Discovering Latent Influence in Online Social Activities via Shared Cascade Poisson Processes

Tomoharu Iwata
University of Cambridge
iwata.tomoharu@lab.ntt.co.jp

Amar Shah
University of Cambridge
as793@cam.ac.uk

Zoubin Ghahramani
University of Cambridge
zoubin@eng.cam.ac.uk

ABSTRACT

Many people share their activities with others through on-line communities. These shared activities have an impact on other users' activities. For example, users are likely to become interested in items that are adopted (e.g. liked, bought and shared) by their friends. In this paper, we propose a probabilistic model for discovering latent influence from sequences of item adoption events. An inhomogeneous Poisson process is used for modeling a sequence, in which adoption by a user triggers the subsequent adoption of the same item by other users. For modeling adoption of multiple items, we employ multiple inhomogeneous Poisson processes, which share parameters, such as influence for each user and relations between users. The proposed model can be used for finding influential users, discovering relations between users and predicting item popularity in the future. We present an efficient Bayesian inference procedure of the proposed model based on the stochastic EM algorithm. The effectiveness of the proposed model is demonstrated by using real data sets in a social bookmark sharing service.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data Mining*; I.2.6 [**Artificial Intelligence**]: Learning; I.5.1 [**Pattern Recognition**]: Model—*Statistical*

General Terms

Algorithms

Keywords

Poisson processes, social community, latent variable models, Bayesian inference

1. INTRODUCTION

Many people share their activities with others through on-line communities, such as social sharing, social networking,

and microblogging services. For example, in social sharing services, users share their favorite items, such as web pages [1], photographs [2], video [3], music [4] and scientific papers [5]. These shared activities have an impact on other users' activities. Users are likely to become interested in favorite items of their friends. A web page that is praised by an influential user in a social network service is likely to get popular.

We can obtain activity data from these online services. However, we cannot generally obtain data that directly describe which activity has had an impact on the following activities. The user interactions in some online communities have been referred to as "ballot box communication" [34], in which users do not directly interact with each other by exchanging messages, but they implicitly interact with one another by expressing their preferences on the shared items.

In this paper, we propose a probabilistic model for discovering latent influence between users in social communities. As activities, we consider adopting items by users over time, such as sharing web pages, voting for music and joining communities. The given data is multiple event sequences, where each event sequence is associated to an item, and contains user and timestamp information for each event of adopting the item. The proposed model can be used for finding influential users, discovering relations between users and predicting trends in item popularity in the future.

With the proposed model, we assume that each user has some latent degree of influence. Each adoption event can trigger the subsequent adoption events for the same item by other users. The possibility of triggering depends on the influence of the users of the preceding events. The proposed model also assumes that the influence differs among pairs of users. Friends can have an impact more than people who have never met. We use a cascade Poisson process [30] for modeling an event sequence of adopting an item. By using multiple cascade Poisson processes for modeling multiple event sequences and sharing their parameters, we can infer influence for each user and relations between users. We call the proposed model *shared cascade Poisson processes* (SCPP).

We present an efficient probabilistic inference procedure for the proposed model based on the stochastic EM algorithm. In the E-step, we sample the latent trigger for each event from the preceding events using collapsed Gibbs sampling while analytically integrating out parameters for influences and relations. By integrating out parameters, we can drastically reduce the memory requirement. In the M-step, we estimate the other model parameters by maximizing the

joint likelihood. By alternately iterating E- and M-steps, we infer the proposed model from the given data.

The paper is organized as follows: In Section 2, we outline related work. In Section 3, we propose the shared cascade Poisson processes for discovering latent influence from multiple sequences of item adoption events. We present its Bayesian inference procedure in Section 4, and discussion on scalability of the inference in Section 5. In Section 6, we describe a sampling procedure for generating data from the inferred model. In Section 7, we demonstrate the effectiveness of the proposed framework by using real data sets in a social bookmark sharing service. Finally, we present concluding remarks in Section 8.

2. RELATED WORK

Great interest is being shown in developing methods for analyzing online social communities. Those methods are used for viral marketing [12, 28, 21, 6], ranking influential users [31, 18] and personalized recommender systems [32, 11, 19, 7]. Learning degree of influence of users is useful for predicting future item adoption and recommendations [31, 11]. However, in many social communities, users may not necessarily declare explicit social ties among themselves even if such functionality exists and the degree of influence of each tie is rarely explicitly declared. Thus, estimating influences and relations among users from their activities are important.

Influences in social communities are often analyzed by using information diffusion models, such as independent cascade models [13, 35, 15] and linear threshold models [16]. They usually assume discrete time although events occur in continuous time. Some information diffusion models in continuous time have been proposed [17, 29]. The proposed model is based on Poisson processes, which are widely used for modeling events in continuous time. Poisson processes have been extensively studied in statistics, and have elegant theoretical properties, such as superposition and subsampling properties [22]. We can use these properties for developing a procedure of generating samples as described in Section 6, and for extending the proposed model. We develop an efficient Bayesian inference procedure, where parameters are analytically integrated out. With continuous time information diffusion models [29], influences are estimated given the network structure of users, where each edge represents the existence of influence. On the other hand, the proposed model can efficiently estimate influences without the network structure because influence parameters, which require memory of the order of the square of user size, integrated out in the Bayesian inference. In recent years, methods for estimating influences without the network structure based on maximum likelihood have been proposed [25, 14, 15].

The proposed model is an extension of the cascade Poisson process [30], which models a sequence of cascading events. We simultaneously model multiple sequences of adoption events for multiple items by sharing parameters, such as influences and relations among users. In [30], parameters are estimated based on maximum likelihood. We take a Bayesian approach that discourages overfitting during inference. Reciprocating Hawkes Processes [9], which are inhomogeneous Poisson processes, are used for discovering social structures from direct interaction time-series data, such as message exchanges. The proposed model does not use interaction data between users, but uses item adoption event

Table 1: Notation.

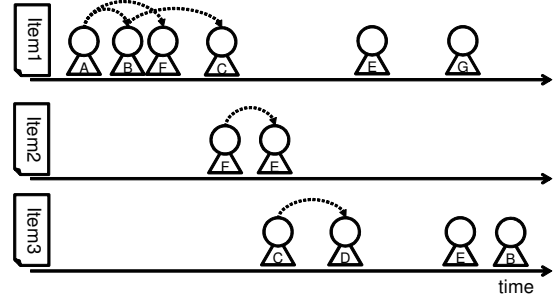| Symbol	Description
U	set of users
I	set of items
i	item, $i \in I$
u_{in}	user of the nth event in item i, $u_{in} \in U$
T	observation period
t_{in}	time of the nth event in item i, $t \in [0, T]$
N_i	number of events for item i in time $[0, T]$
α_u	influence of user u, $\alpha_u \geq 0$
α_i	popularity of item i, $\alpha_i \geq 0$
$\theta_{uu'}$	transition probability from user u to u', $\theta_{uu'} \geq 0$, $\sum_{u' \in U} \theta_{uu'} = 1$
γ	decay parameter, $\gamma \geq 0$

Figure 1: Example of the given data with three items and seven users. Latent triggers are shown by arrows between users.

data to model how users indirectly interact with each other and influence each other in social communities.

