

Who Are Experts Specializing in Landscape Photography?

Analyzing Topic-specific Authority on Content Sharing Services

Bin Bi* Ben Kao† Chang Wan† Junghoo Cho*

*University of California, Los Angeles
Los Angeles, United States
*{bbi, cho}@cs.ucla.edu

†The University of Hong Kong
Pokfulam Road, Hong Kong
†{kao, cwan}@cs.hku.hk

ABSTRACT

With the rapid growth of Web 2.0, a variety of content sharing services, such as *Flickr*, *YouTube*, *Blogger*, and *TripAdvisor* etc, have become extremely popular over the last decade. On these websites, users have created and shared with each other various kinds of resources, such as photos, video, and travel blogs. The sheer amount of user-generated content varies greatly in quality, which calls for a principled method to identify a set of authorities, who created high-quality resources, from a massive number of contributors of content. Since most previous studies only infer global authoritativeness of a user, there is no way to differentiate the authoritativeness in different aspects of life (topics).

In this paper, we propose a novel model of Topic-specific Authority Analysis (TAA), which addresses the limitations of the previous approaches, to identify authorities specific to given query topic(s) on a content sharing service. This model jointly leverages the usage data collected from the sharing log and the favorite log. The parameters in TAA are learned from a constructed training dataset, for which a novel logistic likelihood function is specifically designed. To perform Bayesian inference for TAA with the new logistic likelihood, we extend typical Gibbs sampling by introducing auxiliary variables. Thorough experiments with two real-world datasets demonstrate the effectiveness of TAA in topic-specific authority identification as well as the generalizability of the TAA generative model.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

1. INTRODUCTION

Over the last decade, we have been witnessing the explosion of Web 2.0 applications. In the new era of Web 2.0, web users are participating not only as passive consumers of content provided by websites, but also as contributors creating content collaboratively with fellow users, commonly referred to as *user-generated content*. With the rapid growth

of Web 2.0, a variety of *content sharing services*, such as *Flickr*¹, *YouTube*², *Blogger*³, and *TripAdvisor*⁴ etc, have become tremendously popular over the recent years. These websites enable users to create and share with each other various kinds of *resources*, such as photos, videos, and travel blogs, etc.

The sheer amount of user-generated content made available by the content sharing services can be both a blessing and a curse. From the point of user modeling, richer information content helps to build more accurate user profiles, leading to better services for consumers. On the other hand, the vast quantity of user-generated content available can often complicate the decision making process, as consumers do not have the time or ability to examine all data or compare all options [3]. On a content sharing website, the overwhelming resources vary greatly in quality, which result in confusion, sub-optimum decisions or dissatisfaction with choices made by users [27]. Therefore, it is highly significant to develop a principled method that identifies a set of authorities, who created quality-assured resources, from a massive number of contributors of content.

A lot of work has been done on authority identification in the context of social network and web structure analysis. However, most of these studies, such as typical PageRank, only infer *global* authoritativeness of each user, without assessing the authoritativeness in a particular aspect of life (topics) [24, 21, 11, 30]. It does not make sense for a user to find global authorities on a content sharing website. After all, each user has unique topical interest. For example, on Flickr, a user who is interested in photographing sunsets may look for an photographers expert in this specific topic and learn from her photos about the skill of sunset photography. On the other hand, no one is an authority on every topic. Clearly, *topic-specific* authority analysis provides a more detailed authoritativeness portfolio for a user, which is critical for authority identification on content sharing services.

A common way of distilling latent topics is to build a probabilistic topic model on the usage data collected from a *sharing log*. In a content sharing website, a *sharing log* stores users' posting and tagging history, as illustrated by Figure 1 in Section 3. However, the sharing log does not contain any information about the content quality of resources, based on which authorities are identified. It would be counterintu-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
KDD'14, August 24–27, 2014, New York, NY, USA.
Copyright 2014 ACM 978-1-4503-2956-9/14/08 ...\$15.00.
<http://dx.doi.org/10.1145/2623330.2623752>.

¹<http://www.flickr.com>

²<http://www.youtube.com>

³<http://www.blogger.com>

⁴<http://www.tripadvisor.com>

itive to assume a high sharing frequency for every authority. Therefore, a data source in addition to the sharing log is clearly needed. Luckily, a *favorite log* made available by a content sharing website provides a valuable signal for the derivation of the content quality of resources. On current content sharing services, a resource is often presented with a *favorite* button, which a user clicks if he or she likes the resource. A *favorite click* represents an endorsement of the content quality of the resource by the user. The favorite log records the set of favorite clicks as user feedback, as illustrated by Figure 2 in Section 3.

Despite considerable research on the sharing log for various applications, little is known about the emerging favorite log. It is nontrivial to leverage a favorite log for topic-specific authority analysis in that users do not explicitly specify their topical motivates under the favorite clicks. A statistical model, built upon both the sharing log and the favorite log, is imperative to uncover each user’s authoritativeness on different topics.

In this paper, we propose a novel Bayesian model to identify a list of authorities on given topic(s), which we refer to as Topic-specific Authority Analysis, abbreviated as TAA. The TAA model characterizes each user’s topical authoritativeness by introducing a user-specific random vector over latent topics. To assess the topical authoritativeness, TAA exploits favorite clicks through systematically modeling the associations among users’ interest and authoritativeness as well as the topics of favorited resources. We propose to learn the parameters in the TAA model from a training dataset of observations constructed from both usage logs. To this end, a novel logistic likelihood function specialized for the training set is proposed to relate the parameters to the observations. Bayesian inference for a model with a logistic likelihood has long been recognized as a hard problem. We extend typical collapsed Gibbs sampling by introducing auxiliary variables to overcome this problem. With the inferred parameters, an analysis framework is introduced to produce an ordered list of topic-specific authorities by their authoritativeness degrees that satisfy the user’s query intent.

The major contributions of our work are summarized as follows:

1. We propose a novel Bayesian model, TAA, to address the new problem of topic-specific authority analysis on content sharing services by jointly leveraging the two data sources: *sharing log* and *favorite log*.
2. We propose a principled approach to training dataset construction, in which a novel logistic likelihood function is introduced.
3. We extend classic collapsed Gibbs sampling by *data augmentation* to infer the parameters in the TAA model with the new logistic likelihood.
4. We conducted thorough experiments on the datasets collected from two specific real-world content sharing websites. Experimental results confirm the effectiveness of TAA in topic-specific authority identification as well as the predictive power of the TAA generative model.

The rest of this paper is organized as follows. In Section 2, we describe the prior work related to ours. The problem statement is given in Section 3. Section 4 introduces our

TAA model, for which the inference is depicted in Section 5. In Section 6, we present the experimental results. Finally, we conclude the paper in Section 7.

