ATT: Analyzing Temporal Dynamics of Topics and Authors in Social Media

Nasir Naveed Sergej Sizov Steffen Staab
WeST – Institute for Web Science and Technologies
University of Koblenz-Landau
56070 Koblenz, Germany
{naveed,sizov,staab}@uni-koblenz.de

ABSTRACT

Understanding Topical trends and user roles in topic evolution is an important challenge in the field of information retrieval. In this contribution, we present a novel model for analyzing evolution of user's interests with respect to produced content over time. Our approach Author-Topic-Time model (ATT) addresses this problem by means of Bayesian modeling of relations between authors, latent topics and temporal information. We extend state of the art Latent Dirichlet Allocation (LDA) topic model to incorporate the author and timestamp information for capturing changes in user interest over time with respect to evolving latent topics. We present results of application of the model to the 9 years of scientific publication datasets from CiteSeer showing improved semantically cohesive topic detection and capturing shift in authors interest in relation to topic evolution. We also discuss opportunities of model use in novel mining and recommendation scenarios.

Categories and Subject Descriptors

H.3.1 [**Information Retrieval**]: Content Analysis and Machine Learning—*Abstracting Methods*; H.3.1 [**Models and Principles**]: System and Information Theory—*Value of Information*; I.2.7 [**Data Mining**]: Natural Language Processing—*Text analysis*

General Terms

Algorithms, Experimentation

Keywords

Probabilistic Models, Topic Modeling, Text Categorization

1. INTRODUCTION

The world wide web provides a platform for content sharing activities where people can share views, participate in discussions, publish technical domain specific blogs and research papers, thereby, contribute tremendous online contents related to different domains and subject areas. For better understanding of these text contents, they are often categorized with respect to the subject they discuss. These subjects areas are called as latent topics. Topics discussed in social media vary with respect to their longevity. Some last for a very brief period and some continue to develop over a period of time and thereby enjoy sustained interest from contributors. It has been observed that topics discussed in collaborative social networks exhibit spikes (sudden topics, linked to current events, or enjoying a limited-time interest) and chatters (more recurring, long term topics) indicating strong correlation [7, 8].

WebSci '11, June 14-17, 2011, Koblenz, Germany.

Consider a scenario where a user tries to track back a particular topic for its emergence, growth patterns, popularity and underlying key players contributing to it over a period of time. In such a scenario manual analysis of this tremendous amount of text for finding latent topics, capturing topic evolution, identifying the author's interests and depicting changes in interests is expensive in terms of time and labor. Following this scenario the challenge is to provide a model which is able to capture temporal topic dynamics and provides an insight into changing user interests with respect to evolving topics. One such model can be helpful in finding influential authors at different stages of topic evolution and thus can be helpful in characterizing authors as pioneers, mainstream or laggards in different subject areas.

We tackle the above mentioned problem by Bayesian means and propose Author-Topic-Time (ATT) Model which is a continuation of our previous work [11]. ATT model extends the Latent Dirichlet Allocation model and augment the document contents with author and time at which the contents are generated. We see author and timestamp as metadata attached to contents of documents. Augmenting documents with metadata in ATT document generation process helps it captures the topic dynamics in a given collaborative environment and finds the key authors that are contributing to the specific topics at different stages of topic life cycle. Thereby, ATT provides a deeper understanding of topical shifts and nature of user collaborations in social environment. We evaluate the efficacy of model in predicting authors and capturing lifespan of the topics by running the model on subset of abstracts of scientific publications from CiteSeer dataset and compare the results with the standard LDA. The results obtained can be exploited for social retrieval tasks and recommender systems in recommending specific authors or publications to read for a given user interests.

The rest of the paper is organized as follows. Section 2 describes the related work for topic modeling using probabilistic means and three related models. Section 3 gives the formal definition, generative process and parameter settings of the ATT model and in section 4 we present the results of the application of the model to CiteSeer dataset and compare the results with standard LDA.