3. MODEL

Suppose that we have a set of $|I|$ event sequences $D = \{D_i\}_{i \in I}$, where I is a set of items. An event sequence for item i consists of N_i events, where an event is a pair of time and user, $D_i = \{(t_{in}, u_{in})\}_{n=1}^{N_i}$, which represents that user u_{in} adopts item i at time t_{in}. Without loss of generality, we can assume that events are sorted by their timestamps, $t_{in} \leq t_{in'}$ for $n < n'$. Our notation is summarized in Table 1. An example of the data is shown in Figure 1. Note that triggers that are shown by arrows between users are not observed.

In our model, an event can trigger further events in the future. This property is referred to as cascading, where the occurrence of an event sprouts the possibility of a chain of further events. For example, in Figure 1, adoption of Item-1 by User-A causes adoption of Item-1 by User-B and User-F. In particular, we use the following intensity function for modeling the rate that user u adopts item i at time t influenced by the event that user u' adopts item i at time t',

$$\lambda_{(i,t',u')}(t,u) = \begin{cases} \alpha_{u'} \theta_{u'u} \exp\left(-\gamma(t-t')\right) & \text{if } t' < t \\ 0 & \text{otherwise.} \end{cases}$$
(1)

Here, $\alpha_{u'} \geq 0$ represents the influence of user u' on the other users. Note that this user influence remains the same across

(a) Relation, $\boldsymbol{\Theta}$ (b) Decay, γ

(c) Influence, $\{\alpha_u\}_{u \in \boldsymbol{U}}$ (d) Popularity, $\{\alpha_i\}_{i \in \boldsymbol{I}}$
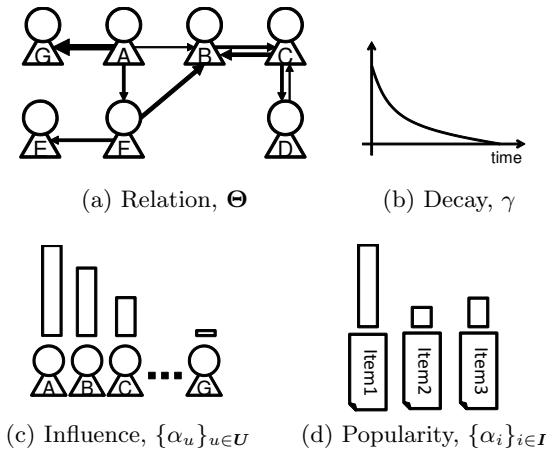
Figure 2: Parameters in the proposed shared cascade Poisson processes.

all items. The second factor $\theta_{u'u} \geq 0$ represents the strength of the relation from user u' to user u, where we assume $\sum_{u \in \boldsymbol{U}} \theta_{u'u} = 1$. This relation matrix is shared across items. The third factor $\exp(-\gamma(t-t'))$ models decay of the influence over time with parameter $\gamma \geq 0$. Since an event cannot have an effect on past events, the intensity is zero when $t' \geq t$. A decaying rate over time is a simple but natural assumption since, in general, trends evolve or are replaced over time. There is no influence from other event sequences, and events of item i do not trigger events of other items $i' \neq i$.

Events can occur without influence from preceding events according to the following background intensity function,

$$\lambda_{(i,0)}(t, u) = \alpha_i \theta_{0u}. \quad (2)$$

Here, $\alpha_i \geq 0$ represents the rate that item i is adopted without being triggered by the preceding events. It represents the general popularity of item i; if it is high, many users adopt the item without copying or influence from peers. $\theta_{0u} \geq 0$ represents the probability that a user who adopts an item without any trigger is in fact user u, and $\sum_{u \in \boldsymbol{U}} \theta_{0u} = 1$. In Figure 1, User-A, User-E and User-G adopt Item-1 independently from the preceding events. The first event for each item always occurs from the background as, by definition, there are no previous users to be influenced by. The background intensity can be seen as influence from outside the social community, such as newspapers, TV programs and online news sites as [26].

Figure 2 shows the parameters in the proposed shared cascade Poisson processes: (a) relations between users, $\boldsymbol{\Theta} = \{\boldsymbol{\theta}\}_{u \in \boldsymbol{U}_+}$, $\boldsymbol{\theta}_u = \{\theta_{uu'}\}_{u' \in \boldsymbol{U}}$, (b) parameter of influence decay over time, γ, (c) influence for each user $\{\alpha_u\}_{u \in \boldsymbol{U}}$, and (d) popularity for each item $\{\alpha_i\}_{i \in \boldsymbol{I}}$. Here, $\boldsymbol{U}_+ \equiv \boldsymbol{U} \cup \{0\}$ represents a set of users contain a virtual user who is used for the background intensity as well as actual users $\boldsymbol{U}$. Notice that influence can occur in both directions amongst a pair of users e.g. in Figure 2 (a) we see that User-B and User-C influence each other.

Let $\boldsymbol{D}_i^{(n)}$ be an event sequence of item i, which consists of events caused by the nth event, $\boldsymbol{D}_i^{(0)}$ be an event sequence without triggers. They are disjoint, $\boldsymbol{D}_i^{(n)} \cap \boldsymbol{D}_i^{(n')} = \emptyset$ for $n \neq n'$. We assume that each sequence is generated by a
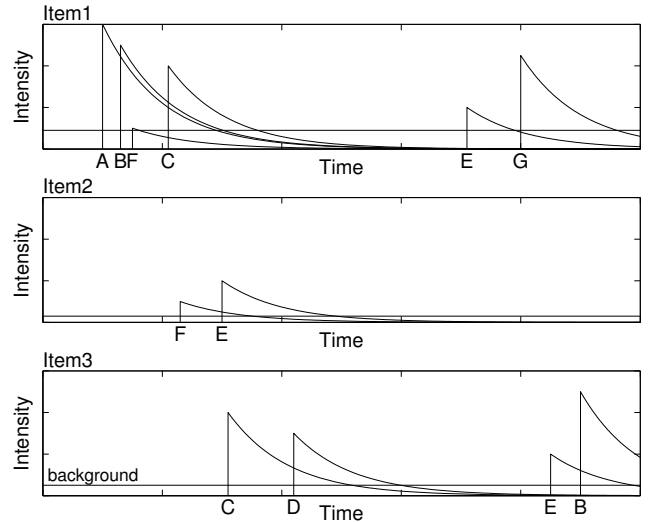

Figure 3: Dynamics of intensity in the shared cascade Poisson processes.

