Selecting Information Diffusion Models over Social Networks for Behavioral Analysis

Kazumi Saito[1], Masahiro Kimura[2], Kouzou Ohara[3], and Hiroshi Motoda[4]

[1] School of Administration and Informatics, University of Shizuoka
52-1 Yada, Suruga-ku, Shizuoka 422-8526, Japan
k-saito@u-shizuoka-ken.ac.jp
[2] Department of Electronics and Informatics, Ryukoku University
Otsu 520-2194, Japan
kimura@rins.ryukoku.ac.jp
[3] Department of Integrated Information Technology, Aoyama Gakuin University
Kanagawa 229-8558, Japan
ohara@it.aoyama.ac.jp
[4] Institute of Scientific and Industrial Research, Osaka University
8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan
motoda@ar.sanken.osaka-u.ac.jp

Abstract. We investigate how well different information diffusion models can explain observation data by learning their parameters and discuss which model is better suited to which topic. We use two models (AsIC, AsLT), each of which is an extension of the well known Independent Cascade (IC) and Linear Threshold (LT) models and incorporates asynchronous time delay. The model parameters are learned by maximizing the likelihood of observation, and the model selection is performed by choosing the one with better predictive accuracy. We first show by using four real networks that the proposed learning algorithm correctly learns the model parameters both accurately and stably, and the proposed selection method identifies the correct diffusion model from which the data are generated. We next apply these methods to behavioral analysis of topic propagation using the real blog propagation data, and show that although the relative propagation speed of topics that are derived from the learned parameter values is rather insensitive to the model selected, there is a clear indication as to which topic better follows which model. The correspondence between the topic and the model selected is well interpretable.

1 Introduction

The growth of Internet has enabled to form various kinds of large-scale social networks, through which a variety of information including innovation, hot topics and even malicious rumors can be propagated in the form of so-called "word-of-mouth" communications. Social networks are now recognized as an important medium for the spread of information, and a considerable number of studies have been made [1–5]. Widely used information diffusion models in these studies are the *independent cascade (IC)* [6–8] and the *linear threshold (LT)* [9, 10]. They have been used to solve such problems as the *influence maximization problem* [7, 11].

J.L. Balcázar et al. (Eds.): ECML PKDD 2010, Part III, LNAI 6323, pp. 180–195, 2010.

These two models focus on different information diffusion aspects. The IC model is sender-centered and each active node *independently* influences its inactive neighbors with given diffusion probabilities. The LT model is receiver-centered and a node is influenced by its active neighbors if their total weight exceeds the threshold for the node. Which model is more appropriate depends on the situation and selecting the appropriate one is not easy. First of all, we need to know how different model behaves differently and how well or badly explain the observation data. Both models have parameters that need be specified in advance: diffusion probabilities for the IC model, and weights for the LT model. However, their true values are not known in practice. This poses yet another problem of estimating them from a set of information diffusion results that are observed as time-sequences of influenced (activated) nodes.

This falls in a well defined parameter estimation problem in machine learning framework. Given a generative model with some parameters and the observed data, it is possible to calculate the likelihood that the data are generated and the parameters can be estimated by maximizing the likelihood. This approach has a thorough theoretical background. In general, the way the parameters are estimated depends on how the generative model is given. To the best of our knowledge, we are the first to follow this line of research. We addressed this problem for the IC model [12] and its variant that incorporates asynchronous time delay (referred to as the AsIC model) [13]. Gruhl et.al. also challenged the same problem of estimating the parameters and proposed an EM-like algorithm, but they did not formalize the likelihood and it is not clear what is being optimized in deriving the parameter update formulas. Goyal et.al attacked this problem from a different angle [14]. They employed a variant of the LT model and estimated the parameter values by four different methods, all of which are directly computed from the frequency of the events in the observed data. Their approach is efficient, but it is more likely ad hoc and lacks in theoretical evidence. Bakshy et.al [15] addressed the problem of diffusion of user-created content (asset) and used the maximum likelihood method to estimate the rate of asset adoption. However, they only modeled the rate of adoption and did not consider the diffusion model itself. Their focus is on data analysis.

In this paper, we first propose a method of learning the parameter values of a variant of the LT model that incorporates asynchronous time delay, similarly to the AsIC model, under the maximum likelihood framework. We refer to this diffusion model as the AsLT model. The model is similar to the one used in [14] but different in that we explicitly model the delay of node activation after the activation condition has been satisfied. Next we propose a method of model selection based on the predictive accuracy, using the two models: AsIC and AsLT.

