Asian Conference on Machine Learning

Estimating Diffusion Probability Changes for AsIC-SIS Model from Information Diffusion Results

Akihiro Koide   J11103@U-SHIZUOKA-KEN.AC.JP
Graduate School of Management and Information of Innovation
University of Shizuoka

Kazumi Saito   K-SAITO@U-SHIZUOKA-KEN.AC.JP
School of Management and Information
University of Shizuoka

Kouzou Ohara   OHARA@IT.AOYAMA.AC.JP
Department of Integrated Information Technology
Aoyama Gakuin University

Masahiro Kimura   KIMURA@RINS.RYUKOKU.AC.JP
Department of Electronics and Informatics
Ryukoku University

Hiroshi Motoda   MOTODA@AR.SANKEN.OSAKA-U.AC.JP
Institute of Scientific and Industrial Research
Osaka University

Abstract

We address the problem of estimating changes in diffusion probability over a social network from the observed information diffusion results, which is possibly caused by an unknown external situation change. For this problem, we focused on the asynchronous independent cascade (AsIC) model in the SIS (Susceptible/Infected/Susceptible) setting in order to meet more realistic situations such as communication in a blogosphere. This model is referred to as the AsIC-SIS model. We assume that the diffusion parameter changes are approximated by a series of step functions, and their changes are reflected in the observed diffusion results. Thus, the problem is reduced to detecting how many step functions are needed, where in time each one starts and how long it lasts, and what the hight of each one is. The method employs the derivative of the likelihood function of the observed data that are assumed to be generated from the AsIC-SIS model, adopts a divide-and-conquer type greedy recursive partitioning, and utilizes an MDL model selection measure to determine the adequate number of step functions. The results obtained using real world network structures confirmed that the method works well as intended. The MDL criterion is useful to avoid overfitting, and the found pattern is not necessarily the same in terms of the number of step functions as the one assumed to be true, but the error is always reduced to a small value.

Keywords: pattern change detection, information diffusion, parameter learning, social networks

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 measures such as degree, betweenness, closeness, etc., are not enough to identify the important nodes (Kimura et al., 2009a, 2010a). Information diffusion is modeled typically by probabilistic models. Most representative and fundamental ones 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). In the IC model the information sender (a node) tries to push the information to the neighboring receivers (child nodes) in a probabilistic way, whereas in the LT model the information receiver (a node) tries to pull the information from the neighoring senders (parents nodes) in a probabilistic way. These models place the constraint that a node is given a single chance to activate the other node, *i.e.*, the same node is not activated multiple times. This setting is called SIR (Susceptible/Infectious/Recovered) in analogy with epidemic disease. Explicit use of these models to solve such problems as the *influence maximization problem* (Kempe et al., 2003; Kimura et al., 2010a) and the *contamination minimization problem* (Kimura et al., 2009a) clearly shows the advantage of the model. They showed that the identified influential nodes and links are considerably different from the ones identified by the standard centrality measures. The SIR setting is simple, but does not model well such communication as in a blogosphere where the same person can post on the same topic multiple times. The SIS (Susceptible/Infectiouse/Susceptible) setting is better suited to this situation, where a node is allowed to activate the other nodes multiple times, *i.e.*, the same node is activated multiple times (Kimura et al., 2009b).

What is common to all the above models is that they are all probabilistic models and have parameters to characterize the information diffusion, and these parameters are assumed to be stationary, *i.e.*, they do not change over time. Evidently, the parameters must be known in advance for the model to be usable for analysis, but it is generally difficult to determine the values of these parameters theoretically. Therefore, attempts have been made to learn these parameter values by the observed information diffusion sequence data (Saito et al., 2009a,b, 2010a,b; Gomez-Rodriguez et al., 2010; Myers and Leskovec; 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 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.

This paper also deals with a parameter learning problem, but addresses a different aspect of information diffusion. We do not assume that the parameter values are stationary, but allow that they change over time. They may change abruptly or gradually depending on the cause of changes which we do not know. Ideally we intend to be able to deal with any shape of changes over time. However, in this paper, we limit the change pattern to those that can be approximated by a series of step functions, and further assume that the change takes place uniformly in space, *i.e.*, the parameters of all nodes change in the same way. We use AsIC-SIS, Asynchronous Independent Cascade model in SIS setting. This is a model in which the original discrete time step IC-SIR model is extended to continuous time model allowing asynchronous time delay (Saito et al., 2009b, 2010a)

as well as allowing multiple activations of the same nodes. We learn the parameter values from an observed sequence of information diffusion under AsIC-SIS model setting, *i.e.*, the problem is reduded to detecting how many step functions are needed, where in time each one starts and how long it lasts, and what the hight of each one is. This is viewed as a generalization of our previous work (Ohara et al., 2011) in which we used the AsIC-SIR model, limited the change pattern to be a single rect-linear shape, and devised an efficient algorithm which searches the optimal window. However, this algorithm works only to this restricted type of the problem.

We extended the parameter optimization algorithm that was developed in Saito et al. (2009b); Kimura et al. (2010b), *i.e.*, the EM-like algorithm for the AsIC-SIR model that iteratively updates the values to maximize the model's likelihood of generating the observed data sequences, to AsIC-SIS. The core part of this paper is how to efficiently search the change pattern. We employed the idea of using the first order derivative of the likelihood with respect to the parameters (Ohara et al., 2011), and newly developed an efficent algorithm that uses a divide-and-conquer type greedy recursive partitioning as a search strategy and an MDL model selection measure as a stopping criterion to determine the most adequate number of step functions. We tested our algorithm to artificially generated change patterns using four real world network structures. The results obtained confirmed that the method works well as intended. The algoritm is efficent because it needs to do expensive parameter optimization only once for each partitioning (which is not that many in many cases). The MDL criterion is useful to avoid overfitting. In many cases it identifies the correct number of step functions, but in some cases the found pattern is not necessarily the same in terms of the number of step functions, but the error is always reduced to a small value.

The paper is organized as follows. After very briefly introducing the AsIC-SIS model in Section 2, we define the problem in Section 3 and derive the liklihood function in Section 4, which is the objective function to be maximized. The parameter estimation algorithm is summarized in Appendix. We then describe how we efficiently search for the change pattern in Section 5 together with the restricted search method used for comparative study. The experimental results are reported in Section 6. We end this paper by summarizing the main result in Section 7.