Poisson process as follows,

$$\boldsymbol{D}_i^{(0)} \sim \text{PoissonProcess}\left(\lambda_{(i,0)}(t, u)\right), \quad (3)$$

for events from the background intensity, and

$$\boldsymbol{D}_i^{(n)} \sim \text{PoissonProcess}\left(\lambda_{(i,t_{in},u_{in})}(t, u)\right), \quad (4)$$

for events triggered by the preceding event $n = 1, \cdots, N_i$.

The whole event sequence of item i is represented by their union, $\boldsymbol{D}_i = \bigcup_{n=0}^{N_i} \boldsymbol{D}_i^{(n)}$. Note that $\boldsymbol{D}_i$ is a union of Poisson processes. The superposition of Poisson processes is also a Poisson process (a proof can be found in [20]) and we use this fact to rewrite the generative process using a Poisson process,

$$\boldsymbol{D}_i \sim \text{PoissonProcess}\left(\lambda_i(t, u)\right), \quad (5)$$

where the recursive intensity function is given by

$$\lambda_i(t, u) = \lambda_{(i,0)}(t, u) + \sum_{(t',u') \in \boldsymbol{D}_{it}} \lambda_{(i,t',u')}(t, u), \quad (6)$$

where $\boldsymbol{D}_{it} = \{(t', u') | (t', u') \in \boldsymbol{D}_i \quad \text{and} \quad t' < t\}$ is the set of events before time t.

Figure 3 shows the dynamics of intensity for each item. The horizontal line in each item represents the background intensity, which is constant over time, but different depending on the item. At the time of an event occurrence, the intensity increases depending on the influence of its user. For example, since User-A has high influence, the event by User-A increases the intensity largely. The intensity decreases over time with parameter γ. The total intensity is the superposition of the background intensity and intensities from the preceding events.

The probability of multiple event sequences $\boldsymbol{D}$ given parameters $\boldsymbol{\alpha} = \{\alpha_m\}_{m \in \boldsymbol{U} \cup \boldsymbol{I}}$, $\boldsymbol{\Theta}$ and γ is described as follows,

$$p(\boldsymbol{D}|\boldsymbol{\alpha}, \boldsymbol{\Theta}, \gamma)$$
$$= \prod_{i=1}^{I} \exp\left(-\int_0^T \sum_{u \in \boldsymbol{U}} \lambda_i(t, u) dt\right) \prod_{n=1}^{N_i} \lambda_i(t_{in}, u_{in}), \quad (7)$$

where we can analytically calculate the integral as follows,

$$\int_0^T \sum_{u \in \boldsymbol{U}} \lambda_i(t, u) dt$$
$$= \alpha_i T + \frac{1}{\gamma} \sum_{n=1}^{N_i} \alpha_u \left(1 - \exp\left(-\gamma(T - t_{in})\right)\right). \quad (8)$$

This value is called the expectation function, which represents the expected number of occurrences in the time period from 0 to T.

4. INFERENCE

We present a Bayesian inference algorithm for the shared cascade Poisson processes based on a stochastic EM algorithm. The algorithm alternates collapsed Gibbs sampling of latent triggers and maximum joint likelihood estimation of hyperparameters, while integrating out parameters for influences, popularities and relations.

We introduce latent variables for simplifying the inference without changing the model. Let $z_{in} \in \{0, 1, \cdots, n' | t_{in'} < t_{in}\}$ be the latent index of the event which caused the arrival of the nth event in item i, where $z_{in} = 0$ indicates that the event was triggered by background intensity. The trigger event must occur before the arrival of this nth event, therefore $t_{in'} < t_{in}$, since cause precedes effect. By using latent variable z, the recursive intensity function (6) can be written as

$$\lambda_i(t, u) = \sum_z \lambda_i(t, u, z), \quad (9)$$

where

$$\lambda_i(t, u, z) = \begin{cases} \lambda_{(i,0)}(t, u) & \text{if } z = 0 \\ \lambda_{(i, t_{iz}, u_{iz})}(t, u) & \text{otherwise.} \end{cases} \quad (10)$$

In words, (9) states that the intensity is a sum of intensities for each possible trigger of the event.

Combining (7)–(10), the joint probability of data $\boldsymbol{D}$ and latent variables $\boldsymbol{Z} = \{\{z_{in}\}_{n=1}^{N_i}\}_{i \in \boldsymbol{I}}$ given parameters $\boldsymbol{\alpha}$, $\boldsymbol{\Theta}$ and γ is as follows,

$$p(\boldsymbol{D}, \boldsymbol{Z} | \boldsymbol{\alpha}, \boldsymbol{\Theta}, \gamma)$$
$$= \prod_{i=1}^{I} \exp\left(-\alpha_i T - \sum_{n=1}^{N_i} \alpha_u \frac{1}{\gamma}\left(1 - \exp(-\gamma(T - t_{in}))\right)\right)$$
$$\times \prod_{n=1}^{N_i} \lambda_i(t_{in}, u_{in}, z_{in}). \quad (11)$$

We assume a Gamma prior for each of user influences and item popularities α_m, $m \in \boldsymbol{U} \cup \boldsymbol{I}$,

$$p(\alpha_m | a, b) = \frac{1}{\Gamma(a)} b^a \alpha_m^{a-1} \exp(-b\alpha_m), \quad (12)$$

and a Dirichlet prior for relations $\boldsymbol{\theta}_u$, $u \in \boldsymbol{U}_+$,

$$p(\boldsymbol{\theta}_u | \beta) = \frac{\Gamma(\beta | \boldsymbol{U} |)}{\Gamma(\beta)^{|\boldsymbol{U}|}} \prod_{u' \in \boldsymbol{U}} \theta_{uu'}^{\beta-1}, \quad (13)$$

where $| \cdot |$ represents the number of elements in the set. The choice of a Gamma and Dirichlet prior permit analytic marginalization over the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\Theta}$, respectively,

