# Exploiting Social Networks for Publication Venue Recommendations

Hiep Luong[1], Tin Huynh[2], Susan Gauch[1], and Kiem Hoang[2]
[1]*University of Arkansas, U.S.A*
[2]*University of Information Technology, Viet Nam*
{*hluong, sgauch*}*@uark.edu;* {*tinhn, kiemhv*}*@uit.edu.vn*

Keywords: Recommender systems, social network analysis, publication history, kNN, machine learning.

Abstract: The impact of a publication venue is a major consideration for researchers and scholars when they are deciding where to publish their research results. By selecting the right conference or journal to which to submit a new paper minimizes the risk of wasting the long review time for a paper that is ultimately rejected. The choice of publication venue is usually based on the author's existing knowledge of publication venues in their research domain or the match of the conference topics with their paper's contents. They may not be aware of new or other more appropriate conference venues to which their paper could be submitted. Traditionally, conferences are recommended to a researcher by analyzing her paper and comparing it to the topics of different conferences using content-based analysis. However, this approach can make errors due to mismatches caused by ambiguity in text comparisons and there is also much more to selecting an appropriate venue than just topical-matching. In our work, we take advantage of relationships within the academic community, as indicated by co-authorship activities, as part of the venue recommendation process. In this paper, we present a new approach allowing researchers to automatically find appropriate publication venues for their research paper by exploring the author's network of related co-authors and other researchers in the same domain. We also recommend appropriate publication venues to a specific user based on her relation with the program committee research activities and with others in her network who have similar paper submission preferences. Experiments with a set of ACM SIG conferences show that our new approach outperforms the content-based approach and provides more accurate recommendations. This works also demonstrates the feasibility of our ongoing work aimed at using social network analysis of researchers and experts in relevant research domains for a variety of recommendation tasks.

## 1 INTRODUCTION

The World Wide Web and its evolving infrastructure have played a significant role in the information explosion. This voluminous amount of unstructured and semi-structured data creates "big data," data sets that grow so large that they become awkward to work with or analyse using existing data management approaches. With the fast growth of digitalized textual data and documents, there is an urgent need for powerful text information management tools to help users find exactly what they are looking for and to help researchers keep abreast of information of whose existence they may be unaware. Recommender systems are one approach to helping users deal with the flood of information. They are tools that automatically filter a large set of items, e.g., movies, books, scientific papers, music, etc., in order to identify those that are most relevant to a user's interest. Recommender systems basically are divided into three categories: (1) content-based filtering; (2) collaborative filtering; and (3) hybrid recommendation systems. Content-based filtering uses actual content features of items, while the collaborative filtering predict new user's preference using other users' rating, assuming the like-minded people tend to have similar choices (Adomavicius and Tuzhilin, 2005).

There are a wide variety of dissemination outlets for research results, e.g., conferences, journals, seminars, scientific forums. When an author has a paper that they want to share, the review cycle can be time consuming and, if the paper is rejected because it is not a good fit, valuable time can be lost. In computer science, in particular, the pace of innovation is high. Selecting the right publication venue the first time is particularly important.

Traditional techniques usually use citation-based metrics with certain bibliometrics, such as the Impact Factor (Garfield, 1955), to measure the reputation and quality of publication. However, these techniques require frequent updates to the bibliometrics in order to maintain an accurate impact factor. However, because conferences are created and discontinued on a regular basis, the quality of a conferences is generally assessed by its reputation within the research community rather than a formal impact factor.

Recently there has been considerable interest in applying social network-based methods for ranking conference quality (Yan and Lee, 2007), seeking research collaborators (Chen et al., 2011) or generating recommendations (Klamma et al., 2009) (OurWorkReference). Recommendation systems are particularly important for researchers and scholars in their professional research activities For some experts in a research domain, or senior researchers who have strong publication records, selecting a conference might be a trivial task since they know well which conferences, journals, or scientific forums are the best places in which to publish their research papers. However, new researchers with less experience may not be able to easily assess conferences and the may not be current on relevant, new publication venues. For these researchers, an automatic system that recommends relevant venues matching their profiles, their professional networks, and research interests could be particularly useful.

