This paper proposes a personalized academic knowledge sharing system that takes advantages of author's initiatives. In our method, we combine the user-level and document-level analysis in the same model, it works in two stages: 1) user-level analysis, which is used to profile users in three dimensions (i.e., research topic relevance, social relation and research quality); and 2) document-level analysis, which calculates the similarity between the target article and reader's publications. The proposed method has been implemented in the ScholarMate, which is a popular academic social network. The experiment results show that the proposed method can effectively promote the academic knowledge sharing, it outperforms other baseline methods.

1. Introduction

The rapid growth of online content is making significant challenge to the knowledge sharing and discovery, especially in the academic area. According to the report of AJE (American Journal Experts)¹, over 2.2 million scientific papers were published in 2016. The proliferation of scientific papers causes the

information overload problem to both the readers and authors. The readers have to spend lots of time to find the relevant papers, while the authors are concerned that their publications may not reach the right readers.

Search engines (e.g., Web of Science and Google Scholar) are developed and used as powerful tools to find scientific articles. Users can enter keywords (e.g., title, author, journal etc.) in the search engine to get answers, but sometimes it is hard for researchers to express their ideas with keywords. They have to revise the keywords continuously until finding the satisfied results. In reality, researchers may have many research areas, but the search engines only return the same results even though the paper belongs to different research areas (or different classifications), it neglects the semantic meaning of articles and the users' research areas. If researchers do not have a clear purpose and are not familiar with a specific research area, find relevant papers can be hard with search engines. Since the 1990s, recommender systems are proposed to improve search efficiencies [1, 2], they are also implemented in real commercial environments, such as electronic commerce, music, movie and document [3-6]. Recent studies are conducted to understand how academic recommender systems facilitate the readers in finding relevant papers[7], and scientific article recommendation has become a hot research topic with great attentions.

Current recommender systems are designed to help readers to eliminate the information overload problems, few studies address the problems from authors' perspectives. The fact is that the authors' papers are not widely cited although they are relevant

¹<http://www.aje.cn/>

to the readers. The numbers of citations are important and common metric to measure their research impacts. Based on the study of Redner [8], the citation distribution is a rapidly power law decay, about 47% of the papers belong to the ISI (Institute for Scientific Information) dataset have no citations and more than 80% of the papers have less than 10 times citations. This report is also consistent with the Adler, Ewing and Taylor [9], the average citation per paper in mathematics and computer disciplines is less than 1 time, while the number of citation is less than 3 in six disciplines. Similarly, 70% of the papers published in the proceeding of the American mathematical society have no citation, 20% have only 1 time citation. The citation distribution reflects a very serious problem, that is most of the academic knowledge is not get attention or not get attention from the right readers.

Previous studies show that the citation is affected by sociological and statistical factors, such as the discipline of the journal, the author's reputation, the type of the journal (review, paper, conference), the measurement window, the cited references and the abstract readability[10-12]. Besides the traditional factors, recent studies[13] prove that open access articles (OA) have the citation advantage compared with non-open access articles (non-OA), moreover, non-OA articles can only get a short period of attention. OA citation advantage indicates that improving accessibility can significantly promote the knowledge sharing.

With the rise of academic social networks, such as Mendeley², CiteULike³, ResearchGate⁴ and ScholarMate⁵, the academic social network provides an online platform for the researchers to upload their publications and share articles with readers. Studies about knowledge sharing indicate trust, outcome expectation and individual benefit [14-16] is the main factor influence the knowledge sharing behavior. Authors are motivated to actively share knowledge to the potential readers who have interests in their work, this share behavior helps the knowledge reach the right person and effectively spread author's knowledge and idea, it is to increase author's reputation and paper citations [14-16]. From the authors' perspective, the author can share their published papers to candidate readers freely and timely. It enables lots of readers who have no paper access to get the paper resources on the online social networks. This is significant for the high school students, the researchers in underdeveloped areas, and the industry people (such as engineer, salesman, manager and so on). Because these kinds of user groups have more difficulties in

locating papers that are relevant to their interests. And our proposed methods can also provide the recommendations based on their user-level analysis. Overall, from the authors' perspective, the knowledge sharing systems can significantly expand the audience and address the cold-start problem. It is a challenge research topic to design a personalized recommendation system to facilitate the article sharing in an academic social network platform.

This paper proposes a novel knowledge sharing system to facilitate the academic knowledge sharing from the authors' perspectives. For a target article, it will be recommended to a list of readers who have the potential interests. We identify the potential readers by combining the user-level analysis and document-level analysis. In the user-level analysis, three dimensions are take into consideration, namely, research topic dimension, social relation dimension, and research quality dimension. Research topic semantic relations are constructed to deal with the sparse of user-topic matrix and capture the rich content information in topics. In social relation dimension, CJC (Local-Community Jaccard) technique is used to measure the trust by social closeness. Research quality dimension evaluate the researcher's performance. In document-level analysis, the document similarity is calculated between target article and reader's publications. Drawing on the knowledge sharing perspective, the first advantage of this work is incorporating the author's subjective initiative and free knowledge sharing activity social network into the article recommendation. It can significantly change the way of academic knowledge sharing and discovery, especially in the areas with frequent new publications. The second advantage is combining the user-level and document-level analysis in the same model, through the user-level analysis, the candidate reader is produced to reduce computational complexity for future steps. Finally, the document-level analysis is added to precisely share the target article to the right hand.

