Scientific Reference Mining using Semantic Information through Topic Modeling

ALI DAUD*, MUHAMMAD AKRAM SHAIKH, AND ALTAF HUSSAIN RAJPAR*****
Received on 26-01-2009 Accepted on 04-03-2009

Abstract—**This paper addresses one time-consuming task in preparing scientific paper of finding appropriate citations (references). In the past this task is usually performed manually or by using links information of papers. Link information only used title of paper and ignored the semantic-based text information present in the paper contents. Due to overlaps between different fields e.g. in computer science only title of a paper cannot be real representative of different other hidden topics of that paper. We think it is necessary to model the semantic information present in papers to provide more appropriate citations by capturing hidden topics. In this paper, we address this issue by modeling citations on the basis of latent topics present in the papers. Latent topics can provide us semantic correlations present between the papers. We propose a topic modeling approach in which each citation of a paper is represented as a probability distribution over latent topics, and each latent topic is represented as a probability distribution over words of paper for that topic, which can provide us more appropriate citations for a given paper. Experimental results on citeceer corpus shows the effectiveness of proposed approach and detailed interpretation of results reveals interesting information about scientific recommendation.**

Keywords-Scientific Recommendation; Information Extraction; Topic Modeling; Unsupervised Learning

I. INTRODUCTION

With the emergence of web world has become a global village and the competition between researchers has become very fast and challenging. Every researcher wants to fasten the process of different tasks related to scientific research including paper writing. In paper writing one very important task is to find most appropriate citations to provide a comprehensive review of literature to make your arguments strong. This task is usually fulfilled manually by using some keywords to search through the academics search engines such as Google scholar, DBLP and Citeseer. Keywords based citation search cannot benefit from semantic-based text information. It is also affected by the human inability to provide accurate keywords for finding accurate citations. Citations found form the search engine may be some time not exactly related to the specific area of researcher due to overlaps between different fields.

Previously, [17] viewed paper recommendation task as an information retrieval problem. They used whole text of paper as an input query and used text similarity and citations relations between the papers to fulfill the task. Differently, collaborative filtering is employed to recommend citations by considering papers as users and citations of papers as items [14]. [18] Recommended papers to users on the basis of network connectivity information present between papers by analyzing researcher's social network. Aforementioned approaches don't investigate the latent themes of the paper, because they ignored correlations between papers text and their citations. While, in real world co-occurrence of words and citations can provide a semantic relationship between the papers.

In this paper, we propose a scientific recommendation (SR) approach which can investigate the correlations between papers on the basis of semantic-based information present between papers text and their citations. SR approach is a variation of Author-topic modeling [16], and in topic modeling topics correspond to short descriptions (semantically related probabilistic clusters of words) in a corpus. Topic modeling is aimed at finding these short descriptions which can be utilized to correlate entities. We discovered latent topics of papers and their citations on the basis of semantic information present in the co-occurrence of text and citations of the papers. We used discovered topics to find appropriate citations for a paper. Solution provided by us for scientific paper recommendation task produced quite intuitive and functional results on real-world corpus.

The novelty of work described in this paper lies in the; formalization of the scientific paper recommendation task, proposal of a SR approach to deal with the task, and experimental verification of the effectiveness of our approach. To the best of our knowledge, we are the first to deal with this task by proposing a SR topic modeling approach.

The rest of the paper is organized as follows. Section II provides related work. In Section III, we formalize the scientific paper recommendation task. Section IV illustrates modeling documents with topics, modeling authors with topics and finally our SR approach with its parameter estimation details to model the citations. In Section V, corpus, experimental setup with empirical studies and discussions about the results are given and section VI brings this paper to conclusions.

II. RELATED WORK

A. *Scientific Recommendation*

Finding suitable citations for a paper has been a hot issue during previous years to reduce paper writing time. Different approaches are used to fulfill the task. Ranking algorithms Page Rank [5] and HITS [12] can be used to do citations analysis on the basis of in and out links in paper on the basis of citations. Bibliographic Coupling [11] is also good tool for citation analysis; on the other hand citation analysis can be used to measure the quality of research paper [7]. The whole text of paper was used as an input query by utilizing text similarity and citations links of papers to fulfill fill the information retrieval task [17], while [4,6,11] employed collaborative filtering to tackle recommendation issue. [11]

Recommend citations by considering papers as users and citations of papers as items. Content-based filtering [2] can be used to recommend items on the basis of correlations between the content of the items and the user's preferences. This method creates a profile for each item or user to characterize their nature. Network connectivity based approach is proposed by [18] recommended papers to users on the basis of citations links information present between papers. Aforementioned approaches were incapable of considering semantic information based correlations between text and citations, however our approach can benefit from it.