2. Information Diffusion Model

An SIS model for the spread of a disease is based on the cycle of disease in a host. A person is first *susceptible* to the disease, and becomes *infected* with some probability and time-delay if he or she has contact with an infected person. The infected person becomes susceptible to the disease again without moving to the immune state. We consider an asynchronous-time SIS model for information diffusion on a network. In this context, infected nodes mean that the nodes have adopted the information, and we call these infected nodes *active* nodes. This can be mapped to realistic situations such as communication in a blogsphere. A typical example would be the following propagation phenomenon of a topic in the blogsphere: A blogger who has not yet posted a message about a certain topic becomes interested in the topic by reading the blog of his or her friend, and posts a message about it with some time-delay from the friend's posting time, *i.e.*, becoming infected (activated) with some time-delay. Right after posting the message, the same blogger can read any other blogs of his or her friends, *i.e.*, becoming susceptible again. The same blogger reads a new message about the topic posted by some other friend, and may post another message, *i.e.*, becoming infected again. This process is repeated.

Let $G = (V, E)$ be a directed network, where V and E stand for the sets of all the nodes and (directed) links, respectively. Here, note that E is a subset of $V \times V$. 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 $B(v) = \{u \in V; (u, v) \in E\}$. We define the AsIC-SIS model for information diffusion on G. In the model, the diffusion process unfolds in continuous-time $t \geq 0$, and it is assumed that the state of a node is either active or inactive. For every link $(u, v) \in E$, we specify a real value $p_{u,v}$ with $0 < p_{u,v} < 1$ in advance. Here, $p_{u,v}$ is referred to as the *diffusion probability* through link (u, v). Given an initial active node v and a time span T, the diffusion process proceeds in the following way. Suppose that node u becomes active at time t ($< T$). Then, node u attempts to activate every $v \in F(u)$, and succeeds with probability $p_{u,v}$. If node u succeeds, then node v will become active at time $t + \delta$. We assume that a delay-time δ is chosen from some probability distribution, and we used the exponential distribution with parameter $r_{u,v}$ for the sake of convenience, but of course other distributions such as power-law and Weibull can be employed. Suppose that u, one of the parent nodes of v, succeeds to activate v at some time after some delay. In our SIS model, when some other parent node also succeeds to activate v before it gets activated by u, we assume that v's activation time is overridden by the one earliest possible. On the other hand, node u gets back inactive right after time t (the time it gets activated) and it can only be reactivated by those parent nodes that have become active after time t[1]. The process terminates if the current time reaches the time limit T.

The AsIC-SIS model is the SIS version of the asynchronous independent cascade (AsIC) model proposed by Saito et al. (2009b) that is an extension of the independent cascade (IC) model studied by Kempe et al. (2003). As mentioned earlier, the AsIC-SIS model was extended to meet more realistic situations.

3. Problem Definition

We address the *problem of estimating diffusion probability changes*. In this problem, we assume that some changes have happened in the way the information diffuses, and we observe the diffusion results of a certain topic in which the changes are embedded, and consider estimating the diffusion probability as a function with respect to time t.

An information diffusion result generated by the AsIC-SIS model is represented as a set of pairs of active nodes and their activation times; *i.e.*, $\{\cdots, (v(\eta), t_{v(\eta)}), \cdots\}$, where $v(\eta)$ indicates v's η-th activation. 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. In fact, our previous experiments (Saito et al., 2009b, 2010a,b) give some evidences which support the validity of this constraint.

Let $p(t)$ be a function of diffusion probability with respect to time t. Here we assume that $p(t)$ is reasonably approximated by combining a number of step functions, *i.e.*,

$$p(t) = p_{i-1} \quad \text{if } t \in [t_{i-1}, t_i), \quad i \in \{1, \cdots, K + 1\}, \tag{1}$$

---

1. In theory we can go back to all the past time points at which the parents of u got activated multiple times in the past, but this is unrealistic and we thought it natural to limit the parents only to those that got activated after time t.

where $t_0 = 0 < \cdots < t_i < \cdots t_{K+1} = T$ and K stands for the number of change points. 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 diffusion probability estimation problem is reduced to detecting the change points $\{t_1, \cdots, t_K\}$ and estimating the associated diffusion probabilities $\{p_0, \cdots, p_K\}$ from the observed diffusion result $\mathcal{D}(0, T)$. For a given integer K, we define the *change point vector* $\boldsymbol{t}_K$ and the *diffusion-probability vector* $\boldsymbol{p}_K$ by $\boldsymbol{t}_K = (t_1, \cdots, t_K)$ and $\boldsymbol{p}_K = (p_0, \cdots, p_K)$, respectively.

4. Model parameter learning

We describe the framework of model parameter learning as a likelihood maximization problem for the AsIC-SIS model.

First, we consider estimating the values of diffusion probability p and time-delay parameter r from an observed diffusion result $\mathcal{D}(0, T) = \{\cdots, (v(\eta), t_{v(\eta)}), \cdots\}$ when there is no change point. 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., $\mathcal{D} = \{v(\eta) \in V; \ (v(\eta), t_{v(\eta)}) \in \mathcal{D}(0, T)\}$. For each node $v(\eta) \in \mathcal{D}$, let $\mathcal{AP}_{v(\eta)}$ be the set of its parent nodes that had a chance to activate it, i.e.,

$$\mathcal{AP}_{v(\eta)} = \{u(\zeta); \ u \in B(v), \ (u(\zeta), t_{u(\zeta)}) \in \mathcal{D}(0, T), \ t_{v(\eta-1)} < t_{u(\zeta)} < t_{v(\eta)}\},$$

and $\mathcal{NC}_{v(\eta)}$ be the set of its child nodes that was not activated by a node $v(\eta)$ within $(t_{v(\eta)}, T)$, i.e.,

$$\mathcal{NC}_{v(\eta)} = \{z \in F(v); \ \neg\exists\, z(\xi), \ \text{s.t.} \ (z(\xi), t_{z(\xi)}) \in \mathcal{D}(0, T), \ t_{v(\eta)} < t_{z(\xi)} < T)\}.$$

Note that from the observed diffusion result, we know that a node v at the η-th activation did not succeed to activate any child node in $\mathcal{NC}_{v(\eta)}$ within the time limit T, and we use this fact for our parameter estimation in order to improve its performance.