2. RELATED WORK

2.1 Topic Modeling

In probabilistic topic modeling a *"Topic"* is seen as a multinomial distribution over a vocabulary that assigns high probability to a set of words that tend to appear in the similar documents. A qualitatively "better topic" is that in which words that have high probability are semantically related to each other and a human subject is able to say that "these words are about X", where X can be any domain like business, computer science, chemistry etc.
There is no consensus in literature on what could be a formal definition of a topic model. So, we see a *"Topic Model"* as a model of the generative process by which documents are created and captures the word co-occurrence patterns in a document corpus to produce semantically coherent topics.

2.2 Probabilistic Topic Models

A variety of statistical models have been proposed for topic-based analysis and modeling of text documents. To name few of them are unigram model, mixture of unigram model [12], latent semantic analysis(pLSA) [10] and Latent Dirichlet Allocation (LDA) [2].

LDA is a Bayesian multinomial mixture model which has become a state of the art and popular method in text analysis due to its ability to produce interpretable and semantically coherent topics. It uses the Dirichlet distribution to model the distribution of the topics for each document. In LDA each word is considered sampled from a multinomial distribution over words specific to this topic. LDA is a well-defined generative model and generalizes easily to new documents without overfitting. Since LDA is highly modular and hierarchical, therefore, it can easily be extended. Many extensions to basic LDA model have been proposed to incorporate document metadata. The simplest method of incorporating the metadata in generative topic models is to generate both the words and the metadata simultaneously given hidden topic variables. In this type of model, each topic has a distribution over words as in the standard model, as well as a distribution over metadata values. Examples of such model includes, Topics over Time model [16] of Wang and McCallum, Continuous Time Dynamic Topic Models [15] of Wand and Blei , the Group-Topic model of Wang, Mohanty and McCallum [17], Author-Topic model [14] of Rosen-Zvi, Griffiths, Steyvers and Smyth, Linked Topic and Interest Model [4] of Cheng and Li.

2.3 Parameter Estimation

Different approaches has been used in the topic-based probabilistic models for parameter estimation. These approaches includes Maximum likelihood estimation (MLE), Maximum a posteriori estimation (MAP) and Bayesian estimation. Expectation-maximization (EM)[9] is used to find the direct estimates of model parameters for MLE and MAP approaches. While variational EM [2], expectation propagation [5], Gibbs Sampling [6] algorithms provide approximate inference of the model parameters in Bayesian estimation. Blei [2] suggested to use approximate methods where parameters θ and ϕ can be integrated out because explicit estimate methods suffer from problem of local maxima in topic models.
In this paper we use Gibbs sampling for approximate inference because it is relatively a simple method for estimating parameters in high-dimensional models.

2.4 Related Models

In this section we briefly introduce the three models (Figure 1) whom capabilities are combined in ATT.

LDA (Figure 1(a)) is a Bayesian network that generates a document using a mixture of topics. In its generative process, for each document *d*, a multinomial distribution θ over topics is randomly sampled from a Dirichlet distribution with parameter α, and then to generate each word, a topic z is chosen from this topic distribution, and a word, *w*, is generated by randomly sampling from a topic-specific multinomial distribution ϕ_z. The robustness of the model is greatly enhanced by integrating out uncertainty about the per-document topic distribution θ.
The Author-Topic model [14] is a similar Bayesian network (Figure 1(b)), in which each author's interests are modeled with a mixture of topics. In its generative process for each document *d*, a set of authors, a_d, is observed. To generate each word, an author *x* is chosen uniformly from this set, then a topic z is selected from a topic distribution θ_x that is specific to the author, and then a word *w* is generated from a topic-specific multinomial distribution ϕ_z.
The Topics over Time (TOT) [16], a topic model that explicitly models time jointly with word co-occurrence pattern (Figure 1(c)). TOT parameterizes a continuous distribution over time associated with each topic, and topics are responsible for generating both observed timestamps as well as words. Parameter estimation is thus driven to discover topics that simultaneously capture word co-occurrence and locality of those patterns in time.