and the joint likelihood is given by

$$p(\boldsymbol{D}, \boldsymbol{Z} | \gamma, \beta, a, b)$$
$$= \int \int p(\boldsymbol{D}, \boldsymbol{Z} | \boldsymbol{\alpha}, \boldsymbol{\Theta}, \gamma) p(\boldsymbol{\alpha} | a, b) p(\boldsymbol{\Theta} | \beta) d\boldsymbol{\alpha} d\boldsymbol{\Theta}$$
$$= \exp\left(-\gamma \sum_{i \in \boldsymbol{I}} \sum_{n: z_{in} \neq 0} (t_{in} - t_{iz_{in}})\right)$$
$$\times \left(\frac{b^a}{\Gamma(a)}\right)^{|\boldsymbol{U}| + |\boldsymbol{I}|} \prod_{m \in \boldsymbol{U} \cup \boldsymbol{I}} \frac{\Gamma(M_m + a)}{(C_m + b)^{M_m + a}}$$
$$\times \left(\frac{\Gamma(\beta | \boldsymbol{U} |)}{\Gamma(\beta)^{|\boldsymbol{U}|}}\right)^{|\boldsymbol{U}| + 1} \prod_{u \in \boldsymbol{U}_+} \frac{\prod_{u' \in \boldsymbol{U}} \Gamma(M_{uu'} + \beta)}{\Gamma(M_u + \beta | \boldsymbol{U} |)}, \quad (14)$$

where $M_{uu'} = \sum_{i \in \boldsymbol{I}} \sum_{n=1}^{N_i} \delta(u_{iz_{in}} = u) \delta(u_{in} = u')$ is the number of events for user u' caused by user u, $M_u = \sum_{u' \in \boldsymbol{U}} M_{uu'}$ is the number of total events caused by user u, and $M_i = \sum_{n=1}^{N_i} \delta(z_{in} = 0)$ is the number of events caused by the background intensity in item i. Here, $\delta(\cdot)$ represents indicator function; $\delta(A) = 1$ if A is true, and $\delta(A) = 0$ otherwise. For each item $i \in \boldsymbol{I}$, $C_i = T$, and for each user $u \in \boldsymbol{U}$,

$$C_u = \frac{1}{\gamma} \sum_{t \in \boldsymbol{D}_u} \left(1 - \exp\left(-\gamma(T - t)\right)\right), \quad (15)$$

where $\boldsymbol{D}_u$ is a set of timestamps of events of user u.

Based on the joint likelihood (14), we develop a stochastic EM procedure for the inference. In the E-step, given the current state of all but one latent assignment z_{in}, a new value for z_{in} is sampled from the following probability,

$$p(z_{in} = y | \boldsymbol{D}, \boldsymbol{Z}_{\backslash in}, \gamma, \beta, a, b)$$
$$\propto \frac{p(\boldsymbol{D}, \boldsymbol{Z}_{\backslash in}, z_{in} = y | \gamma, \beta, a, b)}{p(\boldsymbol{D}_{\backslash in}, \boldsymbol{Z}_{\backslash in} | \gamma, \beta, a, b)}$$
$$= \begin{cases} \frac{M_{i \backslash in} + a}{C_i + b} \frac{M_{0 u_{in} \backslash in} + \beta}{M_{0 \backslash in} + \beta | \boldsymbol{U} |} & \text{if } c = 0 \\ \exp\left(-\gamma(t_{in} - t_{iy})\right) \frac{M_{u_{iy} \backslash in} + a}{C_{u_{iy} \backslash in} + b} \frac{M_{u_{iy} u_{in} \backslash in} + \beta}{M_{u_{iy} \backslash in} + \beta | \boldsymbol{U} |} & \text{otherwise,} \end{cases}$$
$$\quad (16)$$

where $y \in \{0, 1, \cdots, n' | t_{in'} < t_{in}\}$, and subscript $\backslash in$ represents the value or set excluding the nth event in item i.

In the M-step, the decay parameter γ and Dirichlet parameter β are estimated by maximizing the logarithm of the joint likelihood (14). We estimate γ using Newton's method. The objective function to be maximized as a function of γ is given by

$$S(\gamma) = -\gamma \sum_{i \in \boldsymbol{I}} \sum_{n: z_{in} \neq 0} (t_{in} - t_{iz_{in}})$$
$$- \sum_{u \in \boldsymbol{U}} (M_u + a) \log(C_u + b). \quad (17)$$

The update rule is given by

$$\gamma \leftarrow \gamma - \frac{\partial S(\gamma) / \partial \gamma}{\partial^2 S(\gamma) / \partial \gamma^2}, \quad (18)$$

where the first and second derivatives of the joint likelihood with respect to γ are described in Appendix A.

We estimate the Dirichlet parameter β by using the fixed-point iteration method described in [24]. The update rule

is

$$\beta \leftarrow \beta \frac{\sum_{u \in \boldsymbol{U}_+} \sum_{u' \in \boldsymbol{U}} \left(\Psi(M_{uu'} + \beta) - \Psi(\beta) \right)}{\sum_{u \in \boldsymbol{U}_+} \left(\Psi(M_u + \beta|\boldsymbol{U}|) - \Psi(\beta|\boldsymbol{U}|) \right)}, \quad (19)$$

where $\Psi(\cdot)$ is the digamma function defined by $\Psi(x) = \frac{\partial \log \Gamma(x)}{\partial x}$.

By alternately iterating the E-step with (16) and M-step with (18)–(19), we can estimate latent variables $\boldsymbol{Z}$, decay parameter γ and Dirichlet parameter β. Note that because we are able to analytically integrate at all other parameters (influences α_u, item popularities α_i, and the relation matrix $\boldsymbol{\Theta}$), there is very little risk of overfitting. The posteriors for $\boldsymbol{\alpha}$ and $\boldsymbol{\Theta}$ are given by

$$p(\alpha_m|\boldsymbol{D}, \boldsymbol{Z}, a, b, \gamma) = \text{Gamma}(M_m + a, C_m + b), \quad (20)$$

$$p(\boldsymbol{\theta}_u|\boldsymbol{D}, \boldsymbol{Z}, \beta) = \text{Dirichlet}(M_{uu'} + \beta). \quad (21)$$

Thus, we can obtain point estimates of the integrated out parameters as follows,

$$\hat{\alpha}_m = \frac{M_m + a}{C_m + b}, \quad (22)$$

which can be used for finding influential users, and

$$\hat{\theta}_{uu'} = \frac{M_{uu'} + \beta}{M_u + \beta|\boldsymbol{U}|}, \quad (23)$$

which can be used for discovering relations between users. In the experiments, we fixed the hyperparameters as follows: $a = b = 1$.

