

The Research Proposal for Recommender Systems in Academic Domain using Social Network Analysis Approach

Tin Huynh

University of Information Technology - Vietnam,
Km 20, Hanoi Highway, Linh Trung Ward, Thu Duc District, HCMC.

`tinhn@uit.edu.vn`

Abstract. In this paper, we present our research proposal based on social network analysis approach to develop recommender systems in the academic domain. Recommender system is a solution that can help users deal with the flood of information returned by search engines. Recommender systems are widely used nowadays, especially in E-Commerce, but it has not received enough attention in the academic domain. The traditional approaches for recommendation do not mention relationships which can effect to behaviors and interests of individuals. Therefore, we applied the Social Network Analysis approach combining with traditional methods to develop recommender systems.

Keywords: social network analysis, recommender system, collaborative knowledge network.

1 Introduction

The explosive growth and complexity of information that is added to the Web daily challenges all search engines. One solution that can help users deal with flood of information returned by search engines is recommendation. Recommender systems identify user's interests through various methods and provide specific information for users based on their needs. Rather than requiring users to search for information, recommender systems proactively suggest content to users [34]. A well-known statement of Anderson, "We are leaving the age of information and entering the age of recommendation", have been used as a slogan for the RecSys (ACM Conference on Recommender Systems)¹ that is a well-known conference on recommender systems of ACM. It showed that recommender systems have attracted the attention of the research community.

Adomavicius and Tuzhilin provide a survey of the state-of-the-art and possible extensions for recommender systems [3]. Traditional recommender systems are usually divided into three categories: (1) content-based filtering; (2) collaborative filtering and (3) hybrid recommendation systems [3]. Content-based

¹ <http://recsys.acm.org>

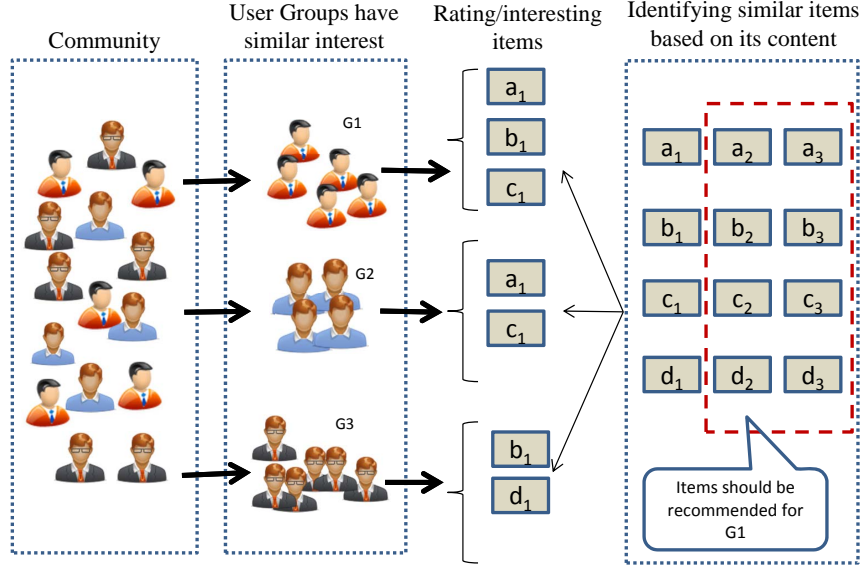


Fig. 1. Content-based filtering

approaches compare the contents of the item to the contents of items in which the user has previously shown interest (figure 1). Collaborative Filtering (CF) determines similarity based on collective user-item interactions, rather than on any explicit content of the items (figure 2). These traditional approaches do not mention relationships which can effect to behaviors and interests of individuals. Combining the social network analysis approach with traditional approaches can help us deal with these disadvantages.

Graphical models, a 'marriage' between probability theory and graph theory, provide a natural tool for dealing with two problems that occur throughout applied mathematics and engineering are uncertainty and complexity [18]. Graphical Models can be considered as expressive tools for analyzing, computing and modeling behaviors, relationships and influence of users in social networks.

