Citation Recommendation via Time-series Scholarly Topic Analysis and Publication Prior Analysis

Zhuoren Jiang
College of Transportation Management
Dalian Maritime University
Dalian, China, 116026
jzr_dlmu@hotmail.com

ABSTRACT

Objective: To improve the performance of citation recommendation, by using innovative citation recommendation methods.

Methodology: By utilizing changes in the topical content of publications over time, citation graph, plus proximity-based citation contexts, I calculate the publication topical importance (prior) over time, and I use the topic prior, publication prior from author relation and Wikipedia along with the language model prioritize a number of scientific publication given their published time and user information need, i.e., a piece of working context or a publication abstract.

Preliminary Experimentation and Result: In the preliminary experience, I employed the classical language model (without priors) and the language model with PageRank priors as the baseline. Preliminary experiment shows publication topic prior is important for citation recommendation given a textual information need.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

General Terms

Algorithms, Measurement, Experimentation

Keywords

Bibliometrics, Citation Recommendation, Supervised Dynamic Topic Modeling, PageRank, Prior Knowledge

1. INTRODUCTION AND MOTIVATION

There are millions of research papers been published every year, the volume of publications has increased dramatically, for example, based on DBLP statistics (as figure 1 shows), computer scientists published 3 times more papers in 2011 than in 2001. Online publications and digital library enable researchers access to many more sources in a significantly shorter timeframe. However, it also makes literature search an arduous task [8, 11]. For instance, it's hard for junior researcher to access publications should be cited.

Typically, researchers rely on some manual retrieval/recommendation mechanisms to discover new studies. These techniques are time-consuming and only allow reaching a limited set of documents in a reasonable time. Current scientific search tools, e.g., Google Scholar and Microsoft Academic, are limited to standardized types of queries to address users' information needs. Meanwhile, measuring the influence of a scientific article is an important and challenging problem; especially when we take the time influence as an impact factor.

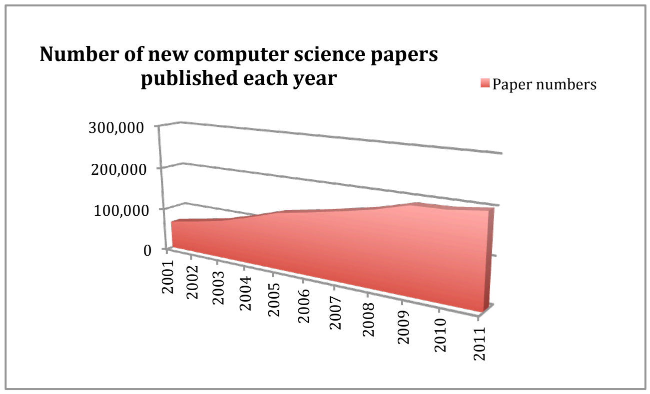

Figure 1. Number of new computer science papers published each year based on DBLP

Hence, researchers need information retrieval (IR), information extraction (IE), and recommendation tools that can recommend a small number of high quality candidates, quickly filter through and locate relevant publications or scientific resources.

Finally, some recent exciting developments have illustrated the possibility of using citation relationship to recommend high quality research publications to users. Citation data is held to be a valid and reliable indicator of scholarly impact because it represents an explicit and objective acknowledgement of influence by expert authors. The traditional method of assessing an article's influence is to count the citations to it. The impact factor of a journal, for example, is based on aggregate citation counts.

However, in most previous works [12], while various methods were used to characterize the citation relationship, the basic assumption was easy and straightforward, and this assumption is oversimplified and may limit the recommendation performance or accuracy.

In the other hand, citation metadata, as important information to tell the relationship between publications, is not always reliable in the repository. The collections, like OCR scans of historical scientific literature, do contain citations, but they are difficult to read in reliable electronic form. Take ACM digital library as an example, 18.5% of publication do not have any citing information.