5. SCALABILITY

The proposed model has relation parameters between users $\boldsymbol{\Theta}$, which require memory of $O(|\boldsymbol{U}|^2)$. The memory requirement increases quadratically with the number of users if we were to explicitly sample the relation parameters in our inference procedure. It makes analyzing data with a huge number of users prohibitively expensive. However, in the proposed inference procedure, we do not need to store these parameters as they are integrated out. Instead, we need to store count $M_{uu'}$ for $u \in \boldsymbol{U}_+$ and $u' \in \boldsymbol{U}$, which represents the number of events of user u' which were caused by user u. In general, most of the counts are zero since the count is incremented only when the two users adopt an item within a short time of one another. In addition, the Bayesian inference procedure exploits sparsity. When $M_{u'u} > M_{u''u}$, the probability of selecting the event of user u' is high, because the probability is proportional to $M_{u'u}$ in (16). The model has "the rich gets richer" effect. In the experiment using the 'science' data set with 4,543 users, which is described in Section 7, 59,590 elements of the influence parameters have nonzero value after the inference, which make up only 13.1 elements per user. Thus, by using the proposed inference, we can analyze data with a huge number of users efficiently.

The most demanding part of the inference is Gibbs sampling in the E-step. The computational complexity of each Gibbs sampling (16) is $O(R)$, where R is the average number of events for each item, since we have R options for each latent variable on average. In total, the computational complexity becomes $O(IR^2)$. The computational time for the inference with the 'science' data set, which consists of activities in four years, was 12 minutes when using a computer with 3.20GHz CPU.

Algorithm 1 Generate an event sequence from shared cascade Poisson Processes from time T to T' for item i.

Input: Parameters, $\boldsymbol{\alpha}, \boldsymbol{\Theta}, \gamma$, and data before T, $\boldsymbol{D} = \{(t_n, u_n)\}_{n=1}^N$
Output: Data from T to T', $\boldsymbol{D}' = \{(t_n, u_n)\}_{n>N}$

1: $\boldsymbol{D}' \leftarrow \emptyset$
2: $n \leftarrow N$
3: **loop**
4:      $s_n \sim \text{Exponential}(1)$
5:      $t_{n+1} \leftarrow \inf\{t| \int_T^t \sum_{u \in \boldsymbol{U}} \lambda_i(t', u)dt' \geq \sum_{n'=N}^n s_{n'}\}$
6:      $u_{n+1} \sim \text{Discrete}(\bar{\lambda}_i(t_{n+1}, u))$
7:      **if** $t_{n+1} > T'$ **then**
8:          **break**
9:      **end if**
10:      $n \leftarrow n + 1$
11:      $\boldsymbol{D}' \leftarrow \boldsymbol{D}' \cup \{(t_{n+1}, u_{n+1})\}$
12: **end loop**

---

6. SIMULATION

By generating sequences, we can simulate when users adopt items in the future. We can generate event sequences based on the shared cascade Poisson processes given parameters and the data up to the present time. Based on Monte Carlo method, we can predict informative values, such as the number of adoptions for each item, by averaging the simulation results. The simulation can be also used for targeting users who are influential and trigger the most events, which is useful for sociology and viral marketing [21, 6].

Algorithm 1 shows the procedure of generating an event sequence for item i from time T to T', which is based on Çinlar's method [10, 27]. The inputs are parameters, $\boldsymbol{\alpha}, \boldsymbol{\Theta}, \gamma$, and the data before T, $\boldsymbol{D} = \{(t_n, u_n)\}_{n=1}^N$. At line 4, an exponential random variable is sampled, which corresponds to a timestamp from a homogeneous Poisson process with rate one. At line 5, we obtain a timestamp of the shared cascade Poisson processes by inverting the expectation function $\int_T^{\cdot} \sum_{u \in \boldsymbol{U}} \lambda_i(t', u)dt'$. The solution is given by

$$t_{n+1} = E + \frac{1}{\gamma} W \left(-E \sum_{n'=1}^n \alpha_{u_n} \exp(\gamma t_n) \right), \quad (24)$$

where $W(\cdot)$ is the Lambert W function, which solves the equation $x = W(x) \exp(W(x))$, and

$$E = \frac{1}{\alpha_i} \left(\sum_{n'=N}^n s_{n'} + \alpha_i T - \frac{1}{\gamma} \sum_{n'=1}^n \alpha_{u_n} \exp(-\gamma(T - t_n)) \right). \quad (25)$$

At line 6, user u_{n+1} is sampled depending on the intensity at time t_{n+1}, where

$$\bar{\lambda}_i(t, u) = \frac{\lambda_i(t, u)}{\sum_{u'} \lambda_i(t, u')}, \quad (26)$$

is the intensity normalized to sum to one. This term represents the probability that the user adopting item i at time t is u given the previous history of user item adoptions. Algorithm 1 can be used to efficiently forecast future event sequences and simulate the effect of various social interventions.

Table 2: Data statistics			
tag	#items	#users	#events
ajax	35,231	9,554	563,605
javascript	47,256	11,258	779,845
java	60,386	6,645	474,325
linux	101,367	13,879	920,352
news	67,188	5,434	324,406
opensource	44,483	6,925	432,379
photography	71,962	11,872	614,143
science	45,986	4,543	268,231
webdesign	79,315	18,037	1,271,192

7. EXPERIMENTS

We evaluated the proposed shared cascade Poisson processes using large data sets collected from a social bookmark sharing service, Delicious [1]. Here, adopting an item corresponds to bookmarking a web page. The original data, described in [33], is distributed by the authors for research purpose. We used nine data sets obtained from the original data, where each data set consist of events of adopting items that are tagged with one of nine frequently used tags. The tags and their statistics are shown in Table 2. The sets of users and items are different depending on each data set. We omit users who appear less than 30 times in each data set. The data was collected between September 2003 to December 2007. Since the data only contain the date on which a certain item was adopted, the timestamp in all our experiments is measured in days.

We evaluated the predictive performance using the logarithm of the likelihood for timestamps at the test period from T to T', which is given by

$$L = \sum_{i \in \boldsymbol{I}} \Big(- \int_T^{T'} \sum_{u \in \boldsymbol{U}} \lambda_i(t,u) dt \Big) \sum_{t \in \boldsymbol{D}_i^{\text{test}}} \log \sum_{u \in \boldsymbol{U}} \lambda_i(t,u), \tag{27}$$

where $\boldsymbol{D}_i^{\text{test}}$ represents the test data for item i. For each data set, we prepared 10 different training and test data splits by changing the test time period from 10 to 100 days. Figure 4 shows the test likelihoods, which are averaged by the test period, $T' - T$. We compared the proposed shared cascade Poisson processes (SCPP) with the homogeneous Poisson processes (PP). The proposed model achieved higher predictive performance than the Poisson processes with all of the data sets except for the 'news' data set. In general, the test likelihood decreases as the test period increases, because long term prediction is more difficult than short term prediction. Many web pages in the 'news' data set are adopted by influence from portal or news sites, not by influence of other users, which may explain why the Poisson processes achieved better performance.

