Random Walk-based Beneficial Collaborators Recommendation Exploiting Dynamic Research Interests and Academic Influence

Xiangjie Kong, Huizhen Jiang, Teshome Megersa Bekele, Wei Wang, Zhenzhen Xu
School of Software, Dalian University of Technology, Dalian 116620, China
xzz@dlut.edu.cn

ABSTRACT
It is laborious for researchers to find proper collaborators who can provide researching guidance besides collaborating. Beneficial Collaborators (BCs), researchers who have a high academic level and relevant topics, can genuinely help researchers to enrich their research. Though many efforts have made to develop collaborator recommendation, most of existing works have mainly focused on recommending most possible collaborators with no intention to recommend specifically the BCs. In this paper, we propose the Beneficial Collaborator Recommendation (BCR) model that considers the dynamic research interest of researcher's and academic level of collaborators to recommend the BCs. First, we run the LDA model on the abstract of researchers' publications in each year for topic clustering. Second, we fix generated topic distribution matrix by a time function to fit interest dynamic transformation. Third, we compute the similarity between the collaboration candidate's feature matrix and the target researcher. Finally, we combine the similarity and influence in collaborators network to fix rank score and mine the candidates with high academic level and academic social impact. BCR generates the topN BCs recommendation. Extensive experiments on a dataset with citation network demonstrate that BCR performs better in terms of precision, recall, F1 score and the recommendation quality compared to baseline methods.

Keywords
Collaborator recommendation; Dynamic research interest; Academic influence

1. INTRODUCTION
Collaborator recommendation is the prediction of links that may appear in the future or recommending links that may be advantageous to academic social networks. In academia, a co-authorship among researchers is modeled as a complex collaboration network. Nowadays, due to a large amount of academic entities (researchers, publications, etc.) and academic relations (co-authorship, inter-citation, etc.) with many complex and heterogeneous academic social networks [1-2], it is challenging to mine useful and effective information from scholarly data. However, these data have also brought opportunities to researchers. Especially, as academic resources are limited, the immense growth of researchers brings more opportunities for collaboration. However, it also creates difficulties for researchers to choose the right potential collaborators who could help them enrich their research skills.

This work presents a collaborator recommendation model that recommends Beneficial Collaborators (BCs) based on the scholarly big data. BCs can not only collaborate with researchers but also give researchers directions and help for their research. With scholarly big data being prevalent, a large amount of researchers' information is accessible online, including their profile, publications, homepages and even personal behaviors [2]. Nevertheless, the large scale of this information and the Rule of 150 [3] create disarray and challenges for researchers to identify the BCs for each researcher personality. To solve these problems and satisfy researcher's desire for collaboration, many collaborator recommendation approaches have been proposed in recent years and played an important role in the academic recommendation. Most of them are designed by mining researchers' preferences history or the appropriate context to help researchers find possible collaborators.

However, academic researching is generally a dynamic process to a certain extent. Hence, considering the variation of researchers' interest with time is of great importance when measuring their academic interest features. In addition, the existing works have mainly focused on recommending most possible collaborators with no intention specifically recommending the beneficial collaborators. Generally, researchers can reap more benefit when collaborating with prolific and active researchers who can provide professional academic guidance. The BCs could bring academic resources and pave the way to make researchers' scientific achievement grow faster [4]. Hence, this study primarily aims to untangle the problem: how can a recommender system help researchers to find proper BCs in order to help researchers benefit more from collaboration.

In this work, we propose a Beneficial Collaborator Recommendation (BCR) model that can recommend researchers the beneficial collaborators. Who are the beneficial collaborators and what is the definition? So they are the researchers who have relevant research interest with the target researcher, being excellent and active so that they can provide academic guidance for the researcher. To this end, in BCR we first divide researchers' publications by year and run the Latent Dirichlet Allocation (LDA) [5] model on the abstracts of researchers' publications to obtain the topic distribution of researchers' research interest in each year, which we consider as the academic feature matrix. Second, we highlight the topics by an increasing time function to fit the interest variation. Third, we compute the similarity between the collaboration candidate's feature matrix and the target researcher. Finally, we combine the impact in collaborators network with the similarity to fix the rank score in order to mine the candidates with high academic level and academic social impact. BCR generates the TopN BCs recommendation.

WWW 2017 Companion, April 3-7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3051154

Extensive experiments on an Arnetminer citation dataset show that BCR has better performance than the baseline methods.

Contributions of this paper are as below.