The widespread use of the Internet allows researchers to create large, distributed academic social networks that can be analyzed to further enhance research productivity. Our interest is in how to use these academic social networks to recommend appropriate publication venues to authors for an unpublished paper. In this work, we present a new method to recommend conferences and journals to researchers based on social network analysis of researchers in computer science. Our approach is empirically evaluated using a dataset of recent ACM conference publications and compared with a baseline content-based.

In this paper, we present a survey of current research on content-based and collaborative filtering recommender systems, and recent trends applying social network analysis in recommender systems in section 2. We will focus mainly on recommendation research for academic activities and digital libraries. In section 3, we present our social netowork-based recommendation. Section 4 presents our content-based recommendation approach. Next, we present and discuss some experimental results for both appoaches in section 5. The final sections present our conclusions and discuss our future work in this area.

# 2 RELATED WORK

Traditional recommender systems are usually classified as content-based, collaborative, or hypbrid based on the type of information that they use and on how they use that information (Adomavicius and Tuzhilin, 2005).

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 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 labelled documents (Yang and Liu, 1999). 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 (Yang and Liu, 1999). 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 (Cunningham and Delany, 2007).

Collaborative Filtering (CF) determines similarity based on collective user-item interactions, rather than on any explicit content of the items. (Su and Khoshgoftaar, 2009) has summarized a detail review of some main CF recommendation techniques. 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. Model-based techniques use the rating data to train a model and then the model is used to derive the recommendations (Pham et al., 2011).

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 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., 2011).

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 [1], ArNetMiner (Tang et al., 2008), and AcaSoNet (Abbasi and Altmann, 2010) 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 (Abbasi and Altmann, 2011), (Miki et al., 2005), (Aleman-Meza et al., 2006), (Matsuo et al., 2007) and (Mika, 2005).

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 (Kirchhoff, 2010). SNA is considered a practical method to improve knowledge sharing (Mller-Prothmann, 2007) and it is being applied in a wide variety of contexts. In particular, (Kirchhoff et al., 2008) and (Gou et al., 2010) apply SNA to enhance an information retrieval (IR) systems.

In order to extracting useful information from an academic social network, (Zhuang et al., 2007) proposed a set of novel heuristics to automatically discover prestigious (and low quality) conferences by mining the characteristics of Program Committee members. (Chen et al., 2011) introduces CollabSeer, a system that considers both the structure of a co-author network and an author's research interests for collaborator recommendation. 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., 2011) proposed a clustering approach based on the social information of users to derive the recommendations. They studied the application of the clustering approach in two scenarios: academic venue recommendation based on collaboration information and trust-based recommendation.

[1] http://academic.research.microsoft.com/

To our best knowledge, we have not seen existing research that exploits the relationships between the authors of an unpublished paper with conference PC members or people in the social network who have previous publications to recommend appropriate publication venues.

# 3 SOCIAL NETWORK-BASED RECOMMENDATION APPROACH

## 3.1 Overview

In this section, we present a new social network-based approach to recommending a list of appropriate conference venues to an author for their unpublished paper. We introduce a new approach to analyse a large-scale network of researchers that we collected from the ACM digital library and Microsoft Academic Search. By analysing this large-scale social network, we recommend publication venues to the unpublished paper's authors based on the 'similarity' they have with conference PC members or with other authors with papers published in the conferences.

