

Detecting Changes in Information Diffusion Pattern over Social Network

KAZUMI SAITO, University of Shizuoka
MASAHIRO KIMURA, Ryukoku University
KOUZOU OHARA, Aoyama Gakuin University
HIROSHI MOTODA, Osaka University

We addressed the problem of detecting the change in behavior of information diffusion over a social network which is caused by an unknown external situation change using a small amount of observation data in a retrospective setting. The unknown change is assumed to be effectively reflected in changes in the parameter values in the probabilistic information diffusion model, and the problem is reduced to detecting where in time and how long this change persisted and how big this change is. We solved this problem by searching the change pattern that maximizes the likelihood of generating the observed information diffusion sequences, and in doing so we devised a very efficient general iterative search algorithm using the derivative of the likelihood which avoids parameter value optimization during each search step. This is in contrast to the naive learning algorithm in that it has to iteratively update the pattern boundaries, each requiring the parameter value optimization and thus is very inefficient. We tested this algorithm for two instances of the probabilistic information diffusion model which has different characteristics. One is of information push style and the other is of information pull style. We chose asynchronous independent cascade (AsIC) model as the former and value-weighted voter (VwV) model as the latter. The AsIC is the model for general information diffusion with binary states and the parameter to detect its change is diffusion probability and the VwV is the model for opinion formation with multiple states and the parameter to detect its change is opinion value. The results tested on these two models using four real world network structures confirmed that the algorithm is robust enough and can efficiently identify the correct change pattern of the parameter values. Comparison with the naive method that finds the best combination of change boundaries by an exhaustive search through a set of randomly selected boundary candidates showed that the proposed algorithm far outperforms the native method both in terms of accuracy and computation time.

Categories and Subject Descriptors: G.2.2 [**Discrete Mathematics**]: Graph Theory—*Network problems*; H.2.8 [**Database Management**]: Database Applications—*Data mining*; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*sociology*

General Terms: Algorithms

Additional Key Words and Phrases: Change point detection, information diffusion, parameter learning, social networks

ACM Reference Format:

Saito, K., Kimura, M., Ohara, K., and Motoda, H. 2011. Detecting changes in information diffusion pattern over social network. *ACM Trans. Intell. Syst. Technol.* V, N, Article A (January YYYY), 23 pages. DOI = 10.1145/0000000.0000000 <http://doi.acm.org/10.1145/0000000.0000000>

This work was partly supported by Asian Office of Aerospace Research and Development, Air Force Office of Scientific Research under Grant No. AOARD-11-4111, and JSPS Grant-in-Aid for Scientific Research (C) (No. 23500194). Author's addresses: K. Saito, School of Administration and Informatics, University of Shizuoka; M. Kimura, Department of Electronics and Informatics, Ryukoku University; K. Ohara, Department of Integrated Information Technology, Aoyama Gakuin University; H. Motoda, Institute of Scientific and Industrial Research, Osaka University.

Permission to make digital or hard copies of part or all 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 show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© YYYY ACM 0000-0003/YYYY/01-ARTA \$10.00

DOI 10.1145/0000000.0000000 <http://doi.acm.org/10.1145/0000000.0000000>

1. INTRODUCTION

Recent technological innovation in the web such as blogosphere and knowledge/media-sharing sites is remarkable, which has made it possible to form various kinds of large social networks, through which behaviors, ideas, rumors and opinions can spread, and our behavioral patterns are to a considerable degree affected by the interaction with these networks and substantial attention has been directed to investigating the spread of information in these networks [Newman et al. 2002; Newman 2003; Gruhl et al. 2004; Domingos 2005; Leskovec et al. 2006; Crandall et al. 2008; Wu and Huberman 2008].

These studies have shown that it is important to consider the diffusion mechanism explicitly and the measures based on network structure alone, *i.e.*, various centrality measure, are not enough to identify the important nodes [Kimura 2009; 2010a]. Information diffusion is modeled typically by probabilistic models. Most representative and fundamental ones for general information diffusion are independent cascade (IC) model [Goldenberg et al. 2001; Kempe et al. 2003], linear threshold (LT) model [Watts 2002; Watts and Dodds 2007] and their extensions that include incorporating asynchronous time delay [Saito et al. 2009b; 2010a]. The IC model is a model of information push style, *i.e.*, the information sender (a node) tries to push the information to the neighboring receivers (child nodes) in a probabilistic way. The LT model is a model of information pull style, *i.e.*, the information receiver (a node) tries to pull the information from the neighboring senders (parents nodes) in a probabilistic way. Since the focus of study is “influence”, these models assume binary states, *i.e.*, nodes are either active (influenced) or inactive (uninfluenced). Explicit use of these models to solve such problems as the *influence maximization problem* [Kempe et al. 2003; Kimura et al. 2010a; Chen et al. 2010a; 2010b] and the *contamination minimization problem* [Kimura et al. 2009] clearly shows the advantage of the model. The identified influential nodes and links are considerably different from the ones identified by the centrality measures. Another type of information diffusion model that is also often used is voter model [Even-Dar and Shapria 2007] and its extensions that include incorporating opinion values [Kimura et al. 2010b], node strength [Yamagishi et al. 2011] and anti-majoritarian tendency [Kimura et al. 2011]. The voter model is a model of information pull style and is used to study the spread of opinions, *i.e.*, opinion formation. It is similar to the LT model in that the opinion of a person is affected by the opinions of his/her neighbors. What is different from the LT model is that it has to have multiple states if it has to deal with multiple opinions¹. This notion is not necessarily limited to opinion. Application such as an on-line competitive service in which a user can choose one from multiple choices/decisions requires a model that handles multiple states. There has been a variety of work on the voter model, too. Dynamical properties of the basic model have been extensively studied including how the degree distribution and the network size affect the mean time to reach consensus from mathematical point of view [Liggett 1999; Sood and Redner 2005]. Several variants of the voter model are also investigated and non equilibrium phase transition is analyzed [Castellano et al. 2009; Yang et al. 2009] from physics point of view. Yet another line of work extends the voter model by combining it with a network evolution model [Holme and Newman 2006; Crandall et al. 2008]. Kimura et al. [2010b] analyzed how the opinion values affect the opinion share dynamics in their recent study.

What is common to all the above models is that they are all probabilistic models and have parameters to characterize the information diffusion. The parameters must be known in advance for the model to be usable for analysis. It is generally difficult

¹The basic voter model has only two opinions but it is straightforward to extend it to handle multiple opinions.

to determine the values of these parameters theoretically, and thus, attempts have been made to learn these parameter values by observing the information diffusion sequence data [Saito et al. 2009a; 2009b; 2010a; 2010b; Gomez-Rodriguez et al. 2010; Myers and Leskovec 2010; Kimura et al. 2010b]. In essence the likelihood of generating the observed data by the model employed is first derived, and then the parameter values are determined such that the likelihood is maximized. In particular, Myers and Leskovec [2010] showed that for a certain class of diffusion models, the problem can effectively be transformed to a convex programming for which a global solution is guaranteed. Another important common assumption made in these studies is that the model is stationary. Since the model is probabilistic, even if the model is stationary, the way information propagates from a particular node is not the same (not deterministic) and each time the diffusion result is different. However, the model parameter values remain the same during the whole course of analysis.

This paper addresses a different aspect of information diffusion, and extends and integrates our recent studies [Saito et al. 2011a; Ohara et al. 2011]. We note that our behavior is affected not only by the behavior of our neighbors but also by other external factors. The model only accounts for the interaction with neighbors. The behavior we observe includes both effects. The problem we address here is to detect the change in the model from a limited amount of observed information diffusion data. If this is possible, this would bring a substantial advantage. For example, we can infer that something unusual happened during a particular period of time by simply analyzing the limited amount of data.

This is in some sense the same, in the spirit, with the work by Kleinberg [2002] and Swan and Allan [2000]. They noted a huge volume of the data stream, tried to organize it and extract structures behind it. This is done in a retrospective framework, *i.e.*, assuming that there is a flood of abundant data already and there is a strong need to understand it. Kleinberg's work is motivated by the fact that the appearance of a topic in a document stream is signaled by a "burst of activity" and identifying its nested structure manifests itself as summarization of the activities over a period of time, making it possible to analyze the underlying content much easier. He used a hidden Markov model in which bursts appear naturally as state transitions, and successfully identified the hierarchical structure of e-mail messages. Swan and Allan's work is motivated by the need to organize huge amount of information in an efficient way. They used a statistical model of feature occurrence over time based on hypotheses testing and successfully generated clusters of named entities and noun phrases that capture the information corresponding to major topics in the corpus, and designed a way to nicely display the summary on the screen (Overview Timelines). Our aim is not exactly the same as theirs. We are interested in detecting changes in the external factors which are hidden/embedded in the data. We also follow the same retrospective approach, *i.e.*, we are not predicting the future, but we are trying to understand the phenomena that happened in the past. There are many factors that bring in changes and evidently the model cannot accommodate all of them. We formalize this as the unknown changes in the parameter value of the diffusion model we employ, and we reduce the problem to that of detecting where in time and how long this change persisted and how big this change is. We call the period where the parameter takes anomalous values as "hot span" and the rest as "normal span".

We have chosen the asynchronous independent cascade (AsIC) model [Saito et al. 2009b; 2010a] as the one that represents the model of information push style, and the value-weighted voter (VwV) model [Kimura et al. 2010b] as the one that represents the

model of information pull style². As explained above, the AsIC is the model for general information diffusion with binary states and the parameter to detect its change is diffusion probability and the VwV is the model for opinion formation with multiple states and the parameter to detect its change is opinion value. These two models are recalled in Section 2. We generalized the parameter optimization algorithm that was first introduced in [Saito et al. 2011a; Ohara et al. 2011] so that it can cover both the models as two different instances and expanded the experiments to verify that the same algorithm works satisfactorily for two different types of information diffusion models. As in our previous work, we limit the form of change to a rect-linear one, that is, the parameter value changes to a new large value, persists for a certain period of time and is restored to the original value and stays the same thereafter³. In this simplified setting, detecting the hot span is equivalent to identifying the time window where the parameter value is anomalous and estimating the parameter values both in the hot and the normal spans.

We use the same parameter optimization algorithm as in [Saito et al. 2009b; Kimura et al. 2010b], *i.e.*, the EM-like algorithm for the AsIC model that iteratively updates the values to maximize the model's likelihood of generating the observed data sequences, and the Newton method for the VwV model that guarantees globally maximizing the likelihood. However, the problem here is more difficult because it has another loop to search for the hot span on top of the above loop. The naive learning algorithm has to iteratively update the pattern boundaries (outer loop) and the value must also be optimized for each combination of the pattern boundaries (inner loop), which is extraordinary inefficient. Our main contribution is that we devised a very efficient general search algorithm which works for probabilistic information diffusion models and avoids the inner loop optimization by using the information of the first order derivative of the likelihood with respect to the parameters. We tested its performance using the structures of four real world networks (Blog, Coauthorship, Enron and Wikipedia), and confirmed that the algorithm can efficiently identify the hot span correctly as well as the parameter values of both the normal and the hot spans. We further compared our algorithm with the naive method that finds the best combination of the hot span boundaries by an exhaustive search from a set of randomly selected boundary candidates, and showed that the proposed algorithm far outperforms the naive method both in terms of accuracy and computation time.