Let $\mathcal{X}_{u(\zeta),v(\eta)}(p, r)$ denote the probability density that a node $u(\zeta) \in \mathcal{AP}_{v(\eta)}$ activates the node $v(\eta)$ at time $t_{v(\eta)}$, that is,

$$\mathcal{X}_{u(\zeta),v(\eta)}(p, r) \;=\; p\, r \exp(-r(t_{v(\eta)} - t_{u(\zeta)})). \tag{2}$$

Let $\mathcal{Y}_{u(\zeta),v(\eta)}(p, r)$ denote the probability that the node $v(\eta)$ is not activated by a node $u(\zeta) \in \mathcal{AP}_{v(\eta)}$ within the time-period $(t_{u(\zeta)}, t_{v(\eta)})$, that is,

$$
\begin{aligned}
\mathcal{Y}_{u(\zeta),v(\eta)}(p, r) \;&=\; 1 - p \int_{t_{u(\zeta)}}^{t_{v(\eta)}} r \exp(-r(t - t_{u(\zeta)}))dt \\
&=\; p \exp(-r(t_{v(\eta)} - t_{u(\zeta)})) + (1 - p).
\end{aligned}
\tag{3}
$$

By using Eqs. (2) and (3), we can obtain the probability density $\phi_{v(\eta)}(p, r)$ that some node $u(\zeta) \in \mathcal{AP}_{v(\eta)}$ succeeds to activate a node $v(\eta)$ at a time $t_{v(\eta)}$,

$$\phi_{v(\eta)}(p, r) \;=\; \sum_{u(\zeta) \in \mathcal{AP}_{v(\eta)}} \mathcal{X}_{u(\zeta),v(\eta)}(p, r) \left(\prod_{z(\xi) \in \mathcal{AP}_{v(\eta)} \setminus \{u(\zeta)\}} \mathcal{Y}_{z(\xi),v(\eta)}(p, r) \right). \tag{4}$$

and the probability $\psi_{v(\eta)}(p, r)$ that a node $v(\eta)$ cannot activate any node $z \in \mathcal{NC}_{v(\eta)}$ within $(t_{v(\eta)}, T)$,

$$\psi_{v(\eta)}(p, r) \;=\; \left(p \exp(-r(T - t_{v(\eta)})) + (1 - p) \right)^{|\mathcal{NC}_{v(\eta)}|}. \tag{5}$$

Then, from Eqs. (4) and (5), the following log likelihood function $\mathcal{L}(p, r; \mathcal{D}(0, T))$ can be obtained for observed data $\mathcal{D}(0, T)$

$$\mathcal{L}(p, r; \mathcal{D}(0, T)) = \sum_{v(\eta) \in \mathcal{D}} \left(\log \phi_{v(\eta)}(p, r) + \log \psi_{v(\eta)}(p, r) \right). \tag{6}$$

The values of parameters p and r can be stably obtained by maximizing Eq. (6) using an EM-like algorithm. (see Appendix A for more details).

Now, we assume that there exist change points specified by the change point vector $\boldsymbol{t}_K$ and the associated diffusion-probability vector $\boldsymbol{p}_K$. For any $v(\eta) \in \mathcal{D}(0, T)$, let $\phi_{v(\eta)}(\boldsymbol{p}_K, r; \boldsymbol{t}_K)$ be the probability density that some node $u(\zeta) \in \mathcal{AP}_{v(\eta)}$ succeeds to activate a node $v(\eta)$ at time $t_{v(\eta)}$, i.e.,

$$\phi_{v(\eta)}(\boldsymbol{p}_K, r; \boldsymbol{t}_K) = \sum_{u(\zeta) \in \mathcal{AP}_{v(\eta)}} \mathcal{X}_{u,v}(p(t_{u(\zeta)}), r) \prod_{z(\xi) \in \mathcal{AP}_{v(\eta)} \setminus \{u(\zeta)\}} \mathcal{Y}_{z,v}(p(t_{z(\xi)}), r) \tag{7}$$

and $\psi_{v(\eta)}(p(t_{v(\eta)}), r; \boldsymbol{t}_K)$ be the probability that a node $v(\eta)$ cannot activate any node $z \in \mathcal{NC}_{v(\eta)}$ within $(t_{v(\eta)}, T]$, i.e.,

$$\psi_{v(\eta)}(p(t_{v(\eta)}), r; \boldsymbol{t}_K) = \left(p(t_{v(\eta)}) \exp(-r(T - t_{v(\eta)})) + (1 - p(t_{v(\eta)})) \right)^{|\mathcal{NC}_{v(\eta)}|}. \tag{8}$$

Using Eqs. (7) and (8), we can define the following objective function $\mathcal{L}(\boldsymbol{p}_K, r; \mathcal{D}(0, T), \boldsymbol{t}_K)$.

$$\mathcal{L}(\boldsymbol{p}_K, r; \mathcal{D}(0, T), \boldsymbol{t}_K) = \sum_{v(\eta) \in \mathcal{D}} \left(\log \phi_v(\eta)(\boldsymbol{p}_K, r; \boldsymbol{t}_K) + \log \psi_{v(\eta)}(p(t_{v(\eta)}), r; \boldsymbol{t}_K) \right). \tag{9}$$

Clearly, $\mathcal{L}(\boldsymbol{p}_K, r; \mathcal{D}(0, T), \boldsymbol{t}_K)$ is expected to be maximized by setting $\boldsymbol{t}_K$ to the true change points vector $\boldsymbol{t}_K^* = (t_1^*, \cdots, t_K^*)$ if a substantial amount of data $\mathcal{D}(0, T)$ is available. Thus, our diffusion probability estimation problem is formalized as the following maximization problem:

$$\hat{\boldsymbol{t}_K} = \arg \max_{\boldsymbol{t}_K} \mathcal{L}(\hat{\boldsymbol{p}}_K(\boldsymbol{t}_K), \hat{r}(\boldsymbol{t}_K); \mathcal{D}(0, T), \boldsymbol{t}_K), \tag{10}$$

where $\hat{\boldsymbol{p}}_K(\boldsymbol{t}_K)$ and $\hat{r}(\boldsymbol{t}_K)$ denote the maximum likelihood estimators for a given $\boldsymbol{t}_K$.

5. Estimation Methods

For a given number of change points, K, in order to obtain the optimal change point vector $\hat{\boldsymbol{t}}_K$ according to Eq. (10), we need to prepare a reasonable set of candidate change poits, denoted by $\mathcal{T}$. One way of doing so is to construct $\mathcal{T}$ by considering all of the observed activation time points.

$$\mathcal{T} = \{t_{v(\eta)}; (v(\eta), t_{v(\eta)}) \in \mathcal{D}(0, T)\} \cup \{T\} = \{\tau_0, \tau_1, \cdots, \tau_N\}, \quad (0 = \tau_0 < \tau_1 < \cdots < \tau_N = T).$$

Here N is equal to the number of activated nodes in a information diffusion result, i.e., $N = |\mathcal{D}(0, T)|$. Hereafter, we denote the model parameter vector by $\boldsymbol{\theta}_K$; i.e., $\boldsymbol{\theta}_K = (\boldsymbol{p}_K, r)$ for the AsIC-SIS model.