To build our dataset, we selected four subdomains of research in Computer Science corresponding to four SIGs (Special Interest Groups). Then, we manually picked four different conferences for each SIG. The total list of selected conferences contained the following 16 conferences:

***SIGBED - Special Interest Group on Embedded System***

1. CASES - Compilers, Architecture, and Synthesis for Embedded Systems
2. CODES+ISSS - International Conference on Hardware/Software Codesign and Systems Synthesis
3. EMSOFT - International Workshop on Embedded Systems
4. SENSYS - Conference On Embedded Networked Sensor Systems

***SIGDA - Special Interest Group on Design Automation***

1. DAC - Design Automation Conference
2. DATE - Design, Automation, and Test in Europe
3. ICCAD - International Conference on Computer Aided Design
4. SBCCI - Annual Symposium On Integrated Circuits And System Design

***SIGIR - Special Interest Group on Information Retrieval***

1. CIKM - International Conference on Information and Knowledge Management
2. JCDL - ACM/IEEE Joint Conference on Digital Libraries
3. SIGIR - Research and Development in Information Retrieval
4. WWW - World Wide Web Conference Series

***SIGPLAN - Special Interest Group on Programming Languages***

1. GPCE - Generative Programming and Component Engineering
2. ICFP - International Conference on Functional Programming
3. OOPSLA - Conference on Object-Oriented Programming Systems, Languages, and Applications
4. PLDI - SIGPLAN Conference on Programming Language Design and Implementation

For each of these 16 conferences, we downloaded all published papers from 2008-2010 from the ACM digital library as well as the list of Program Committee (PC) members for each conference. Since we know the conferences that accepted (and published) each of these papers, we can use this information as the truth-list to evaluate our conference recommendation approaches. For each researcher in this dataset (author or co-author), we also collected the list of all their co-authors and the number of papers co-authored by each pair from Microsoft Academic Search.

## 3.2 Academic Social Network Analysis

In this section, we present recommendations based on a large-scale network of all researchers who have published in our dataset of all papers published over 3 years in 16 ACM conferences. We measure the closeness between authors and conferences by two new methods, the first based on the relationships between the candidate paper authors and the PC members, the second between the candidate paper authors and previously authors previously published in each conference.

This problem can be formalised as a relatedness calculation of vertices in a graph. Each researcher can be viewed as a vertex of the graph, and the graph edges represent co-author relationships with other researchers in the network. Note that a member (a vertex) in the network can be either a paper author, a PC member, or both.

Figure 1 represents part of an academic social network in which A1 and A2 are seeking an appropriate publication venue for their jointly authored, unpublished paper. In this figure, the blue nodes represent the candidate paper authors, the white nodes other researchers, and the orange nodes represent a particular conference's PC member. In our initial work, we weight all edges with 1 for simplicity. A1 and A2 are directly connected to authors B and D because they previously co-authored papers with them and B, for example, has also authored with D and PC member C.

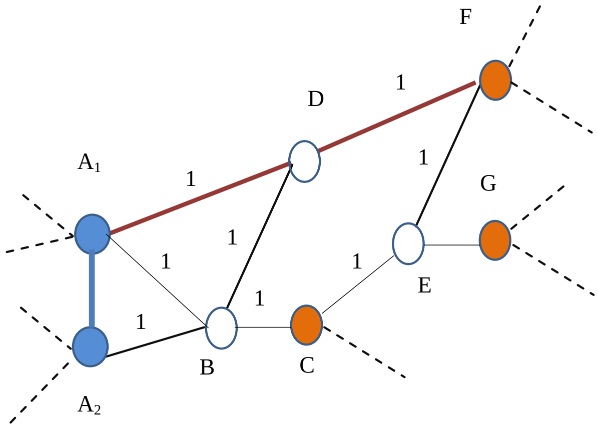


Figure 1: Example of a part of the social network

The closeness between authors is calculated based on the shortest path connecting them. For example, to determine the relation similarity between author A1 and a PC member F, there are several paths between these two nodes such as A1-D-F, A1-B-D-F, A1-B-C-E-F, etc. However, their shortest path, A1-D-F, contains 2 edges. The shorter the path, the closer the nodes are. Thus, closeness between two nodes in a graph, A and B, is the inverse of path length calculated as:

$$Closeness(A,B) = \frac{1}{ShortestPath(A,B)}$$

### 3.2.1 Closeness between paper authors and PC members (Author_PC method)

In this method, we recommend that a paper be submitted to the conference with the strongest relationship between the paper's authors and the conference's program committee. Essentially, this measures the impact that being an insider has to the likelihood of a paper being accepted. The more close relationships that a paper's authors have with PC members, the more likely that conference is to be recommended for the paper. Note that this favors conferences with large program committees and, in future work, we will investigate the effects of normalizing these weights by the number of PC members.