- We propose a novel BCR method that can recommend researchers the beneficial collaborators who could help them pave the way to make achievements rapidly.
- We consider research interest variation with time combined with the topic, and a network-based model to evaluate the academic impact of the candidate BCs.
- We conduct extensive experiments on a citation dataset to evaluate the effectiveness of BCR and get better results.

The remainder of this paper is structured as follows. Section 2 surveys the related work. We discuss the details of our proposed model in Section 3. In Section 4, we discuss our experimental settings and analyze the results. Finally, Section 5 concludes the paper.

2. RELATED WORK

Recommending collaborators require addressing the correlation between research collaboration and productivity. Abramo et al. [6] carried out a number of empirical analyses to measure the effects of extramural collaboration on research performance. Their work together with Lee's paper [7] found that there is a robust and positive correlation between collaboration and productivity. That is to say, in the academic area, collaborating with others can make researchers more fruitful.

In academia, the co-authorship among researchers is portrayed as a complex collaboration network. Generally, collaborator recommendation is to predict the links that may appear in the future [8] or recommend links that may be advantageous to academic social networks. Benchettara et al. [9] tackled the problem of link prediction in collaboration networks by a topological dyadic supervised machine learning approach to realize better collaborator recommendation. In contrast to the traditional link prediction, Wang and Sukthankar [10] proposed a new link prediction framework to avoid the performance of predictors treating all links homogeneously. Dong et al. [11] proposed a ranking factor graph model for predicting links with several social patterns across heterogeneous networks and got better performance.

In general, collaborator recommendation models can be categorized into four categories, namely, CF-based recommendation (Collaborative Filtering), content-based recommendation, social network-based recommendation, and hybrid recommendation [4, 12, 13, 14, 15]. Some of these models give us inspiration on BCs recommendation. Chen et al. [16] introduced an open system to recommend potential research collaborators based on the coauthor network and user's research interest. Sugiyama and Kan [17] proposed a generic model towards recommending scholarly papers relevant to a researcher's interest by capturing their past publications, citations and reference papers. Katz et al. [18] distinguished between collaboration at different levels and found that researchers with a higher level of collaboration tend to be more productive. These studies confirm our hypothesis that it is necessary to recommend high-level collaborators with the relevant research topic.

Researchers' research topics usually go through a period of varying until it gets stable. Similarly, in social media, people's interest change with time. Hence, in collaborator recommendation, the variation of research interest is worth considering. Liang et al. [19] proposed a time-aware topic recommendation on micro-blogs, which considered the temporal dynamics of topics. Duad [20] presented a time-based topic modeling approach and discovered researcher's interest and relationships changing over time. Kanhabua [21] focused on identifying years of interest to a keyword query only based on documents' publication dates.

However, few research works have considered the interest variation during the academic collaborator recommendation. Hence, this work proposes a novel BCR model that considers the topic distribution of research interest, interest variation with time and academic level of collaborators in order to recommend the beneficial collaborators.

3. PROPOSED SCHEME

3.1 Overview of BCR

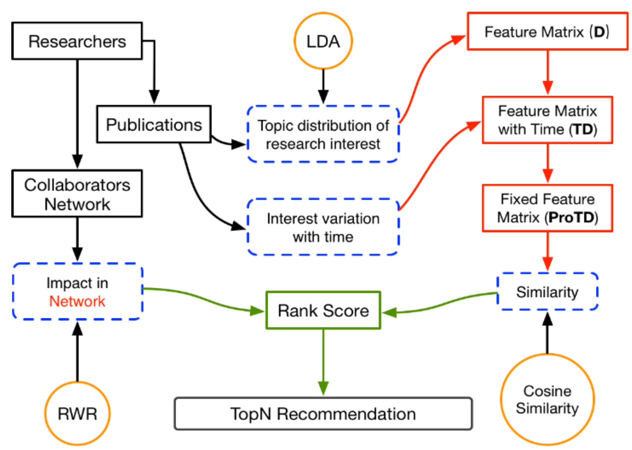

Figure 1. The structure of BCRM.

The recommending approach of BCR consists of the following four steps.

1. To measure accurately researchers' academic interest, we divide their publications by year and run the LDA model on their publications to obtain the academic feature matrix that shows the topic distribution of their research interest in each year.
2. We highlight the topics by an increasing time function to fit the interest variation.
3. We compute the similarity between the collaboration candidate's feature matrix to that of a target researcher. Then we combine the academic impact with the similarity results to fix the rank score to mine the candidates with high academic level.
4. In the end, BCR conducts the TopN BCs recommendation according to the fixed rank score.

The structure of BCR is shown in Figure 1.