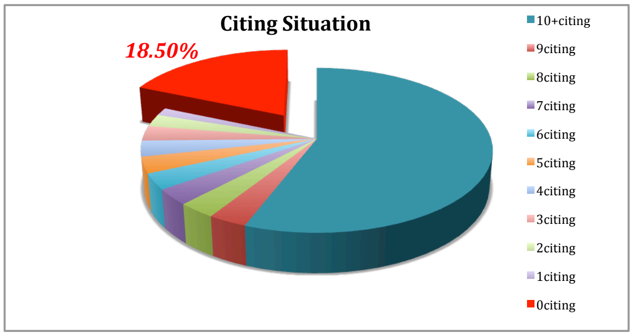

Figure 2. Citing number in ACM digital library

In this study, I propose an innovative citation recommendation method based on a (user) textual working context, *T*, i.e., a publication abstract. By integrating bibliometric network analysis and supervised dynamic topic modeling techniques, for each paper, I will calculate both a time-series publication topical importance matrix based on citation context proximity in the citing paper and a publication topic distribution. In view of the citation relations may be not available, I will use other resource (author relationship and Wikipedia) to improve the publication prior knowledge.

Then I will use the time-series publication topical importance matrix as the topical prior probability together with publication prior probability to enhance the classical language model for citation recommendation. In more detail, I assume that citation context in the citing paper along with citing and cited publication content can provide high-quality topical citation motivation information, and I will employ this information to infer the knowledge transitioning probability between citing and cited papers on a scholarly network. I will use the PageRank with prior algorithm [17] to characterize publication topic importance, extract publication prior knowledge from other resource like author relations and Wikipedia. The time-series publication prior probabilities $P_{Z_t}(pub_i)_t$ with publication prior $P(pub_i)$ could enhance the language model, $P(q|pub_i)_t$, for citation recommendation.

The contribution this study may achieve: 1) innovate method for constructing a bibliometric citation graph using time-series topical importance; 2) create a topic change trend line over time, which enables users to explore the indirectly related but important topics, thus to obtain higher quality recommendation; 3) use other resource to contribute the publication's prior, i.e., Wikipedia or author relations, if the citation metadata is not available or gets a low quality.

2. RELATED WORK

2.1 Citation recommendation

Scientific recommendation is an important new research area, with the rapid development of digital publications. There are some previous efforts on recommending a bibliography list for a manuscript, or recommending papers to reviewers. There are two important approaches:

1) One involves item–item or item–user-based content-sensitive collaborative filtering algorithms. For example, [4] presents a method of recommending scientific papers of potential interest to users by using the ACM Computing Classification System along with hierarchical concept information from both author profiles and paper content. For domain expert recommendation, recommendation methods can also be used, for instance, bag-of-words [1], LSI concept [5].

2) Another important approach is using scholarly or bibliographic networks—i.e., networks based on citation or co-authorship to recommend scientific resources. For instance, Lao and Cohen [10] used both supervised and unsupervised methods with the Random Walk with Restart (RWR) algorithm for citation, author, and venue recommendation.

This study differs from previous research in that I use similarity-based (i.e., the language model) and network-based (i.e., prior probability) methods for citation analysis. Moreover, in the citation network, I use dynamic topic models to characterize each vertex and edge. The citation topical motivation probability distribution is extracted by using the proximity-based citation context, where each topic is a keyword-labeled unigram word probability distribution.

2.2 Full-text citation analysis

Many scholars have focused their research on citation frequency and citation impact as applied in different domains, in classic bibliometrics papers. Harhoff et al. [7] judged the value of patented inventions by citation frequency and concluded, "The higher an invention's economic value estimate was, the more the patent was subsequently cited" (p. 511).

With further study of citation analysis, increasing numbers of researchers have come to doubt the reasonableness of assuming that the raw number of citations reflects an article's influence [13]. While traditional citation analysis treats all citations equally, in reality, not all citations are equal.