In this work, we present our research proposal to do recommendations in the academic domain based on the social network analysis approach. These recommendations aim to support activities of researchers, reviewers while doing research such as research paper recommendation, collaboration recommendation, publication venue recommendation, paper reviewing recommendation, etc.

2 Related work

Recommender systems are widely used nowadays, especially in E-Commerce. Park et al. collected and classified articles on recommender systems from 46 journals published between 2001 and 2010 to understand the trend of recommender

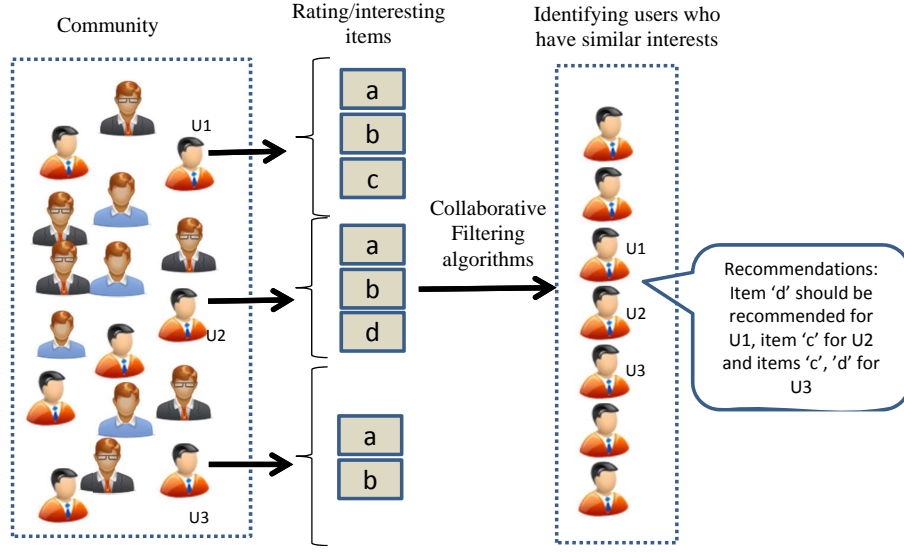


Fig. 2. collaborative filtering

system research and to provide practitioners and researchers with insight and future direction on recommender systems [31]. Their statistical numbers showed that recommender systems have attracted the attention of academics and practitioners. The majority of those research papers relates to movie (53 out of 210 research papers, or 25.2%) and shopping (42 out of 210 research papers, or 20.0%) [31]. In another research, Li et al. said that the utilization of recommender system in academic research itself has not received enough attention [21].

The online world has supported the creation of many research-focused digital libraries such as the Web of Science, ACM Portal, Springer Link, IEEE Xplore, Google Scholar, and CiteSeerX. Initially, these were viewed as somewhat static collections of research literature. These traditional digital libraries and search engines support the discovery of relevant documents but they do not traditionally provide community-based services such searching for people who share similar research interests. Recently, new research is focusing on these as enablers of a community of scholars, building and analyzing social networks of researchers to extract useful information about research domains, user behaviors, and the relationships between individual researchers and the community as a whole.

Microsoft Academic Search, ArNetMiner [36], and AcaSoNet [2] are online, web-based systems whose goal is to identify and support communities of scholars via their publications. The entire field of social network systems for the academic community is growing quickly, as evidenced by the number of other approaches being investigated [1][28][27][6][26].

As we mentioned above, traditional recommender systems are usually divided into three categories: (1) content-based filtering; (2) collaborative filtering

and (3) hybrid recommendation systems [3]. Content-based approaches compare the contents of the item to the contents of items in which the user has previously shown interest. Automated text categorization is considered as the core of content-based recommendation systems. This supervised learning task assigns pre-defined category labels to new documents based on the document's likelihood of belonging to a given class as represented by a training set of labeled documents [39]. Yang et al. reported a controlled study with statistical significance tests on five text categorization methods: Support Vector Machines (SVM), k-Nearest Neighbors (kNN) classifier, neural network approach, Linear Least-squares Fit mapping and a Nave Bayes classifier [39]. Their experiments with the Reuters data set showed that SVM and kNN significantly outperform the other classifiers, while Nave Bayes underperforms all the other classifiers. In other work, kNN was found to be an effective and easy to implement that could, with appropriate feature selection and weighting, outperform SVM [9]. So, kNN was considered as a baseline to compare with our proposed methods for the publication venue recommendation problem [25].