3.2 Topic Clustering

The distribution of researcher's research interest can be extracted from his/her publications. To represent the feature of researcher's research interest, we adopt LDA model to acquire the topic distribution probabilities by clustering researchers' publications.

Based on the context of text modeling, i.e., the topic probabilities provide an explicit representation of a document; the joined researchers' publications can be made for a personal academic document. The topic probabilities of the document generated by LDA model can represent researchers' topic distribution of research.

Comparing to the main body, abstract of a publication is a summary of the main idea of the paper. Therefore, to reduce the computation complexity, we cluster the abstract of publications for each researcher. To measure researcher's dynamic research interest, we first build the "academic documents" for researchers in each year by joining the abstract of researcher's publications published in the same year by space. Then we run LDA model on the generated documents set with a special parameter *k,* which represents the clustered topic number in LDA. After that, we can get the probabilities of each researcher on the *k* topics that represent the research distribution of researchers in each year, which we considered as the researcher's academic feature vector *V*. For example, for researcher Alice, there are five topics A, B, C, D, and E with topic distribution probabilities of 0.3, 0, 0, 0.6 and 0.1 respectively in the year 2010. Hence, Alice's feature vector V_{Alice} is 0.3,0, 0, 0.6, and 0.1. Considering the time dimension, the feature vectors of a researcher in several years can generate a feature matrix *D (Alice's* feature matrix *D* is as shown in Table 1).

3.3 Research Interest Variation

During undertaking scientific research, researchers would like to pay attention to researchers who are working on the similar topics at the same time. On the other hand, the topic distribution probability in recent years can describe researcher's current research interest more accurately. As on behalf of researcher's topics, the feature matrix of the researchers will be used to compute the similarity of researchers, it is reasonable to strengthen the topics near now using an increasing time function. A fixed feature matrix *TD* is shown in Equation 1. An individual researcher's research interest usually changes at the beginning of his/her research, but finally gets stable as time goes by. Based on this notion, we make a hypothesis: the interest variation trend is approximately fitting to the math function *ln*. Thus, we proposed an empirical formula for the time function *T* as $T_i = \ln(t_i - t_0 + 2)$, where t_i is the target year, t_0 is the first year in feature matrix *D*. The constant term 2 is to make sure T_i positive that the function can be normal. Equation 1 is a hypothetical and experiential function which is proved effectual in section 4.

$$TD_{t,i} = D_{t,i}T_i \tag{1}$$

As LDA clusters many topics, there are some weak topics with much smaller value in the feature matrix *TD*. In BCR model, topics with high values in each year is selected as *TopMTopics* and weak topics that are not included in *TopMTopics* set are reset to zero. We modify the value of weak topics, in order to improve the LDA model effectiveness in properly describing a researcher's major interesting topics, and reducing the weak topics will make the feature matrix weighted in favor of major topics and leads BCR to find candidates who are professional at the researcher's major topics. Table 1 shows the *D* of Alice. In the year 2012, topic A seems weaker compared to topic C, D, and E. Therefore, we reset Topic A to 0 to obtain Alice's major topic in the year 2012. We define the set $TopMTopics_t$ as the most professional *M* topics in year *t*. The process of modifying matrix *TD* is shown in Equation 2. If the topic i is part of the set $TopMTopics_t$, the $ProTD_{t,i} = TD_{t,i}$, otherwise, the $ProTD_{t,i}$ is 0. The new feature matrix *ProTD* of Alice with *M* =3 is shown in Table 2.

$$ProTD_{t,i} = \begin{cases} TD_{t,i} & i \in TopMTopics_t \\ 0 & i \notin TopMTopics_t \end{cases} \tag{2}$$

Table 1. Alice's Feature Matrix.

Year	A	B	C	D	E
2010	0.3	0	0	0.6	0.1
2011	0.2	0	0.1	0.5	0.2
2012	0.1	0	0.3	0.4	0.2
2013	0	0	0.1	0.7	0.2
2014	0	0	0	0.7	0.3

Table 2. Alice's New Feature Matrix.