The paper is organized as follows. After very briefly introducing the two diffusion models, AsIC and VwV in Section 2, we define the problem in Section 3 and recall how the parameters can be learned in each model in Section 4. The main part is Section 5 where we explain how we efficiently search for the hot span as well as the parameter values. The results are explained in Section 6, followed by discussion in Section 7. We end this paper by summarizing the main result in Section 8.

2. INFORMATION DIFFUSION MODELS

We focus on two types of information diffusion model on a social network $G = (V, E)$, where V and $E (\subset V \times V)$ are the sets of all the nodes and the links, respectively. One is the asynchronous independent cascade (AsIC) model that is an extension of the independent cascade (IC) model, and the other is the value-weighted voter (VwV) model that is an extension of the standard voter model. They were extended to meet more realistic situations. We recall their definitions below.

²We could have chosen AsLT instead of VwV. There is no specific reason that we cannot handle AsLT. Our aim is to show that our approach is general enough and applicable to a wide variety of diffusion models.

³We discuss that the basic algorithm can be extended to more general change patterns in Section 7, and show that it works for two distinct rect-linear patterns in case of AsIC.

2.1. Asynchronous Independent Cascade (AsIC) Model

The AsIC model we use in this paper incorporates asynchronous time delay into the IC model which does not account for time-delay, noting that each node changes its state asynchronously in reality [Saito et al. 2009b; 2010a]. Here, we consider choosing a delay-time from the exponential distribution for the sake of convenience, but of course other distributions such as power-law and Weibull can be employed.

For the AsIC model, the underlying network $G = (V, E)$ is a directed graph. For any $v \in V$, the set of all the nodes that have links from v (child nodes) is denoted by

$$F(v) = \{u \in V; (v, u) \in E\},$$

and the set of all the nodes that have links to v (parent nodes) is denoted by

$$B(v) = \{u \in V; (u, v) \in E\}.$$

Each node has one of the two states (active and inactive), and the nodes are called *active* if they have been influenced. It is assumed that nodes can switch their states only from inactive to active.

The AsIC model has two types of parameters $p_{u,v}$ and $r_{u,v}$ with $0 < p_{u,v} < 1$ and $r_{u,v} > 0$, where $p_{u,v}$ and $r_{u,v}$ are referred to as the *diffusion probability* through link (u, v) and the *time-delay parameter* through link (u, v) , respectively. We define the *diffusion-probability vector* \mathbf{p} and the *time-delay parameter vector* \mathbf{r} by

$$\mathbf{p} = (p_{u,v})_{(u,v) \in E}, \quad \mathbf{r} = (r_{u,v})_{(u,v) \in E}.$$

The information diffusion process unfolds in continuous-time t , and proceeds from a given initial active node in the following way. When a node u becomes active at time t , it is given a single chance to activate each currently inactive node $v \in F(u)$. A delay-time δ is chosen from the exponential distribution with parameter $r_{u,v}$. The node u attempts to activate the node v if v has not been activated by time $t + \delta$, and succeeds with probability $p_{u,v}$. If u succeeds, v will become active at time $t + \delta$. The information diffusion process terminates if no more activations are possible.

2.2. Value weighted Voter (VwV) Model

The mathematical model we use for the diffusion of opinions is the VwV model with $K (\geq 2)$ opinions [Kimura et al. 2010b]. For the VwV model, the underlying network $G = (V, E)$ is an undirected (bidirectional) graph with self-loops. For a node $v \in V$, let $\Gamma(v)$ denote the set of neighbors of v in G , that is,

$$\Gamma(v) = \{u \in V; (u, v) \in E\}.$$

Note that $v \in \Gamma(v)$ because of the existence of self-loops.

In the VwV model, each node of G is endowed with $(K + 1)$ states; opinions $1, \dots, K$, and *neutral* (i.e., no-opinion state). It is assumed that a node never switches its state from any opinion k back to neutral. The model has a parameter $w_k (> 0)$ for each opinion k , which is called the *opinion value* and must be estimated from observed opinion diffusion data. We define the *opinion-value vector* \mathbf{w} by

$$\mathbf{w} = (w_1, \dots, w_K).$$

Let $f_t : V \rightarrow \{0, 1, 2, \dots, K\}$ denote the opinion distribution at time t , where $f_t(v)$ stands for the opinion of node v at time t , and opinion 0 denotes the neutral state. We also denote by $n_k(t, v)$ the number of v 's neighbors that hold opinion k as the latest one before time t for $k = 1, 2, \dots, K$, i.e.,

$$n_k(t, v) = |\{u \in \Gamma(v); \phi_t(u) = k\}|,$$

where $\phi_t(u)$ is the latest opinion of u before time t .

Given a target time T , and an initial state in which each opinion is assigned to only one distinct node and all other nodes are in the neutral state, the evolution process of the model unfolds in the following way. At time 0, each node v independently decides its update time t according to some probability distribution such as an exponential distribution with parameter r_v , where r_v becomes also a model parameter and then we define the time-delay parameter vector \mathbf{r} by $\mathbf{r} = (r_v)_{v \in V}$. The successive update time is determined similarly at each update time t . Node v changes its opinion at its update time t as follows: If node v has at least one neighbor with some opinion before time t , $f_t(v) = k$ with probability $w_k n_k(t, v) / \sum_{k'=1}^K w_{k'} n_{k'}(t, v)$ for $k = 1, \dots, K$, otherwise, $f_t(v) = 0$ with probability 1. It is noted that since node v is included in its neighbors by definition, its own opinion is also reflected. The process is repeated from the initial time $t = 0$ until the next update-time attains a given final-time T .

3. PROBLEM DEFINITION

We address the *hot span detection problem*. In this problem, we assume that some change has happened in the way the information diffuses, and we observe the diffusion sequences of a certain topic in which the change is embedded, and consider detecting where in time and how long this change persisted and how big this change is. In the following subsections, we describe a specific detection problem by focusing on the above diffusion models, *i.e.*, the AsIC model and the VwV model.

3.1. AsIC Model

An information diffusion result generated by the AsIC model is represented as a set of pairs of active nodes and their activation times; *i.e.*, $\{(u, t_u), (v, t_v), \dots\}$. We consider a diffusion result $\mathcal{D}(0, T)$, where the initial activation time is set to 0 and the final observation time is denoted by T . Since we employ only a single diffusion result $\mathcal{D}(0, T)$, we place a constraint that $p_{u,v}$ and $r_{u,v}$ do not depend on link (u, v) , *i.e.*, $p_{u,v} = p$, $r_{u,v} = r$ ($\forall (u, v) \in E$), which should be acceptable noting that we can naturally assume that people behave quite similarly when talking about the same topic (see Section 7).

Let $[T_1, T_2)$ denote the hot span of the information diffusion, and let p_n and p_h denote the diffusion probability for the normal span and the hot span, respectively. Namely, the diffusion probability p is obtained by $p = p_n$ for the period $[0, T_1)$, $p = p_h$ for the period $[T_1, T_2)$, and $p = p_n$ for the period $[T_2, T)$. Here we assume for simplicity that the time-delay parameter r does not change and takes the same value for the entire period $[0, T)$. Then, the hot span detection problem is reduced to detecting the hot span $[T_1, T_2)$ and estimating p_n and p_h from the observed diffusion result $\mathcal{D}(0, T)$.

Figure 1 shows five examples of diffusion sample with (Fig. 1b) and without (Fig. 1a) a hot span based on the AsIC model, where the parameters are set at $p_n = 0.1$, $p_h = 0.3$, $r = 1.0$, $T_1 = 10$, $T_2 = 20$. The network used is the blog network described later in Subsection 6.1. We plotted the ratio of active nodes (the number of nodes activated at time step t divided by the number of total active nodes over the whole time span) for five independent simulations, each from a randomly chosen initial source node at time $t = 0$. Comparing these two figures, we can clearly see bursty activities around the hot span $[10, 20)$ in Fig. 1b. However, each curve in Fig. 1b behaves differently, *i.e.*, some has its bursty activities only in the first half, some other has them only in the last half, and yet some other has two peaks during the hot span. This means that it is quite difficult to accurately detect the true hot span from only a single diffusion sample. Methods that use only the observed bursty activities, including those proposed by Swan and Allan [2000] and Kleinberg [2002] would not work.

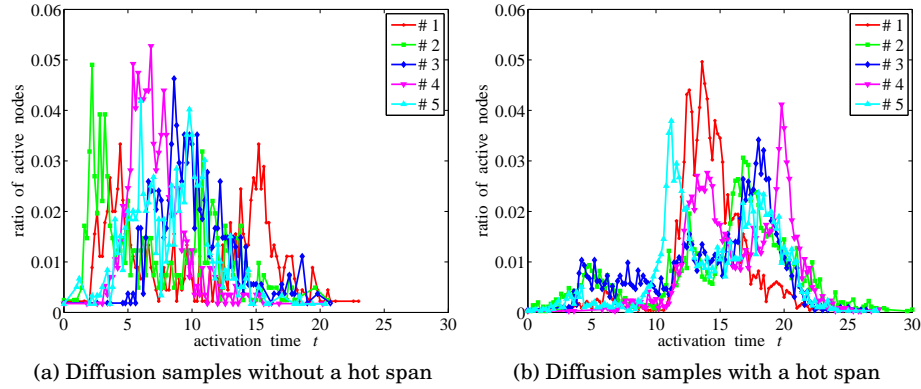


Fig. 1: Information diffusion in the blog network for the AsIC model. Results of five independent runs are shown.