It is indispensable to be able to cope with asynchronous time delay to do realistic analyses of information diffusion because, in the real world, information propagates along the continuous time axis, and time-delays can occur during the propagation asynchronously. In fact, the time stamps of the observed data are not equally spaced. Thus, the proposed learning method has to estimate not only the weight parameters but also the time-delay parameters from the observed data. Incorporating time-delay makes the time-sequence observation data structural, which makes the analyses of diffusion process difficult because there is no way of knowing which node has activated which other node from the observation data sequence. Knowing the optimal parameter values does

not mean that the observation follows the model. We have to decide which model better explains the observation. We solve this problem by comparing the predictive accuracy of each model. We use a variant of hold-out method applied to a set of sequential data, which is similar to the leave-one-out method applied to a multiple time sequence data. Extensive experiments have been performed to evaluate the effectiveness of the proposed method using both artificially generated data and real observation data. Experiments that used artificial data using four real network structures showed that the method can correctly 1) learn the parameters and 2) select the model by which the data have been generated. Experiments that used real diffusion data of topic propagation showed that 1) both AsIC and AsLT models well capture the global characteristics of topic propagations but 2) the predictive accuracy of each model is different for each topic and some topics have clear indication as to which model each better follows.

2 Information Diffusion Models

We first present the *asynchronous independent cascade (AsIC) model* introduced in [13], and then define the *asynchronous linear threshold (AsLT) model*. We mathematically model the spread of information through a directed network $G = (V, E)$ without self-links, where V and E ($\subset V \times V$) stand for the sets of all the nodes and links, respectively. For each node v in the network G, we denote $F(v)$ as a set of child nodes of v, *i.e.*, $F(v) = \{w; (v, w) \in E\}$. Similarly, we denote $B(v)$ as a set of parent nodes of v, *i.e.*, $B(v) = \{u; (u, v) \in E\}$. We call nodes *active* if they have been influenced with the information. In the following models, we assume that nodes can switch their states only from inactive to active, but not the other way around, and that, given an initial active node set S, only the nodes in S are active at an initial time.

2.1 Asynchronous Independent Cascade Model

We first recall the definition of the IC model according to [7], and then introduce the AsIC model. In the IC model, we specify a real value $\kappa_{u,v}$ with $0 < \kappa_{u,v} < 1$ for each link (u, v) in advance. Here $\kappa_{u,v}$ is referred to as the *diffusion probability* through link (u, v). The diffusion process unfolds in discrete time-steps $t \geq 0$, and proceeds from a given initial active set S in the following way. When a node u becomes active at time-step t, it is given a single chance to activate each currently inactive child node v, and succeeds with probability $\kappa_{u,v}$. If u succeeds, then v will become active at time-step $t+1$. If multiple parent nodes of v become active at time-step t, then their activation attempts are sequenced in an arbitrary order, but all performed at time-step t. Whether or not u succeeds, it cannot make any further attempts to activate v in subsequent rounds. The process terminates if no more activations are possible.

In the AsIC model, we specify real values $r_{u,v}$ with $r_{u,v} > 0$ in advance for each link $(u, v) \in E$ in addition to $\kappa_{u,v}$, where $r_{u,v}$ is referred to as the *time-delay parameter* through link (u, v). The diffusion process unfolds in continuous-time t, and proceeds from a given initial active set S in the following way. Suppose that a node u becomes active at time t. Then, u is given a single chance to activate each currently inactive child node v. We choose a delay-time δ from the exponential distribution with parameter

$r_{u,v}$[1]. If v has not been activated before time $t + \delta$, then u attempts to activate v, and succeeds with probability $\kappa_{u,v}$. If u succeeds, then v will become active at time $t + \delta$. Under the continuous time framework, it is unlikely that v is activated simultaneously by its multiple parent nodes exactly at time $t + \delta$. So we ignore this possibility. The process terminates if no more activations are possible.