The remainder of the paper is structured as follows. In section2, related research is reviewed. In section3, a two-stages academic knowledge sharing system was proposed. In section4, we conducted an experiment to validate our approach. Section 5 is the conclusion and limitations.

2. Literature review

² www.mendeley.com

³ www.citeulike.org

⁴ www.researchgate.net

⁵ www.scholarmate.com

This section reviews the reviews the literature related to this study. Firstly, we summarized the influencing factors of knowledge sharing. Secondly, we reviewed the literature on recommender systems. As a powerful personalized technique, recommender systems have been widely used to deal with the problems of information overload and information asymmetry.

The antecedents of knowledge sharing have been investigated in the virtual community, Chiu, Hsu and Wang [17] integrate the social capital and social cognitive theories to understand how could trust and outcome expectation influence the user's motivation to share knowledge. These factors are further validated by Lin, Hung and Chen [18]. In the academic social network, knowledge sharing (i.e., paper sharing) between the authors and the readers are also influenced by two factors. On the one hand, drawing on social learning and social networking theories, Thoms, Garrett, Herrera and Ryan [19] indicate that learner perceives a high levels of value from the trusted users. Sinha and Swearingen [20] point out that the users are more like to accept the recommendations come from the people they trust than some anonymous user who have the similar interests with them. On the other hand, in order to gain a higher impact factor and better reputation in academia, the authors are motivated to share their publications to the potential readers who have interests in their work.

Recommender systems are used to predict user's possible future interests. By extracting the user data (e.g., user's preference, demographic information and social relationship), the system can help to profile the target users. In practice, recommender systems bring various benefits for the service providers. In the online service context, service providers implement the recommender system to improve the user's satisfaction and can significantly gain the user's loyalty, thus attracting more users to their academic platform. From the readers' perspective, scientific article recommendation can facilitate the searching process. Readers can easily find the relevant articles, which are necessary for their rapid progress in research. This is especially useful for junior researchers and PhD students, who are lack of domain knowledge. From the authors' perspective, authors can actively share their publications with the potential readers, making their work more widely available to large audiences and more precisely arrive the right hands. It ensures the academic knowledge continuity and increases the authors' research impacts.

Current research in article recommendation can be generally categorized into three categories: content-based(CB) methods, collaborative filtering(CF) methods and hybrid methods. Content-based approach

tries to recommend the items that users have interest in the past. The CB[21] methods use txt mining and txt processing to deal with the txt information, like news and articles. The txt information comes from the items that have interactions with users. Collaborative filtering approach[22] tries to recommend what target users' neighbors have interests in. The neighbors are selected if they have similar preferences with the target user. The assumption here is if the two users have the similar ratings in some items, then they will also have the similar tastes in other items. The collaborative filtering methods can be further divided into user-based CF and item-based CF. Hybrid approach is not one specific recommender system, but a combination of several techniques. A hybrid recommender system [23] is designed to overcome the disadvantages of a single approach and utilize the advantages of each approach. For example, traditional collaborative filtering method is suffered from the cold-start problem, the new users and new items have no rating data. However, for the content-based method, the prediction is usually generated based on the attributes of the items, but not the rating.

Bogers and Van den Bosch [24] validate three different CF approaches, which are item-based CF approach that uses cosine similarity, item-based CF approach that uses conditional probability, and user-based CF approach that uses cosine similarity. Lee, Lee and Kim [25] use txt similarity to calculate the similarity among different articles, then use collaborative filtering to build their model. To address the sparsity of citation network, Sugiyama and Kan [26] first implement the collaborative filtering to identify the potential citation papers, then used to enhance the traditional citation network.

Due to the cold-start problem and rich content information cannot be utilize, CB has the better performance than CF in document recommender system[27]. Chandrasekaran, Gauch, Lakkaraju and Luong [28] build the user profile based on the user's previously publications, and also introduced a concept-based algorithm to compare with the traditional vector-space model. Silva, Guo, Ma, Jiang and Chen [29] design a user interface by which users can input the topics, and then the system outputs a list of articles. The users are then required to select which articles is relevant to them. Finally, the title, keywords and abstract of the selected articles are crawled to build the user profile. Magalhaes, Souza, Costa and Fechine [30] proposed a novel approach to build user profile, they take user's curriculum vitae into consideration.

With the development of social networks, large amounts of data that reflects user characteristics and research interests are generated from the social

networks, thus social network has been taken into consideration.[31-33].

However, the above methods didn't consider the semantic meaning, thus causing mismatch problem in the recommendation process. Sun, Ma, Liu and Miao [34] use the content-based method and calculate the keywords similarity to analyze the semantic meaning of articles to recommend research papers. Al-Hassan, Lu and Lu [35] propose a semantic enhanced hybrid recommendation approach, which incorporates the ontology into the measurement of the semantic similarity. However, these methods still have their disadvantages. Some among them are using the existing semantic relation, like WordNet, which is not applicable to the academic areas. Some are using a small sub-set of the user's publications as the corpus to obtain the semantic meaning, which is only suitable for their unique task. More important is all this method is conducted in the document-level analysis, it is impossible to implemented in a real academic social network which has millions of users and articles. As the important factors that can influence the readers' acceptance rate, trust and quality are being neglected. Also, few research focuses on knowledge sharing and dissemination from authors' perspectives. Our study tries to address these research gaps by taking advantage of the author's subjective initiative and combine the user-level and document-level analysis in the same model.