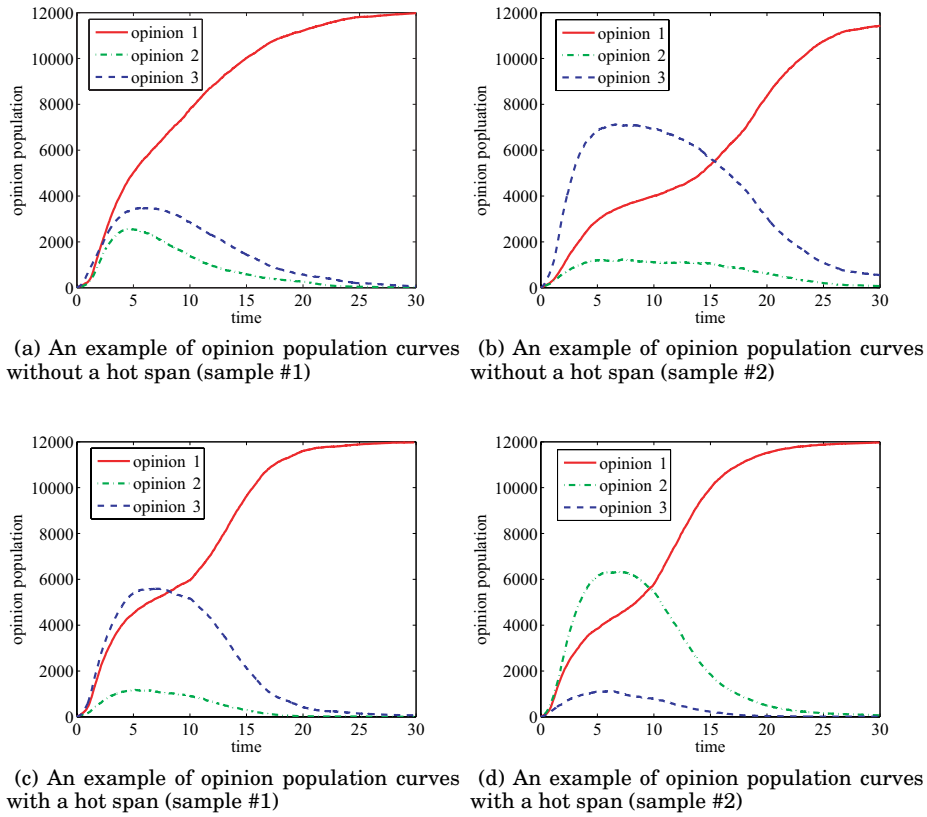


Fig. 2: Information diffusion in the blog network for the VwV model.

3.2. VwV Model

Similarly to the detection problem for the AsIC model, let $[T_1, T_2]$ denote the hot span of the diffusion of opinions under the VwV model. Recall that this implies that the

intervals $[0, T_1)$ and $[T_2, T)$ are the normal spans. We place the same assumption that there is no change in the value of the time-delay parameter vector r for simplicity. Let w_n and w_h denote the opinion-value vectors for the normal span and the hot span, respectively. Note that $w_n/\|w_n\| \neq w_h/\|w_h\|$ since the opinion dynamics under the VwV model is invariant to positive scaling of the opinion-value vector w , where $\|w_n\|$ and $\|w_h\|$ stand for the norm of vectors w_n and w_h , respectively. Then, the change detection problem is formulated as follows: Given the opinion diffusion data $\mathcal{D}(0, T)$ in time-interval $[0, T)$, detect the hot span $[T_1, T_2)$, and estimate the opinion-value vector w_h of the hot span and the opinion-value vector w_n of the normal span. Here, $\mathcal{D}(0, T)$ consists of a sequence of (v, t, k) such that node v changed its opinion to opinion k at time t .

Figure 2 shows two examples of opinion diffusion sample with (Figs. 2c and 2d) and without (Figs. 2a and 2b) a hot span based on the VwV model with $K = 3$ opinions, where the opinion-value vectors are set at $w = (2.0, 1.0, 1.0)$ for Figs. 2a and 2b, and $w_n = (2.0, 1.0, 1.0)$, $w_h = (3.0, 1.0, 1.0)$, $T_1 = 10$ and $T_2 = 20$ for Figs. 2c and 2d. The network used is the same blog network as in Fig. 1. We plotted the population of each opinion k , $|\{v \in V; f_t(v) = k\}|$, as a function of time t . It must be difficult to know the existence of a hot span from only their curves depicted in Figs. 2b and 2d. Moreover, since the VwV model is a stochastic process model, every sample of opinion diffusion can behave differently. Again, this means that it is quite difficult to accurately detect the true hot span from only a single sample of opinion diffusion. We believe that an explicit use of underlying opinion diffusion model is essential to solve this problem. It is crucially important to detect the hot span precisely in order to identify the external factors which caused the behavioral changes.

4. MODEL PARAMETER LEARNING

We describe the framework of model parameter learning as a likelihood maximization problem for the AsIC and the VwV models.

4.1. Parameter Learning for AsIC Model

First, we consider estimating the values of diffusion probability p and time-delay parameter r from an observed diffusion result $\mathcal{D}(0, T) = \{\dots, (v, t_v), \dots\}$ when there is no hot span. Recall that the initial activation time is set to 0 and the final observation time is denoted by T . Let \mathcal{D} be the set of all the activated nodes in $\mathcal{D}(0, T)$, *i.e.*,

$$\mathcal{D} = \{v \in V; (v, t_v) \in \mathcal{D}(0, T)\}.$$

For each node $v \in \mathcal{D}$, let \mathcal{A}_v be the set of its parent nodes that had a chance to activate it, *i.e.*,

$$\mathcal{A}_v = \{u \in B(v); (u, t_u) \in \mathcal{D}(0, T), t_u < t_v\}.$$

Although we place a constraint that $p_{u,v} = p$, $r_{u,v} = r$ ($\forall (u, v) \in E$), we develop a general theory in terms of p and r to be consistent with the description in Subsection 5.2. Let $\mathcal{X}_{u,v}(p_{u,v}, r_{u,v})$ denote the probability density that a node $u \in \mathcal{A}_v$ activates the node v at time t_v , that is,

$$\mathcal{X}_{u,v}(p_{u,v}, r_{u,v}) = p_{u,v} r_{u,v} \exp(-r_{u,v}(t_v - t_u)). \quad (1)$$

Let $\mathcal{Y}_{u,v}(p_{u,v}, r_{u,v})$ denote the probability that the node v is not activated by a node $u \in \mathcal{A}_v$ within the time-period (t_u, t_v) , that is,

$$\begin{aligned} \mathcal{Y}_{u,v}(p_{u,v}, r_{u,v}) &= 1 - p_{u,v} \int_{t_u}^{t_v} r_{u,v} \exp(-r_{u,v}(t - t_u)) dt \\ &= p_{u,v} \exp(-r_{u,v}(t_v - t_u)) + (1 - p_{u,v}). \end{aligned} \quad (2)$$

By using Eqs. (1) and (2), we can obtain the probability density $h_v(\mathbf{p}, \mathbf{r})$ that a node v is activated at time t_v ,

$$h_v(\mathbf{p}, \mathbf{r}) = \sum_{u \in \mathcal{A}_v} \mathcal{X}_{u,v}(p_{u,v}, r_{u,v}) \left(\prod_{z \in \mathcal{A}_v \setminus \{u\}} \mathcal{Y}_{z,v}(p_{z,v}, r_{z,v}) \right), \quad (3)$$

and the probability $\psi_{v,z}(p_{v,z}, r_{v,z})$ that a node z is not activated by a node v within $[0, T)$,

$$\psi_{v,z}(p_{v,z}, r_{v,z}) = p_{v,z} \exp(-r_{v,z}(T - t_v)) + (1 - p_{v,z}). \quad (4)$$

Then, from Eqs. (3) and (4), the following log likelihood function $\mathcal{L}(\mathbf{p}, \mathbf{r}; \mathcal{D}(0, T))$ can be obtained for observed data $\mathcal{D}(0, T)$

$$\mathcal{L}(\mathbf{p}, \mathbf{r}; \mathcal{D}(0, T)) = \sum_{v \in \mathcal{D}} \left(\log h_v(\mathbf{p}, \mathbf{r}) + \sum_{z \in F(v) \setminus \mathcal{D}} \log \psi_{v,z}(p_{v,z}, r_{v,z}) \right). \quad (5)$$

Here, we recall $p_{u,v} = p$, $r_{u,v} = r$ for any $(u, v) \in E$. The values of parameters p and r can be stably obtained by maximizing Eq. (5) using an EM-like algorithm (see Appendix A for more details).

Now, we assume that there exists a hot span $S = [T_1, T_2)$. Let $p(t)$ denote the value of parameter p at time t . According to our problem setting, we consider the parameter switching,

$$p(t) = \begin{cases} p_n & \text{if } t \in [0, T) \setminus S, \\ p_h & \text{if } t \in S. \end{cases}$$

For the hot span S , we split the set of the active nodes \mathcal{D} as follows:

$$\begin{aligned} \mathcal{D}_n(S) &= \{v \in \mathcal{D}; t_v \in [0, T) \setminus S\}, \\ \mathcal{D}_h(S) &= \{v \in \mathcal{D}; t_v \in S\}. \end{aligned}$$

For any $v \in \mathcal{D}$, let $h_v(p_n, p_h, r; S)$ be the probability density that node v is activated at time t_v when there exists hot span S . By using Eqs. (1) and (2), we obtain

$$\begin{aligned} &h_v(p_n, p_h, r; S) \\ &= \sum_{u \in \mathcal{A}_v \cap \mathcal{D}_n(S)} \mathcal{X}_{u,v}(p_n, r) \left(\prod_{z \in \mathcal{A}_v \cap \mathcal{D}_n(S) \setminus \{u\}} \mathcal{Y}_{z,v}(p_n, r) \prod_{z \in \mathcal{A}_v \cap \mathcal{D}_h(S)} \mathcal{Y}_{z,v}(p_h, r) \right) \\ &+ \sum_{u \in \mathcal{A}_v \cap \mathcal{D}_h(S)} \mathcal{X}_{u,v}(p_h, r) \left(\prod_{z \in \mathcal{A}_v \cap \mathcal{D}_n(S)} \mathcal{Y}_{z,v}(p_n, r) \prod_{z \in \mathcal{A}_v \cap \mathcal{D}_h(S) \setminus \{u\}} \mathcal{Y}_{z,v}(p_h, r) \right). \quad (6) \end{aligned}$$

Using Eqs. (4) and (6), we can define an objective function $\mathcal{L}(p_n, p_h, r; \mathcal{D}(0, T), S)$ for the hot span detection problem by adequately modifying Eq. (5) under the switching scheme as follows:

$$\begin{aligned} &\mathcal{L}(p_n, p_h, r; \mathcal{D}(0, T), S) \\ &= \sum_{v \in \mathcal{D}} \log h_v(p_n, p_h, r; S) + \sum_{v \in \mathcal{D}_n(S)} \sum_{z \in F(v) \setminus \mathcal{D}} \log \psi_{v,z}(p_n, r) \\ &+ \sum_{v \in \mathcal{D}_h(S)} \sum_{z \in F(v) \setminus \mathcal{D}} \log \psi_{v,z}(p_h, r). \quad (7) \end{aligned}$$

Clearly, $\mathcal{L}(p_n, p_h, r; \mathcal{D}(0, T), S)$ is expected to be maximized by setting S to the true hot span $S^* = [T_1^*, T_2^*]$ if a substantial amount of data $\mathcal{D}(0, T)$ is available. Thus, our hot span detection problem is formalized as the following maximization problem:

$$\hat{S} = \arg \max_S \mathcal{L}(\hat{p}_n(S), \hat{p}_h(S), \hat{r}(S); \mathcal{D}(0, T), S), \quad (8)$$

where $\hat{p}_n(S)$, $\hat{p}_h(S)$, and $\hat{r}(S)$ denote the maximum likelihood estimators for a given S .