In particular, for conference i, we add together the strength of the relationship between each of the paper's co-authors with the PC as follows:

$$Author\_PC_i = \sum_{m=1}^{A} Closeness(m,i)$$

where A is the number of authors of the unpublished paper and Closeness(m,i) is the closeness between an author m and the program committee for conference i. Specifically, Closeness(m, i) is the sum of the closeness between that author and any member of the PC.

$$Closeness(m,i) = \sum_{j=1}^{N} Closeness(m, PC_{i,j})$$

with m is an author of the paper, PCi,j is the jth program committee member for conference i, and N is the total number of PC members for that conference.

The conference with the highest value is recommended as the publication venue.

#### 3.2.2 Closeness between paper authors and previous conference authors (Author_NetAuthors method)

In this method, we consider how close a paper's authors are to those in the network who have published their paper(s) in a specific conference. Essentially, this is based on the belief that, if papers authors academic colleagues have had their work published by a particular conference in the past, this is a good indication that this paper is also likely to be acceptable. This relation similarity can be defined as following:

$$Author\_NetAuthors_i = \sum_{m=1}^{A} Closeness(m, NetAuthors_i)$$

where A is a number of author(s) of the unpublished paper, and Closeness(m, NetAuthorsi) is the closeness between an author m and all other authors in the network who have published their papers to the conference i. This value Closeness(m, NetAuthorsi) is also calculated as the sum of the closeness between that author with any other author relevant to the conference i.

$$Closeness(m, NetAuthors_i) = \sum_{j=1}^{M} Closeness(m, Authors_{i,j})$$

with $m$ is an author of the paper, $Author_{i,j}$ is the $j^{th}$ author in the network who has paper published in the conference $i$. $M$ is total number of all members that have papers published in the conference $i$.

We recommend the conference with the highest value as a publication venue for the unpublished paper.

## 4 EXPERIMENTS

### 4.1 Dataset

In order to get the publication history of authors, we developed a focused crawler in Java that extracts all co-authors and relevant publications for a given author from the Microsoft Academic Search (MAS) website. This website updates currently more than 27 million publications and 16 million of authors with several thousand new updates every week in natural and life sciences.

As presented in the section 3, our input contains 16 ACM conferences of 4 SIGs. With all papers collected from 2008-2010, we used the papers published in two years 2008 and 2009 as training documents and the ones published in 2010 as test documents for the classification task. Since the number of papers published for each conference varies, we randomly selected 20 documents per year per conference (60 totals). Thus, each conference had 40 training and 20 test documents. With 16 conferences, the total test collection contained 640 training and 320 test documents. We split the 320 test documents into two sets: 160 for tuning and 160 for validation. For each of the 16 conference instances, we also downloaded the names of the program committee members from the conference website for the year 2010.

For each paper in the test collection, we extracted the author names and used a crawler to gather information about each author's publications and co-authors. The co-authors of the co-authors were then recursively collected, until a network 3 levels deep was created. As a result, we collected information about 306,227 authors and 392,878 papers. We also submitted the names of the PC members to MAS to collect their authorship relationship. This information was downloaded and stored in a database. We manually reviewed authors with large numbers of publications to remove publications that were incorrectly attributed to an author due to the ambiguity of the author's name.

Finally, we built a large-scale graph, representing the network of researchers and experts, containing 303,843 vertices (authors) and 1,220,472 edges (co-authorship relationships with an average node degree of 8.03.