3. Academic knowledge sharing system

The proposed academic knowledge sharing system is trying to solve the information overload and information asymmetry problem from the authors' perspective. The system can help the authors effectively spread their publications on the social network. It is a new channel for the authors to disseminate their publications in a free and efficient way. Figure 1 shows the framework of our academic knowledge sharing system, there are two main modules in our system: the user-level analysis and the document-level analysis, the user-level analysis measures the user's research topic relevance, social relationship closeness and research quality. The document-level analysis measures the similarity between author's a specific article and candidate user's publications. The details of our system will be introduced in the following sections.

3.1. User-level analysis

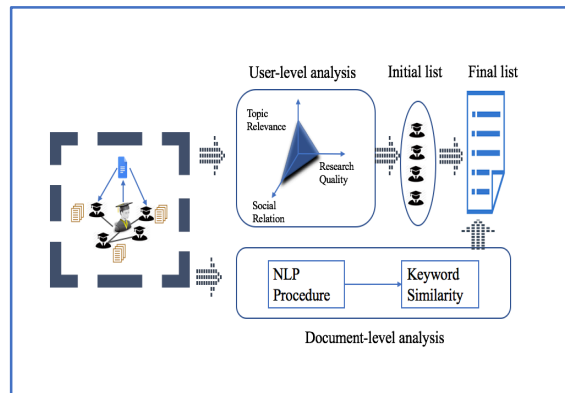


Figure 1. The framework of the academic knowledge sharing system

In the user-level analysis module, we analyze the user characteristics from three dimensions, which is research topic relevance, relation connectivity and user's research quality. In the research topic relevance dimension, we first construct the topic semantic relation, then take the social endorse as the weight to calculate the topic relevance. In social relation dimension, the social relationship is extracted. Then we analyze the social closeness by considering both the friendship relation and co-authorship relation. In the quality dimension, the user's publications and projects are used to evaluate the scholars' research performance. User-level analysis module is essential for our system, not only it can effectively characterize the user, but also it can calculate in advance as the pre-filtering step to get the initial candidate users. This can significantly reduce the calculation time for the online academic social network, which has millions of users and articles.

3.1.1. Research topic relevance analysis. The research topic is the most important and useful factor to determine the users' research interests. Research topic relevance analysis is used to measure the research interests' similarity among the users. We first derive the user-topic matrix from user's academic social network homepage. And then we use the matrix to construct the topic semantic relation. At last, we calculate the users' research topic relevance to identify the candidate users.

Research topic is filled and claimed by the users in their homepages. It is a good reflection of their research interests. In the social network, the user's friends can endorse the user's research topics based on their own judgment and recognition. We extract the user-topic matrix and social endorse to profile their research interests. The initial user-topic matrix is a $n_u \times n_t$ matrix, that is used to indicate the relevance

between users and topics, where the n_u is the number of users and n_t is the number of topics, 1 or 0 is used to represent whether the user have the topics or not respectively. One user usually has several topics and the user-topic matrix is the one-hot representation, thus making the matrix to be very sparse. It leads to the curse of dimensionality, as such, the matrix cannot describe semantic topic relationship. But the reality is that user's topics have the latent semantic relationship, for example, a user has three topics, such as big data, social network and machine learning. the social network can generate big data, and machine learning technique is widely used to deal with the big data problem, as such, the user who focuses on the social network topic may also has some interests in the big data or machine learning topic. Although the users sometimes do not fill their research topics completely in their homepage, there is still possibility to find out all the topics by considering the latent semantic relationship among them.

In this paper, we will take the co-occurrence of topics into consideration to obtain the topics semantic relationship. Specifically, we used the Jaccard similarity method, which is the most popular statistic method used for comparing the semantic relationship [36] to measure the topics relationship. The average semantic relations of all the user's self-filled topics with the target topic is used to represent the potential interest of user to target topic, seen as Equation (1), based on the statistical analysis of millions of the users' data, it can produce a good predict performance to complement the user-topic matrix. The user's self-filled topics have the same weight in the matrix, but users do have different focus, thus we combine the social endorse come from the user's friends own judgement to get the topic weight, see as Equation (2)

$$P(u, y) = \sum \frac{W_x \times J(x, y)}{m} = \sum \frac{W_x \times \sum_i \min(x_i, y_i)}{m \times \sum_i \max(x_i, y_i)} \quad (1)$$

$$W_x = \alpha w_{claimed} + \beta w_{social} = \alpha \frac{\tau}{m} + \beta \frac{s_\tau}{\sum s_\tau} \quad (2)$$

Where, $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ are two topic vectors, m is number of user u 's self-filled topics, $P(u, y)$ are potential interests between user u and topic y , W_x is the weight of topic x , s_τ is the endorse frequency of topic τ . α and β is the weight to adjust the user claimed weight and social weight, here set $\alpha = 0.67, \beta = 0.33$.

After computing the topic semantic relationship and topic weight, the user-topic matrix can be complemented. From one-hot representation changed to distributed representation, such as can effectively address the sparse problem. Then the cosine similarity method is used to calculate the user's topic similarity, see as follows:

$$RTR(A, B) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (3)$$

Where, $A = (A_1, A_2, \dots, A_n)$, $B = (B_1, B_2, \dots, B_n)$ are two user's vectors.