None of the given approaches models documents and author together with the temporal information. In this paper, we propose ATT: a model of topic dynamics in social media which connect the temporal topic dependency with the social actors, thereby, providing an insight into the evolution of topics over time along with capturing the author interests for a given time period.

3. THE ATT MODEL

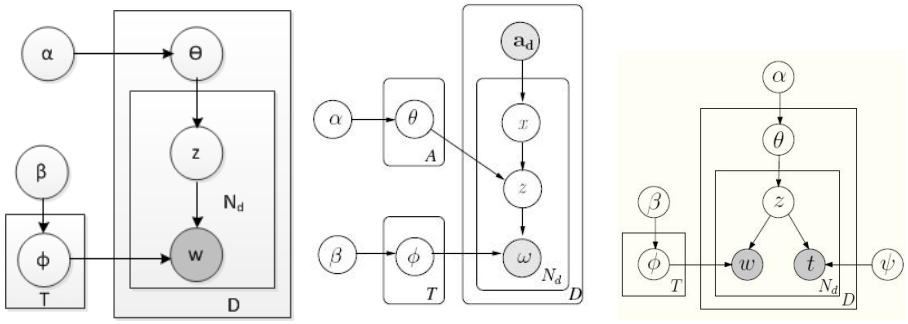
3.1 Model Design

We extend LDA by incorporating document metadata i.e. author and timestamp of the document in ATT model. Figure 3.1 is a graphical representation of ATT model using plate notation. In the model each author is modeled as having distribution overs topics and each topic is modeled as having distribution over words. The document generation process starts by picking each of the N_d words in the document d. Then sampling an author x uniformly at random from the list of authors A_d for document d. Then a topic z is chosen randomly from author specific distribution of topics θ_a. After selecting a topic z, a word w is sampled from the topic specific distribution over words ϕ_z. At the same time a timestamp *t* is generated from topic specific beta distribution ψ_z.

The ATT model has three sets of unknown parameters; the author distribution over topics θ, the topic distribution over words ϕ and the topic distribution over time ψ. Both θ and ϕ have multinomial distributions with symmetric Dirichlet priors having the hyperparameters α and β respectively. To avoid time discretization we use a continuous per-topic parametric Beta distribution ψ as used in [16] over absolute time values in the generative process, which gives a natural distribution of topics over time. We normalize the time-stamps to values between 0 and 1 for parameter estimation. Table 3.1 represents different notations used in the paper.

The generative process of the ATT model which corresponds to the process used in Gibbs sampling for parameter estimation is described as follows.

1. Draw $\theta \sim Dir(\alpha)$



(a) Topic-Word (LDA) (b) Author-Topic (c) Topics over Time (TOT)

Figure 1: Three related Bayesian network models for document content generation

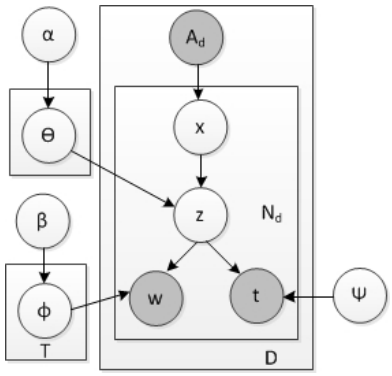


Figure 2: ATT Model for document content generation

Symbol	Description
T	number of topics
D	number of documents
A	number of authors
N_d	number of word tokens in document *d*
θ_a	the multinomial distribution of topics specific to the author *a*
ϕ_z	the multinomial distribution of words specific to topic *z*
ψ_z	the beta distribution of time specific to topic *z*
z_{di}	the topic associated with the ith token in the document *d*
w_{di}	the *i*th token in the document *d*
t_{di}	the timestamp associated with the *i*th token in the document *d*
a_{di}	the author associated with the *i*th token in the document *d*
α , β	Dirichlet priors