Year	A	B	C	D	E
2010	0.21	0	0	0.42	0.07
2011	0.21	0	0	0.55	0.21
2012	0.1	0	0.42	0.56	0.28
2013	0	0	0.16	1.13	0.32
2014	0	0	0	1.25	0.54

3.4 Similarity Calculation

The ProTD describes the academic feature of researchers. The core of BCs recommendation problem is computing the similarity between researchers. In this paper, we use cosine similarity method to calculate the similarity of researchers' feature vector for each year respectively (as shown in Equation 3). Then we compute the arithmetic mean of every years' similarity and take the result as the final similarity between two researchers (as shown in Equation 4).

$$CosSim(V_a, V_b) = \frac{\sum_{i=1}^{k} V_{a,i} * V_{b,i}}{\sqrt{\sum_{i=1}^{k} V^2{}_{a,i}} * \sqrt{\sum_{i=1}^{k} V^2{}_{b,i}}} \tag{3}$$

$$Sim = \frac{\sum_{t=t_0}^{t_n} CosSim(V_a^t, V_b^t)}{n} \tag{4}$$

Where in Equation 3, V_a and V_b respectively represents the feature vector of researcher *a* and *b* in one year. *k* means the dimension of the feature vector, i.e., the number of clustered topics. In Equation 4, V_a^t, V_b^t represent the feature vector of researchers *a* and *b* at year *t* respectively. *n* is the number of academic age span, i.e. the row number of matrix *ProTD*.

3.5 Impact in Collaboration Network

According to the definition of BCs in the introduction section, the recommended collaborators should be able to provide profound directions and help for researchers' work. BCs should be active and relatively important in the academic collaborator's network (academic level). There are several well-known metrics such as h-index, total citations to measure researchers' academic level [22]. From the view of collaborators network, the triadic closures theory describes that two people may get to know each other if they have same friends [23]. It reveals the phenomenon that in the academic social network collaboration can bring new collaboration to some probability and spread researchers' influence along the connections in collaborator network.

Random Walk with Restart model (RWR) [24], a popular network-based model, has been verified that it can describe the delivery process of collaboration along academic collaborator network, and it can be used to measure the impact of researchers on the collaborator network [4]. Compared to the h-index, the RWR can reflect the acceptance of the author in the academic area of the collaborators. The process in RWR is similar to the vote in academic collaborator network. The researcher shows high impact if they have high value in RWR. In this work, we used the RWR value of each researcher to describe researchers' academic impact in collaborator network. To obtain the RankScore of the candidate collaborators, we take the value of RWR as a coefficient and multiply by similarity, i.e. sim, to fix the similarity of researchers as in shown in Equation 5.

$$RankScore = RWR * Sim \tag{5}$$

The RWR is the evaluated value of researchers according to RWR model. The rank score is the recommending evidence of our BCR model.

4. EVALUATION AND ANALYSIS

4.1 Dataset and Settings

Our experiments were conducted on a subset of citation network dataset from Tang [1]. In this dataset, there are 92256 publications and 129354 researchers. It includes the abstract, author name, title, publishing year and the references list of each publication. All experiments were performed on a 64bit Linux-based operation system, Ubuntu 12.04 with a 4-duo and 3.2-GHz 64-bit Intel CPU, 4G Bytes memory. All the programs were implemented with Python. In this paper, we divided the dataset into two parts according to the publishing year: the data before the year 2009 as a training set (86233 researchers), others as a testing set (43121 researchers).

4.2 Metrics

For the experiments evaluation, we employed three metrics that are widely used in recommender systems: precision, recall and F1 [4]. In this work, we conduct a recommendation list for each researcher with several potential collaborators that he/she has never coauthored before the year 2009. When we check the testing dataset, the recommendation will be marked as "accepted" if the recommended collaborators coauthored with the target researcher after the year 2009. Hence, in BCR model, there is a recommendation list and an accepted list for each researcher. The two lists have an intersection. The obtained results after recommendation can be divided into 3 parts, A, B, and C.

- A: recommended and accepted.
- B: recommended and not accepted.
- C: not recommended and accepted.

Therefore, the definition of precision is as in Equation 6:

$$Precision = \frac{A}{A+B} \tag{6}$$

The recall is defined as Equation 7:

$$Recall = \frac{A}{A+C} \tag{7}$$

To get an integrated metric over precision and recall, we can measure the model by F1 score. The definition of F1 is:

$$F1 = \frac{2(Precision*Recall)}{Precision+Recall} \tag{8}$$

The core focus of designing BCR model is to recommend researchers the beneficial collaborators with high academic level, i.e. the quality of recommended items. The higher a recommended quality, the better performance of the recommendation. This work used the average cited number of recommended collaborators to represent the recommending quality. The cited number of a researcher is the total number of its publications' cited times. The formalized definition is shown in Equation 9. V is the set of recommended items. M represents the length of the recommendation list. $CitedNum_v$ is the cited number of researcher v.

$$AcademicLevel = \frac{\sum_{v\in V}^{M} CitedNum_v}{M} \tag{9}$$

4.3 Parameter settings and BCR comparison

To measure the performance of BCR, we compared it with three approaches, i.e. an RWR-based model, a citation-based model, and a content-based model using Tang's citation network dataset. (1) RWR is a kind of popular model widely used in recommender systems. Similar to Chen's comparison approaches [3], we run the RWR-based model on Tang's citation network dataset. The recommendation list is made on the basis of the rank score after several iterations of random walk. (2) The citation-based model directly recommends the researchers who have higher citation number (i.e. the total number of publications' cited times for a researcher). (3) The content-based model is also widely used since the forepart recommender system [25]. It clusters the topics of researchers' publications and computes the similarity of the topic distribution probability (Feature vector) between researchers in order to recommend similar researchers to each other.