3.1.2. Social relation analysis. In academic social network, users build connections with their familiar friends, colleagues, co-authors. It is also important for them to build connections with online users have the similar research interests. Previous studies indicate that users are more likely to take the items that come from users who they are trust [19, 20] but not the anonymous user who have the similar interests with them.

Social relation analysis can quantify the user trust based on their social network closeness. It has the assumption that the higher closeness reflects higher trust. In our research, the social network is first constructed based on multiple relation types, such as friendship, co-author of publications and projects. Social network analysis has attracted a lot of attentions, previous studies mainly focused on the common nodes and their neighbors in the network, Adamic & Adar (AA) and Common Neighbors (CN) is the most used methods. Recently, CJC methods is proposed, which has been proved have better performance in complex network [37], CJC suggests that two nodes have the high closeness when the common first nodes are members of a local-community. Therefore, in our research, the CJC is used to analyze the social relation. The CJC index is defined as follows:

$$CN(A, B) = |\Gamma(A) \cap \Gamma(B)| \quad (4)$$

$$LCL(A, B) = \sum_{s \in \Gamma(A) \cap \Gamma(B)} \frac{|\gamma(s)|}{2} \quad (5)$$

$$CAR(A, B) = CN(A, B) * LCL(A, B) \quad (6)$$

$$SR(A, B) = CJC(A, B) = \frac{CAR(A, B)}{\Gamma(A) \cup \Gamma(B)} \quad (7)$$

where, A and B are nodes in the network, $\Gamma(A)$ refers the neighbors of A , $\gamma(s)$ is the subset of s , which also has the internal connections, $\gamma(s)$ is the local-community degree of s , $SR(A, B)$ is the relation proximity of user A and B .

3.1.3. Research quality analysis. Research quality analysis reflect the user's academic performance [29, 38], different users have multiple needs of article quality. For example, a user published in the top journals may have little interests to the articles which are published in the low impact factor journals, even if they have the same research topics or in the same research areas. Thus, measure the user's research quality and articles' quality is essential for academic

knowledge sharing in online social networks. In our research, user's quality is evaluated by the quantity and quality of their publications and quantity of their funded projects. The journal impactor adopted from the Journal Citation Reports(JCR), which is widely used to evaluate the journal level. Based on the journal rank in the JCR, we divide the journals into three levels, which is level a, level b and level c, different weight is set to different levels according to the experiment. The publication quality is measured by Equation 8. Also, the number of funded projects are another dimension to measure the users' research quality. The publication quality and projects quality are firstly being normalized in order to aggregate to user's research quality, see as Equation 9.

$$Q_{pub_A} = w_a q_{Aa} + w_b q_{Ab} + w_c q_{Ac} \quad (8)$$

$$Qua_A = \gamma Q_{pub_A} + \rho Q_{pro_A} \quad (9)$$

where, the Q_{pub_A} and Q_{pro_A} is user i's publication quality and project quality respectively, the q_{Aa} , q_{Ab} and q_{Ac} is the quantity of the papers published in level a, level b and level c by user i, the w_a , w_b , and w_c are the corresponding weights, in our research, the $w_a = 3.09$, $w_b = 2.17$, $w_c = 0.74$. γ and δ are the parameter to adjust the weight of publication quality and project quality, and $\gamma + \rho = 1$, here in our study, the $\gamma = 0.7$ and $\rho = 0.3$.

3.2. Document-level analysis

After the user-level analysis, for each author, we can obtain their candidate readers from millions of online users. But author's each specific article has the unique and different characteristics, therefor, document-level analysis is essential for making a good knowledge sharing system. In our research, article similarity between target article and candidate readers' publication is used to determine whether the article is sharing to the reader or not.

Traditionally, author-assigned keywords are used to represent the article's main research interest. However, each article only has several keywords and these keywords are not standardized. The limited and unstandardized keywords cannot enough express the content of the article. The abstract and title still contain the important information about the article, moreover, the keywords-article matrix is very spares because the low quantity of the keywords, thus cause the inaccurate match. In our research, not only author-assigned keywords, the abstract and title are also extracted to profile the article, the different weight is assigned respectively. Natural language processing(NLP) techniques is used to process the text,

such as segmentation and filtering the stop words. The article is initially represented by a bag of words, then the term frequency-inverse document frequency(TF-IDF) methods is adopted to process every article. TF-IDF, which is belong to statistical method, is the most widely used technique in information retrieval[39]. TF-IDF can reflect the importance of a word to a document in a corpus. It consists of two parts, term frequency and inverse document frequency. Term frequency(TF) measure the number of times a word occurs in a document, the problem is the common words will have the high TF, but these common words usually have less useful for distinguish the content of documents. Therefore, inverse document frequency(IDF) is used to reduce common words' weight. After calculating the weight of each words in every document, vector space model(VSM) algorithm is selected to represent document into N-dimensional vector. VSM can facilitate the computation of document similarity, the detail is see as follows:

$$TF_{ij} = \frac{n_{ij}}{m_j} \quad (10)$$

$$IDF_i = \log\left(\frac{|D|}{|d:d \ni t_i|}\right) \quad (11)$$

$$TFIDF_{ij} = TF_{ij} * IDF_i \quad (12)$$

where, TF_{ij} is the frequency of word i in document j, n_{ij} is the number of word i occurred in document j, m_j is the total of words in document j. $|D|$ refers to the number of documents in the corpus, $|d:d \ni t_i|$ is the number of documents which own word t_i .