Not until recently, Bernstam et al., [2] used full-text citation context to enhance scientific retrieval performance. They found that the closeness of a word in the citation context provides stronger semantic information about the cited paper, and thus that this closeness could be important to retrieval performance. Their findings motivate me to use the proximity for citation topic inference.

2.3 PageRank

PageRank has become a significant method for evaluating the most important nodes in complex graph analysis. From the point of citation analysis in bibliometrics, PageRank is also an efficient way to evaluate a paper's ranking score in a specific domain. The PageRank algorithm, first proposed by Page et al., [14] and used in Google Search.

White and Smyth [17] first proposed the priors idea in their formalization of a relative-rank extension to PageRank. They demonstrated how the approach could be used to study relative importance in real-world networks. Rodriguez and Bollen [16] used PageRank with priors to compute the publication topic importance score with the node prior and edge transitioning probability vectors.

By computing the topic-sensitive PageRank scores using the topic of the context within which the query appeared, and then generating context-specific importance scores for pages using linear combinations of biased PageRank vectors, this study can generate more accurate rankings compared with a single, generic PageRank vector.

2.4 Topic Modeling

Managing the explosion of electronic document archives requires new tools for automatically organizing, searching, indexing, and browsing large collections. Recent research in machine learning and statistics has developed new techniques for exploring the electronic document archives to finding patterns of words in document collections using hierarchical probabilistic models. These models are called "topic models" because the discovered patterns often reflect the underlying topics, which combined to form the documents.

Blei et al. [3] proposed Latent Dirichlet Allocation (LDA) as a promising unsupervised topic modeling algorithm. LDA employs a generative probabilistic model in the hierarchical Bayesian framework, and extends PLSI [9] by introducing a Dirichlet prior on θ. Labeled LDA (LLDA) [15], which improves topic modeling performance by using scientific metadata, i.e., keywords. [19] developed a dynamic topic model which captures the evolution of topics in a sequentially organized corpus of documents. [20] based their model on dynamic topic models, allowing for multiple threads of influence within a corpus.

What I use in this study is a supervised dynamic topic modeling algorithm. More importantly, as the total number of keywords is stable, I don't need to tune topic number for recommendation performance optimization.

2.5 Publication prior from the contribution of other resource

Low quality scientific metadata [6] threatens the citation based ranking algorithms, i.e., PageRank or HITs. Kittur et.al shows that over 25% of pre-2008 articles in Wikipedia are related to natural or social sciences [21].

Studying the influence of social web on the scholarly community opens a new direction. Evans and Krautham-mer [22] correlate journal citations to Wikipedia and academia.

I would not only use the publication topical importance prior, which was based on citation relation, as the unique resource to compute the publication prior, in addition, the Wikipedia's contribution and other relations such as author relation will be used to fix the publication prior.

3. PROBLEMS AND SOLUTIONS

3.1 How to apply citation recommendation with textual working context

Given a piece of working textual context (from user), *T*, and a candidate citation (cited publication), pub_i, it's hard to estimate the probability that pub_i is relevant and important to the given context *T*, $P(pub_i|T)$ directly. In order to solve this problem, I apply Bayes' rule:

$$P(pub_i|T) = \frac{P(T|pub_i) \cdot P(pub_i)}{P(T)}$$

Classical content-based content recommendation algorithms, simply ignore the publication prior, $P(pub_i)$, which means, without respect to *T*, that all publications have an equal chance to be recommended. As *P(T)* is publication independent, it can be ignored in this ranking function. $P(T|pub_i)$ can be estimated by using the language model, where *T* has a list of words $\{t_1, t_2 \ldots t_m\}$:

$$P(T|pub_i) = \prod_{j=1}^{m} P(t_j|pub_i)$$

From a language model perspective, $P(t_j|pub_i)$ can be calculated by different smoothing techniques [18].