2. RELATED WORK

Much work has been done on authority identification based on a network structure. The two most representative studies are PageRank [24] and HITS [21]. Zhang *et al.* [32] tested PageRank and HITS on a specific online community for expert identification. Jurczyk and Agichtein [19] employed the HITS algorithm to discover authorities in question answer communities. Kempe *et al.* [20] abstracted authority analysis into a influence maximization problem and pioneered the Linear Threshold (LT) Model and Independent Cascade (IC) Model to explain the spread of influence in a social network. Along with subsequent works, such as [11] and [22], all these methods are only after the identification of global authorities instead of authorities for specific topics. Although Barbieri *et al.* [2] extended the LT and IC models to be topic-aware, the topics are obtained based on the network structure, while totally neglecting valuable textual content.

A few studies have been conducted to find topic-level authorities in the context of structure analysis of the web graph and social networks. Given the popularity of PageRank, it is only natural to extend it for topical authority analysis. Topic-Sensitive PageRank (TSPR) [16] was such an extension that computes per-topic PageRank scores for webpages. TSPR biases the computation of PageRank by replacing the classic PageRank’s uniform teleport vector with topic-specific ones. However, it requires an existing manually categorized topic hierarchies to derive per-topic teleport vectors. In [28], Tang *et al.* proposed a Topical Affinity Propagation (TAP) model for topic-level social influence. But, similar to TSPR, TAP requires a separate preprocess to obtain a set of topics. TwitterRank [31] extended TSPR to find topic-level influencers on Twitter. Instead of predefined topic hierarchies, a set of topics is first produced by typical LDA [8] on the tweets. Then TwitterRank applies a method similar to TSPR to compute the per-topic influence rank. Nallapati *et al.* proposed Link-PLSA-LDA [23] on a hyperlink network to estimate the influence of blogs. These studies differ from our TAA model in that they do not exploit the valuable favorite signal to model topic-specific authoritativeness. Although TwitterRank and Link-PLSA-LDA applied to the settings different from ours, we adapted them to the authority identification on content sharing services by building proper graph structures, and compared them with our TAA in empirical studies.

There also exist a few pieces of prior work on finding important users in various applications. Chen *et al.* [9] proposed a latent factor model for rating prediction, based on which reputable users are identified. Zhao *et al.* [33] found topic-level experts on community question answering services, and recommended appropriate experts to answer new questions. In [7], Followship-LDA was proposed to identify topic-specific influencers on microblogs. All these methods find important users under different contexts, with the data different from ours in nature.

In the context of recommender systems, a few topic modeling studies related to our work have been conducted. Several latent factor models were proposed for tag recommendation on social media [5, 6, 4]. Wang and Blei [29] developed the






User ID	Tag	Photo
14829	kingsgate castle	
14829	broadstairs	
14829	beach	
319526	wareham forest	
319526	sunrise	
...

Figure 1: Sample records from the sharing log of a photo sharing website

collaborative topic regression (CTR) model to recommend scientific articles to users of an online community. Agarwal and Chen [1] proposed fLDA, which is a new matrix factorization method integrating LDA priors, to predict ratings in recommender system applications. Despite the relevance of these studies to our work, there are clear differences between them. To make recommendations, CTR utilizes scalar rating responses different from the binary favorite feedback exploited by TAA. fLDA is able to take binary responses, but it aims to predict scalar ratings of users on various items, which is different from the ultimate goal of our work.

3. PROBLEM STATEMENT

In a nutshell, the objective of this paper is developing a statistical model that identifies the authorities on a content sharing website specific to given query topic(s). A topic-specific authority is defined as a user who excels in the specified topic. For example, given *city lights* as a query topic on a photo sharing website, the topic-specific authority model is intended to retrieve a list of users who are expert in city lights shooting at night.

A content sharing website generally logs a massive number of posting and tagging records that reflect every user’s unique interest and taste. These records constitute a *sharing log* that a content sharing service keeps track of. Figure 1 presents a few sample records from the sharing log of a photo sharing website. Each row in the table represents a record indicating that user u assigned tag t to resource r (i.e., a photo) which was posted by herself. For notational convenience, let L denote the total number of unique users in the log, M_u denote the number of resources posted by user u , and N_r denote the number of tags assigned to resource r . The notations used throughout this paper are given in Table 1. Some of the notations will be explained in later sections.

A feasible solution to topic-specific authority identification is adapting the classic topic model *Latent Dirichlet Allocation* (LDA) [8] to historical data in the sharing log. Specifically, we employ typical LDA on the sharing data by regarding a user as a document in a corpus, a tag as a word in a document. By fitting the topic model to observational data collected from the sharing log, we infer the optimal values of parameters θ and φ . The probabilities θ (i.e., $p(z|u)$) give the topic distribution for each user, and the probabilities φ (i.e., $p(t|z)$) give the tag distribution for each topic. As a result, topic-specific authorities can be derived from the distributions $p(z|u)$ and $p(t|z)$ by the standard *query likelihood model*, where each user is scored by the likelihood

Table 1: Notations used throughout this paper

Notation	Description
u	User identity
t	Tag identity
r	Resource identity
z	Topic assignment of a tag
f	Binary favorite feedback
L	Total number of unique users
K	Total number of unique topics
R	Total number of unique resources
V	Total number of unique tags in the vocabulary
M_u	Number of resources posted by user u
N_r	Number of tags assigned to resource r
N_u	Number of tags assigned by user u
θ	Per-user topic distribution
φ	Per-topic tag distribution
α, β	Dirichlet priors on Multinomial distributions
η	Per-user topical authoritativeness

of generating a given query. In particular, given a set of tags as a query q , we compute the likelihood $p(q|u)$ for each user u by:

$$p(q|u) = \prod_{t \in q} p(t|u) = \prod_{t \in q} \sum_{z=1}^K p(t|z)p(z|u). \quad (1)$$

The users with the highest likelihood $p(q|u)$ are then identified as topic-specific authorities.

The LDA-based authority analysis exploits the fact that a user is interested in a particular topic if he or she frequently labels photos with the tags specific to this topic. It further assumes that the more frequently a user uses the tags covering a specific topic, the more authoritative he or she should be on this topic. However, this is an arguable assumption which is not always valid. Tagging frequently on a particular topic does not automatically imply that the user is an authority on this topic. In fact, an authority does not have to tag more than the other users on the topic he or she excels in. For example, on a travel blogging service, a blogger who posts a number of articles tagged with *London travel* may not be an authority on blogging about traveling London, given the unknown quality of these articles. It is likely that he or she is new to blogging, in which case the articles could be at a beginner level in quality. On the other hand, an actual authority may post only a couple of blogs about London travel, but he or she can specialize in this specific topic, leading to the favorable high-quality blogs. By analyzing the usage data from a real-world sharing log, we observed that users’ tag frequency is actually independent of their authoritativeness.

Since the sharing log reports posting and tagging information, but we are looking for the information about the content quality of posted resources, a supplementary data source is needed. Fortunately, a *favorite log* available in most of the content sharing services should help to infer the content quality of resources. A favorite log consists of the records of each user’s favorites. Figure 2 depicts a few sample records from the favorite log of a photo sharing website. Each row in the table represents a record indicating that user u added resource r (i.e., a photo) to his or her favorites. A *favorite click* can be interpreted as the user’s vote in favor of the content quality of the favorited resource. It motivates our modeling the favorite signal to infer the content quality of resources based on which topic-specific authorities are identified.