$$W_i = \varepsilon w_{key} + \mu w_{tit} + \theta w_{abs} \quad (13)$$

$$SVM = D(T_1, W_1; T_2, W_2; \dots, T_n, W_n) \quad (14)$$

$$DLA(i, j) = Cos(D_i D_j) = \frac{D_i D_j}{\|D_i\| \|D_j\|} \quad (15)$$

where, w_{key} , w_{tit} and w_{abs} represent the value of TF-IDF in keywords, title and abstract. α , β and γ are the weight of w_{key} , w_{tit} and w_{abs} , in present research, set $\varepsilon = 0.5$, $\mu = 0.3$, $\theta = 0.2$. Document D is represented by the words and its weight as the vector. $Cos(D_i D_j)$ is the cosine similarity of document D_i and D_j .

3.3. Two-stage academic knowledge sharing system

To address the academic information overload, helping authors effectively spread their publications. we proposed a two-stage knowledge sharing system. The framework is showed in Figure 1.

In the first stage, user-level analysis is conducted to produce the candidate readers for the next stage. The necessity is user-level analysis can be calculated in advance, and in a long period can remain stable, this is essential in the academic social network which has millions of users. Three dimensions are being taking into consideration in user-level analysis, thus we need to aggregate them into one indicator to obtain the final candidate ranking. The research topic relevance and social relation is most important, research quality as the regulator to adjust the ranking list. The aggregate score of two users is represent as Equation 15. Although the potential readers have been calculated, but for target user to share their publications to others, each specific article has the unique characteristics, thus the document-level analysis should be combined with the user-level analysis to get the best performance, see as Equation 16.

$$ULA_{A,B} = (RTR(A, B) + SR(A, B)) \left(1 + \frac{Qua_A - Qua_B}{Qua_B}\right) \quad (15)$$

$$FS(A_i, B) = \varphi ULA_{A,B} + \omega \sum \frac{DLA(i, j)}{n} \quad (16)$$

Where, $ULA_{A,B}$ is the aggregate score of user A and B in user-level analysis, take A as the target author, the other user i who have the high $ULA_{A,i}$ will ranking in front, $(Top(A, B) + Con(A, B)) < \delta$ is defined to maintain the lowest topic relevance and relation connectivity, δ is the threshold. $FS(A_i, B)$ is the final score of author A 's article i to user B , j is the publication of user B , n is the number of publications of user B , φ and ω are parameter set to balance the user-level and document-level weight, and $\varphi + \omega = 1$, here, the $\varphi = 0.41$ and $\omega = 0.59$.

4. Experiment evaluation

To evaluate whether our proposed method can effectively affect the academic knowledge sharing, we conducted an experiment in ScholarMate, which is one of the most popular academic social network in China. ScholarMate has more than 2.2 million users and 18 million publications. The user's homepage contains multiple information, such as their research topics, publications, projects, and social relation. ScholarMate can automatically collect the user's publications and projects, user only need to click and confirm it. This facility the user to share their publications to others, also, in the ScholarMate, users can make connection with each other, comment and share another user's publications.

Our proposed academic knowledge sharing systems has been implemented in the ScholarMate as the value-added services. Thus, the registered user can be used to evaluate the performance of the system. Figure 2 shows the interface of user's homepage and our proposed academic knowledge sharing system on ScholarMate.

In our experiment evaluation, 40 users were randomly selected from the user database as the target users to actively share their publications, that each of them has at least one publication. For each target user, their latest publication as the target article to share is selected, for each target article, top-N reader is generated by different algorithms, which is content-based(CB) method which is using the keywords match, content-based with semantic enhanced method (CBS), algorithm that only consider the research topic relevance(RTR), algorithm that only consider social