B. *Topic Modeling*

Automatic extraction of topics from text is performed by [13,15] to cluster documents into groups based on similar semantic contents. Clustering provides a good way to group similar documents, but clustering is inherently limited by the fact that each document is only associated with one cluster. This is somewhat odd with the multi-topic nature of text documents in many contexts. For this reason soft clustering representation techniques are mandatory, which can allow documents composed of multiple topics to relate to more than one cluster on the basis of their hidden topics. Probabilistic Latent Semantic Indexing (PLSI) [10] was proposed as a probabilistic alternative to projection and clustering methods. In PLSI, topics were modeled as multinomial probability distributions over words, and documents are assumed to be generated by the activation of multiple topics. While PLSI produced impressive results on a number of document modeling problems, the number of parameters in the model grows linearly with the size of the corpus, which leads to serious problems of over fitting and was not clear how to assign a probability to a document outside corpus. In order to overcome the shortcomings of PLSI, a more general probabilistic model LDA was proposed [3]. LDA assumes that each word in the document is generated by a hidden topic and explicitly models the words distribution of each topic as well as the prior distribution over topics in the document. Given these parameters, topics of all words in the same document are assumed to be independent. Later, LDA was extended to Author-Topic model [16], in which an author is represented as a probability distribution over words, and topic is represented as probability distribution over words. It was used for modeling the interests of authors on the basis of latent topics; however we used its variation for scientific recommendation.

III. PROBLEM SETTING

Automatic extraction of useful information from the corpus is a standard problem addressed in information retrieval, statistical natural language processing and machine learning. Our work is mainly focused on finding appropriate papers to cite in a newly written paper; as each paper requires some related work to be cited in it. Each paper contains author names, words and citations. We will utilize semantic dependencies between the text of papers and their citations to discover the latent topics present in the papers to do paper and citation recommendation.

We denote a paper p as a vector of N_p words with C_p citations represented as $(\mathbf{w}_p,\mathbf{c}_p)$ and formulize scientific paper recommendation problem as: Given a paper p with N_p words, find the appropriate citations for that paper. Formally for a paper, we need to predict most suitable citations. To solve this issue, we purpose a SR approach which can smooth the data from semantic level by considering text dependencies and providing us with soft clusters of papers based on the latent themes, which can be utilized to find appropriate citations for a given paper. Fig. 1 pictorially shows the issue investigated in this paper on the basis of latent topics.

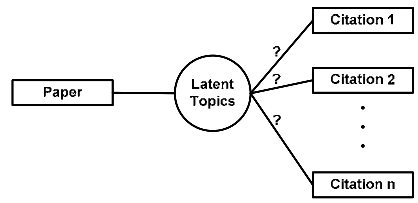

Figure 1: Scientific recommendation issue.

IV. OUR APPROACH

In this section, before describing our Scientific Recommendation (SR) approach, we will first describe how documents and authors are modeled with the latent topics.

A. *Modeling Documents with Topics*

Fundamental topic modeling assumes that there is a hidden topic layer $T = \{z_1, z_2, z_3, \ldots, z_t\}$ between the word tokens and documents. Here, z_i denotes a latent topic and each document d is a vector of N_d words $\mathbf{w}_d$. A collection of D documents is defined by $D = \{\mathbf{w_1}, \mathbf{w_2}, \mathbf{w_3}, \ldots, \mathbf{w_d}\}$ and each word w_{id} is chosen from a vocabulary of size V. For each document, a topic mixture distribution is sampled and a latent topic T is chosen with the probability of topic given document for each word with word having generated probability of word given topic [3,9,10].