<p>For each topic $k = 1, \dots, K$ Sample φ_k from <i>Dirichlet</i>(β)</p> <p>For each user $u = 1, \dots, L$ Sample θ_u from <i>Dirichlet</i>(α) For the n^{th} position of user u, where $n \in \{1, \dots, N_u\}$ Sample a topic z_{un} from <i>Multinomial</i>(θ_u), where $z_{un} \in \{1, \dots, K\}$ Sample a tag t_{un} from <i>Multinomial</i>($\varphi_{z_{un}}$), where $t_{un} \in \{1, \dots, V\}$ Sample authoritativeness vector η_u from <i>MVN</i>(μ, Σ)</p> <p>For each tuple (u, r) Sample a favorite feedback f_{ur} from <i>Bernoulli</i>($\frac{1}{1 + \exp(-\eta_u^T(\hat{z}_u \circ \hat{z}_{u'r}))}$), where $f_{ur} \in \{0, 1\}$ and u' is resource r's owner</p>
--

Figure 4: Generative process for Topic-specific Authority Analysis

set of topics are the ones that the resource r 's owner u' is expert in, and thus should be parameterized by high authoritativeness degrees. In this way, we uncover the hidden topical motivate under each favorite click.

Figure 3 shows the graphical model for our TAA, with the notations described in Table 1. The generative process of a user's tags and favorite feedback is summarized in Figure 4. A favorite feedback is naturally associated with a tuple (u, r) , where r denotes a resource, and u denotes the user who favorited r . To obtain individual user u 's interest distribution over topics, each user is viewed as a mixture of topics from which tags are drawn. More specifically, for each user $u \in \{1, \dots, L\}$, we first pick a topic distribution θ_u from a Dirichlet prior with parameter α . Then, to generate the n^{th} tag in the resources posted by u , a topic z_{un} is sampled from θ_u , after which the tag t_{un} is drawn from the tag distribution $\varphi_{z_{un}}$ for topic z_{un} . With all the obtained topics, we compute individual user u 's topical interest distribution \hat{z}_u by aggregating u 's topic assignments. On the other hand, the topic distribution $\hat{z}_{u'r}$ for individual resource r posted by user u' is obtained in a similar way, except that $\hat{z}_{u'r}$ is computed by counting u' 's topic assignments specific to resource r only.

The topic distributions \hat{z}_u and $\hat{z}_{u'r}$ enable the generation of favorite feedback. In particular, for each tuple (u, r) , the binary favorite feedback f_{ur} is sampled from a Bernoulli distribution with parameter $\frac{1}{1 + e^{-\eta_u^T(\hat{z}_u \circ \hat{z}_{u'r})}}$. More specifically, we compute the likelihoods of $f_{ur} = 1$ and $f_{ur} = 0$ using Equation (3) and Equation (4), respectively. As a result, $f_{ur} \in \{0, 1\}$ is drawn from a Bernoulli distribution of the two likelihoods.

The various parameters we can learn from TAA characterize the different factors that affect the model structure. For a user u , the K -dimensional vector η_u quantifies u 's unique authoritativeness over topics, and the value θ_{uz} gives the probability that u is interested in topic z . For a topic z , the value φ_{zt} indicates the probability of tag t belonging to topic z . The inferred quantities serve as the inputs to our authority analysis framework, which will be described later.

5. INFERENCE FOR TAA

In this section, we present how the parameters of the TAA model are inferred from the usage data collected from the sharing log and the favorite log. More specifically, we first construct a training dataset from the usage data, with which a new Bernoulli likelihood parameterized by a logistic function is specified. Finally, an extension of traditional Gibbs sampling specialized for the logistic likelihood function is proposed to infer the optimal values of the parameters.

5.1 Preference Learning

We learn the parameters of the TAA model from a training set of observations constructed from the usage data. As mentioned above, the favorite log consists of user preferences for resources in a content sharing service. One important fact about the favorite log is that only positive observations are available – each favorite click is viewed as positive feedback for the corresponding tuple (u, r) , i.e., $f_{ur} = 1$. However, there are not such clear conclusions for $f_{ur} = 0$. Considering the non-clicked tuples (u, r) (i.e., user u did not click on the *favorite* button for resource r .) as negative feedback ($f_{ur} = 0$) would misinterpret the signal of these tuples, since there are actually at least two different interpretations for any non-clicked tuple. One possibility is a negative feedback, meaning that the user did not like the resource and did not want to add it to his or her favorites. Another possibility is a missing value, indicating that the user did not even see the resource, in which case whether the user favorited the resource is unknown.

On the other hand, the non-clicked tuples should not be simply ignored, as typical machine learning models are not able to learn anything from the positive observations alone. To overcome the problem of missing negative feedback ($f_{ur} = 0$), we use tuple pairs as training data instead of individual tuples. As opposed to treating non-clicked tuples as negative observations, we assume that users prefer the resources, for which they clicked on the *favorite* buttons, over the other non-clicked resources from the same owner. More specifically, suppose that r_i and r_j represent two resources posted by a user. Given two tuples (u, r_i) and (u, r_j) , user u prefers r_i over r_j if and only if r_i was favorited by u while r_j was not, which is denoted by $r_i \succ_u r_j$. Formally, we create training data \mathcal{D} by including the pairwise preference relations as follows:

$$\mathcal{D} = \{(u, r_i, r_j) | r_i \succ_u r_j\}, \quad (5)$$

where each preference relation $o = (u, r_i, r_j)$ is a training sample representing the fact that user u prefers r_i over r_j . For the resources that are both favorited by a user, we cannot infer any preference. The same is true for two resources either of which a user did not favorite.

As discussed above, we construct the observational dataset \mathcal{D} using the induced preference relations in place of the raw favorite feedback f_{ur} . As a result, the likelihood functions (3) and (4) need to be extended to incorporate the pairwise preference. Therefore, we reformulate the likelihood of a preference relation as:

$$p(r_i \succ_u r_j | \eta_{u'}, \hat{z}_u, \hat{z}_{u'r_i}, \hat{z}_{u'r_j}) = \frac{1}{1 + e^{-\eta_{u'}^T(\hat{z}_u \circ \hat{z}_{u'r_i} - \hat{z}_u \circ \hat{z}_{u'r_j})}}. \quad (6)$$

The probability $p(r_i \succ_u r_j | \eta_{u'}, \hat{z}_u, \hat{z}_{u'r_i}, \hat{z}_{u'r_j})$ gives the likelihood that user u prefers resource r_i over resource r_j , both owned by user u' . Let Θ denote the set of parameters of the TAA model. The likelihood of observing all the preference relations in training data \mathcal{D} is then given by:

$$p(\mathcal{D} | \Theta) = \prod_{(u, r_i, r_j) \in \mathcal{D}} \frac{1}{1 + e^{-\eta_{u'}^T(\hat{z}_u \circ \hat{z}_{u'r_i} - \hat{z}_u \circ \hat{z}_{u'r_j})}}. \quad (7)$$

5.2 Bayesian Inference

Typical LDA-like generative models employ collapsed Gibbs sampling to infer their parameters [15, 17, 26]. However,

Bayesian inference for a model with the logistic likelihood function (6) has long been recognized as a hard problem, due to the analytically inconvenient form of the Gibbs sampler for a logistic likelihood [18, 12, 14]. In this section, we present an extension of traditional collapsed Gibbs sampling to infer the parameters in TAA. Our algorithm takes advantage of the *data-augmentation* idea by introducing auxiliary variables to the posterior distribution. It extends the very recent work on inference for logistic models [25, 10] to learn a Bayesian model for topic-specific authority analysis. Specifically, using the ideas of introducing Pólya-Gamma variables presented in [25, 10], we are able to derive the posterior probabilities for the Gibbs sampler analytically. Part of the derivation is provided in the appendix.