In this research, I would focus on estimating the publication prior, $P(pub_i)$, to enhance citation recommendation performance.

3.2 How to use citation relation as prior

Many researchers constructed a citation network to calculate publication importance by utilizing the PageRank [14] algorithm. Each vertex on the graph is a publication, and each edge is a citation from the citing publication to the cited publication. All the publications and citations were treated equally on the graph. As already mentioned, however, this hypothesis is oversimplified [12], as some citations and publications are more important than others for some scientific topics in a citing paper.

I would use the PageRank-based publication importance as the prior, $P(pub_i)$, along with the language model to calculate the recommendation ranking for a piece of text. Similarly idea has been studied by [10].

3.3 How to improve the statistical citation relations

Statistical citation relations are important but not necessarily accurate means of telling the importance of a publication, in that they ignore the semantic information of the citing/cited publications and the citation motivation itself.

So I plan to use publication topical prior with full-text data to improve the statistical citation relations, I will do citation topic inference with proximity-based citation context; that is, characterize citation relations in terms of two kinds of knowledge: publication (citing or cited paper) topic probability distribution, and citation topic probability distribution. There are three major contributions of this approach. 1) Even with the same citation network topology, different publications can make different contributions to different scientific topics; 2) unlike classical, unsupervised topic modeling algorithms, the topics in this research are associated with scientific keywords (supervised learning), which can help to interpret and evaluate the results; 3) because I utilize full-text citation analysis, one paper can have more than one citation edge with another paper, and the citation topic distributions could be different, resulting in more accurate PageRank random walk modeling.

3.4 How to improve the existing topic modeling algorithm to achieve a better publication topical prior

LDA is the most common topic model method. However, there are three import limitations of LDA: 1) the challenge of interpreting and evaluating topic statistics. For example, it is difficult to assign a label to (provide a semantics for) each topic statistic automatically; 2) arbitrary numbers of topic may not be appropriate for bibliometric studies because, while some topics may be very sparse, others may only focus on quite detailed knowledge of the same scientific topic; 3) the scholarly topics are changing over time, they should not be treated independently, the topics may influence the topic in next time slice.

These limitations motivated me to utilize a supervised or semi-supervised topic modeling algorithm, which employs existing topics from scientific metadata, along with a time-series model of sequential document collections, which can interpret the topic development.

As mentioned before, labeled LDA (LLDA) algorithm [15], counld give both topic labels and topic numbers (the total number of keywords in the metadata repository). And dynamic topic model (DTM)[19] is a time-series model of sequential document collections. Unlike typical language modeling applications, DTM chains the natural parameters of each topic $\beta_{t,k}$ in a state space model that evolves with Gaussian noise:

$$\beta_{t,k}|\beta_{t-1,k} \sim \mathcal{N}(\beta_{t-1,k}, \sigma^2 I)$$

Similarly, the document-specific topic proportions θ in DTM use a logistic normal with mean α to express uncertainty over proportions:

$$\alpha_t|\alpha_{t-1} \sim \mathcal{N}(\alpha_{t-1,k}, \sigma^2 I)$$

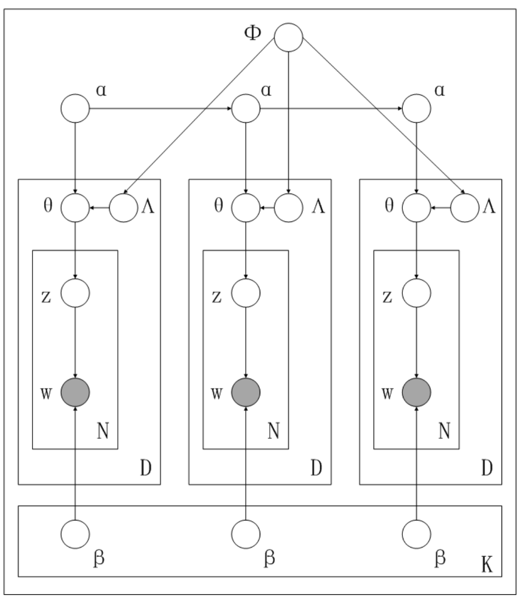

Figure 3. Supervised Dynamic Topic Modeling Algorithm

In this study, I draw lessons from LLDA and DTM, use a supervised dynamic topic modeling algorithm that assumes the availability of topic labels (keywords) and the characterization of each topic in each time slice t, by a multinomial distribution $\beta_{t,key_i,}$ over all vocabulary words.