Table 1: Notation used in the paper

2. Draw $\phi \sim Dir(\beta)$
3. For each document *d*, pick an author from the list of authors a_d and draw a multinomial θ_d from Dirichlet prior α; then for each of the N_d words, w_i,
   - Draw a topic z_{d_i} from multinomial θ_d;
   - Draw a word w_{d_i} from multinomial $\phi_{z_{d_i}}$;
   - Draw a timestamp t_{d_i} from Beta $\psi_{z_{d_i}}$

3.2 Model Parameters

One of the model parameters which needs to be fixed before run is decision upon the the number of topics for the corpus. One can use automated parametric or non parametric methods that decide the correct number of topics for the corpus. Parametric method is to plot model log-likelihood against the number of topics. Select the no. topics for the model for which the log-likelihood reach the maximum. This approach over-fits the training data as for different enough document-term distributions, number of topics = number of documents would give the maximum log-likelihood. The other way is to use log-likelihood of holdout data to avoid over-fitting of the training data, but [3] pointed that better holdout data may infer lesser semantic meanings. In general, held-out likelihood often improves continually with very large number of topics (close to number of documents) or drop off quickly, thus suggesting number of topics that are too small to be useful. Alternate approach is to use non-parametric methods. But the setting of the Dirichlet prior on the topic-word distributions has a strong effect on the number of topics, which appears to be again determining the number of topics with parameter even if we do not realize this. The important question is therefore not "what is the correct number of topics" but rather "what is the model for". As we need more specific topics in order to be useful, therefore, we use manual approach and run the model for different number of topics and qualitatively judged the topics and select the number of topics for the corpus which produce more specific topics.

We found from different runs of the model with varying number of topics that model produced better topics when run with 100 topics and therefore set the number of topics to *K=100* and fix the hyperparameters $\alpha = 50/K$ and $\beta = 0.01$.

In the ATT model three parameters θ, ϕ, ψ are estimated. Exact inference of the parameters of LDA type models is intractable, therefore, we use Gibbs sampling [1] to perform approximate inference. In the model there are three latent variables *z*, *a* and *t*. Each set (z_i, a_i, t_i) of these latent variables is drawn as block conditioned on all other variables. We begin with the joint probability of dataset, and using the chain rule we obtain conditional probability for

$$p\left(z_i = j, x_i = k, t_i = l | w_i = m, z_{-i}, x_{-i}, t_{-i}, w_{-i}, a_d\right) \quad (1)$$

where z_i, x_i, t_i represent topic, author and time assigned to w_i whereas z_{-i}, x_{-i}, t_{-i} are all other assignments of that topic, author and time excluding the current assignment. w_{-i} represents all other words in the document set and a_d is the observed author of the document.
Learning joint probabilities of these three latent variables enables us to query the model conditioned on any combination of these variables using Baye's rule. For example given the author and time find the authors interest in that time period $P(\phi_d|a,t)$ or given the topic and time find the top authors contributing to the topic in that time $P(\theta_d|z,t)$.

3.3 Application Scenarios

The presented approach can be used for variety of applications. For example authors that are assigned high probability for a topic when it starts emerging can be seen as "topic pioneers" who conduct innovative research in that topic. Moreover, active authors that frequently change their topics of interest can be considered as "trend setters" in the respective research community. On the other hand, authors that have high probability at the peak topic activity can be seen as "mainstream" researchers that follow general trends and interests of the community. Finally, authors that have time-independent profiles with stable topics of interest can be recognized as foundational researchers that act independently of fluctuating trends and popular issues. From the application perspective, this knowledge can be exploited in a variety of ways, e.g. for advanced impact ranking, similarity-based contact recommendation for future collaborations, or better summarization of recent research trends and prediction of their further evolution.