Collaborative Filtering (CF) determines similarity based on collective user-item interactions, rather than on any explicit content of the items. Su et al. has summarized a detail review of some main CF recommendation techniques [35]. There are two main methods in CF: (i) memory-based; and (ii) model-based. Memory-based algorithms operate on the entire user-item rating matrix and generate recommendations by identifying a neighborhood for the target user to whom the recommendations will be made, based on the agreement of user's past ratings. Memory-based techniques have some drawback including the sparsity of the user-item rating matrix due to the fact that each user rates only a small subset of the available items and inefficient computation of the similarity between every pair of users (or items) within large-scale datasets. To deal with challenges associated with the sparse and high dimensional dataset in the research paper domain, Lance Parsons et al. presented a survey of the various subspace clustering algorithms. They also compared the two main approaches to subspace clustering and discussed some potential applications where subspace clustering could be particularly useful [32]. Agarwal et al. proposed a scalable subspace clustering algorithm ScuBA which can be applied for research paper recommender systems and for research group collaboration. They took advantage of the unique characteristics of the data in the research paper domain and provided a solution which is fast, scalable and produced high quality recommendations [4][5].

To overcome the weaknesses of memory-based techniques new research focuses on model-based clustering techniques including social network-based or clustering techniques using social information that aim to provide more accurate, yet more efficient, methods. Pham et al. proposed model-based techniques that use the rating data to train a model and then the model is used to derive the recommendations [33]. In another recommendation research using CF, Li et al. proposes a basket-sensitive random walk model for personalized recommendation in the grocery shopping domain. Their proposed method extends the basic random walk model by calculating the product similarities through a weighted

bipartite network and allowing the current shopping behaviors to influence the product ranking scores [22]. In general, the basic idea of the traditional recommendation approaches is to discover users with similar interests or items with similar characteristics or the combination of these. The traditional approaches do not mention the relationship which can effect to the behavior and the interest of individuals.

Social network analysis (SNA) is a quantitative analysis of relationships between individuals or organizations to identify most important actors, group formations or equivalent roles of actors within a social network [19]. SNA is considered a practical method to improve knowledge sharing and it is being applied in a wide variety of contexts [29]. However studies on recommender systems using social network analysis are still deficient. Therefore, developing the recommendation system research using social network analysis will be an interesting area further research [31]. In particular, Kirchhoff et al. [19][20] and Gou et al. [11] apply SNA to enhance an information retrieval (IR) systems. Xu et al and Liu et al applied SNA to detect terrorist crime groups [37][23].

New research recently focuses on SNA approach and also the combination of the traditional approaches and the SNA to bring out better recommendations. Jianming He et al. presented a social network-based recommender system (SNRS) which makes recommendations by considering a user's own preference, an item's general acceptance and influence from friends [12]. They collected data from a real online social network and their analyzing on this dataset reveals that friends have a tendency to review the same restaurants and give similar ratings. Their experiments with the same dataset shown that SNRS outperformed than other methods, such as collaborative filtering (CF), friend average (FA), weighted friends (WVF) and naive Bayes (NB). Yunhong Xu et al. presented using social network analysis as a strategy for E-Commerce Recommendation [38]. Walter Carrer-Neto et al presented a hybrid recommender system based on knowledge and social networks. Their experiments in the movie domain shown promising results compared to traditional methods [7].