By chaining together topics and topic proportion distributions, I have sequentially tied a collection of topic models. Figure 3 visualizes the supervised dynamic topic model generative process.

3.5 How to create a better citation networks

Classical citation networks tend to ignore citation and publication content. In this study, I plan to create a large citation-directed network, $G = (V, E)$, with two kinds of prior knowledge: time-series publication topic priors and a time-series citation topic transitioning probability distribution.

Each vertex, $v \in V$, on the citation graph represents a publication, with the dynamic publication topic prior probability matrix $\{p_{z_{t,key_1}}(v), p_{z_{t,key_2}}(v), \dots p_{z_{t,key_n}}(v)\}$, where $p_{z_{t,key_t}}(v)$ is the prior probability of vertex v for topic z_{t,key_t} in time slice t and $\sum_{i=1}^{|V|} p_{z_{t,key_i}}(v) = 1$.

Each edge, $e \in E$, on the graph represents a citation connecting v_i and v_j (v_i cites v_j). The topic transitioning vector for each edge is $\{ p_{z_{t,key_1}}(v_i|v_j), p_{z_{t,key_2}}(v_i|v_j), \dots p_{z_{kt,eyn}}(v_i|v_j) \}$, where $p_{z_{t,key_t}}(v_i|v_j)$ is the probability of transitioning from vertex v_i to v_j for topic z_{key_t} in time slice t.

For a given publication u, I would use $S_{t,in}(u)$ and $S_{t,out}(u)$ to represent a set of incoming and outgoing edges (citations) to node u in time slice t, with "in" degree $d_{t,in}(u) = |S_{t,in}(u)|$ and "out" degree $d_{t,out}(u) = |S_{t,out}(u)|$. Thus, $\sum_{i=1}^{d_{t,out}(v_j)} p_{z_{t,key_t}}(v_i|v_j) = 1$.

Based on these definitions, I can calculate each vertex's (i.e., each publication's) prior probability:

$$p_{z_{t,key_t}}(v) = \frac{P(z_{t,key_t}|paper_v)}{\sum_{x=1}^{|V|} P(z_{t,key_t}|paper_x)}$$

I can also calculate each edge's (i.e., each citation's) transitioning probability:

$$p_{z_{t,key_t}}(v_i|v_j) = \frac{P(z_{t,key_t}|citation_{j,i})}{\sum_{x=1}^{d_{out}(v_j)} P(z_{t,key_t}|citation_{j,x})}$$

where $P(z_{t,key_t}|paper_v)$ is the publication topic inference score in time slice t and $P(z_{t,key_t}|citation_{j,i})$ is the citation topic inference score in time slice t.

3.6 How to use other resource to improve the publication prior

Though citation relation can be powerful, they may be hard to use in many situations as mentioned before.

So it's important to use other resource (i.e. author relationship or Wikipedia) to calculate the publication prior when the citation relation is not available.

I make an assumption that: the paper importance prior can be extracted from other resource. As a result, I can use other resource to contribute the publication prior. For instance, I plan to create an author citation network and calculate the author's importance score to fix the publications' citation score. In addition, I plan to search the Wikipedia pages, which mention candidate papers' titles or authors, then count the mention times or calculated the weighted score from different Wikipedia pages. Finally, I can get the improved publication prior.

$$P(pubi) = \frac{p_{z_{t,key_t}}(pub_k) + \alpha\left(p_{reource_i}(pub_k)\right)}{1 + \alpha}$$

4. EXPERIMENTATION PLAN AND PRELIMINARY RESULT

The whole workflow of experimentation is illustrated in Figure 4.