4.4 Results and Analysis

4.4.1 Influence of clustered topics number

This is the first group of experiments where we evaluated the influence of clustered topics number. We run BCR model by randomly choosing 100 researchers as the target nodes and observed the average value of metrics for the 100 times experiments. We repeatedly executed such experiments with different size of recommendation lists to evaluate the influence of clustered topics number on the result. We set four candidate values for the number of clustered topics k in LDA model, i.e. {10, 50,100,200}. To make sure the k is the only variable in these of experiments, we set the number of *TopMTopics* to k=5. Four independent experiments were conducted with different topic numbers. Figure 2 (a), (b) and (c) represent the performance of BCR in terms of precision, recall, and F1 score respectively during the four experiments.

As can be seen in Figure 2 precision shows downtrend with the recommendation list increases while recall shows overall upward trend and flattens out in the last. The F1 score performs the upper

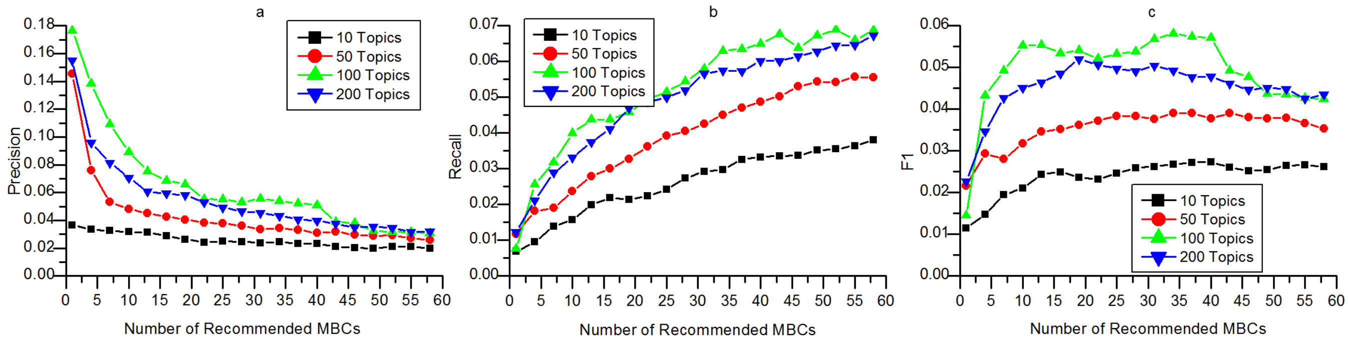

Figure 2. Influence of clustered of topics number.

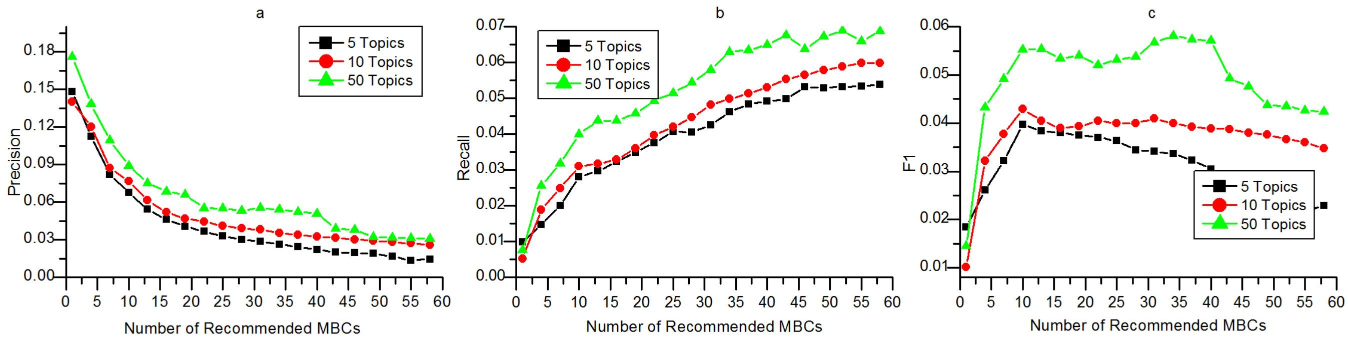

Figure 3. Influence of setting-zero topics number.