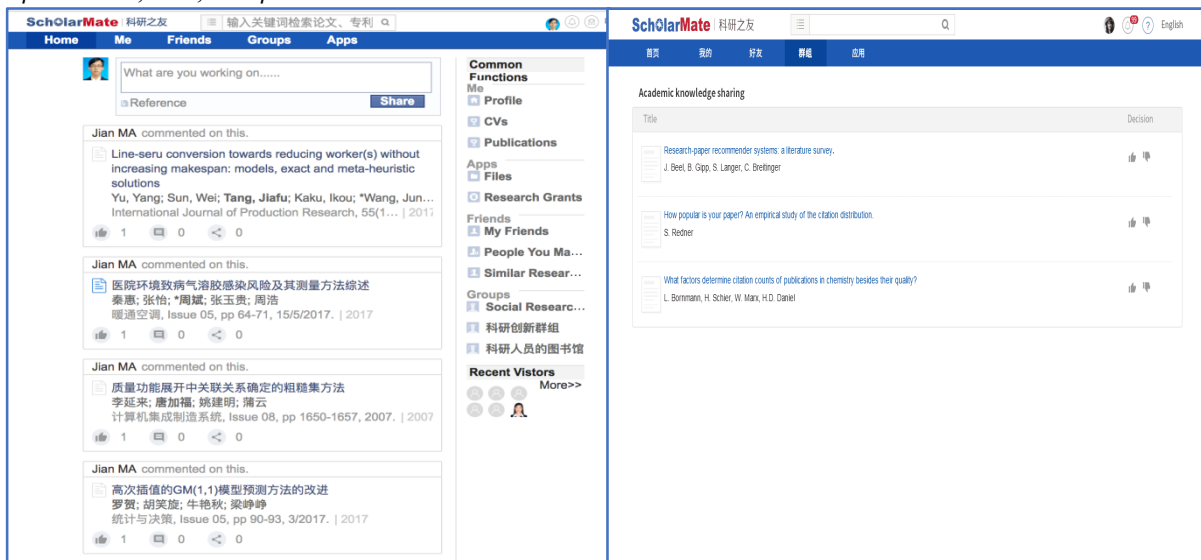


Figure 2. Interface of user homepage and academic knowledge sharing system in ScholarMate

relation(SR), algorithm that based on user-level analysis(ULA) and our proposed method(ULA-DLA). After the reader received the target article, the decision should be made whether accept it or not based on their judgment, when the reader consider that the recommended article is relevance, trust or meets their quality needs, they can click to accept it, otherwise, they can click to reject it. The reader’s feedback is the most authoritative result to evaluate the performance of our algorithm, the average of response rate was about 69% for each algorithm.

Academic knowledge sharing systems provide a list of candidate readers for target articles, the readers can choose to accept the target articles or not. The accuracy is the most important indicator to evaluate the recommender systems, accuracy is usually measured by precision and mean average precision(MAP)[40], the detail of the evaluation metrics is showed as Equation 17 and 18.

$$\text{Pre} - @K = \frac{N_{tp}}{K} \quad (17)$$

$$\text{MAP} = \frac{1}{U} \sum_{i=1}^u \frac{\sum_{k=1}^{n_i} P_{ik}}{n_i} \quad (18)$$

Where, K is the number of generated readers for each target article, here set, K = 5, K = 10 and K = 30, N_{tp} is the number of readers who accept the articles; U is the number of target articles, n_i is the number of recommended readers with regard to article i, P_{ik} is the precision of a list of ranked retrieval results from the top results until the reader k.

The precision results of five methods is showed in Figure 3-5, as we can see, in Figure3, when k=5, the precision of ULA-DLA is 0.54, ULA is 0.43, SR is 0.39, CBS is 0.37, RTR is 0.32 and CB is 0.25. When k=10 and k=30, we can reach the same conclusion. The rank performance of our proposed method and baselines is also evaluated. MAP measures the effectiveness of our rank. Top-ranked readers accept the target article indicate the algorithm has a good performance. Figure 6 shows the MAP scores of 5 methods. The MAP score of our proposed method is 0.50, while CB is 0.23, there has more than 100% improvement. The MAP of ULA, SR, CBS and RTR is 0.41, 0.37, 0.33 and 0.31 respectively.

The results prove that our proposed method has the best performance compared with all the baselines. The ULA algorithm which is only consider the user-level analysis have more than 40% acceptance rate, this means target article can be effectively shared to new users that has no publications. SR has the better performance than the RTR, further confirm the importance of trust in recommendation system. Although CB is the most widely used method in

document information retrieval, but there has a poor performance in our research. The CBS method take advantage of semantic meaning of keywords, has a better performance than CB, but worse than SR, the reason maybe is co-author and online friend have the similar research interests. The ULA only worse than ULA-DLA, better than other four methods, the reason maybe is user-level analysis combine three dimensions to profile users, research topic relevance, social relation and research quality is the main factors affect the readers’ decisions.

Furthermore, a paired-sample T test is conducted to verify whether there is a significant improvement than the baseline methods. As we can see in Table1, that our proposed methods in Pre-@5, Pre-@10 and Pre-@30 are all significantly better than the baseline.