4.2. Parameter Learning for VwV Model

We also consider estimating the value of opinion-value vector w from an observed opinion diffusion data $\mathcal{D}(0, T)$ in time interval $[0, T)$ (a single example) when there is no hot span⁴. From the evolution process of the model, we can obtain the following log likelihood function

$$\mathcal{L}(w; \mathcal{D}(0, T)) = \log \prod_{(v, t, k) \in \mathcal{C}(0, T)} \frac{n_k(t, v)w_k}{\sum_{k'=1}^K n_{k'}(t, v)w_{k'}}, \quad (9)$$

where

$$\mathcal{C}(0, T) = \{(v, t, f_t(v)) \in \mathcal{D}(0, T); |\{u \in \Gamma(v); f_t(u) \neq 0\}| \geq 2\}.$$
⁵

Thus, our estimation problem is formulated as a maximization problem of the log likelihood function $\mathcal{L}(w; \mathcal{D}(0, T))$ with respect to w . We find the optimal value of w by employing a standard Newton method (see Appendix B for more details).

Now, we assume that there exists a hot span $S = [T_1, T_2)$. Let $w(t)$ denote the value of opinion-value vector w at time t . We also consider the following parameter vector switching:

$$w(t) = \begin{cases} w_n & \text{if } t \in [0, T) \setminus S, \\ w_h & \text{if } t \in S. \end{cases}$$

For $\forall T_s, T_e$ with $0 \leq T_s < T_e \leq T$, we denote by $\mathcal{D}(T_s, T_e)$ the opinion diffusion data in time interval $[T_s, T_e)$; *i.e.*,

$$\mathcal{D}(T_s, T_e) = \{(v, t, k) \in \mathcal{D}(0, T); t \in [T_s, T_e)\}. \quad (10)$$

Then, similarly to the case of the AsIC model, an objective function $\mathcal{L}(w_n, w_h; \mathcal{D}(0, T), S)$ can be defined for the hot span detection problem by adequately modifying Eq. (9) under this switching scheme as follows:

$$\mathcal{L}(w_n, w_h; \mathcal{D}(0, T), S) = \mathcal{L}(w_n; \mathcal{D}(0, T_1) \cup \mathcal{D}(T_2, T)) + \mathcal{L}(w_h; \mathcal{D}(T_1, T_2)). \quad (11)$$

Again, the extended objective function is expected to be maximized by setting S to be the true span $S^* = [T_1^*, T_2^*)$, provided that $\mathcal{D}(0, T)$ is generated by the VwV model with hot span S^* and is sufficiently large. Therefore, our hot span detection problem is formalized as the following maximization problem:

$$\hat{S} = \arg \max_S \mathcal{L}(\hat{w}_n(S), \hat{w}_h(S); \mathcal{D}(0, T), S), \quad (12)$$

where $\hat{w}_n(S)$ and $\hat{w}_h(S)$ denote the maximum likelihood estimators for a given S .

⁴The time-delay parameter vector r can simply be estimated by averaging the time intervals for each node, and thus excluded from the estimation problem.

⁵We use only those observed data in which there is at least one neighbor that has an opinion.

5. CHANGE DETECTION METHODS

We propose a general method of detecting a hot span that is applicable to both the AsIC model and the VwV model. In order to obtain the optimal hot span \hat{S} according to either Eq. (8) or Eq. (12), we need to prepare a reasonable set of candidate hot spans, denoted by \mathcal{H} . One way of doing so is to construct \mathcal{H} by considering all pairs of observed activation (or opinion change) time points. In general, let \mathcal{T} denote the set of all the observed activation (or opinion change) time points,

$$\mathcal{T} = \{t_0, t_1, \dots, t_N\}, \quad (0 = t_0 < t_1 < \dots < t_N < T).$$

Then, we can construct a set of candidate hot spans by

$$\mathcal{H} = \{S = [T_1, T_2]; T_1 < T_2, T_1 \in \mathcal{T}, T_2 \in \mathcal{T}\}.$$

Hereafter, we denote the model parameter vector by θ ; *i.e.*, $\theta = (p, r)$ for the AsIC model and $\theta = w$ for the VwV model. Since the parameter vector θ is a function of time t in our problem setting, we denote by $\theta(t)$ the value of θ at time t . Given a hot span $S = [T_1, T_2)$, we consider the following parameter vector switching:

$$\theta(t) = \begin{cases} \theta_n & \text{if } t \in [0, T) \setminus S, \\ \theta_h & \text{if } t \in S. \end{cases}$$

Let $S^* = [T_1^*, T_2^*)$ be the true hot span. We assume that observed data $\mathcal{D}(0, T)$ is generated by using the parameter vector $\theta^*(t)$ of hot span S^* . In what follows, after introducing a naive method, we describe our proposed detection method.

5.1. Naive Method

Both Eq. (8) and Eq. (12) can be solved by a naive method which has two iterative loops. In the inner loop we first obtain the maximum likelihood estimators, $\hat{\theta}_n$ and $\hat{\theta}_h$, for each candidate S by maximizing the objective function $\mathcal{L}(\theta_n, \theta_h; \mathcal{D}(0, T), S)$ using either the EM-like algorithm or the Newton method. In the outer loop we select the optimal \hat{S} which gives the largest $\mathcal{L}(\hat{\theta}_n, \hat{\theta}_h; \mathcal{D}(0, T), S)$ value. However, this method can be extremely inefficient when the number of candidate spans is large. Thus, in order to make it work with a reasonable computational cost, we consider restricting the number of candidate time points to a smaller value, denoted by J , *i.e.*, we construct $\mathcal{T}_J \subset \mathcal{T}$ by selecting J points from \mathcal{T} ; then we construct a restricted set of candidate spans by

$$\mathcal{H}_J = \{S = [T_1, T_2); T_1 < T_2, T_1 \in \mathcal{T}_J, T_2 \in \mathcal{T}_J\}.$$

Note that $|\mathcal{H}_J| = J(J-1)/2$, which is large when J is large.

5.2. Proposed Method

It is easily conceivable that the naive method can detect the hot span with a reasonably good accuracy when we set J large at the expense of the computational cost, but the accuracy becomes poorer when we set J smaller to reduce the computational load. We propose a novel detection method below which alleviates this problem and can efficiently and stably detect a hot span from the diffusion result $\mathcal{D}(0, T)$.

We first obtain the maximum likelihood estimators, $\hat{\theta}$, based on the original objective function of either Eq. (5) or Eq. (9). Next, we focus on the first-order derivative of the objective function $\mathcal{L}(\theta; \mathcal{D}(0, T))$ with respect to the parameter vector θ in each observation interval $[t_{j-1}, t_j)$. More specifically, we define a function $\tilde{\mathcal{L}}(\theta_1, \dots, \theta_N; \mathcal{D}(0, T))$ of $\theta_1, \dots, \theta_N$ by

$$\tilde{\mathcal{L}}(\theta_1, \dots, \theta_N; \mathcal{D}(0, T)) = \mathcal{L}(\tilde{\theta}(t); \mathcal{D}(0, T)),$$

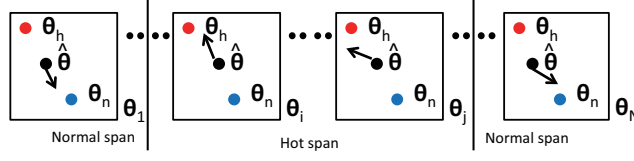


Fig. 3: Direction of the gradient vector at $\hat{\theta}$ in the normal and the hot span.

where $\tilde{\theta}(t) = \theta_j$ if $t \in [t_{j-1}, t_j)$, ($j = 1, \dots, N$). Since $\mathcal{L}(\theta; \mathcal{D}(0, T)) = \tilde{\mathcal{L}}(\theta, \dots, \theta; \mathcal{D}(0, T))$ and $\hat{\theta}$ is the maximum likelihood estimator based on $\mathcal{L}(\theta; \mathcal{D}(0, T))$, *i.e.*, when no change in θ is assumed, we have

$$0 = \frac{\partial \mathcal{L}(\hat{\theta}; \mathcal{D}(0, T))}{\partial \theta} = \sum_{j=1}^N \frac{\partial \tilde{\mathcal{L}}(\hat{\theta}, \dots, \hat{\theta}; \mathcal{D}(0, T))}{\partial \theta_j} \quad (13)$$

Note that $\tilde{\mathcal{L}}(\theta_1, \dots, \theta_N; \mathcal{D}(0, T))$ can be expected to attain the maximum when each θ_j is given as follows: $\theta_j = \theta_h$ if $[t_{j-1}, t_j)$ is included in the hot span and $\theta_j = \theta_n$ if $[t_{j-1}, t_j)$ is included in the normal span, *i.e.*, $\tilde{\mathcal{L}}(\theta_1, \dots, \theta_N; \mathcal{D}(0, T)) = \mathcal{L}(\theta_n, \theta_h; \mathcal{D}(0, T), S^*)$. Thus, we introduce modification vectors, $\vartheta_1, \dots, \vartheta_N$, defined by $\vartheta_j = \theta_h - \hat{\theta}$ if $[t_{j-1}, t_j)$ is included in the hot span and $\vartheta_j = \theta_n - \hat{\theta}$ if $[t_{j-1}, t_j)$ is included in the normal span. Let $\Delta \mathcal{L}$ be $\mathcal{L}(\theta_n, \theta_h; \mathcal{D}(0, T), S^*) - \tilde{\mathcal{L}}(\hat{\theta}, \dots, \hat{\theta}; \mathcal{D}(0, T)) = \mathcal{L}(\hat{\theta} + \vartheta_1, \dots, \hat{\theta} + \vartheta_N; \mathcal{D}(0, T), S^*) - \tilde{\mathcal{L}}(\hat{\theta}, \dots, \hat{\theta}; \mathcal{D}(0, T))$. Then, we can obtain the following first-order Taylor expansion:

$$\begin{aligned} \Delta \mathcal{L} &\approx \sum_{j=1}^N \frac{\partial \tilde{\mathcal{L}}(\hat{\theta}, \dots, \hat{\theta}; \mathcal{D}(0, T))}{\partial \theta_j} \vartheta_j \\ &= \sum_{j; [t_{j-1}, t_j) \subset S^*} \frac{\partial \tilde{\mathcal{L}}(\hat{\theta}, \dots, \hat{\theta}; \mathcal{D}(0, T))}{\partial \theta_j} (\theta_h - \hat{\theta}) + \sum_{j; [t_{j-1}, t_j) \not\subset S^*} \frac{\partial \tilde{\mathcal{L}}(\hat{\theta}, \dots, \hat{\theta}; \mathcal{D}(0, T))}{\partial \theta_j} (\theta_n - \hat{\theta}). \end{aligned}$$

Moreover, by noting Eq. (13), we obtain the following result:

$$\Delta \mathcal{L} \approx \sum_{j; [t_{j-1}, t_j) \subset S^*} \frac{\partial \tilde{\mathcal{L}}(\hat{\theta}, \dots, \hat{\theta}; \mathcal{D}(0, T))}{\partial \theta_j} (\theta_h - \theta_n). \quad (14)$$

Here note that we can naturally assume that each gradient vector with respect to θ_j is likely to be parallel to $(\theta_h - \theta_n)$, as shown by arrows in Fig. 3. Therefore, from Eq. (14), by considering the following partial sum for a candidate hot span $S = [T_1, T_2) \in \mathcal{H}$:

$$g(S) = \sum_{j; [t_{j-1}, t_j) \subset S} \frac{\partial \tilde{\mathcal{L}}(\hat{\theta}, \dots, \hat{\theta}; \mathcal{D}(0, T))}{\partial \theta_j}. \quad (15)$$

we can expect that $\|g(S)\|$ is maximized when $S \approx S^*$.