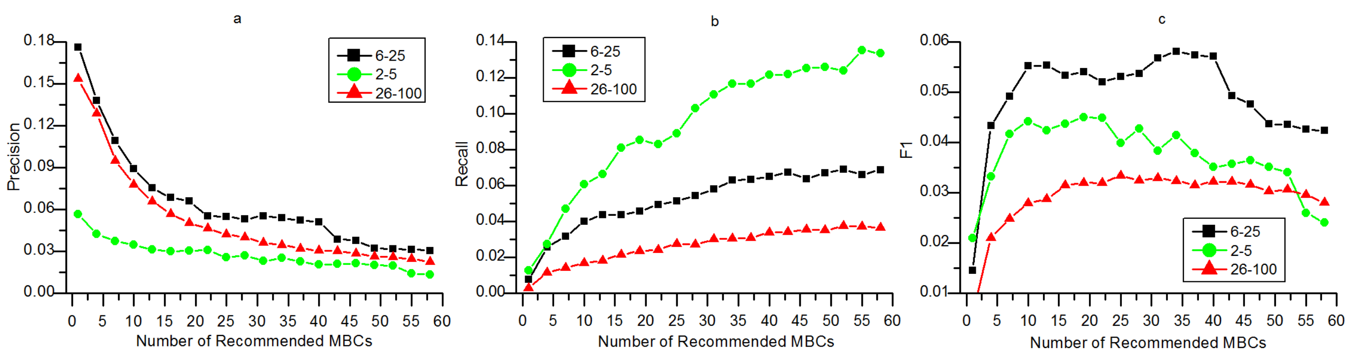

Figure 4. The performance of BCR on researchers with different academic level.

convex curve, rapidly rising and then slightly decline. A close view of the lines, the model performs better when the number of clustered topics is 100 compared to other candidate *k* values. The model achieves the highest F1 score, i.e. 5.82%, at the point 34.

4.4.2 Influence of setting-zero topics number

In the second group of experiments, we evaluated the influence of the number of *TopMTopics*. We used 100 target nodes and other experimental settings similarly to the first group of experiments. We set *x* topics with low values as zero in these experiments, that is, the length of *TopMTopics* is *k-x*. The variate *x* has three candidate values {5, 10, 50}. The experimental results are shown in Figure 3. The variation tendency of precision, recall, and F1 score performs roughly consistent as in Figure 3. Overall, the model performs a little better with 50 setting-zero topics. The highest F1 score in these experiments, i.e. 5.82 %, is achieved when recommending 34 BCs.

4.4.3 Target researcher's academic level Influence

In the third group of experiments, we evaluated the performance of BCR model on the researchers with different academic level. In this paper, we use the total cited times of researcher's publications to represent its academic level. Based on the cited number distribution of researchers in the dataset, we clustered researchers into three sets with different citation number: {2-5, 6-25, 26-100}. We described the three sets of researchers as primary level, intermediate level, and advanced level respectively. We randomly chose 100 researchers as the target researchers from each set of researcher respectively, and run the BCR model to recommend BCs for them. The experimental settings are the same with the second group of experiments. The clustered topics number is 100, and the number of setting-zero topics is 50. Figure 4 shows the experimental results.

In terms of precision, BCR performs better on recommending BCs for intermediate and advanced level researchers while it shows relatively low precision for the primary level researchers. However, BCR is good at recommending BCs for primary level researchers in terms of recall as in Figure 4 (b) while its performance for advanced level researchers is worse. Considering the F1 score, BCR performs better on recommending BCs for the intermediate level researchers compared to the primary and advanced level researchers as in Figure 4 (c). These experiments demonstrated that the academic level of target researchers has a great impact on the performance of BCR. Taking a look at the comprehensive metric F1 score, BCR is better at recommending BCs for intermediate level researchers, i.e., researchers who have entered the academic field for a considerable period of time and achieved a good performance. This is consistent with our original intention, finding out the BCs who could help researchers to improve academic level and achievement.

4.5 Performance of BCR and Baseline Models

In the fourth group of experiments, we compared the performance of BCR with three models: the content based model, citation-based model, and RWR-based model. We also explored the academic level of recommended BCs, i.e. the metric AcademicLevel in Equation 9. The experimental settings were the same with the last group experiments: 100 cluster topics and 50 setting-zero topics. The 100 target researchers were chosen from the intermediate level set. The series of results are shown in Figure 5.