Let us first familiarize ourselves with a new family of Pólya-Gamma distributions [25].

Definition A random variable X has a Pólya-Gamma distribution with parameters $b > 0$ and $c \in \mathcal{R}$, denoted by $X \sim PG(b, c)$, if

$$X \stackrel{d}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k-1/2)^2 + c^2/(4\pi^2)}, \quad (8)$$

where the $g_k \sim \text{Gamma}(b, 1)$ are independent Gamma random variables; the notation $\stackrel{d}{=}$ denotes equality in distribution.

The Pólya-Gamma family has been carefully constructed to yield a simple Gibbs sampler for the Bayesian logistic model. Let $\delta_{ur_{ij}}$ denote a Pólya-Gamma variable specific to (u, r_i, r_j) . With the introduction of the auxiliary random variable $\delta_{ur_{ij}}$, the likelihood function (6) can be represented as mixtures of Gaussians with respect to a Pólya-Gamma distribution, which is rewritten as:

$$p(r_i \succ_u r_j | \boldsymbol{\eta}_{u'}, \hat{\mathbf{z}}_u, \hat{\mathbf{z}}_{u'r_i}, \hat{\mathbf{z}}_{u'r_j}) = \frac{1}{2} e^{\frac{\boldsymbol{\eta}_{u'}^\top \hat{\mathbf{z}}_{ur_{ij}}}{2}} \int_0^\infty e^{-\frac{\delta_{ur_{ij}} (\boldsymbol{\eta}_{u'}^\top \hat{\mathbf{z}}_{ur_{ij}})^2}{2}} p(\delta_{ur_{ij}} | 1, 0) d\delta_{ur_{ij}}, \quad (9)$$

where $\hat{\mathbf{z}}_{ur_{ij}} = \hat{\mathbf{z}}_u \circ \hat{\mathbf{z}}_{u'r_i} - \hat{\mathbf{z}}_u \circ \hat{\mathbf{z}}_{u'r_j}$.

As a result, the collapsed posterior distribution of TAA augmented with the variables δ is given by:

$$p(\mathbf{z}, \delta, \boldsymbol{\eta} | \mathbf{t}, \boldsymbol{\alpha}, \beta, \mu, \Sigma) \propto \prod_{u=1}^L \frac{\prod_{k=1}^K \Gamma(c_{ku} + \alpha_k)}{\Gamma(\sum_{k=1}^K c_{ku} + \alpha_k)} \times \prod_{k=1}^K \frac{\prod_{t=1}^V \Gamma(g_{kt} + \beta_t)}{\Gamma(\sum_{t=1}^V g_{kt} + \beta_t)} \times p(\boldsymbol{\eta} | \mu, \Sigma) \prod_{(u, r_i, r_j) \in \mathcal{D}} e^{\frac{\boldsymbol{\eta}_{u'}^\top \hat{\mathbf{z}}_{ur_{ij}} - \delta_{ur_{ij}} (\boldsymbol{\eta}_{u'}^\top \hat{\mathbf{z}}_{ur_{ij}})^2}{2}} p(\delta_{ur_{ij}} | 1, 0) \quad (10)$$

where c_{ku} is the number of user u 's tags assigned to topic k , and g_{kt} is the total number of times tag t is assigned to topic k over the dataset. The detailed derivation of Equation (10) is provided in the appendix.

The univariate conditionals for a Gibbs sampler are then given as follows. The notation \bullet represents all the variables other than the one to be sampled.

$[p(\boldsymbol{\eta}_x | \bullet)]:$

We impose a zero-mean isotropic Gaussian prior on the K -dimensional random vector $\boldsymbol{\eta}_x$ which characterizes user

x 's topical authoritativeness:

$$p(\boldsymbol{\eta}_x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\sum_k \eta_{xk}^2}{2\sigma^2}}. \quad (11)$$

Thanks to the invariance property of the conjugate prior, the posterior distribution of $\boldsymbol{\eta}_x$ is also a Multivariate Gaussian:

$$p(\boldsymbol{\eta}_x | \bullet) \propto p(\boldsymbol{\eta}_x) \prod_{r_i \in R(x) \wedge r_j \in R(x)} e^{\frac{\boldsymbol{\eta}_x^\top \hat{\mathbf{z}}_{ur_{ij}} - \delta_{ur_{ij}} (\boldsymbol{\eta}_x^\top \hat{\mathbf{z}}_{ur_{ij}})^2}{2}} = \text{MVN}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) \quad (12)$$

where $r_i \in R(x)$ represents that resource r_i is posted by user x . The posterior mean $\boldsymbol{\mu}_x$ and posterior covariance $\boldsymbol{\Sigma}_x$ are given by:

$$\boldsymbol{\mu}_x = \boldsymbol{\Sigma}_x \left(\sum_{r_i \in R(x) \wedge r_j \in R(x)} \frac{1}{2} \hat{\mathbf{z}}_{ur_{ij}} \right) \boldsymbol{\Sigma}_x = \left(\frac{1}{\sigma^2} I + \sum_{r_i \in R(x) \wedge r_j \in R(x)} \delta_{ur_{ij}} \hat{\mathbf{z}}_{ur_{ij}} \hat{\mathbf{z}}_{ur_{ij}}^\top \right)^{-1}$$

$[p(z_{un} | \bullet)]:$

The posterior distribution of \mathbf{z} is:

$$p(\mathbf{z} | \bullet) \propto \prod_{u=1}^L \frac{\prod_{k=1}^K \Gamma(c_{ku} + \alpha_k)}{\Gamma(\sum_{k=1}^K c_{ku} + \alpha_k)} \times \prod_{k=1}^K \frac{\prod_{t=1}^V \Gamma(g_{kt} + \beta_t)}{\Gamma(\sum_{t=1}^V g_{kt} + \beta_t)} \times \prod_{(u, r_i, r_j) \in \mathcal{D}} e^{\frac{\boldsymbol{\eta}_{u'}^\top \hat{\mathbf{z}}_{ur_{ij}} - \delta_{ur_{ij}} (\boldsymbol{\eta}_{u'}^\top \hat{\mathbf{z}}_{ur_{ij}})^2}{2}} \quad (13)$$

The univariate conditional distribution of one variable z_{un} given all the other variables is then given by:

$$p(z_{un} = k | \bullet) \propto \frac{(c_{ku}^{-(un)} + \alpha_k)(g_{kt}^{-(un)} + \beta_{tun})}{\sum_{t=1}^V g_{kt}^{-(un)} + \sum_{t=1}^V \beta_t} \times \prod_{(u, r_i, r_j) \in \mathcal{D}} p(r_i \succ_u r_j | \boldsymbol{\eta}_{u'}, \mathbf{z}_{-(un)}, z_{un} = k) \quad (14)$$

where $c_{ku}^{-(un)}$ bears the same meaning of c_{ku} only with the n th tag of user u excluded; similarly $g_{kt}^{-(un)}$ is defined in the same way as g_{kt} only without the count for the n th tag of user u , and $\mathbf{z}_{-(un)}$ denotes the topics for all tags except z_{un} .