5.1. Proposed Method

Our proposed method employs a greedy strategy. Clearly, we can obtain the parameter vector $\boldsymbol{\theta}_0$ from the original objective function of Eq. (6). Now, under the condition that we have obtained the K chage point(s), we consider selecting the next $(K + 1)$-th change point. Of course, we can obtain the maximum likelihood estimators, $\hat{\boldsymbol{\theta}}_K$, from the extended objective function of Eq. (9). Then, we focus on the first-order partial derivative of the objective function $\mathcal{L}(\hat{\boldsymbol{\theta}}_K; \mathcal{D}(0, T))$ with respect to a new parameter $p_{v(\eta)}$ introduced by considering as if each node $v \in V$ has an individual diffusion probability $p_{v(\eta)}$ at each activation time $t_{v(\eta)}$. Note that under this situation, by posing the restriction of parameter sharing setting, defined by $p_{v(\eta)} = p_i$ if $t_{v(\eta)} \in [t_i, t_{i+1})$, we obtain each maximum likelihood estimator by $\hat{p}_{v(\eta)} = \hat{p}_i$. Thus, from the optimal necessary condition of the maximum likelihood estimation, we have

$$0 = \frac{\partial \mathcal{L}(\hat{\boldsymbol{\theta}}_K; \mathcal{D}(0, T))}{\partial p_i} = \sum_{t_{v(\eta)} \in [t_i, t_{i+1})} \frac{\partial \tilde{\mathcal{L}}(\hat{\boldsymbol{\theta}}_K; \mathcal{D}(0, T))}{\partial p_{v(\eta)}}. \tag{11}$$

Now we assume that there exists an undetected change point $t_j \in [t_i, t_{i+1})$. Then the estimated parameter $\hat{p}_i$ for the time span $[t_i, t_{i+1})$ is nothing but a compromised value between diffusion probabilities of $[t_i, t_j)$ and $[t_j, t_{i+1})$. Thus, we can expect that the following relation holds for the product of the partial derivatives between many pairs of $p_{u(\zeta)}$ and $p_{v(\eta)}$ if both $t_{u(\zeta)}$ and $t_{v(\eta)}$ are included in either before the change point $[t_i, t_j)$ or after the change point $[t_j, t_{i+1})$.

$$\frac{\partial \tilde{\mathcal{L}}(\hat{\boldsymbol{\theta}}_K; \mathcal{D}(0, T))}{\partial p_{u(\zeta)}} \frac{\partial \tilde{\mathcal{L}}(\hat{\boldsymbol{\theta}}_K; \mathcal{D}(0, T))}{\partial p_{v(\eta)}} > 0 \tag{12}$$

Here, we consider the following partial sum for the derivatives:

$$g(\tau_n) = \sum_{t_{v(\eta)} < \tau_n} \frac{\partial \tilde{\mathcal{L}}(\hat{\boldsymbol{\theta}}_K; \mathcal{D}(0, T))}{\partial p_{v(\eta)}}, \quad n = 1, \cdots, N, \tag{13}$$

where $g(\tau_n) = 0$ if $\tau_n = t_i$. By Eqs. (11) to (13), we can expect that $|g(n)|$ is locally maximized at each undetected change point $\tau_n = t_j$. This is because the sign of the product of the partial derivatives $\partial \mathcal{L}(\hat{\boldsymbol{\theta}}_K; \mathcal{D}(0, T))/\partial p_{u(\zeta)}$ and $\partial \mathcal{L}(\hat{\boldsymbol{\theta}}_K; \mathcal{D}(0, T))/\partial p_{v(\eta)}$ changes at the boundaries of the undetected change points $\{t_j\}$. Therefore, we propose the method of detecting the next change point by

$$\hat{\tau}_n = \arg\max_{\tau_n \in \mathcal{T}} |g(\tau_n)|. \tag{14}$$

Here note that we can incrementally calculate $g(\tau_n)$. More specifically, we can obtain the following formula by $t_{v(\eta)} = \tau_{n+1}$:

$$g(\tau_{n+1}) = g(\tau_n) + \frac{\partial \mathcal{L}(\hat{\boldsymbol{\theta}}_K; \mathcal{D}(0, T))}{\partial p_{v(\eta)}} \tag{15}$$

for any $\tau_n, \tau_{n+1} \in \mathcal{T}$.

Thus far, we assumed that the number of change points, K, is known. However, since this assumption does not hold in many applications, we need to obtain an adequate K from a given

diffusion result. For this purpose, we can utilize some statistical measure such as MDL (Minimum Description Length). Note that due to a time-series nature of our observation data, we cannot straightforwardly apply a resampling technique such as k-fold cross-validation for this model selection. In our experiments, we employed the following MDL value.

$$MDL(\boldsymbol{\theta}_K) \quad = \quad -\mathcal{L}(\hat{\boldsymbol{\theta}}_K; \mathcal{D}(0, T)) + (K + 1)\log M, \quad M \quad = \quad \sum_{v(\eta) \in \mathcal{D}} |F(v(\eta))|, \qquad (16)$$

where $K + 1$ and M correspond to the number of parameters and the number of coin-flips performed by the AsIC-SIS model, respectively. Note that we regard M as the number of samples for our learning. Then we can summarize our proposed method below.

1. Set $K = 0$ and $\boldsymbol{t}_0$ to an empty list, and intialize $\boldsymbol{\theta}_0$ adequately.
2. Maximize $\mathcal{L}(\boldsymbol{\theta}_K; \mathcal{D}(0, T))$ by using the parameter estimation method, and calculate $MDL(\boldsymbol{\theta}_K)$.
3. If $K > 0$ and $MDL(\boldsymbol{\theta}_K) > MDL(\boldsymbol{\theta}_{K-1})$, output $\boldsymbol{t}_{K-1}$ and $\boldsymbol{\theta}_{K-1}$.
4. Detect the change point $\hat{\tau}_n$ by Eq. (14), construct $\boldsymbol{t}_{K+1}$ by adding $\hat{\tau}_n$ to $\boldsymbol{t}_K$, set $K = K + 1$, and return to step 2.