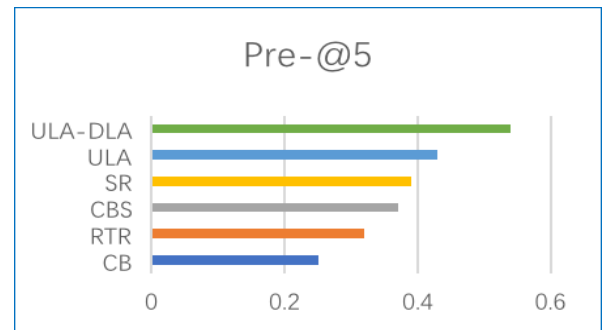


Figure 3. The precision rate of K=5

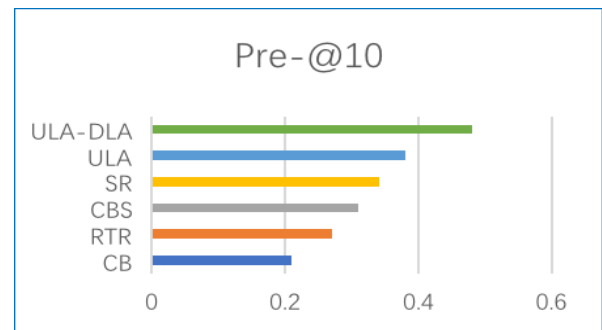


Figure 4. The precision rate of K=10

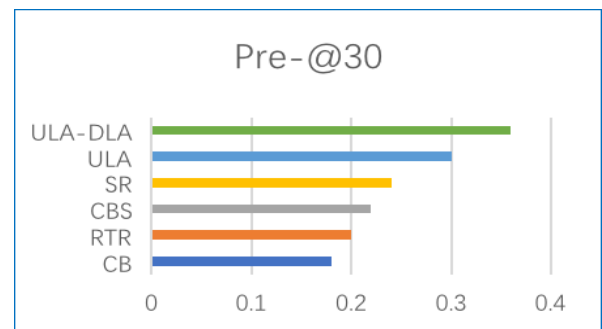


Figure 5. The precision rate of K=30

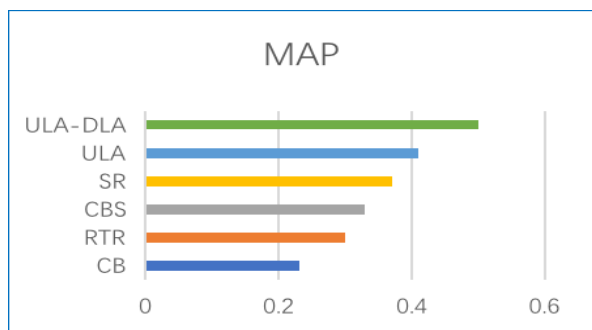


Figure 6. The MAP scores

Table 1. Result of T-test

	ULA-DLA			
	Pre-@5	Pre-@10	Pre-@30	MAP
CB	0.027**	0.000*	0.000*	0.000*
RTR	0.000**	0.001**	0.028**	0.014**
CBS	0.002**	0.018**	0.013**	0.004**
SR	0.003**	0.026**	0.034**	0.000**
ULA	0.000**	0.001**	0.015**	0.000**

*P significant at $\alpha \leq 0.01$

**P significant at $\alpha \leq 0.001$

5. Conclusion

The rapid growth of online content has making significant challenges for knowledge sharing and discovery, especially in the academic area. With the rise of social networks, authors can actively share their publications to potential readers so as to facilitate the knowledge sharing and to promote research innovation. In this paper, an academic knowledge sharing system is proposed to address the academic information overload problem from author's perspectives. Taking advantages of the author's subjective initiative and knowledge sharing capabilities in the social network, the proposed method can significantly change the way of academic knowledge sharing and discovery. In the proposed method, we combine the user-level and document-level analysis in the same model. research topic relevance, social relation and research quality are take into consideration in user-level analysis. The three dimensions are calculated in advance so as to provide efficient recommendation results in the academic social network with millions of users. Moreover, for a new user (e.g., a PhD student who has no publications), the user-level analysis can successfully solve the cold-start problem and recommend the most relevant articles to them. For the user's which has the publications, by combine the document-level analysis, target articles can be recommended with ease.

There are also some limitations in our research. First, in user-level analysis, research topic, social relation and research quality are take into consideration, but there are other factors maybe influence the knowledge sharing, like user's personal motivation. In future, user's personal preference can be surveyed and combined into the analysis framework. Second, user's research interests are a dynamic factor, using publications to measure it may generate a delay. Therefore, the user's online behavior (e.g. like, comment and download) should be considered to profile the user's recent interests.

References

- [1] S.S. Anand, B. Mobasher, Intelligent Techniques for Web Personalization, *Acm Transactions on Internet Technology*, 7 (2007) 18.
- [2] D. Goldberg, D. Nichols, B.M. Oki, D. Terry, Using collaborative filtering to weave an information tapestry, *Communications of the ACM*, 35 (1992) 61-70.
- [3] Y.M. Li, C.T. Wu, C.Y. Lai, A social recommender mechanism for e-commerce: Combining similarity, trust, and relationship, *Decision Support Systems*, 55 (2013) 740-752.
- [4] S. Deng, D. Wang, X. Li, G. Xu, Exploring user emotion in microblogs for music recommendation ☆, *Expert Systems with Applications*, 42 (2015) 9284-9293.
- [5] W. Carrer-Neto, M.L. Hernández-Alcaraz, R. Valencia-García, F. García-Sánchez, Social knowledge-based recommender system. Application to the movies domain, *Expert Systems with Applications*, 39 (2012) 10990-11000.
- [6] T.P. Liang, Y.F. Yang, D.N. Chen, Y.C. Ku, A semantic-expansion approach to personalized knowledge recommendation ☆, *Decision Support Systems*, 45 (2008) 401-412.
- [7] J. Beel, B. Gipp, S. Langer, C. Breiteringer, Research-paper recommender systems: a literature survey, *International Journal on Digital Libraries*, 17 (2016) 305-338.
- [8] S. Redner, How popular is your paper? An empirical study of the citation distribution, *The European Physical Journal B*, 4 (1998) 131-134.
- [9] R. Adler, J. Ewing, P. Taylor, Citation Statistics: A Report from the International Mathematical Union (IMU) in Cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS), *Statistical Science*, 24 (2009) 1-14.
- [10] M. Amin, M.A. Mabe, Impact factors: use and abuse, *Medicina*, 63 (2003) 347-354.
- [11] L. Bornmann, H. Schier, W. Marx, H.D. Daniel, What factors determine citation counts of publications in chemistry besides their quality?, *Journal of Informetrics*, 6 (2012) 11-18.
- [12] F. Didegah, M. Thelwall, Which factors help authors produce the highest impact research? Collaboration, journal and document properties, *Journal of Informetrics*, 7 (2013) 861-873.