$[p(\delta_{ur_{ij}} | \bullet)]:$

By definition, the posterior distribution of the auxiliary variable $\delta_{ur_{ij}}$ turns out to be a Pólya-Gamma distribution:

$$p(\delta_{ur_{ij}} | \bullet) \propto e^{-\frac{\delta_{ur_{ij}} (\boldsymbol{\eta}_{u'}^\top \hat{\mathbf{z}}_{ur_{ij}})^2}{2}} p(\delta_{ur_{ij}} | 1, 0) = \text{PG}(1, \boldsymbol{\eta}_{u'}^\top \hat{\mathbf{z}}_{ur_{ij}}) \quad (15)$$

The above posterior univariate distributions create a Markov chain for Gibbs sampling. It has been shown that the stationary distribution of the Markov chain is just the sought-after posterior joint distribution [13]. Specifically, the Gibbs sampler iteratively draws samples from $p(\boldsymbol{\eta}_x | \bullet)$, $p(z_{un} | \bullet)$ and $p(\delta_{ur_{ij}} | \bullet)$ using Equations (12), (14) and (15), respectively. After the Gibbs sampler has run for an appropriate number of iterations (until the chain has converged to a stationary distribution), we draw a sample $\boldsymbol{\eta}_x$ for each user x , which

quantifies x 's topical authoritativeness, and obtain the estimates for the distributions θ and φ via the following equations:

$$\theta_{uz} = \frac{c_{zu} + \alpha_z}{\sum_{k=1}^K c_{ku} + \sum_{k=1}^K \alpha_k} \quad (16)$$

$$\varphi_{zt} = \frac{g_{zt} + \beta_t}{\sum_{t=1}^V g_{zt} + \sum_{t=1}^V \beta_t} \quad (17)$$

5.3 Authority Analysis Framework

With the inferred parameters, we introduce an analysis framework for topic-specific authority identification. The analysis framework allows a user to issue a query q reflecting the topic(s) on which authorities are to be identified. The query q consists of a list of tags, where multi-occurrences of a tag are allowed to reflect its importance to the query topic(s). The analysis framework subsequently produces an ordered list of authorities by their authoritativeness degrees that satisfy the user's query intent.

To rank a list of authorities, the analysis framework requires (a) every user's topical authoritativeness: η , and (b) the topic(s) of query q : \mathbf{z}_q . When the TAA model is used as the underlying topic-specific authority analysis method, the topical authoritativeness η is produced as part of the results. To derive q 's topic(s) \mathbf{z}_q , we use the folding-in technique on TAA by treating the query as a new user, and perform the sampling for only the tags of the pseudo user. Given the derived topical authoritativeness $\boldsymbol{\eta}_u$ and the query topic(s) \mathbf{z}_q , we obtain the final authoritativeness $\Psi(u, q) = \sum_{i=1}^{N_q} \eta_{uz_{qi}}$ for a user u with respect to the query q , where N_q denotes the number of tags in q . Finally, the users are returned in decreasing order of their authoritativeness $\Psi(u, q)$.

6. EMPIRICAL EVALUATION

In this section, we report the experimental results of the TAA model on real-world data collected from two specific content sharing services: *Flickr*⁵ and *500px*⁶. We quantitatively compare the results of TAA with those of several competitors on both datasets. We also give real examples of Flickr authorities identified by TAA. Analysis and discussion of the experimental results are presented in this section.

6.1 Data Collections

Although TAA is a generic Bayesian model which is applicable to topic-specific authority identification on various kinds of content sharing services, we conduct experiments on the real-world datasets collected from two specific websites *Flickr* and *500px* to evaluate the quality of identified authorities. *Flickr* is one of the most popular photo sharing website, which allows users to store, share, tag and organize their photos. The huge number of Flickr users calls for a topic-specific authority model to identify the best photographers for a specified query topic. As opposed to Flickr's general user base, *500px* is a photo sharing platform catered to professional photographers. A distinct feature of *500px* is the *Editors' Choice* page⁷ which shows the finest photos hand-picked by the professional editors employed by *500px*. These high-quality photos are used to derive the ground truth for our empirical evaluation.

⁵<http://www.flickr.com>

⁶<http://www.500px.com>

⁷<http://www.500px.com/editors>

Table 2: Statistics of Experimental Datasets

Data	#users	#photos	#tag asgmts	#fav. clicks
Flickr	21,054	204,335	3,014,813	1,562,805
500px	33,581	318,906	3,520,179	1,837,049

We collected the sharing logs and the favorite logs from both Flickr and 500px. The usage data obtained from the collected logs were processed to create training data \mathcal{D} , on which a TAA model was built. Extra usage information was collected to derive the ground truth for both datasets, which will be described in the next subsection. The basic statistics of the Flickr dataset and the 500px dataset are given in Table 2.

6.2 Evaluation Strategy

Quantitatively evaluating the quality of topic-specific authority analysis is a difficult task, since a content sharing service generally does not explicitly specify real authorities given a topic. Luckily, the abundant information embedded in the databases of Flickr and 500px helps to derive ground truth of topic-specific authorities.

Flickr has a large number of user-created groups that allow people who have similar interests to get together and share their photos reflecting these interests. Each of the groups is generally dedicated to a certain topic, such as food, animals, certain photo techniques, or creative commons, etc. Every group has one or more administrators which can be viewed as the real authorities specific to the group topic. On the other hand, 500px organizes photos by category, such as wedding, underwater, concert, or transportation, etc. We rank the users for each category according to their numbers of photos get selected by the editors by category. The ranked list of users for each category is instead viewed as ground truth, since unlike a Flickr group, a 500px category has no administrators specific to the category topic.