B. *Modeling Authors with Topics*

Following topic modeling basic idea of modeling words and documents, words and authors are modeled by considering latent topics to discover the interests of authors [16]. In Author-Topic model, each author is represented by the probability distribution over topics and each topic is represented as a probability distribution over words for that topic. It can successfully discover the latent topics with respect to author's relationships by explicitly representing the topics in terms of the words of the documents. In this model, each topic is associated with a multinomial distribution Φ_z over words. Each author from a set of R authors is associated with a multinomial distribution θ_a over topics. Both θ_a and Φ_z have symmetric Dirichlet prior with hyper parameters α and β. For each word in the document, an author r is uniformly sampled from set of coauthors $\mathbf{a}_d$, then topic z is sampled from the multinomial distribution θ_a associated with author r and word w is sampled from multinomial topic distribution Φ_z associated with topic z.

C. *Modeling Citations with Topics*

The basic idea presented in [16], that words and authors can be modeled by considering latent topics became the intuition of modeling words and citations. We consider that a citation is responsible for generating some semantically related cluster of words (latent topics) which can be utilized to recommend citations for the papers. Each topic is associated with a multinomial distribution Φ_z over words. Each citation from a set of k citations of a paper p is associated with a multinomial distribution θ_c over topics. Both θ_c and Φ_z have symmetric Dirichlet prior with hyper parameters α and β. For each word in a paper p, a citation c is uniformly chosen from a set of citations $\boldsymbol{p_d}$, then topic z is sampled from the multinomial distribution θ_c associated with citation c and word w is sampled from multinomial topic distribution Φ_z associated with topic z.

The generative process is as follows:

- For each topic $z = 1,\ldots, T$
  Choose Φ_z from Dirichlet (β)
- For each citation $c = 1,\ldots, C$ of paper p
  Choose θ_c from Dirichlet (α)
- For each word $w = 1,\ldots, N_p$ of paper p
  Choose a citation c uniformly from all citations $\boldsymbol{p_d}$
  Choose a topic z from multinomial (θ_c) conditioned on c
  Choose a word w from multinomial (Φ_z) conditioned on z

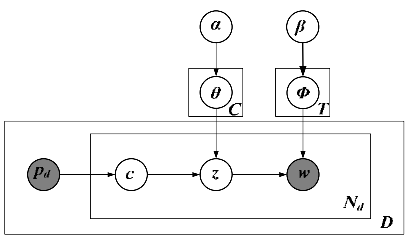

Figure 2: Scientific recommendation approach.

We utilize Gibbs sampling [1] for parameter estimation in SR approach. In SR approach, we have two latent variables z and c; the conditional posterior distribution for latent topic z and citation c is given by:

$$P(z_i = j, c_i = k | w_i = m, \mathbf{z}_{-i}, \boldsymbol{c}_{-i}, \boldsymbol{p_d}) \propto \frac{n_{-i,j}^{(wi)} + \beta}{n_{-i,j}^{(.)} + W\beta} \frac{n_{-i,j}^{(ci)} + \alpha}{n_{-i,.}^{(ci)} + T\alpha} \quad (1)$$

where $z_i = j$ and $c_i = k$ represent the assignments of the i^{th} word in a document to a topic j and citation k respectively, $wi = m$ represents the observation that i^{th} word is the m^{th} word in the lexicon, and z_{-i} and $\boldsymbol{c}_{-i}$ represents all topics and citations assignments not including the i^{th} word, respectively. Furthermore, $n_{-i,j}^{(wi)}$ is the total number of words associated with topic j, excluding the current instance, and $n_{-i,j}^{(ci)}$ is the number of times citation k is assigned to topic j, excluding the current instance, and W is the size of the lexicon. "." Indicates summing over the column where it occurs and $n_{-i,j}^{(.)}$ stands for number of all words that are assigned to topic z excluding the current instance.

During parameter estimation, the algorithm only needs to keep track of W x T (word by topic) and T x C (topic by citation) count matrices. From these count matrices, topic-word distribution Φ_z and citation-topic distribution θ_c can be calculated by:

$$\emptyset_{zw} = \frac{n_{-i,j}^{(wi)} + \beta}{n_{-i,j}^{(.)} + W\beta} \quad (2)$$

$$\theta_{cz} = \frac{n_{-i,j}^{(ci)} + \alpha}{n_{-i,.}^{(ci)} + T\alpha} \quad (3)$$

Where, $\emptyset_{zw}$ is the probability of word w in topic z and θ_{cz} is the probability of topic z for citation k. These values correspond to the predictive distributions over new words w and new topics z conditioned on w and z. To find T x P (topic by paper) count matrix we calculated the probability distribution of topic given paper as:

$$p(z|p) = \sum_{c \epsilon C_p} p(z|c)p(c|p) = \frac{1}{|C_p|} \sum_{c \epsilon C_p} p(z|c) \quad (4)$$

where, c_p is the number of citations belongs to a paper p.