4. EXPERIMENTS AND RESULTS

4.1 Experimental Settings

To show the effectiveness of our approach in capturing the topic evolution and finding the main contributors for different topics we run the model on subset of the CiteSeer publications. The dataset consists of abstracts and titles of research papers published in computer science domain by authors having more than 150 publications from 2001 to 2009. Total of 18 authors were selected at random for collecting documents. The minimum limit of 150 publications is applied to overcome the sparsity in data and to have sufficient text for capturing author interest over time. The dataset is divided into test set and training set. Test set contains 3054 documents and training set contains 950 documents. Dataset is preprocessed to remove stop words and noise be removing highly frequent terms and terms occurring in less than 10 documents. We used Porter stemmer [13] to reduce the word inflection to their stems.

We visualize each topic by showing the top K terms in descending order of the probability values assigned to terms as being the most representative terms of that topic. We show topic representation of 8 topics determined by ATT in Table5 and 6 topics captured by LDA in Table6 and 4 topics with authors that are assigned high probability in that topic and deemed as most influential authors along with beta PDF representing the topic evolution in Table2. The results shown in Table2, Table5 and Table6 are obtained by sampling from the 2000th iteration of Gibbs Sampler.
From topic visualization in Table5 and in Table6 we see that top k terms that are assigned high probability in that topic produced by ATT are semantically more cohesive than the terms in topics captured by LDA. Table3 presents average KL divergence between topic produced by ATT and LDA. Higeher values for ATT shows that topic produced by ATT are more distinct than topics produced by LDA. Table4 presents symmetric KL divergence of 4 sample topics. High KL divergence values show that topics are distinct to each other.

Model	Average KL Divergence
ATT	14.5934
LDA	8.4524

Table 3: Average KL divergence between topics for ATT and LDA

Topic Pair	KL Divergence
Image Analysis - Grid Computing	15.4372
Grid Computing - Semantic Web	14.942225
Semantic Web - Database Systems	14.4469
Database Systems - Image Analysis	14.000675

Table 4: Symmetric KL divergence for pairs of topics shown in Table 2

Table2 shows the top 5 terms and the top 4 authors for each topic and respective beta distribution capturing the topic activity. The interesting observation from qualitative analysis of the results is that the activities in the Semantic Web and Database System topics are correlated. As one topic starts gaining, the activity in other topic starts decreasing. Top authors for "Semantic Web" and "Database Systems" topics are well known authors in this field in our dataset. For example S. Staab started the carrier in database field and then shifted the interest to semantic web area later on appears to be producing maximum contents in these areas. This is also evident from the results in Table2 being the top author in both the topics. Results also shows that as the topic of semantic web started to emerge, influential authors in the database systems topic shifted to semantic web topic.

Authors that are assigned high probability for a topic when it starts emerging can be seen as "topic pioneers" who conduct innovative research in that topic. Moreover, active authors that frequently change their topics of interest can be considered as "trend setters" in the respective research community. On the other hand, authors that have high probability at the peak topic activity can be seen as "mainstream" researchers that follow general trends and interests of the community. Finally, authors that have time-independent profiles with stable topics of interest can be recognized as foundational researchers that act independently of fluctuating trends and popular issues.

4.2 Evaluation and Application

Quantitative evaluation of probabilistic models is done by using perplexity and it is the standard measure for estimating the performance of probabilistic models. Perplexity is defined as the ability of the model to predict words to new documents. It gives a measure of how much the model is surprised when it sees data which is unseen previously. While the qualitative evaluation is done by looking at top topic terms produced by the model that if they are semantically related to each other and a human subject is able to infer the topic by looking at the top terms.