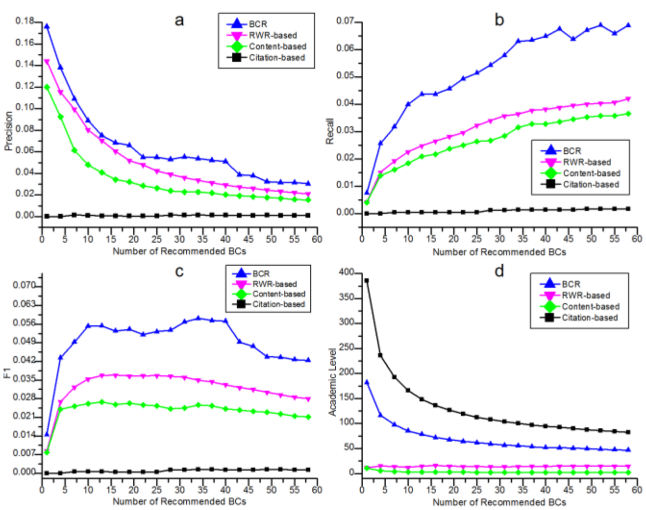

Figure 5. Performance of BCR and three comparison models.

In the Figure 5 (a), we can see that all models show a similar downwards trend with the recommendation list increases. BCR and the RWR-based model perform better than others on precision. BCR can get the highest precision value. In addition, the curve that represents the citation-based model changes gently and performs low values all the time. The citation-based model is an essentially popularity-based model, which seems not good at recommending BCs. As shown in Figure 5 (b) BCR shows a great performance on recall comparing with other models. Similarly, with the precision, the RWR-based model is at the second place over recall. From Figure 5 (c), the rank of the four models show invariant. BCR performs the best in terms of F1 score comparing to others. The curve rapidly rises at first and then comes down gradually. BCR can get the highest F1 score 5.82% when recommending 34 BCs. Figure 5 (d) shows differences compared to Figure 5 (b), (c) and (d). The citation-based model can easily find out the high academic level researchers. As the RWR-based model is able to rank nodes in the networks, the important nodes will be recommended. The RWR-based model is better at finding high academic level researchers compared to the content-based model. Moreover, BCR also performs obviously better in comparison to the RWR-based and content-based model. This group of experiments demonstrated that our BCR model could generate a high-quality recommendation. The recommended researchers with high academic level can provide professional academic guidance.

To summarize, overall, BCR shows better performance in terms of precision, recall, and F1 score. Moreover, it can identify the high-level researchers and make high-quality recommendation list. The three academic features we explored are indeed effective to make the beneficial collaborators.

5. CONCLUSIONS

In this paper, we proposed a Beneficial Collaborator Recommendation (BCR) model to help researchers find beneficial collaborators who can provide academic guidance for their research besides collaboration. BCR runs a topic model on researcher's publications and associates three academic features: topic distribution of research interest, interest variation with time and academic level of researchers. Extensive experiments on a dataset with citation network demonstrate that BCR performs better in terms of precision, recall, F1 score and the recommendation quality compared to baseline approaches.

Nonetheless, there is still room for future study in this direction. As a future work, we will exploit more factors to evaluate the researchers' academic level. There are also some cases such as that same research topic may exists in different research communities. The cross community recommendation will also be considered. Besides, more experiments should be performed on academic datasets considering more academic features.

6. ACKNOWLEDGMENTS

This work is partially supported by the Fundamental Research Funds for Central Universities under Grants No. DUT15QY06.

7. REFERENCES

[1] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, Z. Su. 2008. Arnetminer: extraction and mining of academic social networks. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge discovery and data mining, Las Vegas, Nevada, 990-998. DOI: 10.1145/1401890.1402008

[2] Z. Wu, J. Wu, M. Khabsa, K. Williams, H.-H. Chen, W. Huang, S. Tuarob, S. R. Choudhury, A. Ororbia, P. Mitra, et al. 2014. Towards building a scholarly big data platform: Challenges, lessons, and opportunities. In IEEE/ACM Joint Conference on Digital Library, 117-126. DOI: 10.1109/JCDL.2014.6970157

[3] A. Hernando, D. Villuendas, C. Vesperinas, M. Abad, A. Plastino. 2010. Unravelling the size distribution of social groups with information theory in complex networks. The European Physical Journal B , vol.76, no.1, 87-97. DOI: 10.1140/epjb/e2010-00216-1

[4] F. Xia, Z. Chen, W. Wang, J. Li, L. T. Yang. 2014. Mvcwalker: Random walk-based most valuable collaborators recommendation exploiting academic factors. IEEE Transactions on Emerging Topics in Computing, vol. 2, no.3, 364-375. DOI: 10.1109/TETC.2014.2356505

[5] D. M. Blei, A. Y. Ng, M. I. Jordan. 2003. Latent Dirichlet allocation, the Journal of Machine Learning Research, vol.3, 993-1022.