V. EXPERIMENTS

A. Corpus

Citeseer corpus is well maintained and contains richer information about citations. In citeseer corpus each paper consists of title, abstract and its citations. We selected a subset of 3,335 papers from a total of 3,37,090 papers for conducting our experiments. We preprocessed corpus by a) removing stop-words, punctuations and numbers b) down casing the obtained words, and c) removing papers having citations less than 4. We divided selected papers into training corpus (80%) which consists of 2,668 papers, 4,345 citations and 9,642 words and testing data set (20%) which consists of 667 papers, 2,210 citations and 4,601 words. In testing data set, we only kept papers with at least one citation contained in the training data set.

B. Experimental Setup

One can estimate the optimal values of hyper-parameters α and β (fig. 2) by using variational-EM method [3] or Gibbs sampling algorithm [9]. In our experiments, for 200 topics T the hyper-parameters α and β were set at 50/T and .01 respectively, by following the values used in [16]. We ran five independent Gibbs sampling chains for 2000 iterations each. All experiments were carried out on a machine running Windows XP 2002 with AMD Athlon I Dual Core Processor

(1.90 GHz) and 1 GB memory. The run time per each chain was 2.1 hours.

C. Results and Discussions

This section reports results and discuss in detail the interpretation of results disclosing useful information about scientific recommendation.

Paper Recommendation: We extracted papers related to specific area of research on the basis of latent topics by using SR approach. Fig. 3 illustrates 3 different topics out of 200, discovered from the 2000th iteration of particular Gibbs sampler run. Each topic in fig. 3 shows 10 words that are most likely to be produced if the topic is activated, and 5 papers that are most likely to be related to that topic. The words associated with each topic are quite intuitive and precise in the sense of conveying semantic summary of a specific are of research. The papers associated with each topic are also quite representative, here it is obligatory to mention that top 5 papers associated with a topic are not necessarily the most cited papers in that area, but rather are the papers that tend to produce most words for that topic in the corpus.

Topic 91 “Machine Learning”		**Topic 88 “Data Mining”**		**Topic 16 “XML Databases”**	
Word	Probability	Word	Probability	Word	Probability
Learn	0.079376	Rule	0.060032	Data	0.126691
Class	0.025823	Data	0.057327	Queries	0.118686
Classifier	0.024956	Mining	0.043971	Database	0.063856
Train	0.024739	Association	0.040421	Structure	0.034439
Machine	0.023871	Algorithm	0.026052	Language	0.023633
Classification	0.023004	Databases	0.021825	View	0.019231
Method	0.021703	Set	0.016584	Xml	0.018630
Sample	0.017801	Pattern	0.016246	Schema	0.015429
Data	0.016066	Large	0.014725	semistructure	0.014228
Error	0.014548	Problem	0.012527	Relations	0.014028

Topic 91 “Machine Learning”	
Recommended Papers	Probability
Bagging Predictors	0.043777
A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting	0.033427
Information-Based Objective Functions for Active Data Selection	0.019697
Text Categorization with Support Vector Machines: Learning with Many Relevant Features	0.018218
Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods	0.014205
Topic 88 “Data Mining”	
Recommended Papers	Probability
Mining Association Rules between Sets of Items in Large Databases	0.139509
Fast Algorithms for Mining Association Rules	0.079713
Discovery of Multiple-Level Association Rules from Large Databases	0.061328
Efficient Algorithms for Discovering Association Rules	0.040457
Mining Sequential Patterns	0.038801
Topic 16 “XML Databases”	
Recommended Papers	Probability
The Lorel Query Language for Semistructured Data	0.061386
Object Exchange Across Heterogeneous Information Sources	0.043024
Answering Queries Using Views	0.027201
Querying Semi-Structured Data	0.023685
Lore: A Database Management System for Semistructured Data	0.020755

Figure 3: An illustration of 3 discovered topics from a 200-topic solution for the corpus. Each topic is shown with the top 10 words and top five papers that have highest probability conditioned on that topic. The titles are our interpretation of the topics.