Recently, it has emerged some researches applied social network analysis in the academic area such as building a social network system for analyzing publication activities of researchers [2], research paper recommendation [16][30][21][10], collaboration recommendation [8][24], publication venue recommendation [25][33]. In order to extracting useful information from an academic social network Zhuang et al. proposed a set of novel heuristics to automatically discover prestigious (and low quality) conferences by mining the characteristics of Program Committee members [40]. Chen et al. introduces CollabSeer, a system that considers both the structure of a co-author network and an author's research interests for collaborator recommendation [8]. CollabSeer suggests a different list of collaborators to different users by considering their position in the co-authoring network structure. In work related to publication venues recommendation, Pham et al. proposed a clustering approach based on the social information of users to derive the recommendations [33]. They studied the application of the clustering

approach in two scenarios: academic venue recommendation based on collaboration information and trust-based recommendation.

In summary, traditional approaches for recommendation do not mention the users' relationship which can effect to the behavior and the interest of individuals. So, we are going to apply the Social Network Analysis approach combine with traditional methods to develop recommender systems in the academic domain which has not received enough attention.

3 Research Procedures

3.1 Overview of our research

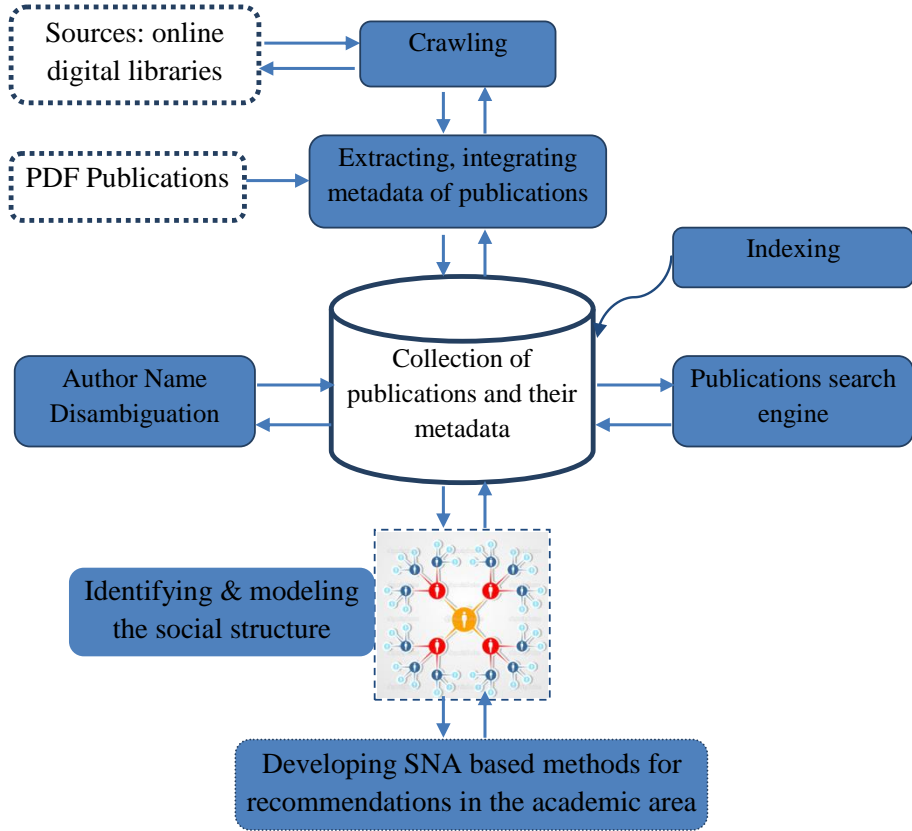


Fig. 3. A framework for SNA based recommender systems in the academic area

In order to develop SNA based methods used for recommendations in academic research field, we need to do some prepared steps or to solve some sub

problems such as extracting, integrating metadata of publications from many various sources, identifying and modeling the social structure from this collection. The overview of these tasks is shown in the picture 3.

3.2 Research methodology

There are many various research methodologies, but we have applied research methods such as quantitative and qualitative analyzing methods, trial-and-error methods, modeling methods, and experiment-and-evaluation methods.