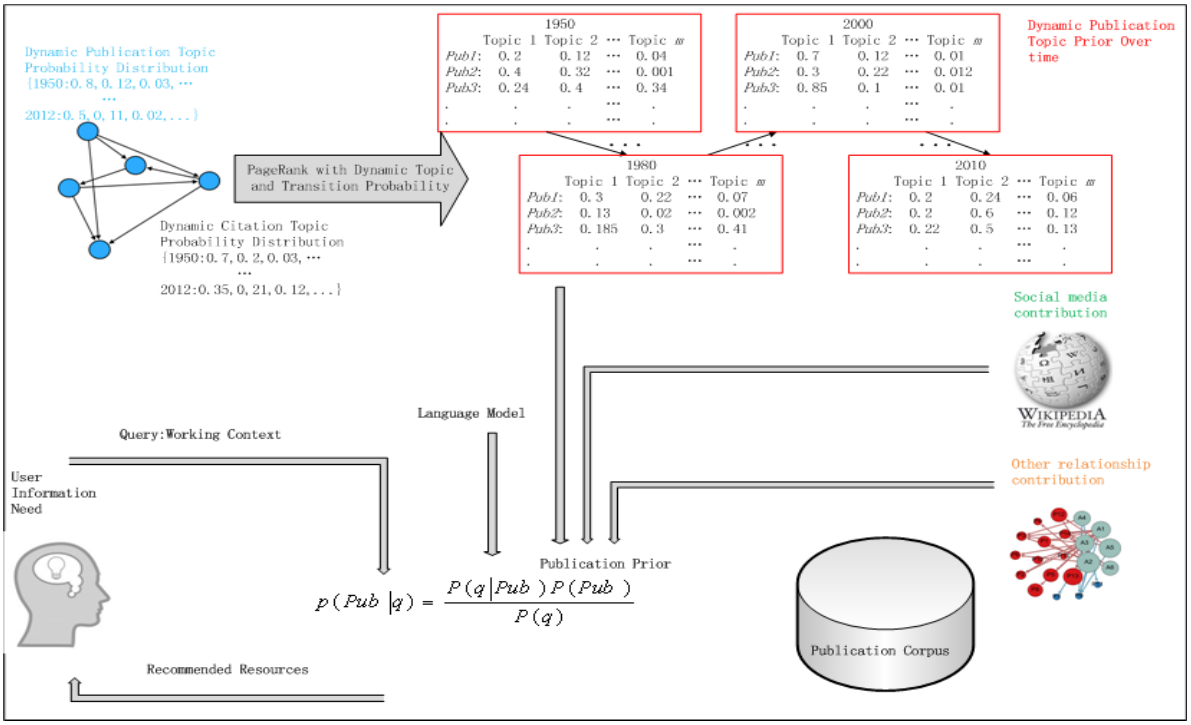

Figure 4. Citation recommendation experimentation workflow sketch

In preliminary experimentation, I just trained supervised static topic model, I used 41,370 publications (as candidate citation collection) from 111 journals and 1,442 conference proceedings or workshops on computer science for the experiment (mainly from the ACM digital library), where full text and citations were extracted from the PDF files. The selected papers were published between 1951 and 2011. From these I extracted 28,013 publication texts (accounting for 67.7% of all the sampled publications), including titles, abstracts, and full text. For the other publications, I used the title, the abstract, and keyword information from a metadata repository to represent the content of the paper. I sampled 10,000 publications (with full text) to train the supervised topic model. Author-provided keywords were used as topic labels. I use nDCG to evaluate the degree of citation importance.