2.2 Asynchronous Linear Threshold Model

Similarly to the above, we first define the LT model. In this model, for every node $v \in V$, we specify a *weight* ($\omega_{u,v} > 0$) from its parent node u in advance such that $\sum_{u \in B(v)} \omega_{u,v} \leq 1$. The diffusion process from a given initial active set S proceeds according to the following randomized rule. First, for any node $v \in V$, a *threshold* θ_v is chosen uniformly at random from the interval $[0, 1]$. At time-step t, an inactive node v is influenced by each of its active parent nodes, u, according to weight $\omega_{u,v}$. If the total weight from active parent nodes of v is no less than θ_v, that is, $\sum_{u \in B_t(v)} \omega_{u,v} \geq \theta_v$, then v will become active at time-step $t + 1$. Here, $B_t(v)$ stands for the set of all the parent nodes of v that are active at time-step t. The process terminates if no more activations are possible.

Next, we define the AsLT model. In the AsLT model, in addition to the weight set $\{\omega_{u,v}\}$, we specify real values r_v with $r_v > 0$ in advance for each node $v \in V$. We refer to r_v as the *time-delay parameter* on node v. Note that r_v depends only on v unlike $r_{u,v}$ of the AsIC model, which means that it is the node v's decision when to receive the information once the activation condition has been satisfied[2]. The diffusion process unfolds in continuous-time t, and proceeds from a given initial active set S in the following way. Suppose that the total weight from active parent nodes of v became no less than θ_v at time t for the first time. Then, v will become active at time $t + \delta$, where we choose a delay-time δ from the exponential distribution with parameter r_v. Further, note that even if some other non-active parent nodes of v has become active during the time period between t and $t + \delta$, the activation time of v, $t + \delta$, still remains the same. The other diffusion mechanisms are the same as the LT model.

3 Learning Algorithms

We define the time-delay parameter vector $\boldsymbol{r}$ and the diffusion parameter vector $\boldsymbol{\kappa}$ by $\boldsymbol{r} = (r_{u,v})_{(u,v)\in E}$ and $\boldsymbol{\kappa} = (\kappa_{u,v})_{(u,v)\in E}$ for the AsIC model and the parameter vectors $\boldsymbol{\omega}$ and $\boldsymbol{r}$ by $\boldsymbol{\omega} = (\omega_{u,v})_{(u,v)\in E}$ and $\boldsymbol{r} = (r_v)_{v\in V}$ for the AsLT model. We next consider an observed data set of M independent information diffusion results, $\{D_m;\ m = 1, \cdots, M\}$. Here, each D_m is a set of pairs of active nodes and their activation times in the mth diffusion result, $D_m = \{(u, t_{m,u}),\ (v, t_{m,v}),\ \cdots\}$. For each D_m, we denote the observed initial time by $t_m = \min\{t_{m,v};\ (v, t_{m,v}) \in D_m\}$, and the observed final time by $T_m \geq \max\{t_{m,v};\ (v, t_{m,v}) \in D_m\}$. Note that T_m is not necessarily equal to the final activation time. Hereafter, we express our observation data by $\mathcal{D}_M = \{(D_m, T_m);\ m = 1, \cdots, M\}$. For any $t \in [t_m, T_m]$, we set $C_m(t) = \{v;\ (v, t_{m,v}) \in D_m,\ t_{m,v} < t\}$. Namely, $C_m(t)$ is the set

[1] Similar formulation can be derived for other distributions such as power-law and Weibull.

[2] It is also possible to adopt the same edge time-delay model as in the AsIC model, in which case, for example, r_v in Equation (2) in Section 3 is replaced with $r_{u,v}$.

of active nodes before time t in the mth diffusion result. For convenience sake, we use C_m as referring to the set of all the active nodes in the mth diffusion result. Moreover, we define a set of non-active nodes with at least one active parent node for each by $\partial C_m = \{v;\ (u,v) \in E,\ u \in C_m,\ v \notin C_m\}$. For each node $v \in C_m \cup \partial C_m$, we define the following subset of parent nodes, each of which has a chance to activate v.

$$\mathcal{B}_{m,v} = \begin{cases} B(v) \cap C_m(t_{m,v}) & \text{if } v \in C_m(t_{m,v}), \\ B(v) \cap C_m & \text{if } v \in \partial C_m. \end{cases}$$

Note that the underlying model behind the observed data is not available in reality. Thus, we investigate how the model affects the information diffusion results, and consider selecting a model which better explains the given observed data from the candidates, i.e., AsIC and AsLT models. To this end, we first have to estimate the values of $\boldsymbol{r}$ and $\boldsymbol{\kappa}$ for the AsIC model, and the values of $\boldsymbol{r}$ and $\boldsymbol{\omega}$ for the AsLT model for the given $\mathcal{D}_M$. For the former, we adopt the method proposed in [13], which is only briefly explained here. For the latter, we propose a novel method of estimating those values.