Therefore, we propose the method of detecting the hot span by

$$\hat{S} = \arg \max_{S \in \mathcal{H}} \|g(S)\|. \quad (16)$$

In case of the AsIC model,

$$\|g(S)\|^2 = \left| \sum_{(u,v) \in E; u \in \mathcal{D}_h(S)} \frac{\partial \mathcal{L}(\hat{p}, \hat{r}; \mathcal{D}(0, T))}{\partial p_{u,v}} \right|^2$$

(see Eq. (5)), and in case of the VwV model,

$$g(S) = \frac{\partial \mathcal{L}(\hat{w}; \mathcal{D}(T_1, T_2))}{\partial w}$$

(see Eqs. (9)).

Here note that we can incrementally calculate $g(S)$. More specifically, we can obtain the following formula:

$$g([t_i, t_j]) = g([t_i, t_{j-1}]) + \frac{\partial \tilde{\mathcal{L}}(\hat{\theta}, \dots, \hat{\theta}; \mathcal{D}(0, T))}{\partial \theta_j} \quad (17)$$

for any $t_i, t_{j-1}, t_j \in \mathcal{T}$ with $t_i < t_{j-1} < t_j$. The computational cost of the proposed method for examining each candidate span is much smaller than the naive method described above. When $|\mathcal{T}|$ is very large, we construct a restricted set of candidate spans \mathcal{H}_J as explained above. We summarize our proposed method below.

1. Maximize $\mathcal{L}(\theta; \mathcal{D}(0, T))$ by using the parameter estimation method.
2. Construct the candidate time set \mathcal{T} and the candidate hot span set \mathcal{H} .
3. Detect the hot span \hat{S} by Eq. (16) and output \hat{S} .
4. Maximize $\mathcal{L}(\theta_n, \theta_h; \mathcal{D}(0, T), \hat{S})$ by using the parameter estimation method, and output $(\hat{\theta}_n, \hat{\theta}_h)$.

Here note that the proposed method requires likelihood maximization by using the parameter estimation method only twice.

6. EXPERIMENTAL EVALUATION

We experimentally investigated how accurately the proposed method can estimate both the hot span and the diffusion parameters for the hot and the normal spans, as well as its efficiency, by comparing it with the naive method using four real world networks.

6.1. Datasets

We used four real large networks which are all bidirectionally connected⁶. The first one is a trackback network of Japanese blogs used in [Kimura et al. 2009]. It has 12,047 nodes and 79,920 directed links (the blog network). The second one is a coauthorship network used in [Palla et al. 2005], which has 12,357 nodes and 38,896 directed links (the Coauthorship network). The third one is a network derived from the Enron Email Dataset [Klimt and Yang 2004] by extracting the senders and the recipients and linking those that had bidirectional communications. It has 4,254 nodes and 44,314 directed links (the Enron network). The fourth one is a network of people that was derived from the “list of people” within Japanese Wikipedia, used in [Kimura et al. 2008], and has 9,481 nodes and 245,044 directed links (the Wikipedia network).

6.2. Experimental Settings

We generated diffusion results using both the AsIC model (for information diffusion evaluation) and the VwV model (for opinion population diffusion evaluation) for each of the above networks under the following settings. As for the AsIC model, we considered $p = 1/\bar{d}$ as the base value of the diffusion probability of each link, where \bar{d} is the mean out-degree of the network. With this base value, for an arbitrary node in the

⁶We wanted to use the real data measured in the real network where there is a known external change, but unfortunately we were not able to find such data. We are still looking for a good dataset that can be used to validate our approach.

network, the expected number of its child nodes that it succeeds to activate becomes approximately equal to one at least at an early phase of the information diffusion. If the diffusion probability is much smaller than the base value, the diffusion process would end up with only a small number of active nodes on the average. On the other hand, if it is much larger, the information rapidly spreads out the entire network and the process finishes at an early phase of the diffusion. Both cases are not appropriate to our aim of investigating the hot span detection, *i.e.*, we need a fair amount of information diffusion taking place around the hot span. Thus, in our experiments, we set the diffusion probability for the normal span, p_n^* , to be a value slightly smaller than the base value, and set the diffusion probability for the hot span, p_h^* , to be three times as large as the p_n^* . As a result, p_n^* and p_h^* are 0.1 and 0.3 for the blog network, 0.2 and 0.6 for the Coauthorship network, 0.05 and 0.15 for the Enron network and 0.02 and 0.06 for the Wikipedia network, respectively. As explained in 3.1, we assumed that the time delay parameter does not change, and fixed its value to be 1 ($r^* = 1$) for all the networks because changing r^* works only for scaling the time axis of the diffusion results. We set the observation period to be $[0, T = 30)$ and the hot span to be $[T_1^* = 10, T_2^* = 20)$ based on the observation on the preliminary experiments. In all we generated 10 information diffusion results using these parameter values, each starting from a randomly selected initial active node for each network.

As for the VwV model, for each of the above networks, we generated opinion diffusion results according to the model for three different values of K (the number of opinions), *i.e.*, $K = 2, 4$, and 8 , by choosing the top K nodes with respect to node degree ranking as the initial K nodes. We assumed that the value of all the opinions were initially 1.0, *i.e.*, the value-parameters for all the opinions are 1.0 for the normal span, and further assumed that only the value of the first opinion changed to double for the hot span, *i.e.*, the value-parameter of the first opinion is 2.0 and the value-parameters of all the other opinions are 1.0 for the hot span. Again, based on the observation on the preliminary experiments, we set the observation period and the hot span to be $[0, T = 25)$ and $[T_1^* = 10, T_2^* = 15)$, respectively, and generated 10 opinion diffusion results for each network.

We then estimated the hot span $[T_1^*, T_2^*)$ and the diffusion parameters of each model, *i.e.*, the diffusion probabilities p_n^* (for the normal span) and p_h^* (for the hot span) for the AsIC model, and the opinion-value vectors w_n^* (for the normal span) and w_h^* (for the hot span) for the VwV model by the two methods (the proposed and the naive), and compared them in terms of 1) the accuracy of the estimated hot span $\hat{S} = [\hat{T}_1, \hat{T}_2)$, 2) the accuracy of the estimated diffusion parameters, $\hat{p}_n, \hat{p}_h, \hat{w}_n$, and \hat{w}_h , 3) their integrated estimation error, and 4) the computation time. The accuracy of the estimated hot span is measured in the absolute error $\mathcal{E}_S = |\hat{T}_1 - T_1^*| + |\hat{T}_2 - T_2^*|$ for both the AsIC and VwV models. The accuracy of the estimated diffusion parameters is evaluated in the mean relative error, *i.e.*, $\mathcal{E}_p = |\hat{p}_n - p_n|/p_n + |\hat{p}_h - p_h|/p_h$ for the AsIC model, and $\mathcal{E}_w = \sum_{i=1}^K (|\hat{w}_{n_i} - w_{n_i}^*|/w_{n_i}^* + |\hat{w}_{h_i} - w_{h_i}^*|/w_{h_i}^*)/K$ for the VwV model, where $w_{n_i}^*$ and $w_{h_i}^*$ are values of opinion i for the normal and the hot spans, respectively, and $\hat{w}_{n_i}^*$ and $\hat{w}_{h_i}^*$ are their estimated values. Integrating their estimation errors by $\mathcal{E}_{\theta(t)} = \int_0^T \|\hat{\theta}(t) - \theta^*(t)\|_{L1} dt$ allows us to evaluate the estimation ability of each method in a comprehensive manner, where $\theta^*(t)$ and $\hat{\theta}(t)$ is the diffusion parameter vector to be assumed true and its estimation at time t for the corresponding model, respectively. For the proposed method, we adopted 1,000 as the value of J (the number of candidate time points) for the VwV model, while we used all the possible time points, *i.e.*, $J = N$ for the AsIC model because the number of time points for opinion changes in the VwV model is observed to be much larger than the number of node activation for the AsIC

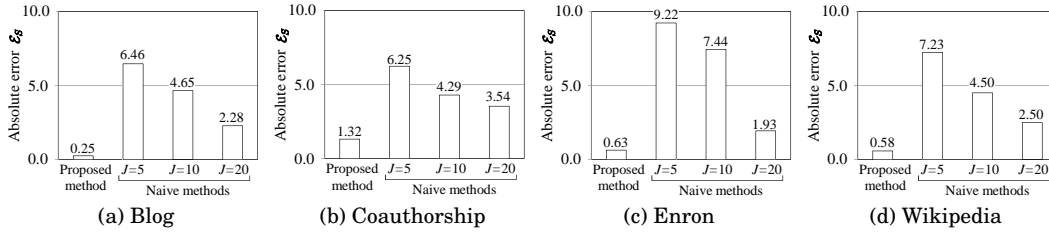


Fig. 4: Comparison of the accuracy in the estimated hot span for the AsIC model.

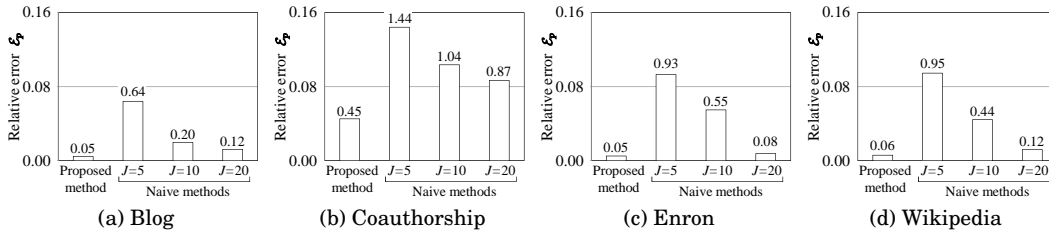


Fig. 5: Comparison of the accuracy in the estimated diffusion probability for the AsIC model.