## 4.2 Baseline

In order to evaluate our new recommendation approach using academic social networks, we compared the methods described in section 3 with a baseline of content-based recommendations. In the baseline, conferences in the dataset are treated as different categories into which we classify a candidate paper. Each paper is represented by a vector of terms weighted by TF-IDF. The similarity between two papers is calculated using the cosine similarity measure.

Papers are classified into conferences using a k-Nearest Neighborhood (kNN) algorithm (Gauch et al., 2004) trained using the 640 training documents. In order to identify the best k value for our experiment, we varied k, the number of classifier results used for to select the conference, from 3-40. Based in previous testing (OurWorkReference), we report our best performing k (i.e., k = 25) at which the content-based approach reached the highest classification precision.

We also compare the performance of our new research to our previous work, i.e., PubHistory that recommends conferences to authors based on their own publication histories, rather than the publication histories of researchers in their academic social network (OurWorkReference).

## 4.3 Evaluation Metrics

We evaluated each method's performance using precision, i.e., the percentage of the time that the recommender system recommends the conference and/or SIG in which the test paper actually appeared:

Conference Precision: measures the percentage of time that the recommender recommends the true conference. This is reported at various cut-offs, i.e., Top1 means that the correct conference was the top-ranked recommendation, Top2 that the correct conference was within the two highest-ranked conferences, etc.

SIG Precision: measures the percentage of time that the recommender recommends a conference from the correct SIG. Since there are fewer SIGS, and they differ more than the conferences do, this is an easier task. Thus, we only report the SIG precision for the top-ranked conference's SIG.

## 4.4 Results and Discussion

Table 1 summarizes the conference precision results for each of the four methods using the 160 tuning documents. The conference precision results are also shown graphically in Figure 2.

Table 1. Conference precision results using four different methods.

| | ***Conference Precision*** | | | |
|---|---|---|---|---|
| | **Top1** | **Top2** | **Top3** | **Top4** |
| Content-Based | *48.8%* | *68.8%* | *80.6%* | *90.6%* |
| PubHistory | 66.2% | 83.8% | 95.5% | 96.8% |
| Author_PC | 43.6% | 59.6% | 69.9% | 77.6% |
| **Author_Net Authors** | **74.5%** | **91.0%** | **94.3%** | **96.8%** |

These results show that the Author_NetAuthors method is by far the most accurate method for recommending a conference. It recommends the correct conference as the top choice 74.5% of the time and the correct conference is within the top 4 choices almost 97% of the time. Authors tend to submit to, and be accepted by, conferences in which their co-authors (direct or indirect) have previously been published. Interestingly, this method outperforms the PubHistory method (66.2% Top1 conference precision) that recommends conferences based only on the author's own publication history. The difference is most evident in the Top1 and Top2 conference precision. Clearly, information from the author's academic social network helps identify good publication venues.

The two methods based on publication history, Author_NetAuthors and PubHistory, both outperform the content-based baseline method by a wide margin. The content-based baseline achieved a Top1 conference precision of only 49%. Conferences within a SIG cover similar topics, but attract researchers from somewhat different communities. This information is exploited by the Author_NetAuthors method, but unavailable to the baseline that looks only at content.

The Author_PC method (43.6% Top1 conference precision) is the only method that underperforms the Content-Based method. This is actually a very positive result for the integrity of the research community. It demonstrates that relationships between authors and PC members does not predict acceptance of a paper at a conference. It is more important that the paper be of interest to the community, and on topic, than that the author has worked with someone on the PC in the past.