Given the different kinds of ground truth for Flickr and 500px, we used different evaluation metrics to measure the quality of the results from compared algorithms. Let Q denote a set of queries. For each query $q \in Q$, each algorithm returns an ordered list of users by their authoritativeness. For the Flickr dataset, we employed the standard Mean Reciprocal Rank (MRR). The Reciprocal Rank of a ranked list is the multiplicative inverse of the rank of the first hit in the list. The MRR score of an algorithm is the average reciprocal rank obtained by the ranked lists given by the algorithm with respect to the query set Q . Formally,

$$MRR = \frac{1}{|Q|} \sum_{q \in Q} \frac{1}{rank_q} \quad (18)$$

where $rank_q$ is the rank of the first real authority in the ranked list for query q . By definition, a higher MRR score indicates a better algorithm. For the 500px dataset, on the other hand, we employed the Spearman's rank correlation coefficient to assess the correlation between ground truth and a ranked list of users given by each algorithm. The Spearman's coefficient ρ_q for query q can take a range of values from -1 to +1 ($\rho_q < 0$ for a negative correlation, $\rho_q > 0$ for a positive correlation). The Spearman's coefficient ρ of an algorithm is the average Spearman's coefficient over the query set Q given by the algorithm. Formally,

$$\rho = \frac{1}{|Q|} \sum_{q \in Q} \rho_q \quad (19)$$

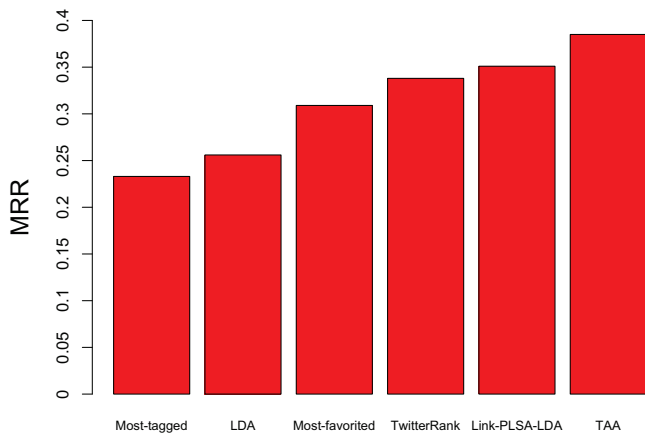


Figure 5: MRR for the Flickr dataset

6.3 Quality of Authority Analysis

In our experiments, we evaluated the quality of the authorities identified by the six algorithms, *Most-tagged*, *LDA*, *Most-favorited*, *TwitterRank*, *Link-PLSA-LDA*, and *TAA*. Given a set of tags as a query, the *Most-tagged* approach first identifies relevant photos by lexical matches against the query tags. The number of relevant photos of each user is viewed as his or her authoritative degree, by which *Most-tagged* produces a ranked list of users as a final result. By contrast, *LDA* identifies relevant photos using probabilistic topic modeling [8]. As a result, users are ranked in descending order of the query likelihoods given by Equation (1). Note that both *Most-tagged* and *LDA* utilize observational data from the sharing log while neglecting the valuable signal from the favorite log. On the contrary, *Most-favorited* leverages both the sharing log and the favorite log in a way that produces an ordered list of users by the numbers of times their relevant photos are favorited. As opposed to the previous three approaches, *TwitterRank* and *Link-PLSA-LDA* both build upon the graph structure constructed from the favorite log. Specifically, we construct the graph by creating a node for each user. There exists a link from node u to node v if the user corresponding to u favorited any photo of the user corresponding to v . A user’s tags are associated with the corresponding node. The *TwitterRank* algorithm was originally proposed to find topic-level key influencers on Twitter [31]. It extends typical Topic-Sensitive PageRank [16] to compute per-topic influence scores. This requires a separate preprocess to create topics by running LDA on the text content associated with the nodes. The transition probability between two nodes in *TwitterRank* is defined based on the topical similarity between the corresponding users. Given the similar nature of the Twitter network and our constructed graph, we employ the *TwitterRank* algorithm to find topic-level authorities on a content sharing service. On the other hand, *Link-PLSA-LDA* is a probabilistic topic model on a hyperlink/citation network, which jointly models text and citations to estimate the influence of blogs/publications [23]. We adapt it to our constructed graph for topic-specific authority analysis. In our experiments, for every topic-sensitive algorithm, we set the number of topics to 100. We set all symmetric priors as 0.1 for every model with Dirichlet priors. For our TAA, we ran Gibbs sampling for 500 iterations. These settings are fairly typical and their tuning is beyond the scope of this paper.

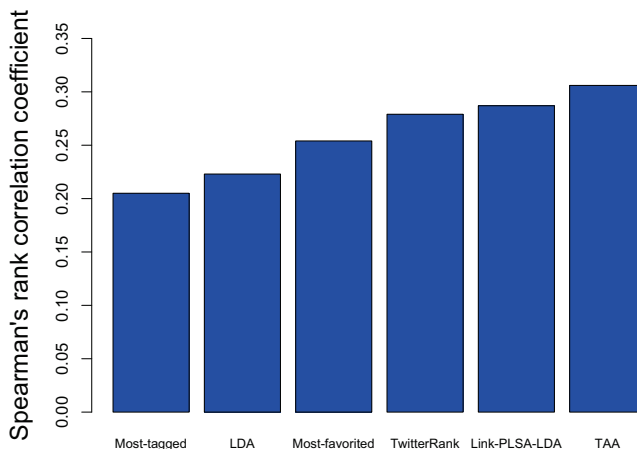


Figure 6: Spearman’s rank correlation coefficient for the 500px dataset

To compute MRRs on the Flickr dataset, we randomly selected 200 Flickr groups, whose administrators were treated as the real authorities on the respective group topics. The *Top Tags* generated by Flickr for each group were fed as a query to each algorithm. Figure 5 shows the MRR score of each algorithm on the Flickr dataset. It is observed that *Most-tagged* and *LDA* were inferior to the other algorithms, as neither of them models the valuable favorite signal. On the contrary, by exploiting the favorite data, the algorithms *Most-favorited*, *TwitterRank*, *Link-PLSA-LDA* and the proposed *TAA* produced higher MRR scores. In particular, *TwitterRank* underperformed *Link-PLSA-LDA* and *TAA*, due to its separation between topic modeling and authority analysis. To further measure the improvement of *TAA* over the runner-up *Link-PLSA-LDA*, we performed a paired t -test between them, which gave p -value < 0.05 . It indicated that the improvement of *TAA* over *Link-PLSA-LDA* was *statistically significant*. This is not surprising because *Link-PLSA-LDA* as well as *TwitterRank* fail to uncover the latent topical motivate under each favorite click. Instead, they establish a link on the graph as long as a user favorited any photo of another, disregarding the identity of the photo as well as its underlying topics.

For 500px, we plot the Spearman’s coefficient for each algorithm in Figure 6. From this figure, we observe the pattern similar to that of Figure 5. *TAA* outperformed all the other algorithms, thanks to its unified framework of topic modeling and authority analysis. In addition, *TAA* benefited from its ability to identify users’ topical authoritative by uncovering each favorite click’s underlying topical motivate and learning from pairwise resource preference.