The result shows below:

Table 1: Different publication and citation inference methods (nDCG)

	LM	LM+ PageRank	LM+PageRank-With-Topic-Prior·
nDCG@10	0.1183	0.1740	**0.2032**
nDCG@30	0.1424	0.1951	**0.2317**
nDCG@50	0.1586	0.2091	**0.2460**
nDCG@100	0.1774	0.2297	**0.2648**
nDCG@300	0.2022	0.2539	**0.2907**
nDCG@500	0.2100	0.2615	**0.3007**
nDCG@1000	0.2202	0.2711	**0.3108**
nDCG@3000	0.2326	0.2847	**0.3241**
nDCG@5000	0.2375	0.2890	**0.3290**
nDCG@all	0.2647	0.3157	**0.3470**

· In preliminary experiment, I just used the supervised static topic model for publications & citation topic inference

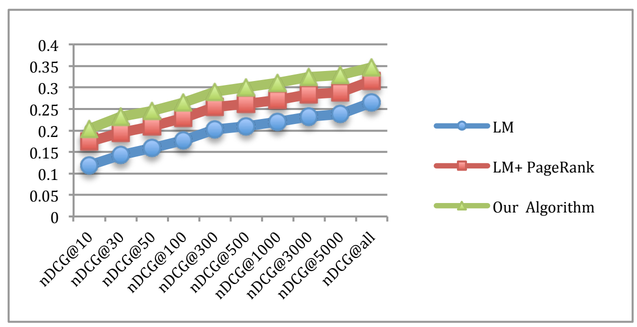

Figure 5: Different publication and citation inference methods (nDCG)

It is clear that, compared with "LM", the language model without priors; and "LM + PageRank", the language model with PageRank priors, "LM+PageRank-With-Topic-Prior", the language model combined with PageRank with topic priors has a better performance in the preliminary experiment. Preliminary experiment shows publication topic prior is important for citation recommendation given a textual information need.

REFERENCES

[1] Basu, C., Hirsh, H., Cohen, W.W., and Nevill-Manning, C. 2001. Technical paper recommendation: a study in combining multiple information sources. *J. Artif. Intell. Res. 14*, 1 (Jan. 2001), 231-252.

[2] Bernstam, E.V., Herskovic, J.R., Aphinyanaphongs, Y., Aliferis, C.F., Sriram, M.G., and Hersh, W.R. 2006. Using citation data to improve retrieval from MEDLINE. *J. Am. Med. Inform. Assn. 13*, 1 (Jan. 2006), 96-105. DOI= http://dx.doi.org/10.1197/jamia.M1909.

[3] Blei, D.M., Ng, A.Y., and Jordan, M.I. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res. 3* (Mar. 2003), 993-1022.

[4] Chandrasekaran, K., Gauch, S., Lakkaraju, P., and Luong, H.P. 2008. Concept-based document recommendations for CiteSeer authors. In *Proceedings of the 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems* (Hannover, Germany, July 29 - August 1, 2008), Springer-Verlag, Berlin, 83-92. DOI= http://dx.doi.org/10.1007/978-3-540-70987-9_11.

[5] Dumais, S.T. and Nielsen, J. 1992. Automating the assignment of submitted manuscripts to reviewers. In *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Copenhagen, Denmark, June 21 - 24, 1992), ACM, New York, NY, USA, 233-244. DOI= http://dx.doi.org/10.1145/133160.133205.

[6] Guo, C., Zhang, J., and Liu, X. 2013. Scientific metadata quality enhancement for scholarly publications. In *Proceedings of the 2013 iConference* (Fort Worth, Texas, February 12-15, 2013).

[7] Harhoff, D., Narin, F., Scherer, F.M., and Vopel, K. 1999. Citation frequency and the value of patented inventions. *Rev. Econ. Stat. 81*, 3 (Aug. 1999), 511-515. DOI= http://dx.doi.org/10.2307/2646773.