3.1 Learning Parameters of AsIC Model

To estimate the values of $\boldsymbol{r}$ and $\boldsymbol{\kappa}$ from $\mathcal{D}_M$ for the AsIC model, We derived the following likelihood function $\mathcal{L}(\boldsymbol{r}, \boldsymbol{\kappa}; \mathcal{D}_M)$ to use as the objective function [13],

$$\mathcal{L}(\boldsymbol{r}, \boldsymbol{\kappa}; \mathcal{D}_M) = \prod_{m=1}^{M} \prod_{v \in C_m} \left(h_{m,v} \prod_{w \in F(v) \setminus C_m} g_{m,v,w} \right), \tag{1}$$

where $h_{m,v}$ is the probability density that the node v such that $v \in D_m$ with $t_{m,v} > 0$ for the mth diffusion result is activated at time $t_{m,v}$, and $g_{m,v,w}$ is the probability that a node w is not activated by a node v within the observed time period $[t_m, T_m]$ when there is a link $(v,w) \in E$ and $v \in C_m$ for the mth diffusion result. Then, we derived an iterative algorithm to stably obtain the values of $\boldsymbol{r}$ and $\boldsymbol{\kappa}$ that maximize equation (1). Please refer to [13] for more details. Hereafter, we refer to this method as the *AsIC model based method*.

3.2 Learning Parameters of AsLT Model

Likelihood function. To estimate the values of $\boldsymbol{r}$ and $\boldsymbol{\omega}$ from $\mathcal{D}_M$ for the AsLT model, we first derive the likelihood function $\mathcal{L}(\boldsymbol{r}, \boldsymbol{\omega}; \mathcal{D}_M)$ with respect to $\boldsymbol{r}$ and $\boldsymbol{\omega}$ in a rigorous way to use as the objective function. For the sake of technical convenience, we introduce a slack weight $\omega_{v,v}$ for each node $v \in V$ such that $\omega_{v,v} + \sum_{u \in B(v)} \omega_{u,v} = 1$. Here note that we can regard each weight $\omega_{*,v}$ as a multinomial probability since a threshold θ_v is chosen uniformly at random from the interval $[0,1]$ for each node v.

Suppose that a node v became active at time $t_{m,v}$ for the mth result. Then, we know that the total weight from active parent nodes of v became no less than θ_v at the time when one of them, $u \in \mathcal{B}_{m,v}$, became first active. However, in case of $|\mathcal{B}_{m,v}| > 1$, there is no way of exactly knowing the actual nodes due to the asynchronous time-delay. Suppose that a node v was actually activated when a node $\zeta \in \mathcal{B}_{m,v}$ became activated. Then

θ_v is between $\sum_{u\in B(v)\cap C_m(t_{m,\zeta})}\omega_{u,v}$ and $\omega_{\zeta,v}+\sum_{u\in B(v)\cap C_m(t_{m,\zeta})}\omega_{u,v}$. Namely, the probability that θ_v is chosen from this range is $\omega_{\zeta,v}$. Here note that such events with respect to different active parent nodes are mutually disjoint. Thus, the probability density that the node v is activated at time $t_{m,v}$, denoted by $h_{m,v}$, can be expressed as

$$h_{m,v}=\sum_{u\in\mathcal{B}_{m,v}}\omega_{u,v}r_v\exp(-r_v(t_{m,v}-t_{m,u})). \tag{2}$$

Here we define $h_{m,v}=1$ if $t_{m,v}=t_m$.

Next, we consider any node $w\in V$ belonging to $\partial C_m=\{w;(v,w)\in E\wedge v\in C_m(T_m)\wedge w\notin C_m(T_m)\}$ for the mth result. Let $g_{m,v}$ denote the probability that the node v is not activated within the observed time period $[t_m,T_m]$. We can calculate $g_{m,v}$ as

$$\begin{aligned}g_{m,v}&=1-\sum_{u\in\mathcal{B}_{m,v}}\omega_{u,v}\int_{t_{m,u}}^{T_m}r_v\exp(-r_v(t-t_{m,u}))dt=1-\sum_{u\in\mathcal{B}_{m,v}}\omega_{u,v}(1-\exp(-r_v(T_m-t_{m,u})))\\&=\omega_{v,v}+\sum_{u\in B(v)\setminus\mathcal{B}_{m,v}}\omega_{u,v}+\sum_{u\in\mathcal{B}_{m,v}}\omega_{u,v}\exp(-r_v(T_m-t_{m,u})).\end{aligned} \tag{3}$$