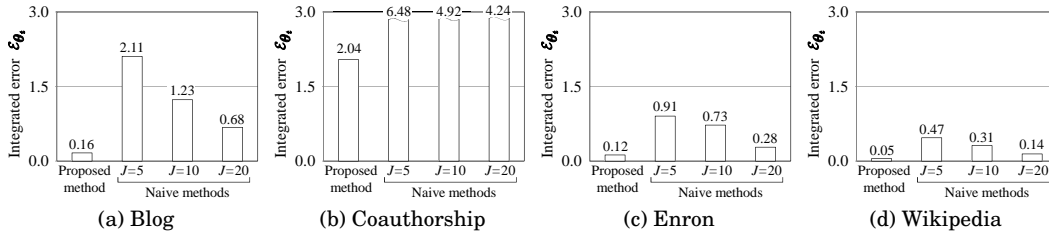


Fig. 6: Comparison of the integrated estimation error for the AsIC model.

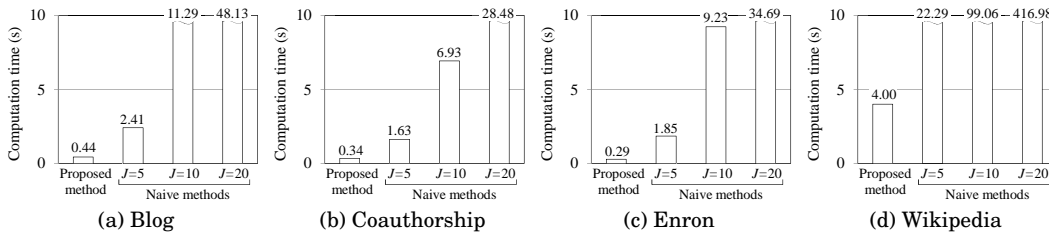


Fig. 7: Comparison of the computation time for the AsIC model.

model for the same period of time. For the naive method, we tested three cases of $J = 5, 10,$ and 20 for both the models. Both the proposed and the naive methods were tested on each diffusion result for each model mentioned above on a PC with Intel Core 2 Duo 3GHz, and the results were averaged over the ten independent trials for each network.

6.3. Results for AsIC Model

Figures 4 to 7 summarize the results for the AsIC model. Figure 4 shows the accuracy of \hat{S} in the absolute error \mathcal{E}_S defined above. We see that the proposed method achieves a good accuracy, much better than the naive method for every network. As expected, \mathcal{E}_S for the naive method decreases as J becomes larger. But, even in the best case ($J = 20$), its average error is about 3 to 10 times larger than that of the proposed method. Figure 5 shows the accuracy of \hat{p}_n and \hat{p}_h in the relative error \mathcal{E}_p . Here again, the average relative error for the naive method decreases as J becomes larger. However, even in the best case ($J = 20$), it is about 2 to 3 times larger than that of the proposed method. We note that the average errors for the Coauthorship network are relatively large. This is because the number of active nodes within the normal span was relatively small for this network. Figure 6 shows the integrated estimation error given by $\mathcal{E}_{\theta(t)}$, which supplements our insights derived from the above results. For example, although the relative error of the estimated diffusion probabilities of the naive method ($J = 20$) is less than twice as big as the proposed method for the Enron network, its value of $\mathcal{E}_{\theta(t)}$ becomes more than twice of the proposed method by considering the estimation error of the hot span. Overall, even in the best case of the naive method ($J = 20$), its integrated estimation error is about 2 to 4 times larger than that of the proposed method. Figure 7 shows the computation time. It is clear that the proposed method is much faster than the naive method. The significant difference is attributed to the difference in the number of runs of the EM-like algorithm. The proposed method executes the EM-like algorithm only twice: steps 1 and 4 in the algorithm (see Section 5.2). On the other hand, the naive method has to execute the EM-like algorithm once for every single candidate hot span $S \in \mathcal{H}_J$ which is $|\mathcal{H}_J| = J(J-1)/2$ times (see Section 5.1). Indeed, the computation time of the naive method for $J = 5$ is about 5 times larger for every network, which is consistent with the fact that $|\mathcal{H}_5| = 10$. This relation roughly holds also for the other two cases ($J = 10$ and $J = 20$). This means that even if the naive method could achieve a good accuracy by setting J to a sufficiently large value, it would require unacceptable computation time for such a large J . Overall, the proposed algorithm is about 3 times more accurate in the fastest case for the naive method (in the case of the Coauthorship network under $J = 5$) and about 100 times faster in its most accurate case (in the case of the Wikipedia network under $J = 20$). Finally, we illustrate the actual behavior of $\|g(S)\|$ derived from an information diffusion result for the blog network under the AsIC model in Fig. 8a, where $\|g(S)\|$ is depicted as a function of the ending point t_j of S when its starting point is fixed to a certain value. We can see the blue broken curve showing $\|g([0, t_j])\|$ has two peaks at around $t_j = 10$ and $t_j = 20$, which are the starting and ending points of the true hot span, respectively. This means that the sign of $\partial \mathcal{L}(\hat{p}, \hat{r}; \mathcal{D}(0, T)) / \partial p_{u,v}$ reversed at these time points as explained in Section 5.2⁷. Thus, the red solid curve showing $\|g([10, t_j])\|$ has only one peak at around $t_j = 20$, which is the global maximum among all the possible $\|g(S)\|$. Thanks to Eq.(17), the proposed method can efficiently calculate the behavior of $\|g(S)\|$, and thus can find out the hot span more accurately and more efficiently than the naive method does.

In summary, we can say that the proposed method can detect and estimate the hot span and diffusion probabilities for the AsIC model much more accurately and efficiently compared with the naive method.

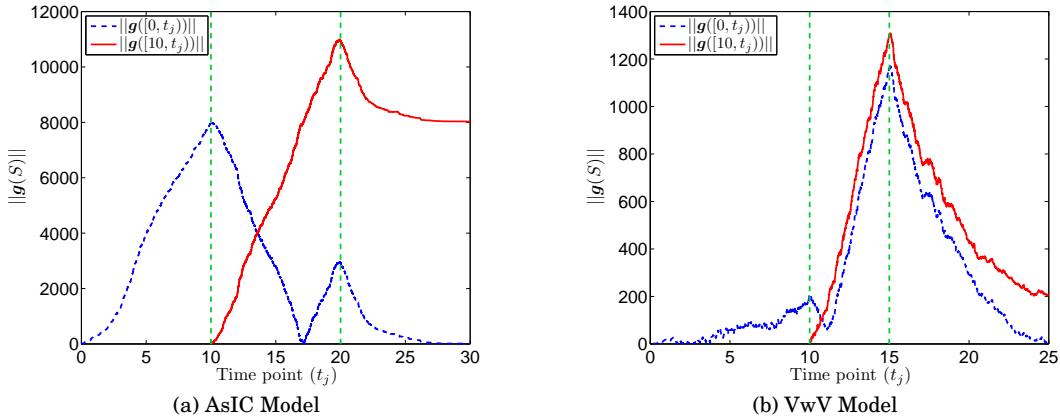


Fig. 8: Change of $\|g(S)\|$ when given a fixed starting point of a time span S for a diffusion result retrieved from the blog network under the respective experimental setting for each information diffusion model.

6.4. Results for Voter Model

Figures 9 to 12 show the experimental results for the VwV model. Similarly to the results for the AsIC model, from these results, we can find that the proposed method is much more accurate than the naive method for every network. Again, the average error for the naive method decreases as J becomes larger. But, even in the best case for the naive method ($J = 20$), its average error in the estimation of the hot span is maximum about 30 times larger than that of the proposed method (in the case of the Enron network under $K = 2$) as shown in Fig. 9, and it is maximum about 6 times larger in the estimation of opinion-values (in the case of the Coauthorship network under $K = 2$) as shown in Fig. 10. Figure 11 shows that the proposed method is better than the naive method in the integrated estimation accuracy for every case. It is noted that the naive method needs much longer computation time to achieve these best accuracies than the proposed method as shown in Fig. 12 despite that the number of candidate time points for the naive method is 50 times smaller. Indeed, it is about 20 times longer in the case of the Enron network under $K = 2$, about 13 times longer in the case of the Coauthorship network under $K = 2$, and maximum about 95 times longer for the whole results (in the case of the Enron network under $K = 8$). Overall, the proposed method is about 7 times more accurate in the fastest case for the naive method (in the case of the blog network under $K = 2$ and $J = 5$) and about 13 times faster in its most accurate case (in the case of the Coauthorship network under $K = 2$ and $J = 20$). Figure 8b shows the behavior of $\|g(S)\|$ derived from an opinion diffusion result for the blog network under the VwV model. Similarly to the case of the AsIC model, it is found that the blue broken curve showing $\|g([0, t_j])\|$ has two peaks at around $t_j = 10$ and $t_j = 15$, which are the starting and ending points of the true hot span, respectively. In this case, the red solid curve starting from $t_j = 10$ has only one peak at around $t_j = 15$, which becomes the global maximum among all the possible $\|g(S)\|$. The proposed method can find out the time span that results in the global maximum from a set of the candidate time points efficiently for the VwV model, too.

⁷Since in this case the partial derivative is a scalar, it suffices to say its sign.

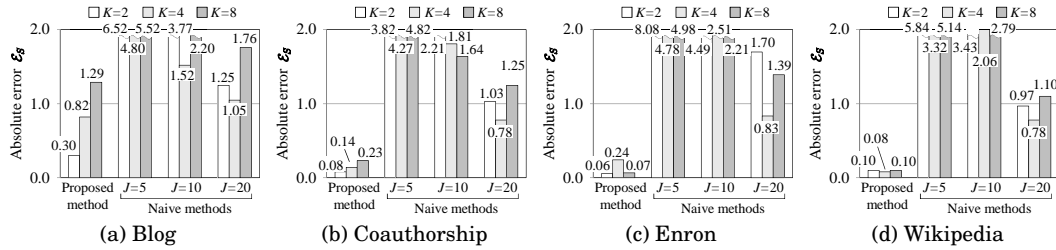


Fig. 9: Comparison of the accuracy in the estimated hot span for the VwV model.