Here note that the proposed method requires likelihood maximization by using the parameter estimation method only $(K + 1)$ times.

5.2. Comparison Method

As mentioned earlier, we have already proposed a hot span detection method for the AsIC model in the SIR (Susceptible/Infected/Recover) setting, although this method is only applicable to a restricted form of the change pattern expressed by a pair of $\boldsymbol{t}_2 = (t_1, t_2)$ and $\boldsymbol{p}_2 = (p_0, p_1, p_0)$ (Ohara et al., 2011). The results reported are good. Thus, we extend this method to the SIS (Susceptible/Infected/ Susceptible) setting, and use the extended method for performance comparison, knowing that the method is intended to a single rect-linear pattern change. In what follows, we outline this method.

The comparison method also utilizes a modified version of Eq. (13) as the measure of interval selection, expressed by

$$[\hat{\tau}_m, \hat{\tau}_n) \quad = \quad \arg\max_{\tau_m, \tau_n \in \mathcal{T}} \left| \sum_{t_{v(\eta)} \in [\tau_m, \tau_n)} \frac{\partial \mathcal{L}(\hat{\boldsymbol{\theta}}_K; \mathcal{D}(0, T))}{\partial p_{v(\eta)}} \right|. \qquad (17)$$

However, this method can be extremely inefficient when the number of candidate time points N 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 randomly selecting J points from $\mathcal{T}$; then we construct a restricted set of candidate spans by

$$\mathcal{H}_J \quad = \quad \{S = [\tau_i, \tau_j); \ \tau_i < \tau_j, \tau_i \in \mathcal{T}_J, \tau_j \in \mathcal{T}_J\}.$$

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

6. Experimental Evaluation

We experimentally evaluated, given an observed diffusion result, how accurately the proposed method can estimate diffusion probability changes underlying it by investigating the difference between the estimated change pattern and the one that is assumed true using four real world networks.

6.1. Datasets

Here we adopted four large networks in the real world, all of which are bidirectional. The first one is a trackback network of Japanese blogs used in Kimura et al. (2009a), where there are $12,047$ nodes and $79,920$ directed links (the blog network). The second one is a network representing the co-occurrence relation extracted from the "list of people" within Japanese Wikipedia that is used in Kimura et al. (2008), which has $9,481$ nodes and $245,044$ directed links (the Wikipedia network). The third one is a network derived from the Enron Email Dataset (Klimt and Yang, 2004) where the sender and the recipient extracted from the dataset were linked if they had bidirectional communications. It contains $4,254$ nodes and $44,314$ directed links (the Enron network). The last one is a coauthorship network employed in Palla et al. (2005). It has $12,357$ nodes and $38,896$ directed links (the coauthorship network).

6.2. Experimental Setting

We generated diffusion results using the AsIC-SIS model for each of the above networks under the following setting. We considered $p = 1/\bar{d}$ as the base value of the diffusion probability of each link in a network, where $\bar{d}$ is the mean out-degree of the network. For an arbitrary node in the network, the expected number of its children that it succeeds to activate is approximately one at least at an early phase of the information diffusion for this base value. If the diffusion probability is much smaller than the base value, the diffusion process could terminate soon resulting in only few active nodes on the average. If it is much larger, the information rapidly spreads out the entire network and the majority of nodes could be active at any time point in the process, which would also be unrealistic. As a result, too little or too much amount of information diffusion is inappropriate to our aim of investigating the diffusion probability change estimation. Thus, we set the initial diffusion probability, p_0, to be a value slightly smaller than the base value, which is 0.10 for the blog network, 0.02 for the Wikipedia network, 0.05 for the Enron network, and 0.20 for the Coauthorship network, respectively. We considered two kinds of change pattern: one is a rect-linear pattern that has two change points, which is the same as the one used in Ohara et al. (2011) and can be regarded as the most fundamental; and the other is a two-step pattern having three change points, which represents a situation where an event that caused an increase in the diffusion probability of a certain topic occurred, followed by an even bigger event that futher increased the probability, and then the probability returned back to the normal value due to the cease of the event. As for the former pattern, we set the diffusion probability during the second period, p_1, to be three times as large as p_0, and the probability during the third period, p_2, to be the same as p_0. Table 1 summarizes the diffusion probability $\boldsymbol{p}_2^*$ that is assumed true. For all the networks we used the same $\boldsymbol{t}_2^* = (10, 15)$ as the change point vector that is assmued true and $T = 20$ as the final observation time. As for the latter pattern, we set the second and the third diffusion probability, p_1 and p_2, to be twice and three times as large as p_0, respectively, and the last one, p_3 to be the same as p_0. Table 2 summarizes the diffusion probability $\boldsymbol{p}_3^*$ that is assumed true. We used $\boldsymbol{t}_3^* = (10, 15, 20)$ and $T = 25$ for all

Table 1: The diffusion probability $\boldsymbol{p}_2^*$ that is assumed true for each of the networks .

diffusion probability ($\boldsymbol{p}_2^*$)	Blog	Wikipedia	Enron	Coauthorship
p_0	0.10	0.02	0.05	0.20
p_1	0.30	0.06	0.15	0.60
p_2	0.10	0.02	0.05	0.20

Table 2: The diffusion probability $\boldsymbol{p}_3^*$ that is assumed true for each of the networks .

diffusion probability ($\boldsymbol{p}_3^*$)	Blog	Wikipedia	Enron	Coauthorship
p_0	0.10	0.02	0.05	0.20
p_1	0.20	0.04	0.10	0.40
p_2	0.30	0.06	0.15	0.60
p_3	0.10	0.02	0.05	0.20

the networks. As we mentioned in Section 3, we assumed that the time delay parameter does not change, and fixed its value to be 1 ($r = 1$) for every network as changing r works only for scaling the time axis of the diffusion results. In all we generated 100 information diffusion results for each pattern, using the above parameter values, each starting from a randomly selected initial active node for each network.