We also evaluated the predictive performance of users by using the logarithm of the likelihood for users at the test period as follows,

$$L' = \sum_{i \in \boldsymbol{I}} \sum_{(t,u) \in \boldsymbol{D}_i^{\text{test}}} \log \bar{\lambda}_i(t,u). \tag{28}$$

We compared the proposed model with latent Dirichlet allocation (LDA) [8], and multinomial distribution as the baseline. LDA is a widely used probabilistic model for discrete data, such as text documents, where each document consists
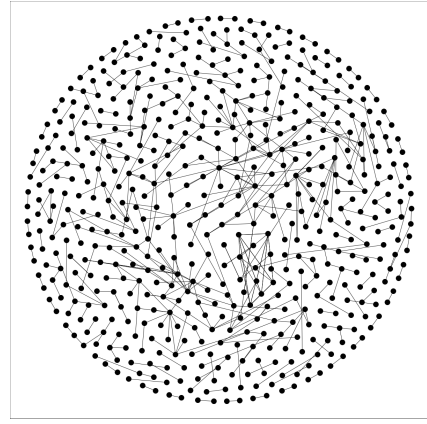

Figure 7: Relations between users estimated by the proposed model in the 'science' data set. Each node represents a user, and users who are influenced are connected. We omit isolated users who are not connected with others.

of a set of words. In our data sets, a word corresponds to a user, and a document corresponds to an event sequence of an item. Figure 5 shows the result. The proposed model achieved higher user predictive performance than the other methods with all of the data sets except for the 'news' data set. These results indicate that modeling influences and relations is important for analyzing social activity data sets.

Figure 6 shows the event sequences in the 'science' data set with the estimated results of influence between events. Some events occur in a short time interval, and they were learned to have been triggered by the precedent events in the sequence. Events which occurred in isolation were estimated as being generated from the background intensity.

Figure 7 shows relations between users estimated by the proposed model in the 'science' data set. The relation graph is visualized using the method in [23] so that connected nodes are located closely. Most users have a small number of edges, but a few users have a lot of edges; these would be the main influencers or most influenced depending on the direction. Some users form clusters, where users in each cluster are connected to each other, representing communities of influence. Overall the influence graph for this data set is quite sparse.

Figure 8 shows the values of influences and popularities estimated by the proposed model in the 'science' data set, where user and item indices were sorted by their values. Only a small number of users and items have high influence or popularity, whilst most have very small values.

8. CONCLUSION

In this paper, we have proposed a probabilistic model for discovering latent influence in social activities. Our model is based on multiple inhomogeneous Poisson processes, in which events are triggered by preceding events, and parameters are shared among the Poisson processes. We also presented its efficient Bayesian inference procedure based on the stochastic EM algorithm. We have confirmed experimentally that the proposed model can appropriately model sequences of item adoption events in social bookmark sharing data sets.
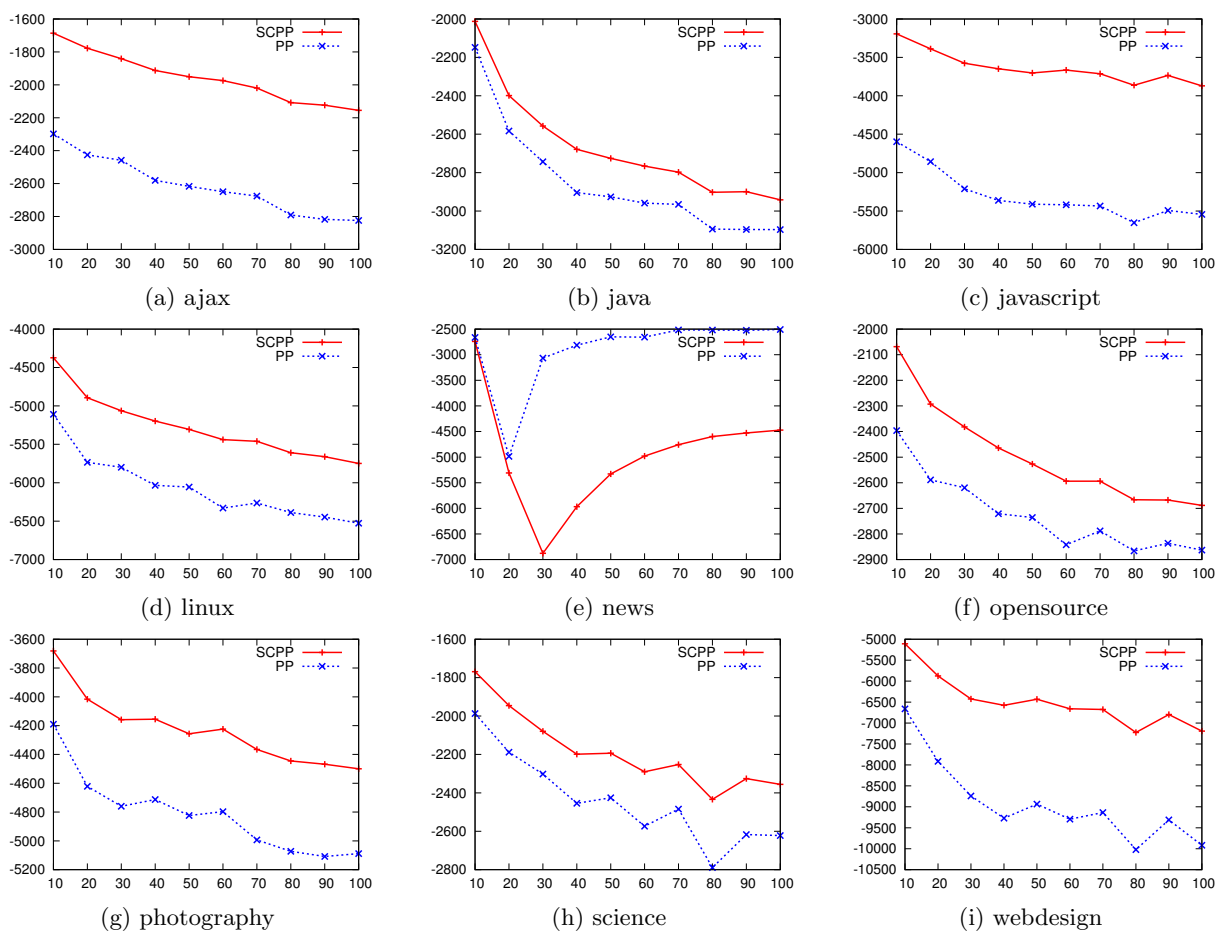
Figure 4: Average test likelihood for timestamps with different test periods in nine social bookmark data sets. The vertical axis is the test likelihood, and the horizontal is the test time period.

9. REFERENCES

[1] http://www.delicious.com/.

[2] http://http://www.flickr.com/.

[3] http://www.youtube.com/.