- [13] X. Wang, C. Liu, W. Mao, Z. Fang, The open access advantage considering citation, article usage and social media attention, *Scientometrics*, 103 (2015) 555-564.
- [14] D. Constant, L. Sproull, S. Kiesler, The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice, *Organization Science*, 7 (1996) 119-135.
- [15] M. Tampoe, Motivating knowledge workers—The challenge for the 1990s, *Long Range Planning*, 26 (1993) 49-55.
- [16] M.L. Wasko, S. Faraj, “It is what one does”: why people participate and help others in electronic communities of practice, *Journal of Strategic Information Systems*, 9 (2000) 155-173.
- [17] C.-M. Chiu, M.-H. Hsu, E.T. Wang, Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories, *Decision support systems*, 42 (2006) 1872-1888.
- [18] M.-J.J. Lin, S.-W. Hung, C.-J. Chen, Fostering the determinants of knowledge sharing in professional virtual communities, *Computers in Human Behavior*, 25 (2009) 929-939.
- [19] B. Thoms, N. Garrett, J.C. Herrera, T. Ryan, Understanding the roles of knowledge sharing and trust in online learning communities, *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual, IEEE*, 2008, pp. 3-3.
- [20] R.R. Sinha, K. Swearingen, Comparing Recommendations Made by Online Systems and Friends, *DELOS workshop: personalisation and recommender systems in digital libraries*, 2001.
- [21] M. Pazzani, D. Billsus, Content-based recommendation systems, *The adaptive web*, DOI (2007) 325-341.
- [22] J. Schafer, D. Frankowski, J. Herlocker, S. Sen, Collaborative filtering recommender systems, *The adaptive web*, DOI (2007) 291-324.
- [23] R. Burke, Hybrid web recommender systems, *The adaptive web*, DOI (2007) 377-408.
- [24] T. Bogers, A. Van den Bosch, Recommending scientific articles using citeulike, *Proceedings of the 2008 ACM conference on Recommender systems*, ACM, 2008, pp. 287-290.
- [25] J. Lee, K. Lee, J.G. Kim, Personalized academic research paper recommendation system, *arXiv preprint arXiv:1304.5457*, DOI (2013).
- [26] K. Sugiyama, M.-Y. Kan, A comprehensive evaluation of scholarly paper recommendation using potential citation papers, *International Journal on Digital Libraries*, 16 (2015) 91-109.
- [27] T.-P. Liang, Y.-F. Yang, D.-N. Chen, Y.-C. Ku, A semantic-expansion approach to personalized knowledge recommendation, *Decision Support Systems*, 45 (2008) 401-412.
- [28] K. Chandrasekaran, S. Gauch, P. Lakkaraju, H.P. Luong, Concept-based document recommendations for citeseer authors, *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, Springer, 2008, pp. 83-92.
- [29] T. Silva, Z. Guo, J. Ma, H. Jiang, H. Chen, A social network-empowered research analytics framework for project selection, *Decision Support Systems*, 55 (2013) 957-968.
- [30] J. Magalhaes, C. Souza, E. Costa, J. Fechine, Recommending scientific papers: Investigating the user curriculum, *The Twenty-Eighth International Flairs Conference*, 2015.
- [31] Y.-M. Li, C.-T. Wu, C.-Y. Lai, A social recommender mechanism for e-commerce: Combining similarity, trust, and relationship, *Decision Support Systems*, 55 (2013) 740-752.
- [32] N.Y. Asabere, F. Xia, Q. Meng, F. Li, H. Liu, Scholarly paper recommendation based on social awareness and folksonomy, *International Journal of Parallel, Emergent and Distributed Systems*, 30 (2015) 211-232.
- [33] Y. Xu, X. Guo, J. Hao, J. Ma, R.Y. Lau, W. Xu, Combining social network and semantic concept analysis for personalized academic researcher recommendation, *Decision Support Systems*, 54 (2012) 564-573.
- [34] J. Sun, J. Ma, Z. Liu, Y. Miao, Leveraging content and connections for scientific article recommendation in social computing contexts, *The Computer Journal*, 57 (2014) 1331-1342.
- [35] M. Al-Hassan, H. Lu, J. Lu, A semantic enhanced hybrid recommendation approach: A case study of e-Government tourism service recommendation system, *Decision Support Systems*, 72 (2015) 97-109.
- [36] S. Niwattanakul, J. Singthongchai, E. Naenudorn, S. Wanapu, Using of Jaccard coefficient for keywords similarity, *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 2013, pp. 6.
- [37] C.V. Cannistraci, G. Alanis-Lobato, T. Ravasi, From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks, *Scientific reports*, 3 (2013) 1613.
- [38] J. Sun, W. Xu, J. Ma, J. Sun, Leverage RAF to find domain experts on research social network services: A big data analytics methodology with MapReduce framework, *International Journal of Production Economics*, 165 (2015) 185-193.
- [39] J. Ramos, Using tf-idf to determine word relevance in document queries, *Proceedings of the first instructional conference on machine learning*, 2003.
- [40] C.C. Yang, *Search engines information retrieval in practice*, Wiley Online Library, 2010.