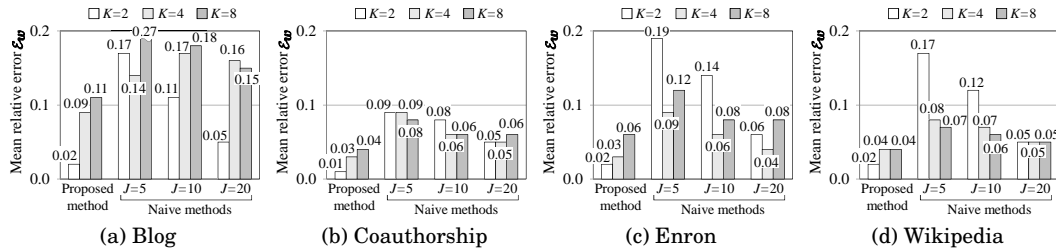


Fig. 10: Comparison of the accuracy in the estimated opinion-value vector for the VwV model.

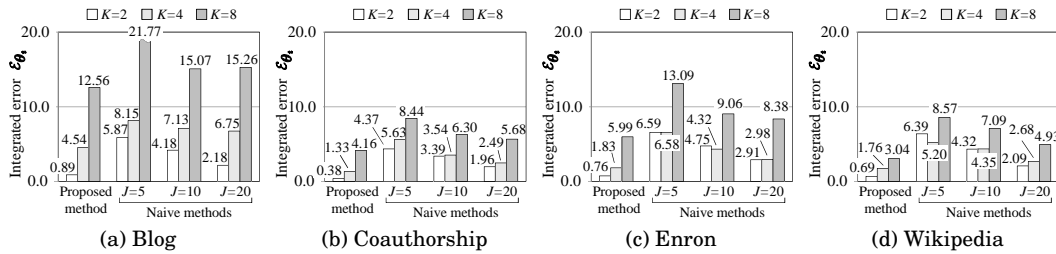


Fig. 11: Comparison of the integrated estimation error for the VwV model.

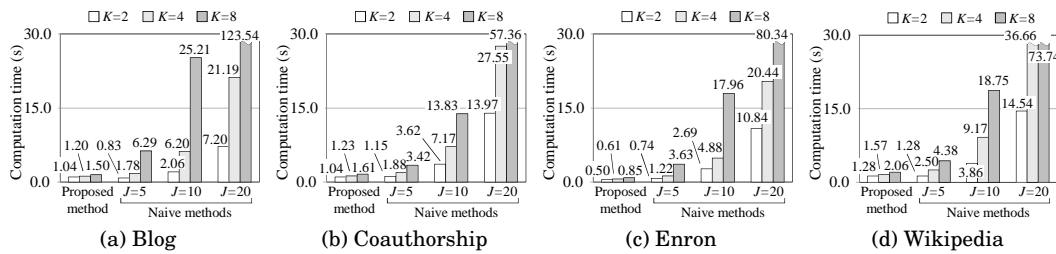


Fig. 12: Comparison of the computation time for the VwV model.

From these results, it can be concluded that the proposed method is able to detect and estimate the hot span and opinion-values for the VwV model much more accurately and efficiently compared with the naive method.

7. DISCUSSION

The results in the previous section indicates that the proposed approach works as intended to both AsIC and VwV diffusion models. Although we believe that the approach is generic, it has yet to be verified whether the approach is applicable to any other model in so far as it is formulated as a probabilistic diffusion model.

We placed a simplifying constraint that the parameters $p_{u,v}$ and $r_{u,v}$ of the AsIC model do not depend on link (u, v) , *i.e.*, $p_{u,v} = p$, $r_{u,v} = r$ ($\forall (u, v) \in E$), by focusing on single topic diffusion sequences. Our previous experiments [Saito et al. 2009b; 2010a; 2010b] give some evidences which support the validity of this constraint. We examined 7,356 diffusion sequences for a real blogroll network containing 52,525 bloggers and 115,552 blogroll links, and have experimentally confirmed that the diffusion and time-delay parameters that were learned from different diffusion sequences belonging to the same topic were quite similar for most of the topics. This observation naturally suggests that people behave quite similarly for the same topic. On the other hand, our recent study indicates that these parameters can be learned by assuming their functional dependency on the neighboring node attributes [Saito et al. 2011b]. We can extend this approach to augment the attributes to include the node independent external factor. This way the uniformity assumption can be removed. We have considered only the AsIC as a model of general information diffusion, but it is straightforward to apply the same technique to the AsLT model [Saito et al. 2010b] and the SIS (susceptible/infectious/susceptible) versions of both the models in which each node is allowed to be activated multiple times.

The change pattern we used is also very simple. We assumed that the parameters of all nodes and links change instantaneously and simultaneously in the same degree and stay the same during a given hot span. We can assume a more intricate problem setting such that $p_{u,v}$ (for AsIC), w_u (for VwV) and $r_{u,v}$ (for both) change for multiple distinct hot spans and the shape of change pattern $p_{u,v}$ and w_u are not necessarily rect-linear. One possible extension is to approximate the pattern of any shape by J pairs of time interval each with its corresponding $p_{u,v,j}$ and $w_{u,j}$, *i.e.*, $Z_J = \{(p_j, [t_{j-1}, t_j]); j = 1, \dots, J\}$ ($t_0 = 0, t_J = \infty$) and use a divide-and-conquer type greedy recursive partitioning, still employing the derivative of the likelihood function as the main measure for search. For brevity we drop the u, v dependency and consider only the AsIC model. More specifically, we first initialize $Z_1 = \{(\hat{p}_1, ([0, \infty)))\}$ where \hat{p}_1 is the maximum likelihood estimator, and search for the first change time point t_1 , which we expect to be the most distinguished one, by maximizing $\|g(S)\|$ that uses \hat{p}_1 as $\hat{\theta}$ for the whole span $[0, \infty)$.⁸ We recursively perform this operation J times by fixing the previously determined change points. When to stop can be determined by a statistical criterion such as AIC or MDL. This algorithm requires parameter optimization J times. Figure 13 is one of the preliminary results obtained for two distinct rect-linear patterns using five sequences in case of the blog network. MDL is used as the stopping criterion. The change pattern of p is almost perfectly detected with respect to both p_j and t_j ($J = 5$). We might further want to introduce some stochastic natures into the model for some external factors that affect parameter changes reflecting the fact that each individual's response to the external factors is different, *i.e.*, some people respond quickly and others slowly.

The change we considered is only in the time domain and we assumed that there is no spatially local change. We can consider a more general setting, *i.e.*, spatio-temporal change in parameter values. We need a more elaborate algorithm to cope with this extension but the basic approach of using the first derivative of the likelihood function

⁸Note that the total sum of $g = 0$.

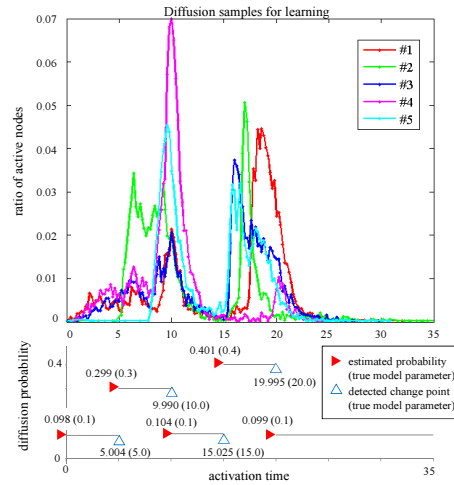


Fig. 13: Information diffusion in the blog network with two hot spans for the AsIC model.

remains valid. We assumed that the network structure is stationary although we introduced the change in the parameter value. The model we used does not account for the structure change by itself. However, once the structure change is known, *i.e.*, addition/deletion of nodes and links at each time instance, it is straightforward to apply the proposed algorithm to these changes because the dynamics of a node is determined by the interaction with its neighbors, *i.e.*, local structure of the network.

8. CONCLUSION

In this paper, we addressed the problem of detecting changes in behavior of information diffusion over a social network which is caused by changes in unknown external factors from a limited amount of observed diffusion sequences in a retrospective setting. The information diffusion process is described by a probabilistic model with some parameters that characterize the behavior, and the change in unknown external factors is assumed to be effectively reflected in changes in the parameter values in the model. We called the period where the parameter takes anomalous values as “hot span” and the rest as “normal span”, and the problem is reduced to detecting the hot span, *i.e.*, identifying the time window where the parameter value is anomalous and estimating the parameter values both in the hot and normal spans. We solved this problem by searching the time window that maximizes the likelihood of generating the observed information diffusion sequences. Our main contribution is that we devised a very efficient general iterative search algorithm which is robust and applicable to a wide class of probabilistic information diffusion models. The algorithm uses the first derivative of the likelihood with respect to the parameters, uses it in the window search (outer loop) and avoids parameter value optimization during the search (inner loop). It only needs to estimate the parameter value twice (at the first and the final steps of the search). This is in contrast to the naive learning algorithm which has to iteratively update the pattern boundaries (outer loop), each requiring the parameter value optimization to maximize the likelihood for the candidate window (inner loop), which is very inefficient and totally unacceptable. We showed that the algorithm works satisfactorily well for two instances of the probabilistic information diffusion model which has different characteristics: asynchronous independent cascade (AsIC) model as a model of information push style and value-weighted voter (VwV) model as a model of informa-

tion pull style. The AsIC is the model for general information diffusion with binary states and the parameter to detect its change is diffusion probability and the VwV is the model for opinion formation with multiple states and the parameter to detect its change is opinion value. The results tested on these two models using four real world network structures and a single rect-linear change confirmed that the algorithm is robust enough and can efficiently identify the correct change pattern of the parameter values. Comparison with the naive method that finds the best combination of change boundaries by an exhaustive search through a set of randomly selected boundary candidates showed that the proposed algorithm far outperforms the native method both in terms of accuracy (about 3 times more accurate for the AsIC model and about 7 times accurate for the VwV model in the fastest case for the naive method) and computation time (about 100 times faster for the AsIC model and about 13 times faster for the VwV model in the most accurate case for the naive method). The problem setting we assumed in this paper is very simple, but we expect that the proposed method can be easily extended to solve more intricate problems. We showed one possible direction and the preliminary result obtained for two rect-linear shape hot spans was very promising. Our immediate future work is to evaluate our method using real world information diffusion samples with hot spans, as well as to deal with spatio-temporal hot span detection problems using more appropriate stochastic models under a similar problem solving framework.

APPENDIX

A. ESTIMATION ALGORITHM FOR ASIC MODEL

We briefly describe the estimation algorithm of parameters p and r for the AsIC model from a sequence of observed data $\mathcal{D}(0, T)$ (see [Saito et al. 2009b; 2010a] for more details).