6.4 Predictive Power Analysis

As generative models, our *TAA*, as shown in Figure 3, and the competitor *Link-PLSA-LDA* [23] are able to generate and predict unseen new data. We evaluated the predictive power and generalizability of both models using the standard *perplexity* metric [8]. The perplexity is monotonically decreasing in the likelihood of the unseen test data. Hence, a lower perplexity score indicates stronger predictive power. Formally, the perplexity is defined as:

$$\text{perplexity}(F_{\text{test}}) = \exp \left\{ -\frac{\sum_{f \in F_{\text{test}}} \log p(f)}{|F_{\text{test}}|} \right\}, \quad (20)$$

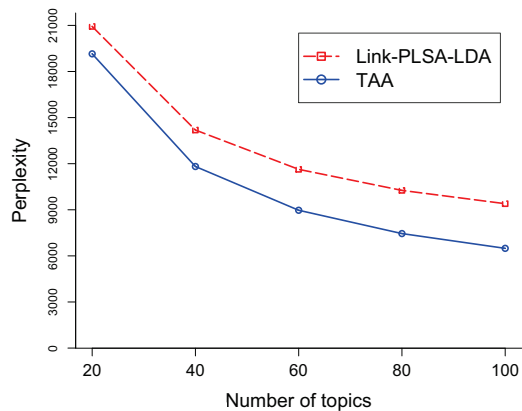


Figure 7: Perplexity for the Flickr dataset

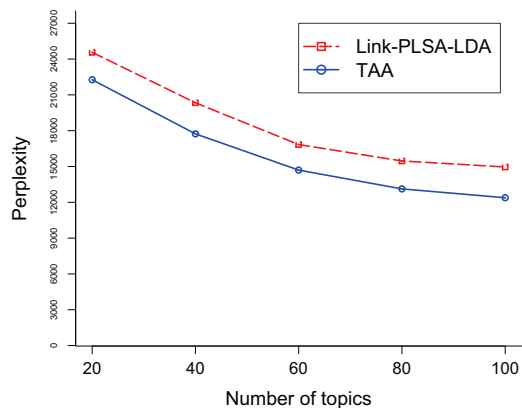


Figure 8: Perplexity for the 500px dataset

where F_{test} denotes the test set of favorites. For both Flickr and 500px, we held out 10% of the data for test purposes and trained the models on the remaining 90%.

Figure 7 and Figure 8 present the perplexity as a function of the numbers of topics for both models on Flickr data and 500px data, respectively. It is clear that the *TAA* consistently produced lower perplexity scores than *Link-PLSA-LDA* for both Flickr and 500px, indicating that our *TAA* model has stronger predictive power and better generalizability. Moreover, *TAA* predicted unseen favorites even better as the number of topics increases.

6.5 Case Visualization

For the visualization of the *TAA* model, we performed searches on Flickr data for a list of photographers who are expert in two specific topics. Figure 9 shows the examples of photographers identified by *TAA* together with their ranks in the lists. To illustrate their expertise in photography, photos on the query topics are presented as well. For the first query topic: *winter snow landscape*, we see from the photos that the first user in the ranked list demonstrated the expertise in shooting snow landscape in winter. By contrast, the user in rank 100 seemed to have broader interests, not specializing in this specific topic. The last user looked even irrelevant to the query topic. For the second query topic: *waterscape*, the user at the top was clearly superior to the others in waterscape shooting, although some photos from the last two users were somewhat related to the water topic.

7. CONCLUSION

This paper addresses the problem of authority analysis specific to given query topic(s) for users on a content shar-

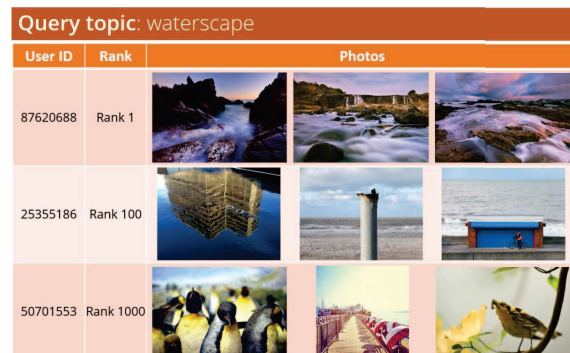
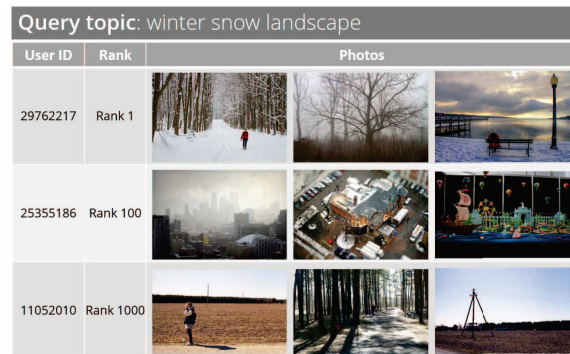


Figure 9: Examples of the ranked lists of photographers identified by *TAA* on Flickr data

ing service. To model topic-specific authoritativeness, we introduce a novel method of Topic-specific Authority Analysis (*TAA*), which properly captures the associations among users' interest and authoritativeness as well as the topics of favorited resources to exploit the signal of favorite clicks. The parameters in the *TAA* model are learned from a training set of observations constructed from two data sources: *sharing log* and *favorite log*. To overcome the limitation of missing negative feedback, we propose a preference learning technique embedding a new logistic likelihood function. An extension of typical collapsed Gibbs sampling is further proposed for Bayesian inference with the logistic likelihood. With the inferred parameters, our analysis framework produces a ranked list of authorities by their authoritativeness specific to given query topic(s).

We conducted thorough experiments on the datasets collected from two specific real-world content sharing websites, Flickr and 500px. Experimental results demonstrate that the *TAA* model outperforms the competitors, confirming its effectiveness in topic-specific authority analysis and its generalizability to unseen data.

8. ACKNOWLEDGMENT

This research is supported by Hong Kong Research Grants Council grant HKU712712E.

9. REFERENCES

- [1] D. Agarwal and B.-C. Chen. fda: Matrix factorization through latent dirichlet allocation. In *Proc. of WSDM '10*, pages 91–100, New York, NY, USA, 2010.
- [2] N. Barbieri, F. Bonchi, and G. Manco. Topic-aware social influence propagation models. In *Proc. of ICDM '12*, pages 81–90, Washington, DC, USA, 2012.
- [3] S. Bellman, E. J. Johnson, G. L. Lohse, and N. Mandel. Designing marketplaces of the artificial with consumers in mind: Four approaches to

understanding consumer behavior in electronic environments. *J. Interactive Marketing*, 20(1), 2006.

[4] B. Bi and J. Cho. Automatically generating descriptions for resources by tag modeling. In *Proc. of CIKM '13*, pages 2387–2392, 2013.

[5] B. Bi, S. D. Lee, B. Kao, and R. Cheng. Cubels: An effective and efficient method for searching resources in social tagging systems. In *ICDE*, pages 27–38, 2011.

[6] B. Bi, L. Shang, and B. Kao. Collaborative resource discovery in social tagging systems. In *Proc. of CIKM '09*, pages 1919–1922, 2009.