The initial values for p_0 and r were set to a reasonably small random value and a random value around 1, respectively. The termination condition of our parameter learning was as follws:

$$\max_{\theta_i \in \boldsymbol{\theta}_K} |\partial \mathcal{L}(\boldsymbol{\theta}_K; \mathcal{D}(0, T))/\partial \theta_i| < 10^{-4}.$$

We then estimated both the change point vector $\hat{\boldsymbol{t}}_K$ and the model parameter vector $\hat{\boldsymbol{\theta}}_K$, and evaluated their accuracy by integrating the absolute error of the estimated diffusion probability with respect to time t, $i.e.$,

$$\mathcal{E} = \int_0^T |p^*(t) - \hat{p}(t; \hat{\boldsymbol{t}}_K, \hat{\boldsymbol{\theta}}_K)| dt,$$

where $p^*(t)$ is the diffusion probability that is assumed true at time t and $\hat{p}(t; \hat{\boldsymbol{t}}_K, \hat{\boldsymbol{\theta}}_K)$ is its estimation. The estimation with a smaller $\mathcal{E}$ is a better approximation of the true change pattern. In this regards it is not essential that the estimated number of change points, $\hat{K}$, is identical to K^*, the number of change points used to generate the diffusion result. What matters is how close is the estimated pattern as a whole to the true patteren. In fact, K^* is unknown in reality.

6.3. Experimental Results

Table 3 summarizes the results for the first (rect-linear) change pattern, where the integrated estimation errors are the average over independent 100 trials for distinct 100 diffusion results. Here we executed our method until $K = 10$ ignoring the stopping condition at Step 3 of the algorithm shown in Section 5.1, and investigated how the estimation error $\mathcal{E}$ changes over K. The value in the parenthes is the number of trials where the MDL value defined by Eq. (16) took the minimal at that K, which is what the proposed method outputs as the optimal pattern. The row indicated by "MDL estimation" contains the averaged integral error of such optimal patterns. In addition, we showed

Table 3: Integral error $\mathcal{E}$ of the proposed method averaged over 100 trials to estimate a rect-linear change pattern (the value in parentheses is the number of trials where the obtained pattern took the minimal MDL value at K).

♯change points (K)	Blog		Wikipedia		Enron		Coauthorship	
0	1.296	(0)	0.273	(5)	0.692	(0)	3.494	(0)
1	1.610	(0)	0.348	(0)	0.575	(0)	3.575	(0)
2 (= K^*)	0.126	(64)	0.150	(25)	0.025	(74)	0.614	(7)
3	0.130	(12)	0.108	(41)	0.029	(12)	0.176	(31)
4	0.134	(16)	0.099	(10)	0.032	(6)	0.162	(29)
5	0.136	(4)	0.084	(7)	0.036	(4)	0.156	(12)
6	0.139	(1)	0.081	(4)	0.037	(4)	0.153	(6)
7	0.139	(2)	0.075	(3)	0.039	(0)	0.155	(9)
8	0.139	(0)	0.070	(4)	0.041	(0)	0.155	(1)
9	0.140	(1)	0.070	(1)	0.044	(0)	0.157	(5)
MDL estimation	0.122		0.060		0.022		0.117	
Comparison method	0.120		0.047		0.028		0.117	

Table 4: Integral error $\mathcal{E}$ of the proposed method averaged over 100 trials to estimate a two-step change pattern (the value in parentheses is the number of trials where the obtained pattern took the minimal MDL value at K).

♯change points (K)	Blog		Wikipedia		Enron		Coauthorship	
0	1.500	(0)	0.358	(1)	0.750	(0)	3.837	(0)
1	1.725	(0)	0.379	(0)	0.420	(0)	3.721	(0)
2	0.871	(0)	0.213	(18)	0.324	(0)	1.889	(0)
3 (= K^*)	0.133	(95)	0.138	(37)	0.128	(12)	0.279	(32)
4	0.135	(3)	0.116	(18)	0.057	(18)	0.157	(37)
5	0.135	(2)	0.113	(10)	0.052	(20)	0.149	(15)
6	0.135	(0)	0.107	(8)	0.046	(29)	0.154	(9)
7	0.135	(0)	0.107	(4)	0.047	(11)	0.155	(4)
8	0.135	(0)	0.107	(2)	0.046	(12)	0.162	(0)
9	0.135	(0)	0.107	(2)	0.047	(8)	0.169	(3)
MDL estimation	0.133		0.103		0.038		0.123	
Comparison method	0.845		0.180		0.321		2.043	

the estimation error for the comparison method described in Section 5.2 in the row indicated by "Comparison method" as a reference value for evaluation, where J was set to $1,000$.

From these results, we see that the esitmation error drastically drops down at $K = 2 (= K^*)$ for every network, which means that the proposed method succeeded in detecting the correct change points and estimating the diffusion probabilities in good accuracy. In fact, the errors of the optimal patterns obtained by the proposed method (shown in the row indicated by "MDL estimation") are

very favorably comparable to those obtained by the comarison method that is optimized solely to a single rect-linear pattern used here. Further, the comparison method explicitly uses the constraint $p_0 = p_2$, but the proposed method does not use this constraint and estimates p_2 independenly of p_0. This implies that the pattern obtained by the proposed method can be a good approximation of the changes of the diffusion probability underlying the observed diffusion result. The number of trials where the MDL reaches the minimum is largest either at $K = 2$ or 3, which means that the MDL criteron works well to avoid an over-fitting that could be attained by introducing many change points. There are some differences in the performace over the networks. We observe that there are more trias that the MDL criterion gives a larger K than the correct K^* for Wikipedia and Coauthorship netowrks. This is mainly attiributed to the diffusion data we used. However, more deeper alanysis is needed to understand what causes this difference, but it is true to say that the error is always small enough for the MDL results on the average.

Table 4 show the results for the second (two-step) pattern. The results are qualitatively the same as in the first pattern. The estimation error drops down drastically at $K = 3(= K^*)$ and the MDL value takes the minimum at around $K = 3$ in most of the cases. For every network, the estimation errors of the optimal patterns obtained by the proposed method are about the same to those for the first pattern, and are much better than those obtained by the corresponding comparison method. In fact, it is unfair to compare the results with the comparison method because the latter is not designed to detect patterns other than the rect-linear shape, It simply shows that the comparison method cannot approximate the correct pattern by any means. The proposed method can estimate the underlying diffusion probability change in good accuracy, and the MDL based criterion to select an optimal K works well as intended also for the case of two-step pattern.