Since there are 16 conferences that overlap in topics, but only 4 SIGs that cover different research areas, predicting the correct SIG should be an easier task than predicting the correct conference. Table 2, which summarizes the SIG precision results for the four methods, confirms this hypothesis. Looking only at the SIG for the top-ranked conference, we see that

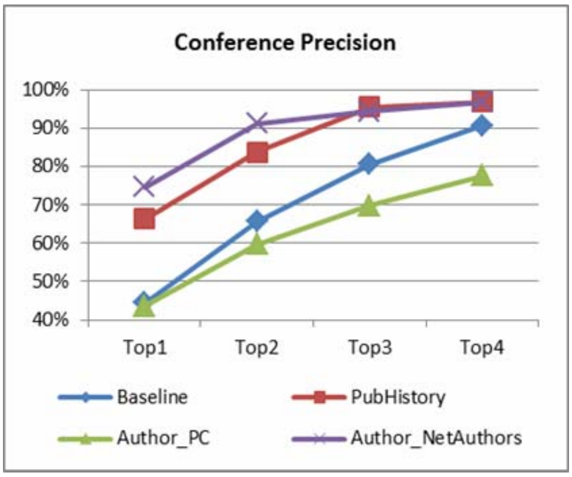


Figure 2: Comparison of classification precisions

the Author_NetAuthors method recommends the correct SIG 93.0%. Once again, PubHistory is the second best method, 87.6% accurate, followed by 80.6% for the content-based baseline. Author_PC is again last, only 74.4% accurate in recommending a conference from the correct SIG.

Table 2. SIG precision using four different methods.

| Methods | SIG Precision for the top-ranked |
|---|---|
| Content-Based | 80.63% |
| PubHistory | 87.66% |
| Author_PC | 74.36% |
| **Author_NetAuthors** | 93.00% |

These results confirm our hypothesis that a publication venue recommendation system can benefit from social network analysis instead of, or in addition to, traditional content-based approaches.

# 5 CONCLUSION

The goal of this research is to implement and evaluate a new approach to recommend publication venues for an unpublished article. Our approach takes advantage of information analysed from an academic social network of researchers linked by their co-authorship relationships.

We compared two social network approaches to recommending publication venues to authors of new papers. The first, Author_NetAuthors, recommended conferences based on the paper authors' relationships with authors previously published in a conference. The second, Author_PC, recommended conferences based on the paper authors' relationships with members of the conference's program committee. These new approaches were compared to a baseline content-based recommendation approach and also to recommendations based on the paper authors' own publication history. Our approaches were empirically tested using a dataset of a large-scale network of more than 306,000 published authors and evaluated using papers from 16 ACM conferences from 4 SIGs.

The results show that the Author_NetAuthors approach that incorporates relationships between a paper's authors' academic social network and each conference's network of previously published authors is the best performing result. The top-ranked result was the correct conference 74.5% of the time and it was from the correct SIG 93.0% of the time. Recommendations based only on the paper's authors own publication history was second best, (66.2%; 87.6%) followed by the content-based approach (48.8%; 80.6%). The approach based on relationships between the authors and program committee members, Author_PC, actually underperformed the content-based method (43.6%; 74.4%).

Overall, we conclude that social network-based approaches can outperform content-based approaches when recommending publication venues. They work well even when deciding between conferences that overlap in topics, a task that is very difficult for content-based recommender systems. We also showed that relationships with the community of authors who publish in specific conferences is more important than relationships with members of the conference's program committee members.

Our main tasks in the future are to enhance the publication venue recommendation system by developing algorithms that take into account more sophisticated graph relationships and different kinds of links in the network such as citation and other indications of research collaboration (e.g., researchers from the same institution). We also want to incorporate weighted links within the network and investigate the effect of incorporating more (or fewer) levels of co-authorship links in the recommendations. Finally, we wish to apply this algorithm within the CiteSeerX digital library.

# ACKNOWLEDGEMENTS

This research is partially supported by the NSF grant number 0958123 - Collaborative Research: CI-ADDO-EN: Semantic CiteSeer$^{X}$.

# REFERENCES


Abbasi, A. and Altmann, J. (2010). A social network system for analyzing publication activities of researchers. TEMEP Discussion Papers 201058, Seoul National University; Technology Management, Economics, and Policy Program (TEMEP).