3.3 Planing Specific Procedures

Table 1. The list of research procedures

Specific tasks	Research methodology
Studying the overview of recommender systems and approaches for recommendation.	Survey, the quantitative and qualitative analyzing methods
Studying the fundamentals of graphical models and its application in social network analysis.	Survey, the quantitative and qualitative analyzing methods
Crawling science publications from various online.	Experiment-and-evaluation methods
Analyzing, extracting the bibliographical data of science publications.	Trial-and-error method; experiment-and-evaluation methods
Building the collaborative network from the collection of publications.	The quantitative and qualitative analyzing methods, the modeling methods
Modeling and analyzing collaborative behaviours of the research community by using probability graphical approach.	The quantitative and qualitative analyzing methods, the modeling methods
Developing measures, algorithms, methods based on probabilistic inference in the collaborative network to improve the recommendation results in the academic domain. (Focus on the recommendation problems such as research paper recommendation, collaboration recommendation, publication venue recommendation.)	The quantitative and qualitative analyzing methods, trial-and-error methods, experiment-and-evaluation methods.

4 Our initial results

We have solved subproblems which mentioned in the picture 3 for our research objective. We set focus on computer science publications. We proposed methods and developed tools used for extracting and integrating metadata of computer science publication from online digital libraries. We used JAPE grammar of

GATE to define rules, patterns for extracting metadata from PDF publications [13][14]. In order to have a rich collection of computer science publications, we developed tools and methods for integrating bibliographical data of these publications from various online digital libraries [17].

To identify and model social structure from the collection of these papers, we proposed a collaborative knowledge model that based on graph theory and probability measures [15]. The model and measures can be used to identify users or groups that have same interest in the network. It is useful information for recommendation. We also developed and improved methods based on the collaborative network analysis approach for research paper recommendation [16] and publication venue recommendation [25].

5 Conclusion and future work

In this paper, we presented our research proposal based on social network analysis approach to develop recommender systems in the academic domain. We did the literature review related to recommender systems, social network analysis: methods and applications. Our research problem is a interesting problem which has attracted the attention of the research community in many different fields such as computer science, social science. The proposed approach can be used to fill the gap of traditional approaches. With our initial results, we believe that the approach based on social network analysis is a potential approach. For the future work, we have developed methods based on the science collaborative network analysis for recommender systems in the academic area.

References

1. Abbasi, A., Altmann, J.: On the correlation between research performance and social network analysis measures applied to research collaboration networks. In: Proceedings of the 2011 44th Hawaii International Conference on System Sciences. pp. 1–10. HICSS '11, IEEE Computer Society, Washington, DC, USA (2011)
2. Abbasi, A., Altmann, J.: A social network system for analyzing publication activities of researchers. TEMEP Discussion Papers 201058, Seoul National University; Technology Management, Economics, and Policy Program (TEMEP) (2010)
3. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowl. and Data Eng.* 17, 734–749 (June 2005)
4. Agarwal, N., Haque, E., Liu, H., Parsons, L.: Research paper recommender systems: a subspace clustering approach. In: Proceedings of the 6th international conference on Advances in Web-Age Information Management. pp. 475–491. WAIM'05, Springer-Verlag, Berlin, Heidelberg (2005), http://dx.doi.org/10.1007/11563952_42
5. Agarwal, N., Haque, E., Liu, H., Parsons, L.: A subspace clustering framework for research group collaboration. *IJITWE* pp. 35–58 (2006)
6. Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A.P., Arpinar, I.B., Joshi, A., Finin, T.: Semantic analytics on social networks:

- experiences in addressing the problem of conflict of interest detection. In: Proceedings of the 15th international conference on World Wide Web. pp. 407–416. WWW '06, ACM, New York, NY, USA (2006)
7. Carrer-Neto, W., Hernandez-Alcaraz, M.L., Valencia-Garca, R., Garca-Snchez, F.: Social knowledge-based recommender system. application to the movies domain. *Expert Systems with Applications* 39(12), 10990 – 11000 (2012), <http://www.sciencedirect.com/science/article/pii/S0957417412004952>
 8. Chen, H.H., Gou, L., Zhang, X., Giles, C.L.: Collabseer: a search engine for collaboration discovery. In: Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries. pp. 231–240. JCDL '11, ACM, New York, NY, USA (2011), <http://doi.acm.org/10.1145/1998076.1998121>
 9. Cunningham, P., Delany, S.J.: k-nearest neighbour classifiers. Tech. Rep. UCD-CSI-2007-4, School of Computer Science and Informatics, University College Dublin, Ireland (2007)
 10. Ekstrand, M.D., Kannan, P., Stemper, J.A., Butler, J.T., Konstan, J.A., Riedl, J.T.: Automatically building research reading lists. In: Proceedings of the fourth ACM conference on Recommender systems. pp. 159–166. RecSys '10, ACM, New York, NY, USA (2010), <http://doi.acm.org/10.1145/1864708.1864740>
 11. Gou, L., Zhang, X.L., Chen, H.H., Kim, J.H., Giles, C.L.: Social network document ranking. In: Proceedings of the 10th annual joint conference on Digital libraries. pp. 313–322. JCDL '10, ACM, New York, NY, USA (2010)
 12. He, J.: A social network-based recommender system. Ph.D. thesis, Los Angeles, CA, USA (2010), aAI3437557
 13. Huynh, T., Hoang, K.: Automatic metadata extraction from scientific papers. In: Proceeding of IT@EDU. University of Information Technology, Phan Thiet, Viet-Nam (2010)
 14. Huynh, T., Hoang, K.: Gate framework based metadata extraction from scientific papers. In: Proceedings of the Education and Management Technology (ICEMT), 2010 International Conference on. pp. 188 – 191. Cairo, Egypt (2-4 Nov 2010 2010)
 15. Huynh, T., Hoang, K.: Modeling collaborative knowledge of publishing activities for research recommendation. In: ICCCI 2012 (2012)
 16. Huynh, T., Luong, H., Hoang, K., Gauch, S., Do, L., Tran, H.: Scientific publication recommendations based on collaborative citation networks. In: Proceedings of the 3rd International Workshop on Adaptive Collaboration (AC 2012) as part of The 2012 International Conference on Collaboration Technologies and Systems (CTS 2012). pp. 316 – 321. Denver, Colorado, USA (21-25 May 2012 2012)
 17. Huynh, T., Luong, H.P., Hoang, K.: Integrating bibliographical data of computer science publications from online digital libraries. In: ACIIDS (3). pp. 226–235 (2012)
 18. Jordan, M.I. (ed.): Learning in graphical models. MIT Press, Cambridge, MA, USA (1999)
 19. Kirchhoff, L.: Applying Social Network Analysis to Information Retrieval on the World Wide Web. Ph.D. thesis, the University of St. Gallen, Graduate School of Business Administration, Economics, Law and Social Sciences (HSG) (2010)
 20. Kirchhoff, L., Stanoevska-Slabeva, K., Nicolai, T., Fleck, M., Stanoevska, K.: Using social network analysis to enhance information retrieval systems. In: Applications of Social Network Analysis (ASNA) (Zurich). vol. 7, pp. 1–21 (2008)
 21. Li, C.P.W.: Research paper recommendation with topic analysis. In: Computer Design and Applications (ICCD), 2010 International Conference. pp. 264–268. IEEE (2010)