In order to analyze our experimental results more closely, we examined the diffusion probability patterns obtained by our proposed method. Figure 1 shows typical examples of desirable and undesirable cases for Wikipedia network by which a relatively large number of undesirable ones were observed. Here we simply denoted our obtained result $\hat{p}(t; \hat{t}_k, \hat{\theta}_k)$ as $p(t; K = k)$ for a notational convenience. From this figure, we observe that for both cases, similar change points were detected until $K \leq 2$, but their results are drastically different in the optimal number of change points, $K = K^* = 3$. In the desirable case, an almost accurate change point around $t = 10$ was detected at $K = 3$, and after that, several change points that bring about over-fitting results were detected. Actually, in terms of the MDL criterion, we could obtain the optimal number of change points and a reasonably accurate diffusion probability pattern in this case. On the other hand, in the undesirable case, a change point that brings about over-fitting results was detected at $K = 3$. At $K = 5$, a change point between $t = 10$ and 15 was detected, but this point is not so accurate compared to the point detected in the desirable case. The main reason why such undesirable cases happen for Wikipedia network is that for a relatively large number of diffusion results generated by using this network, the numbers of active nodes at an early period before $t = 10$ was quite small due to our setting of the diffusion probability $p_0 = 0.02$, which is small. As for the comparison method shown in case of the rect-linear shape in Table 3, we consider that this problem caused by small numbers of active nodes at an early period was alleviated by the imposed constraint $p_0 = p_2$.

In summary, we can say that the proposed method can approximate the changes of diffusion probability underlying the observed diffusion result in good accuracy, and the MDL criteron helps avoid the over-fitting.
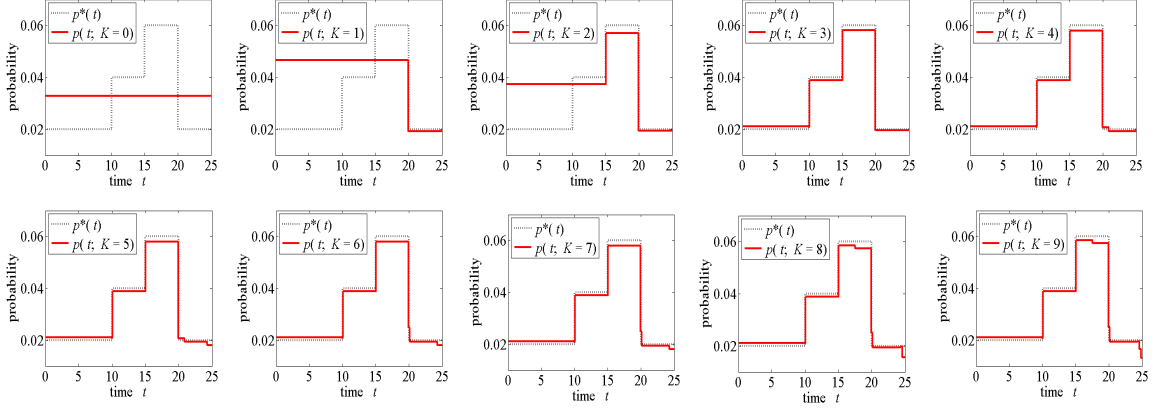
(*a*) Desirable Case for Wikipedia Network
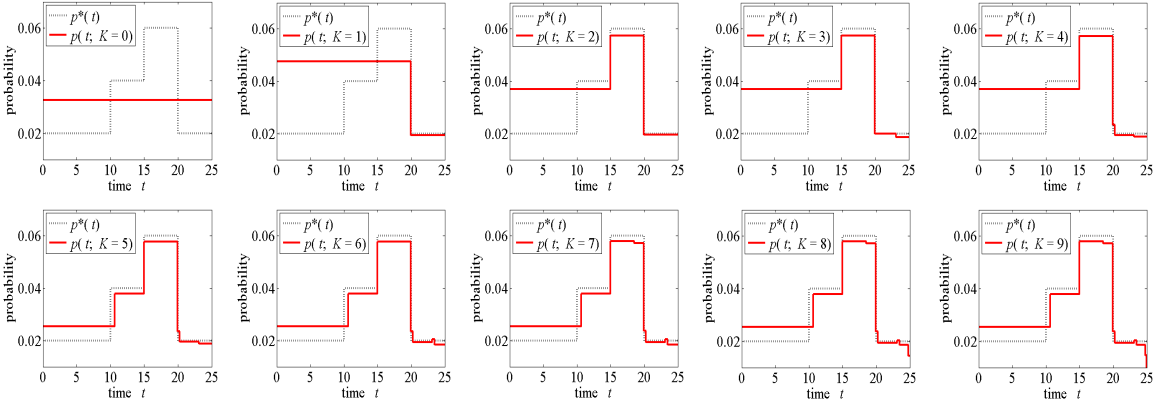
(*b*) Undesirable Case for Wikipedia Network

Figure 1: Examples of Diffusion Probability Functions Obtained by Varying *K*.

7. Conclusion

We addressed the problem of estimating diffusion probability changes, which are caused by changes in unknown external factors, for the AsIC-SIS (Asynchronous Indepdendent Cascade - Susceptible/Infectious/Susceptible) model over a social network from an observed information diffusion sequence. Here, the AsIC-SIS model is an information diffusion model in which the well-known discrete time IC-SIR (Independent Cascade - Susceptible/Infectious/Recovered) model is extended to contiuous time model allowing asynchronous time-delay as well as allowing multiple activations of the same nodes. We assumed that the change pattern of diffusion parameter for the AsIC-SIS model is approximated by a series of step functions, and proposed a method for detecting how many step functions are needed, where in time each one starts and how long it lasts, and what the height of each one is, from an observed sequence of information diffusion under the AsIC-SIS model. The proposed method employs "model parameter learning" by maximizing the likelihood function of the observed data (which is embedded inside the pattern search loop) and "efficient search" that uses the first order derivative of the likelihood function with respect to the parameters as a primary guide to search. The search algorithm adopts a divide-and-conquer type greedy recursive partioning that requires the expensive parameter learning only once for each partitioning, and utilizes an MDL

selection measure to determine the adequate number of step functions, *i.e.*, when to stop the search. Using four real world network structures, we confirmed the effectiveness of the proposed method. We evaluated the performance of the proposed method in terms of the L^1 norm of the difference between the true and the estimated diffusion probability patterns. We tested two kinds of artificially generated change pattern: One is a rect-linear pattern having two change points, and the other is a two-step pattern having three change points. For the rect-linear pattern, the perfomance of the proposed method was very close to that of the existing method which was devised soley for this restricted change pattern and known to work well. The performance of the proposed method for the two-step pattern did not degrade and the errors were comparable to those for the rect-linear pattern. The MDL criterion was useful to decide when to stop the search in order to avoid overfitting, and it identified the correct number of step functions in many cases. It returned a slightly large number in some cases, but the the L^1 norm of the difference between the two patterns which we use as a measure for the goodness of the found pattern was always small. Since the diffusion probability may change abruptly or gradually over time, our immediate future work is to evaluate the proposed method for a wide range of change patterns over time and reenforce the results obtained in this paper. Another immediate future work is to do a deeper analysis about why different networks give different results and understand the key factors to explain this.