Prediction power of the ATT model can be used in variety of ways. One such scenario is to recommend target authors whose research paper to read given user's interest at a given time point. That is, given some terms which describes an author's interest the task is to generate a ranked list of target authors whose interest are highly

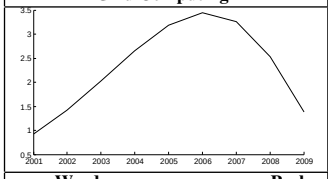
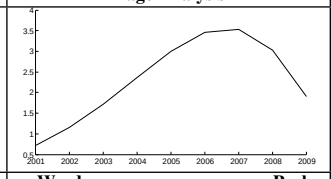
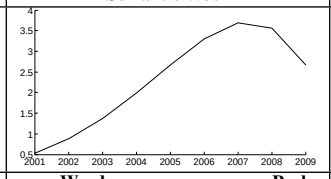
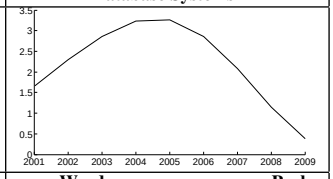
Grid Computing[1]		Image Analysis[2]		Semantic Web[3]		Database Systems[4]	
							
Word	**Prob.**	**Word**	**Prob.**	**Word**	**Prob.**	**Word**	**Prob.**
grid	0.0297	image	0.0185	resource	0.0380	database	0.0223
framework	0.0190	spectral	0.0150	web	0.0375	query	0.0190
dynamic	0.0175	test	0.0130	metadata	0.0298	sequence	0.0114
resource	0.0175	cluster	0.0120	rdf	0.0211	control	0.0100
integration	0.0131	statistic	0.0120	semantic	0.0195	search	0.0095
Author	**Prob.**	**Author**	**Prob.**	**Author**	**Prob.**	**Author**	**Prob.**
L. Tong	0.0746	X. Gu	0.1201	I. Horrocks	0.0892	S. Staab	0.1000
E. Gold	0.0207	C. Chan	0.0047	S. Staab	0.0686	I. Horrocks	0.0510
W. Zhao	0.0186	H. Lin	0.0025	A. Lin	0.0042	A. Joshi	0.0478
H. Lin	0.0134	J. Gao	0.0013	W. Nejdl	0.0011	W. Nejdl	0.0328

Table 2: Four topics captured by ATT model with influential authors and beta PDF showing topic activity

Topic 9		Topic 26		Topic 21	
Word	**Prob.**	**Word**	**Prob.**	**Word**	**Prob.**
storag	0.027969	ontolog	0.04307	protocol	0.039171
disk	0.023475	web	0.037838	control	0.020078
failur	0.018981	semant	0.028983	rout	0.019588
reliabl	0.017483	languag	0.022543	packet	0.018609
select	0.015985	rdf	0.018518	wireless	0.01763
server	0.014986	knowledg	0.015298	access	0.016651
fault	0.014986	schema	0.014896	servic	0.014692
cach	0.012988	servic	0.012481	util	0.014692
Topic 79		**Topic 93**		**Topic 63**	
Word	**Prob.**	**Word**	**Prob.**	**Word**	**Prob.**
file	0.028499	sensor	0.026113	peer	0.018658
metadata	0.022959	channel	0.025711	control	0.015105
secur	0.019793	access	0.023703	cach	0.013329
analysi	0.019002	protocol	0.020088	resourc	0.010664
safeti	0.015045	schedul	0.016874	manag	0.01022
share	0.014253	optim	0.013661	search	0.009775
storag	0.012671	wireless	0.012858	server	0.009775
express	0.012671	resourc	0.012054	dynam	0.009331
Topic 84		**Topic 90**		**Topic 62**	
Word	**Prob.**	**Word**	**Prob.**	**Word**	**Prob.**
access	0.024221	cluster	0.021389	delay	0.020169
traffic	0.022284	gene	0.017015	circuit	0.01274
sensor	0.018411	transact	0.015557	pair	0.011679
rang	0.017442	array	0.012155	optim	0.010618
load	0.017442	estim	0.011669	gate	0.010087
composit	0.0126	studi	0.011183	fault	0.010087
dynam	0.011631	construct	0.010697	function	0.009026
bank	0.011631	express	0.010697	axiom	0.009026

Table 5: Representation of 8 topics from a 100-topic run of Gibbs Sampler for CiteSeer dataset discovered by ATT model

likely to be similar with the user interest. This is achieved by computing the topic assignments to given user using the query terms from the posterior distributions of trained model. Then highly likely similar authors are found by computing the similarity between user and existing authors topic distributions in the model using KL divergence as distance measure as defined in Equation2. The small values of KL divergence between a pair of authors means both authors are similar. The target authors are then ranked based on the values of KL divergence between the user and target authors.