[6] G. Abramo, C. A. D'Angelo, F. Di Costa. 2009. Research collaboration and productivity: is there correlation? Higher Education, vol.57,no.2, pp. 155-171.

[7] S. Lee, B. Bozeman. 2005. The impact of research collaboration on scientific productivity, Social studies of science, vol. 35, no.5, pp.673-702.

[8] L. Lü, T. Zhou. 2011. Link prediction in complex networks: A survey, Physica A: Statistical Mechanics and its Applications, vol.390, no.6, pp.1150-1170.

[9] N. Benchettara, R. Kanawati, C. Rouveirol. 2010. A supervised machine learning link prediction approach for academic collaboration recommendation, in Proc. of the fourth ACM Conf. on Recommender systems, pp. 253-256.

[10] X. Wang, G. Sukthankar. 2013. Link prediction in multi-relational collaboration networks, in Proc. of the 2013 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining, pp. 1445-1447.

[11] Y. Dong, J. Tang, S. Wu, J. Tian, N. V. Chawla, J. Rao, H. Cao. 2012. Link prediction and recommendation across heterogeneous social networks, in 2012 IEEE 12th Int. Conf. on Data Mining, pp. 181-190.

[12] J. Li, F. Xia, W. Wang, Z. Chen, N. Y. Asabere, H. Jiang. 2014. Acrec: a co-authorship based random walk model for academic collaboration recommendation, in Proc. of the companion publication of the 23rd Int. Conf. on World wide web companion, Int. World Wide Web Conf. Steering Committee, pp. 1209-1214.

[13] X. Kong, H. Jiang, Z. Yang, Z. Xu, F. Xia, A. Tolba. 2016. Exploiting publication contents and collaboration networks for collaborator recommendation, PloSone, vol.11, no.2. e0148492. doi:10.1371/journal.pone.0148492

[14] G. R. Lopes, M. M. Moro, L. K. Wives, J. P. M. De Oliveira. 2010. Collaboration recommendation on academic social networks, LNCS, vol. 6413, pp. 190-199.

[15] M. C. Pham, Y. Cao, R. Klamma, M. Jarke. 2011. A clustering approach for collaborative filtering recommendation using social network analysis, J. UCS, vol.17, no.4, pp.583-604.

[16] H.-H. Chen, L. Gou, X. Zhang, C. L. Giles. 2011. Collabseer: a search engine for collaboration discovery, in Proc. of the 11th annual Int. ACM/IEEE joint Conf. on Digital libraries, pp. 231-240.

[17] K. Sugiyama, M.-Y. Kan. 2010. Scholarly paper recommendation via user's recent research interests, in Proceedings of the 10th annual joint conference on Digital libraries, ACM, pp. 29-38.

[18] J. S. Katz, B. R. Martin. 1997. What is research collaboration? Research policy, vol.26, no. 1, pp.1-18.

[19] H. Liang, Y. Xu, D. Tjondronegoro, P. Christen. 2012. Time-aware topic recommendation based on micro-blogs, in Proc. of the 21st ACM Int. Conf. on Information and Knowledge Management pp. 1657-1661.

[20] A. Daud. 2012. Using time topic modeling for semantics-based dynamic research interest finding, Knowledge-Based Systems, vol. 26, 154-163.

[21] N. Kanhabua, K. Nørvåg. 2010. Determining time of queries for re-ranking search results, in Research and Advanced Technology for Digital Libraries, pp. 261-272.

[22] G. Saad. 2006. Exploring the h-index at the author and journal levels using bibliometric data of productive consumer scholars and business-related journals respectively, Scientometrics, vol.69, no.1, 117-120.

[23] T. Opsahl. 2013. Triadic closure in two-mode networks: Redefining the global and local clustering coefficients, Social Networks, vol. 35, no.2, pp. 159-167.

[24] H. Tong, C. Faloutsos, J.-Y. Pan. 2006. Fast random walk with restart and its applications, in Sixth Int. Conf. on Data Mining (ICDM'06), pp. 613-622.

[25] M. J. Pazzani, D. Billsus. 2007. Content-based recommendation systems, the adaptive web, LNCS 4321, pp. 325-341.