[4] http://www.last.fm/.

[5] http://www.citeulike.org/.

[6] N. Agarwal and H. Liu. Blogosphere: research issues, tools, and applications. *ACM SIGKDD Explorations Newsletter*, 10(1):18–31, 2008.

[7] C. Au Yeung and T. Iwata. Capturing implicit user influence in online social sharing. In *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, pages 245–254. ACM, 2010.

[8] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[9] C. Blundell, K. Heller, and J. Beck. Modelling reciprocating relationships with Hawkes processes. In *Advances in Neural Information Processing Systems 25*, pages 2609–2617, 2012.

[10] E. Çinlar. *Introduction to stochastic processes*, volume 3. Prentice-Hall Englewood Cliffs, NJ, 1975.

[11] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri. Feedback effects between similarity and social influence in online communities. In *KDD '08*, pages 160–168, 2008.

[12] P. Domingos and M. Richardson. Mining the network value of customers. In *KDD '01*, pages 57–66, 2001.

[13] J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters*, 12(3):211–223, 2001.

[14] M. Gomez-Rodriguez, D. Balduzzi, and B. Schölkopf. Uncovering the temporal dynamics of diffusion networks. In *Proceedings of the 29th International Conference on Machine Learning*, 2011.

[15] M. Gomez-Rodriguez, J. Leskovec, and B. Schölkopf. Structure and dynamics of information pathways in online media. In *Sixth ACM International Conference on Web Search and Data Mining (WSDM)*, pages 23–32, 2013.

[16] M. Granovetter. Threshold models of collective behavior. *American journal of sociology*, pages 1420–1443, 1978.

[17] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *Proceedings of the 13th international conference on World Wide Web*, WWW '04, pages 491–501, 2004.
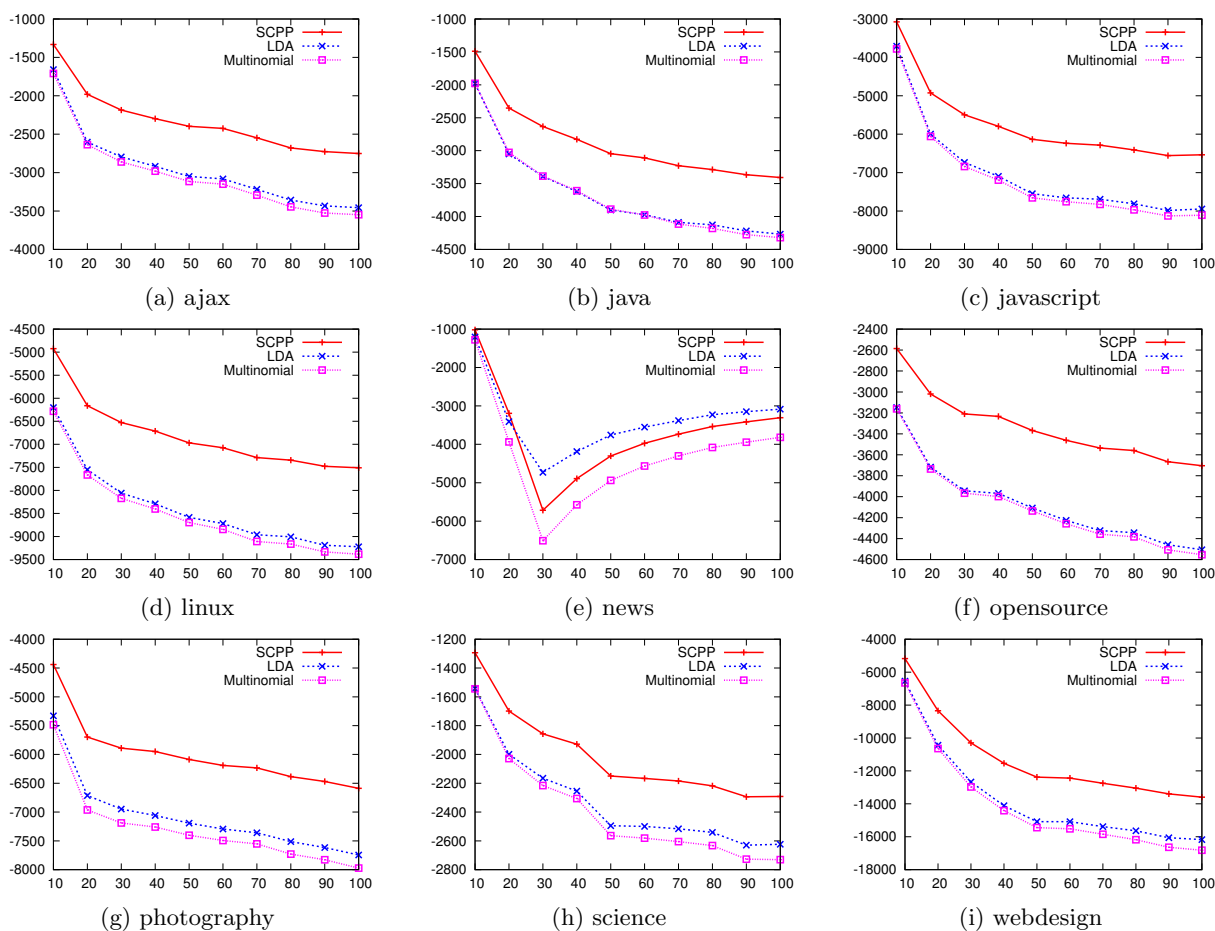
Figure 5: Average test likelihood for users with different test periods in nine social bookmark data sets. The vertical axis is the test likelihood, and the horizontal is the test time period.
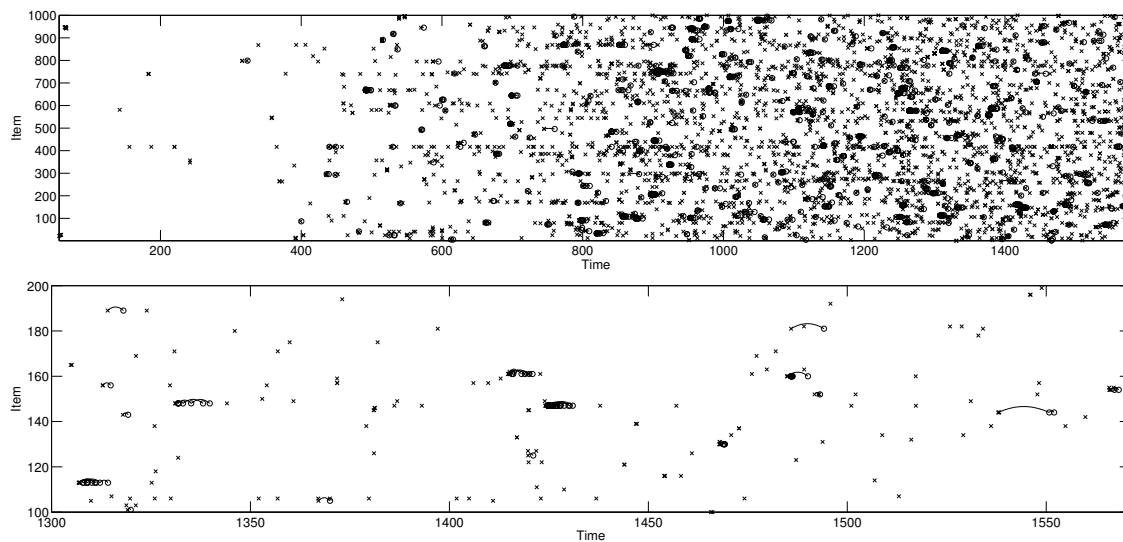