[8] He, Q., Kifer, D., Pei, J., Mitra, P., and Giles, C.L. 2011. Citation recommendation without author supervision. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining* (Hong Kong, China, February 09 - 12, 2011), ACM, New York, NY, USA, 755-764. DOI= http://dx.doi.org/10.1145/1935826.1935926.

[9] Hofmann, T. 1999. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (Berkeley, California, USA, August 15 - 19, 1999), ACM, New York, NY, USA, 50-57. DOI= http://dx.doi.org/10.1145/312624.312649.

[10] Lao, N. and Cohen, W.W. 2010. Relational retrieval using a combination of path-constrained random walks. *Mach. Learn. 81*, 1 (Oct. 2010), 53-67. DOI= http://dx.doi.org/10.1007/s10994-010-5205-8.

[11] Liu, X. 2012. Generating Metadata for Cyberlearning Resource through Information Retrieval and Meta-search. *J. Am. Soc. Inf. Sci. Tec.*2012).

[12] Liu, X., Zhang, J., and Guo, C. 2012. Full-text citation analysis: enhancing bibliometric and scientific publication ranking. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management* (Maui, Hawaii, USA, October 29 - November 02, 2012), ACM, 1975-1979. DOI= http://dx.doi.org/10.1145/2396761.2398555.

[13] Macroberts, M.H. and Macroberts, B.R. 1989. Problems of citation analysis: a critical review. *J. Am. Soc. Inform. Sci. 40*, 5 (Sep. 1989), 342-349. DOI= http://dx.doi.org/10.1002/(sici)1097-4571(198909)40:5%3C342::aid-asi7%3E3.0.co;2-u.

[14] Page, L., Brin, S., Motwani, R., and Winograd, T. 1999. *The PageRank Citation Ranking: Bringing Order to the Web.* Technical Report. Stanford InfoLab.

[15] Ramage, D., Hall, D., Nallapati, R., and Manning, C.D. 2009. Labeled LDA: a supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1* (Singapore, August 6-7, 2009), Association for Computational Linguistics, Stroudsburg, PA, USA, 248-256.

[16] Rodriguez, M.A. and Bollen, J. 2006. Simulating network influence algorithms using Particle-Swarms: PageRank and PageRank-Priors.

[17] White, S. and Smyth, P. 2003. Algorithms for estimating relative importance in networks. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Washington, D.C., August 24 - 27, 2003), ACM, New York, NY, USA, 266-275. DOI= http://dx.doi.org/10.1145/956750.956782.

[18] Zhai, C. and Lafferty, J. 2001. A study of smoothing methods for language models applied to Ad Hoc information retrieval. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (New Orleans, Louisiana, USA, September 09 - 12, 2001), ACM, New York, NY, USA, 334-342. DOI= http://dx.doi.org/10.1145/383952.384019.

[19] David M. Blei, John D. Lafferty. 2006. Dynamic topic models. *In Proceedings of ICML '06 Proceedings of the 23rd international conference on Machine learning* (New York, NY, USA, June 25-29, 2006), ACM, 113-120. DOI= http://dx.doi.org/10.1145/1143844.1143859.

[20] Sean M. Gerrish, David M. Blei. 2010. A Language-based Approach to Measuring Scholarly Impact. *In Proceedings of the 27th International Conference on Machine Learning* (Haifa, Israel, June 21-24, 2010), ACM, 375-382. DOI= http://dx.doi.org/10.1.1.167.8277

[21] A. Kittur, E. H. Chi, and B. Suh. What's in wikipedia?: mapping topics and conflict using socially annotated category structure. *In Proceedings of CHI '09 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2009), ACM, 1509–1512. DOI= http://dx.doi.org/10.1145/1518701.1518930

[22] P. Evans and M. Krauthammer. Exploring the use of social media to measure journal article impact. *AMIA ... Annual Symposium proceedings / AMIA Symposium*, Vol. 2011 (2011), 374-381.