We employ an EM-like algorithm. Let \bar{p} and \bar{r} be the current estimates of p and r . Using Eqs. (1) and (2), we define $\bar{\alpha}_{u,v}$ and $\bar{\beta}_{u,v}$ as follows:

$$\begin{aligned}\bar{\alpha}_{u,v} &= \frac{\mathcal{X}_{u,v}(\bar{p}, \bar{r})/\mathcal{Y}_{u,v}(\bar{p}, \bar{r})}{\sum_{z \in \mathcal{A}_v} \mathcal{X}_{z,v}(\bar{p}, \bar{r})/\mathcal{Y}_{z,v}(\bar{p}, \bar{r})} \\ \bar{\beta}_{u,v} &= \frac{\bar{p} \exp(-\bar{r}(t_v - t_u))}{\mathcal{Y}_{u,v}(\bar{p}, \bar{r})}\end{aligned}$$

The update formulas of p and r are as follows:

$$\begin{aligned}p &= \frac{\sum_{v \in \mathcal{D}} \sum_{u \in \mathcal{A}_v} (\bar{\alpha}_{u,v} + (1 - \bar{\alpha}_{u,v})\bar{\beta}_{u,v})}{|\{(u, v) \in E; u \in \mathcal{D}\}|} \\ r &= \frac{\sum_{v \in \mathcal{D}} \sum_{u \in \mathcal{A}_v} \bar{\alpha}_{u,v}}{\sum_{v \in \mathcal{D}} \sum_{u \in \mathcal{A}_v} (\bar{\alpha}_{u,v} + (1 - \bar{\alpha}_{u,v})\bar{\beta}_{u,v}) (t_v - t_u)}.\end{aligned}$$

B. ESTIMATION ALGORITHM FOR VWV MODEL

We briefly describe the estimation algorithm of parameter vector w for the VwV model from an observed data $\mathcal{D}(0, T)$ (see [Kimura et al. 2010b] for more details). As mentioned in Subsection 3.2, the opinion dynamics for the VwV model is invariant to positive scaling of w . Thus, we transform the parameter vector w by $w = w(z)$, where

$$w(z) = (\exp(z_1), \dots, \exp(z_{K-1}), 1), \quad (z = (z_1, \dots, z_{K-1}) \in \mathbf{R}^{K-1}).$$

Namely, our problem is to estimate the value of z that maximize $\mathcal{L}(w(z); \mathcal{D}(0, T))$.

Then, for any $i, j \in \{1, \dots, K-1\}$, we obtain

$$\frac{\partial \mathcal{L}(\mathbf{w}(\mathbf{z}); \mathcal{D}(0, T))}{\partial z_i} = \sum_{(v, t, k) \in \mathcal{C}(0, T)} (\delta_{k, i} - q_i(t, v)),$$

$$\frac{\partial^2 \mathcal{L}(\mathbf{w}(\mathbf{z}); \mathcal{D}(0, T))}{\partial z_i \partial z_j} = \sum_{(v, t, k) \in \mathcal{C}(0, T)} (q_i(t, v) q_j(t, v) - \delta_{i, j} q_i(t, v)),$$

where $\delta_{i, j}$ is the Kronecker's delta, and

$$q_i(t, v) = \frac{n_i(t, v) \exp(z_i)}{n_K(t, v) + \sum_{\ell=1}^{K-1} n_\ell(t, v) \exp(z_\ell)}.$$

We can show that the Hessian matrix $(\partial^2 \mathcal{L}(\mathbf{w}(\mathbf{z}); \mathcal{D}(0, T)) / \partial z_i \partial z_j)$ is negative semi-definite. Hence, by solving the equations $\partial \mathcal{L}(\mathbf{w}(\mathbf{z}); \mathcal{D}(0, T)) / \partial z_i = 0$, ($i = 1, \dots, K-1$), we can find the value of \mathbf{z} that maximizes $\mathcal{L}(\mathbf{w}(\mathbf{z}); \mathcal{D}(0, T))$. We employed a standard Newton Method in our experiments.

REFERENCES

- CASTELLANO, C., MUNOZ, M. A., AND PASTOR-SATORRAS, R. 2009. Nonlinear q -voter model. *Physical Review E* 80, 041129.
- CHEN, W., WANG, C., AND WANG, Y. 2010a. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2010)*. 1029–1038.
- CHEN, W., YUAN, Y., AND ZHANG, L. 2010b. Scalable influence maximization in social networks under the linear threshold model. In *Proceedings of the 10th IEEE International Conference on Data Mining (ICDM 2010)*. 88–97.
- CRANDALL, D., COSLEY, D., HUTTENLOCHNER, D., KLEINBERG, J., AND SURI, S. 2008. Feedback effects between similarity and social influence in online communities. In *Proceedings of KDD 2008*. 160–168.
- DOMINGOS, P. 2005. Mining social networks for viral marketing. *IEEE Intelligent Systems* 20, 80–82.
- EVEN-DAR, E. AND SHAPIRA, A. 2007. A note on maximizing the spread of influence in social networks. In *Proceedings of WINE 2007*. 281–286.
- GOLDENBERG, J., LIBAI, B., AND MULLER, E. 2001. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters* 12, 211–223.
- GOMEZ-RODRIGUEZ, M., LESKOVEC, J., AND KRAUSE, A. 2010. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2010)*. 1019–1028.
- GRUHL, D., GUHA, R., LIBEN-NOWELL, D., AND TOMKINS, A. 2004. Information diffusion through blogspace. *SIGKDD Explorations* 6, 43–52.
- HOLME, P. AND NEWMAN, M. E. J. 2006. Nonequilibrium phase transition in the coevolution of networks and opinions. *Physical Review E* 74, 056108.
- KEMPE, D., KLEINBERG, J., AND TARDOS, E. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2003)*. 137–146.
- KIMURA, M., SAITO, K., AND MOTODA, H. 2008. Minimizing the spread of contamination by blocking links in a network. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI-08)*. 1175–1180.
- KIMURA, M., SAITO, K., AND MOTODA, H. 2009. Blocking links to minimize contamination spread in a social network. *ACM Transactions on Knowledge Discovery from Data* 3, 9:1–9:23.
- KIMURA, M., SAITO, K., NAKANO, R., AND MOTODA, H. 2010a. Extracting influential nodes on a social network for information diffusion. *Data Min. Knowl. Disc.* 20, 70–97.
- KIMURA, M., SAITO, K., OHARA, K., AND MOTODA, H. 2010b. Learning to predict opinion share in social networks. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI-10)*. 1364–1370.
- KIMURA, M., SAITO, K., OHARA, K., AND MOTODA, H. 2011. Detecting anti-majority opinionists using value-weighted mixture voter model. In *Proceedings of the 14th International Conference on Discovery Science (DS 2011)*. LNAI 6926, 150–164.

- KLEINBERG, J. 2002. Bursty and hierarchical structure in streams. In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2002)*. 91–101.
- KLIMT, B. AND YANG, Y. 2004. The enron corpus: A new dataset for email classification research. In *Proceedings of the 2004 European Conference on Machine Learning (ECML'04)*. 217–226.
- LESKOVEC, J., ADAMIC, L. A., AND HUBERMAN, B. A. 2006. The dynamics of viral marketing. In *Proceedings of the 7th ACM Conference on Electronic Commerce (EC'06)*. 228–237.
- LIGGETT, T. M. 1999. *Stochastic interacting systems: contact, voter, and exclusion processes*. Springer, New York.
- MYERS, S. A. AND LESKOVEC, J. 2010. On the convexity of latent social network inference. In *Proceedings of Neural Information Processing Systems (NIPS)*.
- NEWMAN, M. E. J. 2003. The structure and function of complex networks. *SIAM Review* 45, 167–256.
- NEWMAN, M. E. J., FORREST, S., AND BALTHROP, J. 2002. Email networks and the spread of computer viruses. *Physical Review E* 66, 035101.
- OHARA, K., SAITO, K., KIMURA, M., AND MOTODA, H. 2011. Efficient detection of hot span in information diffusion from observation. In *Proceedings of the IJCAI Workshop on Link Analysis in Heterogeneous Information Networks (HINA 2011)*. arXiv: 1110.2659.
- PALLA, G., DERÉNYI, I., FARKAS, I., AND VICSEK, T. 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435, 814–818.
- SAITO, K., KIMURA, M., NAKANO, R., AND MOTODA, H. 2009a. Finding influential nodes in a social network from information diffusion data. In *Proceedings of the International Workshop on Social Computing and Behavioral Modeling (SBP09)*. 138–145.
- SAITO, K., KIMURA, M., OHARA, K., AND MOTODA, H. 2009b. Learning continuous-time information diffusion model for social behavioral data analysis. In *Proceedings of the 1st Asian Conference on Machine Learning (ACML2009)*. LNAI 5828, 322–337.
- SAITO, K., KIMURA, M., OHARA, K., AND MOTODA, H. 2010a. Behavioral analyses of information diffusion models by observed data of social network. In *Proceedings of the 2010 International Conference on Social Computing and Behavioral Modeling (SBP10)*. 149–158.
- SAITO, K., KIMURA, M., OHARA, K., AND MOTODA, H. 2010b. Selecting information diffusion models over social networks for behavioral analysis. In *Proceedings of the 2010 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2010)*. LNAI 6323, 180–195.
- SAITO, K., KIMURA, M., OHARA, K., AND MOTODA, H. 2011a. Detecting changes in opinion value distribution for voter model. In *Proceedings of the 4th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction (SBP2011)*. LNAI 6389, 89–96.
- SAITO, K., OHARA, K., YAMAGISHI, Y., KIMURA, M., AND MOTODA, H. 2011b. Learning diffusion probability based on node attributes in social networks. In *Proceedings of the 19th International Symposium on Methodologies for Intelligent Systems (ISMI S2011)*. LNAI 6804, 153–162.
- SOOD, V. AND REDNER, S. 2005. Voter model on heterogeneous graphs. *Physical Review Letters* 94, 178701.
- SWAN, R. AND ALLAN, J. 2000. Automatic generation of overview timelines. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2000)*. 49–56.
- WATTS, D. J. 2002. A simple model of global cascades on random networks. *Proceedings of National Academy of Science, USA* 99, 5766–5771.
- WATTS, D. J. AND DODDS, P. S. 2007. Influence, networks, and public opinion formation. *Journal of Consumer Research* 34, 441–458.
- WU, F. AND HUBERMAN, B. A. 2008. How public opinion forms. In *Proceedings of WINE 2008*. 334–341.
- YAMAGISHI, Y., SAITO, K., OHARA, K., KIMURA, M., AND MOTODA, H. 2011. Learning attribute-weighted voter model over social networks. In *Proceedings of the 3rd Asian Conference on Machine Learning (ACML 2011)*, to appear. JMLR Workshop and Conference Proceedings.
- YANG, H., WU, Z., ZHOU, C., ZHOU, T., AND WANG, B. 2009. Effects of social diversity on the emergence of global consensus in opinion dynamics. *Physical Review E* 80, 046108.

Received June 2011; revised November 2011; accepted December 2011