22. Li, M., Dias, B.M., Jarman, I., El-Deredy, W., Lisboa, P.J.: Grocery shopping recommendations based on basket-sensitive random walk. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1215–1224. KDD '09, ACM, New York, NY, USA (2009), <http://doi.acm.org/10.1145/1557019.1557150>
23. Liu, Q., Tang, C., Qiao, S., Liu, Q., Wen, F.: Mining the core member of terrorist crime group based on social network analysis. In: Proceedings of the 2007 Pacific Asia conference on Intelligence and security informatics. pp. 311–313. PAIST'07, Springer-Verlag, Berlin, Heidelberg (2007), <http://dl.acm.org/citation.cfm?id=1763599.1763644>
24. Lopes, G.R., Moro, M.M., Wives, L.K., De Oliveira, J.P.M.: Collaboration recommendation on academic social networks. In: Proceedings of the 2010 international conference on Advances in conceptual modeling: applications and challenges. pp. 190–199. ER'10, Springer-Verlag, Berlin, Heidelberg (2010), <http://dl.acm.org/citation.cfm?id=1927973.1928011>
25. Luong, H.P., Huynh, T., Gauch, S., Do, L., Hoang, K.: Publication venue recommendation using author network's publication history. In: ACIIDS (3). pp. 426–435 (2012)
26. Matsuo, Y., Mori, J., Hamasaki, M., Nishimura, T., Takeda, H., Hasida, K., Ishizuka, M.: An advanced social network extraction system from the web. *Journal of Web Semantics* 5 (4), 262–278 (2007)
27. Mika, P.: Flink: Semantic web technology for the extraction and analysis of social networks. *Journal of Web Semantics* 3, 211–223 (2005)
28. Miki, T., Nomura, S., Ishida, T.: Semantic web link analysis to discover social relationships in academic communities. In: Proceedings of the The 2005 Symposium on Applications and the Internet. pp. 38–45. IEEE Computer Society, Washington, DC, USA (2005)
29. Mller-Prothmann, T.: Social network analysis: A practical method to improve knowledge sharing. In: Hands-On Knowledge Co-Creation and Sharing. pp. 219–233 (2007)
30. Ohta, M.; Hachiki, T.T.A.: Related paper recommendation to support online-browsing of research papers. In: Applications of Digital Information and Web Technologies (ICADIWT), 2011 Fourth International Conference. pp. 130–136 (2011)
31. Park, D.H., Kim, H.K., Choi, I.Y., Kim, J.K.: A literature review and classification of recommender systems research. *Expert Syst. Appl.* 39(11), 10059–10072 (Sep 2012), <http://dx.doi.org/10.1016/j.eswa.2012.02.038>
32. Parsons, L., Haque, E., Liu, H.: Subspace clustering for high dimensional data: a review. *SIGKDD Explorations* 6(1), 90–105 (2004)
33. Pham, M.C., Cao, Y., Klamma, R., Jarke, M.: A clustering approach for collaborative filtering recommendation using social network analysis. *J. UCS* 17(4), 583–604 (2011)
34. Pudhiyaveetil, A.K., Gauch, S., Luong, H.P., Eno, J.: Conceptual recommender system for citeseerx. In: RecSys. pp. 241–244 (2009)
35. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. *Adv. in Artif. Intell.* 2009, 4:2–4:2 (Jan 2009), <http://dx.doi.org/10.1155/2009/421425>
36. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., Su, Z.: Arnetminer: extraction and mining of academic social networks. In: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 990–998. KDD '08, ACM, New York, NY, USA (2008)

37. Xu, J.J., Chen, H.: Crimenet explorer: a framework for criminal network knowledge discovery. *ACM Trans. Inf. Syst.* 23(2), 201–226 (Apr 2005), <http://doi.acm.org/10.1145/1059981.1059984>
38. Xu, Y., Ma, J., Sun, Y.H., Hao, J., Sun, Y., Zhao, Y.: Using social network analysis as a strategy for e-commerce recommendation. In: *PACIS*. p. 106. *AISel* (2009), <http://dblp.uni-trier.de/db/conf/pacis/pacis2009.html#XuMSHSZ09>
39. Yang, Y., Liu, X.: A re-examination of text categorization methods. In: *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. pp. 42–49. *SIGIR '99*, ACM, New York, NY, USA (1999)
40. Zhuang, Z., Elmacioglu, E., Lee, D., Giles, C.L.: Measuring conference quality by mining program committee characteristics. In: *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*. pp. 225–234. *JCDL '07*, ACM, New York, NY, USA (2007), <http://doi.acm.org/10.1145/1255175.1255220>