Abbasi, A. and Altmann, J. (2011). 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*, HICSS '11, pages 1–10, Washington, DC, USA. IEEE Computer Society.

Adomavicius, G. and Tuzhilin, A. (2005). 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.

Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A. P., Arpinar, I. B., Joshi, A., and Finin, T. (2006). 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*, WWW '06, pages 407–416, New York, NY, USA. ACM.

Chen, H.-H., Gou, L., Zhang, X., and Giles, C. L. (2011). Collabseer: a search engine for collaboration discovery. In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*, JCDL '11, pages 231–240, New York, NY, USA. ACM.

Cunningham, P. and Delany, S. J. (2007). k-nearest neighbour classifiers. Technical Report UCD-CSI-2007-4, School of Computer Science and Informatics, University College Dublin, Ireland.

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, 122(3159):108–111.

Gauch, S., Madrid, J. M., Induri, S., Ravindran, D., and Chadlavada, S. (2004). Keyconcept: A conceptual search engine. Technical Report ITTC-FY2004-TR-8646-37, Information and Telecommunication Technology Center, University of Kansas.

Gou, L., Zhang, X. L., Chen, H.-H., Kim, J.-H., and Giles, C. L. (2010). Social network document ranking. In *Proceedings of the 10th annual joint conference on Digital libraries*, JCDL '10, pages 313–322, New York, NY, USA. ACM.

Kirchhoff, L. (2010). *Applying Social Network Analysis to Information Retrieval on the World Wide Web*. PhD thesis, the University of St. Gallen, Graduate School of Business Administration, Economics, Law and Social Sciences (HSG).

Kirchhoff, L., Stanoevska-Slabeva, K., Nicolai, T., Fleck, M., and Stanoevska, K. (2008). Using social network analysis to enhance information retrieval systems. In *Applications of Social Network Analysis (ASNA) (Zurich)*, volume 7, pages 1–21.

Klamma, R., Cuong, P. M., and Cao, Y. (2009). You never walk alone: Recommending academic events based on social network analysis. In *Complex Sciences*, volume 4 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 657–670. Springer Berlin Heidelberg.

Matsuo, Y., Mori, J., Hamasaki, M., Nishimura, T., Takeda, H., Hasida, K., and Ishizuka, M. (2007). An advanced social network extraction system from the web. *Journal of Web Semantics*, 5 (4):262–278.

Mika, P. (2005). Flink: Semantic web technology for the extraction and analysis of social networks. *Journal of Web Semantics*, 3:211–223.

Miki, T., Nomura, S., and Ishida, T. (2005). Semantic web link analysis to discover social relationships in academic communities. In *Proceedings of the The 2005 Symposium on Applications and the Internet*, pages 38–45, Washington, DC, USA. IEEE Computer Society.

Mller-Prothmann, T. (2007). Social network analysis: A practical method to improve knowledge sharing. In *Hands-On Knowledge Co-Creation and Sharing*, pages 219–233.

Pham, M. C., Cao, Y., Klamma, R., and Jarke, M. (2011). A clustering approach for collaborative filtering recommendation using social network analysis. *J. UCS*, pages 583–604.

Su, X. and Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Adv. in Artif. Intell.*, 2009:4:2–4:2.

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., and Su, Z. (2008). Arnetminer: extraction and mining of academic social networks. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, pages 990–998, New York, NY, USA. ACM.

Yan, S. and Lee, D. (2007). Toward alternative measures for ranking venues: a case of database research community. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, JCDL '07, pages 235–244, New York, NY, USA. ACM.

Yang, Y. and Liu, X. (1999). A re-examination of text categorization methods. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '99, pages 42–49, New York, NY, USA. ACM.

Zhuang, Z., Elmacioglu, E., Lee, D., and Giles, C. L. (2007). Measuring conference quality by mining program committee characteristics. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, JCDL '07, pages 225–234, New York, NY, USA. ACM.