Appendix A. Estimation Algorithm for AsIC-SIS Model

We briefly describe the estimation algorithm of parameters p and r for the AsIC-SIS model from an observed data $\mathcal{D}(0, T)$ (see Saito et al. (2009b, 2010a) for more details about the parameter learning algorithm of the AsIC model).

We employ an EM-like algorithm. Let $\bar{p}$ and $\bar{r}$ be the current estiamtes of p and r. Using Eqs. (2) and (3), we define $\bar{\alpha}_{u(\zeta),v(\eta)}$ and $\bar{\beta}_{u(\zeta),v(\eta)}$ as follows:

$$
\alpha_{u(\zeta),v(\eta)} = \frac{\mathcal{X}_{u(\zeta),v(\eta)}(\bar{p},\bar{r})/\mathcal{Y}_{u(\zeta),v(\eta)}(\bar{p},\bar{r})}{\sum_{z(\xi)\in\mathcal{AP}_{v(\eta)}}\mathcal{X}_{z(\xi),v(\eta)}(\bar{p},\bar{r})/\mathcal{Y}_{z(\xi),v(\eta)}(\bar{p},\bar{r})}
$$

$$
\beta_{u(\zeta),v(\eta)} = \frac{\bar{p}\exp(-\bar{r}(t_{v(\eta)}-t_{u(\zeta)}))}{\mathcal{Y}_{u(\zeta),v(\eta)}(\bar{p},\bar{r})}
$$

The update formulas of p and r are as follows:

$$
p = \frac{\sum_{v(\eta)\in\mathcal{D}}\sum_{u(\zeta)\in\mathcal{AP}_{v(\eta)}}\left(\bar{\alpha}_{u(\zeta),v(\eta)}+(1-\bar{\alpha}_{u(\zeta),v(\eta)})\bar{\beta}_{u(\zeta),v(\eta)}\right)}{\sum_{u(\zeta)\in\mathcal{D}}|F(u(\zeta))|}
$$

$$
r = \frac{\sum_{v(\eta)\in\mathcal{D}}\sum_{u(\zeta)\in\mathcal{AP}_{v(\eta)}}\bar{\alpha}_{u(\zeta),v(\eta)}}{\sum_{v(\eta)\in\mathcal{D}}\sum_{u(\zeta)\in\mathcal{AP}_{v(\eta)}}\left(\bar{\alpha}_{u(\zeta),v(\eta)}+(1-\bar{\alpha}_{u(\zeta),v(\eta)})\bar{\beta}_{u(\zeta),v(\eta)}\right)(t_{v(\eta)}-t_{u(\zeta)})}.
$$

Acknowledgments

References

D. Crandall, D. Cosley, D. Huttenlocner, J. Kleinberg, and S. Suri. Feedback effects between similarity and sociai infiuence in online communities. In *Proceedings of KDD 2008*, pages 160–168, 2008.

P. Domingos. Mining social networks for viral marketing. *IEEE Intelligent Systems*, 20:80–82, 2005.

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*, 12:211–223, 2001.

M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2010)*, pages 1019–1028, 2010.

D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. *SIGKDD Explorations*, 6:43–52, 2004.

D. Kempe, J. Kleinberg, and E. Tardos. 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)*, pages 137–146, 2003.

M. Kimura, K. Saito, and H. Motoda. Minimizing the spread of contamination by blocking links in a network. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI-08)*, pages 1175–1180, 2008.

M. Kimura, K. Saito, and H. Motoda. Blocking links to minimize contamination spread in a social network. *ACM Transactions on Knowledge Discovery from Data*, 3:9:1–9:23, 2009a.

M. Kimura, K. Saito, and H. Motoda. Efficient estimation of influence functions fot sis model on social networks. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI-09)*, 2009b.

M. Kimura, K. Saito, R. Nakano, and H. Motoda. Extracting influential nodes on a social network for information diffusion. *Data Min. Knowl. Disc.*, 20:70–97, 2010a.

M. Kimura, K. Saito, K. Ohara, and H. Motoda. Learning to predict opinion share in social networks. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI-10)*, pages 1364–1370, 2010b.

B. Klimt and Y. Yang. The enron corpus: A new dataset for email classification research. In *Proceedings of the 2004 European Conference on Machine Learning (ECML'04)*, pages 217–226, 2004.

J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. In *Proceedings of the 7th ACM Conference on Electronic Commerce (EC'06)*, pages 228–237, 2006.

S. A. Myers and J. Leskovec. On the convexity of latent social network inference. In *Proceedings of Neural Information Processing Systems (NIPS)*.

M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

M. E. J. Newman, S. Forrest, and J. Balthrop. Email networks and the spread of computer viruses. *Physical Review E*, 66:035101, 2002.

K. Ohara, K. Saito, M. Kimura, and H. Motoda. Efficient detection of hot span in information diffusion from observation (to appear). In *Proceedings of the IJCAI Workshop on Link Analysis in Heterogeneous Information Networks (HINA2011)*, 2011.

G. Palla, I. Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435:814–818, 2005.

K. Saito, M. Kimura, R. Nakano, and H. Motoda. Finding influential nodes in a social network from information diffusion data. In *Proceedings of the International Workshop on Social Computing and Behavioral Modeling (SBP09)*, pages 138–145, 2009a.

K. Saito, M. Kimura, K. Ohara, and H. Motoda. Learning continuous-time information diffusion model for social behavioral data analysis. In *Proceedings of the 1st Asian Conference on Machine Learning (ACML2009)*, pages 322–337. LNAI 5828, 2009b.

K. Saito, M. Kimura, K. Ohara, and H. Motoda. 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)*, pages 149–158, 2010a.

K. Saito, M. Kimura, K. Ohara, and H. Motoda. 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)*, pages 180–195. LNAI 6323, 2010b.

D. J. Watts. A simple model of global cascades on random networks. *Proceedings of National Academy of Science, USA*, 99:5766–5771, 2002.

D. J. Watts and P. S. Dodds. Influence, networks, and public opinion formation. *Journal of Consumer Research*, 34:441–458, 2007.

F. Wu and B. A. Huberman. How public opinion forms. In *Proceedings of WINE 2008*, pages 334–341, 2008.