Table7 shows author pairs and their symmetric KL divergence when both author are present in top K authors of the same topic. The subscript numbers with author names indicate topics for which comparisons are made and are taken from Table2. Small values of KL divergence show that both author share similar interests. To mention S. Staab and I. Horrocks are well known authors in Semantic

Topic 3		Topic 4		Topic 9	
Word	**Prob.**	**Word**	**Prob.**	**Word**	**Prob.**
node	0.013466	peer	0.019961	ontolog	0.024649
optim	0.012448	queri	0.015684	logic	0.021779
cluster	0.012109	databas	0.014258	role	0.017559
test	0.011543	metadata	0.011407	knowledg	0.017052
graph	0.010298	resourc	0.008944	languag	0.015871
rout	0.008827	view	0.008944	descript	0.015026
structur	0.008148	grid	0.008426	web	0.015026
point	0.008035	search	0.008037	reason	0.014351
Topic 11		**Topic 6**		**Topic 19**	
Word	**Prob.**	**Word**	**Prob.**	**Word**	**Prob.**
ontolog	0.021625	queri	0.015773	node	0.015152
learn	0.013778	build	0.013414	traffic	0.013258
logic	0.010639	optim	0.011498	peer	0.013123
delay	0.009767	function	0.010909	optim	0.009606
tree	0.009593	size	0.010909	imag	0.00947
reason	0.009418	databas	0.008845	rout	0.009335
function	0.009244	oper	0.008403	watermark	0.008659
power	0.008895	logic	0.007813	symbol	0.008253

Table 6: Representation of 6 topics from CiteSeer dataset discovered by LDA

Web area and therefore share similar interest as shown in our results also. Table8 shows author pairs and their symmetric KL divergence when one author has high probability assigned in one topic and the other has high probability assigned in another topic. The large KL divergence values show that both authors have dissimilar interests.

$$sKL(i,j) = \sum_{t=1} \left[\theta_{it} log \frac{\theta_{it}}{\theta_{jt}} + \theta_{jt} log \frac{\theta_{jt}}{\theta_{it}} \right] \quad (2)$$

Author Pair	KL Divergence
$S.Staab_4 - W.Nejdl_4$	1.82
$S.Staab_3 - I.Horrocks_3$	2.40
$C.Chan_2 - H.Lin_2$	2.19
$H.Lin_2 - J.Gao_2$	2.06

Table 7: Intra-topic symmetric KL divergence for different pairs of authors

5. CONCLUSION

In this paper we have proposed a probabilistic approach that models text, authors and timestamps in a given set of documents thus enabling us to capture temporal topic activity and finding out influ-

Author Pair	KL Divergence
$S.Staab_4 - L.Tong_1$	8.87
$I.Horrocks_4 - X.Gu_2$	8.69
$S.Staab_3 - X.Gu_2$	8.64
$L.Tong_1 - X.Gu_2$	7.64

Table 8: Inter-topic symmetric KL divergence for different pairs of authors

ential authors for the captured topics. Joint modeling and learning posterior probabilities of text, author and time allows us to query model for any combination of these variables conditioned on each other for finding information about how author's interests change over time and how activity in topics changes with emergence of new topics. Results from the application of this model to the CiteSeer dataset show the applicability of the model to arbitrary document collections with author and temporal information for detecting topics trends, topic evolution and author's interests.