Fig. 3 first topic 91 “Machine learning” related top five papers are very informative for learning major classification approaches and are very useful for a new researcher or a group looking for new direction of research. It is also true for top five papers discovered for Topic 88 “Data Mining” and topic 16 “XML databases”, as they can also provide a comprehensive overview of specific research area. Proposed approach discovers several other more specific topics related to data mining such as neural networks, multi-agent systems and pattern matching, also other topics that span a full range of areas encompassed in the corpus. A fraction of non-research topics, perhaps 10-15%, are also discovered that are not directly related to a specific area of research, as the words present in those topics were usually used as a glue between scientific terms.

Citation Recommendation: One can predict the citations for papers which are contained in the testing data set by learning the model on the training data set. For prediction purpose we apply eq. 1 only on the word tokens in the new papers each time temporarily updating the count matrices of (word by topic) and (topic by citation). The resulting assignments of words to topics can be saved after a few iterations (20 in our simulations). Then we used eq. 4 to calculate count matrix of (topic by papers) to predict citations for new papers. Tab. 1 shows this type of inference. To show predictive power of SR approach we treated 667 papers as test papers, by using trained model on the remaining 2668 papers to discover latent topics. Discovered topics are then used to predict the citations for the papers contained in the testing data set.

Table 1: An illustration of top 10 predicted citations for three papers.

Paper 1: Compatible Genericity with Run-time Types for the Java Programming Language
Recommended Citations
Type Inclusion Constraints and Type Inference
How to Make Ad-Hoc Polymorphism Less Ad Hoc
An Extended Calculus of Constructions
Implementing Regular Tree Expressions
Compiling Polymorphism Using Intensional Type Analysis
A Theory of Qualified Types
Revised^4 Report on the Algorithmic Language Scheme
A Syntactic Approach to Type Soundness
Typed Concurrent Objects
Dynamic Typing in a Statically Typed Language
Paper 2: Message Logging: Pessimistic, Optimistic, and Causal
Recommended Citations
Preserving and Using Context Information in Interprocess Communication
Uniform Reliable Multicast in a Virtually Synchronous Environment
Sender-Based Message Logging
Atomic Broadcast: From Simple Message Diffusion to Byzantine Agreement
Fast Message Ordering and Membership Using a Logical Token-Passing Ring
The Weakest Failure Detector for Solving Consensus
Membership Algorithms for Multicast Communication Groups
The Transis Approach to High Availability Cluster Communication
Reaching Agreement on Processor Group Membership in Synchronous Distributed Systems
Totem: A Fault-Tolerant Multicast Group Communication System (s)
Paper 3: A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols
Recommended Citations
Implementing Regular Tree Expressions
Inferring Web Communities from Link Topology
The Complexity of Set Constraints
Symbolic Model Checking: 10 20 States and Beyond
Randomness in Interactive Proofs
Balanced Allocations
TCP Extensions for High Performance
Discovering Generalized Episodes Using Minimal Occurrences
Atomic Decomposition by Basis Pursuit
Extending Planning Graphs to an ADL Subset

Predicted citations associated with each paper are quite intuitive, as they cover highly specific area of papers on the basis of semantic information and provide true literature review for the papers to cite. For example paper 1 in tab. 1, "Compatible Generosity with Run-time Types for the Java Programming Language" related citations are quite precise and are in fact strong candidate to be cited in this paper. Citations predicted for paper 2 "Message Logging: Pessimistic, Optimistic, and Causal" are also intuitive and precise, as they are very much semantically related to the paper contents.

In addition, by doing analysis we have found that for each of the three papers, citations include at least five heavily cited citations of the specific area of the paper. For example citation 1 of paper 1, "Type Inclusion Constraints and Type Inference" was published in 1893 and in total cited by 163 papers and citation 2 of paper 1, "How to Make Ad-Hoc Polymorphism Less Ad Hoc" was published in 1998 and has been cited by 178 papers so far according to citeseer [8].

Citations predicted for each paper also include at least one top class paper which can be viewed as key citation for that paper. For example citation 10 of paper 1, "Dynamic Typing in a Statically Typed Language" was published in 1989 and it has in total 311 papers cited it that shows its importance in dynamic programming field. Some discovered citations and not heavily cited in the past but they are selected due to similarity of words on the basis of semantic information presented in the text.