Figure 6: Event sequences in the 'science' data set with the estimated results of influence between events. (a) item indices 1 to 1000 and time 100 to 1559, and (b) its zoom of item indices 100 to 200 and time 1300 to 1559. Here, 'x' represents an event that is caused by the background intensity, 'o' represents an event that is triggered by the preceding event, and in that case the triggered event is indicated by a curved line connected between them.
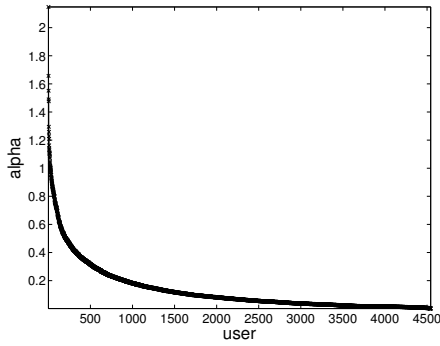
(a) Influence $\{\alpha_u\}_{u \in U}$
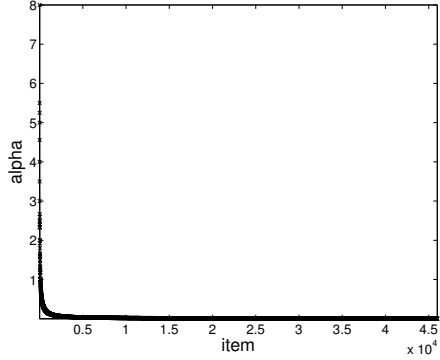
(b) Popularity $\{\alpha_i\}_{i \in I}$

Figure 8: Estimated user influence (a), and item popularity (b) in the 'science' data set. User and item indices are sorted by their values.

[18] M. Katagiri and M. Etoh. Social influence modeling on smartphone usage. *Advanced Data Mining and Applications*, pages 292–303, 2011.

[19] N. Kawamae, H. Sakano, and T. Yamada. Personalized recommendation based on the personal innovator degree. In *Proceedings of the 3rd ACM Conference on Recommender systems*, pages 329–332. ACM, 2009.

[20] J. F. C. Kingman. *Poisson processes*, volume 3. Oxford University Press, USA, 1993.

[21] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1):5, 2007.

[22] D. Lin, E. Grimson, and J. Fisher. Construction of dependent Dirichlet processes based on Poisson processes. In *Advances in Neural Information Processing Systems*, 2010.

[23] T. Matsubayashi and T. Yamada. A force-directed graph drawing based on the hierarchical individual timestep method. *International Journal of Electronics, Circuits and Systems*, 1(2):116–121, 2007.

[24] T. Minka. Estimating a Dirichlet distribution, 2000.

[25] S. A. Myers and J. Leskovec. On the convexity of latent social network inference. In *Advances in Neural Information Processing Systems*, 2010.

[26] S. A. Myers, C. Zhu, and J. Leskovec. Information diffusion and external influence in networks. In *KDD '12*, pages 33–41, 2012.

[27] R. Pasupathy. Generating Poisson arrival streams. *Wiley Encyclopedia of Operations Research and Management Science*, 2009.

[28] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *KDD '02*, pages 61–70, 2002.

[29] K. Saito, M. Kimura, K. Ohara, and H. Motoda. Learning continuous-time information diffusion model for social behavioral data analysis. *Advances in Machine Learning*, pages 322–337, 2009.

[30] A. Simma and M. I. Jordan. Modeling events with cascades of Poisson processes. In *UAI: Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*, pages 546–555, 2010.

[31] X. Song, Y. Chi, K. Hino, and B. L. Tseng. Information flow modeling based on diffusion rate for prediction and ranking. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 191–200, 2007.

[32] X. Song, B. L. Tseng, C.-Y. Lin, and M.-T. Sun. Personalized recommendation driven by information flow. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '06, pages 509–516, New York, NY, USA, 2006. ACM.

[33] R. Wetzker, C. Zimmermann, and C. Bauckhage. Analyzing social bookmarking systems: A del. icio. us cookbook. In *Proceedings of the ECAI 2008 Mining Social Data Workshop*, pages 26–30, 2008.

[34] M. Xia, Y. Huang, W. Duan, and A. B. Whinston. Ballot box communication in online communities. *Commun. ACM*, 52(9):138–142, Sept. 2009.

[35] J. Yang and J. Leskovec. Modeling information diffusion in implicit networks. In *IEEE 10th International Conference on Data Mining (ICDM)*, pages 599–608. IEEE, 2010.

APPENDIX

A. ESTIMATING DECAY PARAMETER

This appendix describes the first and second derivatives of the joint likelihood (14) with regard to decay parameter γ, which are required for estimating the decay parameter based on Newton's method. The first derivative of $S(\gamma)$ is given by

$$\frac{\partial S(\gamma)}{\partial \gamma} = -\sum_{i \in I} \sum_{n:z_{in} \neq 0} (t_{in} - t_{iz_{in}}) - \sum_{u \in U} \frac{M_u + a}{A_u + \gamma b} \left(-\frac{A_u}{\gamma} + A'_u \right), \quad (29)$$

and the second derivative of S is given by

$$\frac{\partial^2 S(\gamma)}{\partial \gamma^2} = -\sum_{u \in U} \frac{M_u + a}{A_u + \gamma b} \left(A''_u + \left(-\frac{A_u}{\gamma} + A'_u \right) \times \left(\frac{A'_u + b}{A_u + \gamma b} + \frac{1}{\gamma} \right) \right), \quad (30)$$

where $A_u = \gamma C_u$, $A'_u = -\sum_{t \in D_u} \exp(-\gamma(T-t))(T-t)$, and $A''_u = \sum_{t \in D_u} \exp(-\gamma(T-t))(T-t)^2$.