[7] B. Bi, Y. Tian, Y. Sismanis, A. Balmin, and J. Cho. Scalable topic-specific influence analysis on microblogs. In *Proc. of WSDM*, pages 513–522, 2014.

[8] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *JMLR*, 3:993–1022, Mar. 2003.

[9] B.-C. Chen, J. Guo, B. Tseng, and J. Yang. User reputation in a comment rating environment. In *Proc. of KDD '11*, pages 159–167, New York, USA, 2011.

[10] N. Chen, J. Zhu, F. Xia, and B. Zhang. Generalized relational topic models with data augmentation. In *Proc. of IJCAI '13*, pages 1273–1279, 2013.

[11] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *Proc. of KDD '09*, pages 199–208, New York, NY, USA, 2009.

[12] S. Fruhwirth-Schnatter and R. Fruhwirth. Data augmentation and mcmc for binary and multinomial logit models. In *Sta Mod Reg Str*, pages 111–132. 2010.

[13] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. November 2013.

[14] R. B. Gramacy and N. G. Polson. Simulation-based regularized logistic regression. *Bayesian Analysis*, 7(3):567–590, September 2012.

[15] T. L. Griffiths and M. Steyvers. Finding scientific topics. *PNAS*, 101(Suppl. 1):5228–5235, April 2004.

[16] T. Haveliwala. Topic-sensitive pagerank: a context-sensitive ranking algorithm for web search. *IEEE TKDE*, 15(4):784–796, July 2003.

[17] G. Heinrich. Parameter estimation for text analysis. Technical report, University of Leipzig, 2008.

[18] C. C. Holmes and L. Held. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, 1(1):145–168, March 2006.

[19] P. Jurczyk and E. Agichtein. Discovering authorities in question answer communities by using link analysis. In *Proc. of CIKM '07*, pages 919–922, New York, 2007.

[20] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proc. of KDD '03*, pages 137–146, New York, 2003.

[21] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *JACM*, 46(5):604–632, 1999.

[22] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *KDD*, pages 420–429, 2007.

[23] R. M. Nallapati, A. Ahmed, E. P. Xing, and W. W. Cohen. Joint latent topic models for text and citations. In *Proc. of KDD '08*, pages 542–550, 2008.

[24] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. In *Proc. of WWW '98*, pages 161–172, Brisbane, 1998.

[25] N. G. Polson, J. G. Scott, and J. Windle. Bayesian inference for logistic models using pólya-gamma latent variables. *JASA*, 108(504):1339–1349, 2013.

[26] I. Porteous, D. Newman, A. Ihler, A. Asuncion, P. Smyth, and M. Welling. Fast collapsed gibbs sampling for latent dirichlet allocation. In *KDD*, 2008.

[27] D. Smith, S. Menon, and K. Sivakumar. Online peer and editorial recommendations, trust, and choice in virtual markets. *J. Interactive Marketing*, 19(3), 2005.

[28] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *Proc. of KDD '09*, pages 807–816, New York, NY, USA, 2009.

[29] C. Wang and D. M. Blei. Collaborative topic modeling for recommending scientific articles. In *KDD*, 2011.

[30] Y. Wang, G. Cong, G. Song, and K. Xie. Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In *Proc. of KDD '10*, pages 1039–1048, New York, 2010.

[31] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twiterrank: Finding topic-sensitive influential twitterers. In *Proc. of WSDM '10*, pages 261–270, 2010.

[32] J. Zhang, M. S. Ackerman, and L. Adamic. Expertise networks in online communities: Structure and algorithms. In *WWW '07*, pages 221–230, 2007.

[33] T. Zhao, N. Bian, C. Li, and M. Li. Topic-level expert modeling in community question answering. In *SDM '13*, pages 776–784. SIAM, 2013.

APPENDIX

Let us derive the collapsed posterior distribution of TAA augmented with the variables δ , as follows:

$$\begin{aligned}
& p(\mathbf{z}, \delta, \eta | \mathbf{t}, \mathbf{o}, \alpha, \beta, \mu, \Sigma) \\
& \propto p(\mathbf{t}, \mathbf{z} | \alpha, \beta) p(\mathbf{o}, \delta | \mathbf{z}, \eta) p(\eta | \mu, \Sigma) \\
& = \int \int p(\mathbf{t}, \mathbf{z}, \theta, \varphi | \alpha, \beta) d\theta d\varphi \times p(\eta | \mu, \Sigma) p(\mathbf{o}, \delta | \mathbf{z}, \eta) \\
& = \int p(\mathbf{z} | \theta) p(\theta | \alpha) d\theta \times \int p(\mathbf{t} | \varphi, \mathbf{z}) p(\varphi | \beta) d\varphi \\
& \quad \times p(\eta | \mu, \Sigma) p(\mathbf{o}, \delta | \mathbf{z}, \eta) \\
& = \prod_{u=1}^L \int \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_{uk}^{\alpha_k - 1} \prod_{n=1}^{N_u} \theta_{uz_{un}} d\theta_u \\
& \quad \times \prod_{k=1}^K \int \frac{\Gamma(\sum_{t=1}^V \beta_t)}{\sum_{t=1}^V \Gamma(\beta_t)} \prod_{t=1}^V \varphi_{kt}^{\beta_t - 1} \prod_{u=1}^L \prod_{n=1}^{N_u} \varphi_{z_{un} t_{un}} d\varphi_k \\
& \quad \times p(\eta | \mu, \Sigma) p(\mathbf{o}, \delta | \mathbf{z}, \eta) \\
& \quad \text{(Expand out Dirichlet and Multinomial distributions)} \\
& = \prod_{u=1}^L \int \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_{uk}^{\alpha_k + c_{ku} - 1} d\theta_u \\
& \quad \times \prod_{k=1}^K \int \frac{\Gamma(\sum_{t=1}^V \beta_t)}{\sum_{t=1}^V \Gamma(\beta_t)} \prod_{t=1}^V \varphi_{kt}^{\beta_t + g_{kt} - 1} d\varphi_k \\
& \quad \times p(\eta | \mu, \Sigma) p(\mathbf{o}, \delta | \mathbf{z}, \eta) \\
& \propto \prod_{u=1}^L \frac{\prod_{k=1}^K \Gamma(c_{ku} + \alpha_k)}{\Gamma(\sum_{k=1}^K c_{ku} + \alpha_k)} \times \prod_{k=1}^K \frac{\prod_{t=1}^V \Gamma(g_{kt} + \beta_t)}{\Gamma(\sum_{t=1}^V g_{kt} + \beta_t)} \\
& \quad \times p(\eta | \mu, \Sigma) \prod_{(u, r_i, r_j) \in \mathcal{D}} e^{\frac{\boldsymbol{\eta}_{ur_{ij}}^\top \mathbf{z}_{ur_{ij}} - \delta_{ur_{ij}} (\boldsymbol{\eta}_{ur_{ij}}^\top \mathbf{z}_{ur_{ij}})^2}{2}} p(\delta_{ur_{ij}} | 1, 0)
\end{aligned}$$