Most of the citations inferred for paper 1 and 2, can also be judged appropriate by matching citation titles with paper titles. However, citations inferred for paper 3 can't be judged as appropriate citations for that paper, while predicted citations are semantically related to the contents of that paper. It also shows that only keyword based search by using search engines or only link based search [5,12] by utilizing input and output links between papers cannot provide accurate citations in situations, where no proper keywords are appearing in the title of papers and links information is not rich enough.

VI. CONCLUSIONS

This study deals with the problem of scientific paper recommendation. Initially we defined this problem and then introduced a SR approach that can automatically recommend semantically related citations, as its generative process links citations to latent topics. We demonstrated how this approach can be used to recommend citations for newly written papers. Some potential applications of acquired results are; citation recommendation for papers and suggestion of papers about different research areas to new researchers and research groups. Even though SR approach is quite simple, nonetheless it provides functional information by providing semantic-based citations. Possible future direction of this work can be use of links on the basis of citations by also considering time periods, in addition to already used semantic-based information, as one can say for different time periods different citations can be recommended.

REFERENCES

[1] C. Andrieu, N. D. Freitas, A. Doucet, and M. Jordan. An introduction to MCMC for machine learning. Journal of Machine Learning, 2003, vol. 50, pp. 5–43.

[2] M. Balabanovic and Y. Shoham. Content-Based Collaborative Recommendation. Commun. ACM, 1997, vol. 40(3), 1997.

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

[4] J. Breese, D. Heckerman, and C. Kadie. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In Proceedings of the 14th conference on Uncertainty in Artificial Intelligence (UAI), Madison, Wisconsin, USA, July 24-26, 1998, pp. 43-52.

[5] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems, 1998, vol. 30(1-7), p.107–117.

[6] M. Deshpande and G. Karypis. Item-based Top-n Recommendation Algorithms. ACM Transactions on Information Systems, 2004, vol. 22(1), pp.143-177.

[7] E. Garfield. Citation analysis as a tool in journal evaluation. Reprinted Science, 1972, vol. 178, pp.471-479.

[8] C. Giles, K. Bollacker, and S. Lawrence. Citeseer: An automatic citation indexing system. In Proceedings of 3rd ACM Conference on Digital Libraries, Pittsburgh, PA, USA, June 23-26, 1998, pp. 89-98.

[9] T. L. Griffiths and M. Steyvers. Finding scientific topics. In Proceedings of the National Academy of Sciences, 2004, pp. 5228-5235.

[10] T. Hofmann. Probabilistic latent semantic analysis. In Proceedings of the 15th Annual Conference on Uncertainty in Artificial Intelligence (UAI), Stockholm, Sweden, July 30-August 1, 1999.

[11] M. M. Kessler. Bibliographic coupling between scientific papers. American Documentation, 1963, vol. 14, pp:10–25.

[12] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. Journal of the ACM, 1999, vol. 46(5), pp.604–632.

[13] A. McCallum, K. Nigam, and L. H. Ungar. Efficient clustering of high-dimensional data sets with application to reference matching. In Proceedings of the 6th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Boston, MA, USA, August 20-23, 2000, pp. 169-178.

[14] S. M. McNee, I. Albert, D. Cosley, P. Gopalkrishnan, S. K. Lam, A. M. Rashid, J. A. Konstan, and J. Riedl. On the recommending of citations for research papers. In Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW), New Orleans, Louisiana, USA, November 16-20, 2002.

[15] A. Popescul, G. W. Flake, S. Lawrence, L.H. Ungar, and C.L. Giles. Clustering and identifying temporal trends in document databases. IEEE Advances in Digital Libraries (ADL) 2000, pp. 173-182.

[16] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth. The author-topic model for authors and documents. In Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI), Banff, Canada, July 7-11, 2004.

[17] T. Strohman, W. B. Croft, and D. Jensen. Recommending citations for academic papers. (Poster) In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, Netherlands, July 23-27, 2007, pp. 705-706.

[18] J. Zhang, J. Tang, B. Liang, Z. Yang, S. Wang, J. Zuo, and J. Li. Recommendation over a Heterogeneous Social Network. In Proceedings of the 9th International Conference on Web-Age Information Management (WAIM), ZhangJiaJie, China, July 20-22, 2008.