Because JAGS is not scalable to large document corpus, in future we are planning to implement Gibbs sampler specific to ATT to make it scalable for larger document corpus. We also plan to apply the model to twitter and facebook to investigate how changing user interests produces changes or correlate with underlying community structure evolution.

6. ACKNOWLEDGMENTS

This work was supported by the project WeGov (www.wegov-project.eu) funded by the European Commission under EC Grant number 248512 in the 7th Framework Programme, ICT2009.7.3 and Higher Education Commission(HEC); Government of Pakistan in collaboration with German Academic Exchange Service (DAAD) with in the framework of HEC-DAAD Fellowship Programme for higher education of Universities and College Teachers of Pakistan in Germany.

7. REFERENCES

[1] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan. An introduction to mcmc for machine learning. *Machine Learning*, pages 5–43, 2003.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003.

[3] J. Chang, J. Boyd-Graber, C. Wang, S. Gerrish, and D. M. Blei. Reading tea leaves: How humans interpret topic models. In *Neural Information Processing Systems*, 2009.

[4] V. Cheng and C. H. Li. Linked topic and interest model for web forums. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, pages 279–284, Washington, DC, USA, 2008. IEEE Computer Society.

[5] T. M. Department, T. Minka, and J. Lafferty. Expectation-propagation for the generative aspect model. In *In Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, pages 352–359. Morgan Kaufmann, 2002.

[6] T. L. Griffiths and M. Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101 Suppl 1:5228–5235, 2004.

[7] D. Gruhl, R. Guha, R. Kumar, J. Novak, and A. Tomkins. The predictive power of online chatter. In *KDD '05: Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 78–87, New York, NY, USA, 2005. ACM.

[8] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 491–501, New York, NY, USA, 2004. ACM.

[9] T. Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '99, pages 50–57, New York, NY, USA, 1999. ACM.

[10] T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, 42(1-2):177–196, 2001.

[11] N. Naveed, S. Sizov, and S. Staab. Attention: Understanding authors and topics in context of temporal evolution. In *Advances in Information Retrieval - 33rd European Conference on IR Research, ECIR 2011, Proceedings*, pages 733–737. Springer, 2011.

[12] K. Nigam, A. K. McCallum, S. Thrun, and T. Mitchell. Text classification from labeled and unlabeled documents using em. *Mach. Learn.*, 39(2-3):103–134, 2000.

[13] M. F. Porter. An algorithm for suffix stripping. *Readings in Information Retrieval*, 14(3):130–137, 1980.

[14] 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 '04, pages 487–494, Arlington, Virginia, United States, 2004. AUAI Press.

[15] C. Wang, D. M. Blei, and D. Heckerman. Continuous time dynamic topic models. In *UAI'08*, pages 579–586, 2008.

[16] X. Wang and A. McCallum. Topics over time: a non-markov continuous-time model of topical trends. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '06, pages 424–433, New York, NY, USA, 2006. ACM.

[17] X. Wang, N. Mohanty, and A. McCallum. Group and topic discovery from relations and text. In *Proceedings of the 3rd international workshop on Link discovery*, LinkKDD '05, pages 28–35, New York, NY, USA, 2005. ACM.

APPENDIX

A. ATT MODEL SCRIPT FOR JAGS

```
model{
 for( k in 1 : K ){
  phi[k , 1:V]~ddirch(beta[])
 }
 for ( ii in 1 : A ) {
  theta[ii,1:K]~ddirch(alpha[])
 }
 for( j in 1 : K ){
  alphab[j]~dunif (1,10)
  betab[j]~dunif (1,10)
 }
 for( m in 1:M ){
  for( n in  1:wdim[m] ){
   x[m,n]~dcat(a[m*A-A+1:m*A])
   z[m,n]~dcat(theta[x[m,n],1:K])
   w[wstart[m]+n-1]~dcat(phi[z[m,n] , 1:V])
   t[wstart[m]+n-1]~dbeta(alphab[z[m,n]],betab[z[m,n
  }
```

```
  }
}
```