Therefore, by using Equations (2) and (3), and the independence properties, we can define the likelihood function $\mathcal{L}(\boldsymbol{r},\boldsymbol{\omega};\mathcal{D}_M)$ with respect to $\boldsymbol{r}$ and $\boldsymbol{\omega}$ by

$$\mathcal{L}(\boldsymbol{r},\boldsymbol{\omega};\mathcal{D}_M)=\prod_{m=1}^{M}\left(\prod_{v\in C_m}h_{m,v}\right)\left(\prod_{v\in\partial C_m}g_{m,v}\right). \tag{4}$$

Thus, our problem is to obtain the time-delay parameter vector $\boldsymbol{r}$ and the diffusion parameter vector $\boldsymbol{\omega}$, which together maximize Equation (4).

Learning Algorithm. For the above learning problem, we can derive an estimation method based on the Expectation-Maximization algorithm in order to stably obtain its solutions. Hereafter, we refer to this proposed method as the *AsLT model based method*. By the following formulas, we define $\phi_{m,u,v}$ for each $v\in C_m$ and $u\in\mathcal{B}_{m,v}$, $\varphi_{m,u,v}$ for each $v\in\partial C_m$ and $u\in\{v\}\cup B(v)\setminus\mathcal{B}_{m,v}$, and $\psi_{m,u,v}$ for each $v\in\partial C_m$ and $u\in\mathcal{B}_{m,v}$, respectively.

$$\begin{aligned}\phi_{m,u,v}&=\omega_{u,v}r_v\exp(-r_v(t_{m,v}-t_{m,u}))\,/\,h_{m,v},\quad \varphi_{m,u,v}=\omega_{u,v}\,/\,g_{m,v},\\ \psi_{m,u,v}&=\omega_{u,v}\exp(-r_v(T_m-t_{m,u}))\,/\,g_{m,v}.\end{aligned}$$

Let $\bar{\boldsymbol{r}}=(\bar{r}_v)$ and $\bar{\boldsymbol{\omega}}=(\bar{\omega}_{u,v})$ be the current estimates of $\boldsymbol{r}$ and $\boldsymbol{\omega}$, respectively. Similarly, let $\bar{\phi}_{m,u,v}$, $\bar{\varphi}_{m,u,v}$, and $\bar{\psi}_{m,u,v}$ denote the values of $\phi_{m,u,v}$, $\varphi_{m,u,v}$, and $\psi_{m,u,v}$ calculated by using $\bar{\boldsymbol{r}}$ and $\bar{\boldsymbol{\omega}}$, respectively.

From equations (2), (3), (4), we can transform $\mathcal{L}(\boldsymbol{r},\boldsymbol{\omega};\mathcal{D}_M)$ as follows:

$$\log\mathcal{L}(\boldsymbol{r},\boldsymbol{\omega};\mathcal{D}_M)=Q(\boldsymbol{r},\boldsymbol{\omega};\bar{\boldsymbol{r}},\bar{\boldsymbol{\omega}})-H(\boldsymbol{r},\boldsymbol{\omega};\bar{\boldsymbol{r}},\bar{\boldsymbol{\omega}}), \tag{5}$$

where $Q(\boldsymbol{r},\boldsymbol{\omega};\bar{\boldsymbol{r}},\bar{\boldsymbol{\omega}})$ is defined by

$$Q(\boldsymbol{r},\boldsymbol{\omega};\bar{\boldsymbol{r}},\bar{\boldsymbol{\omega}})=\sum_{m=1}^{M}\left(\sum_{v\in C_m}Q_{m,v}^{(1)}+\sum_{v\in\partial C_m}Q_{m,v}^{(2)}\right), \tag{6}$$

$$Q^{(1)}_{m,v} = \sum_{u \in \mathcal{B}_{m,v}} \bar{\phi}_{m,u,v} \log(\omega_{u,v} r_v \exp(-r_v(t_{m,v} - t_{m,u})))$$

$$Q^{(2)}_{m,v} = \sum_{u \in \{v\} \cup B(v) \setminus \mathcal{B}_{m,v}} \bar{\varphi}_{m,u,v} \log(\omega_{u,v}) + \sum_{u \in \mathcal{B}_{m,v}} \bar{\psi}_{m,u,v} \log(\omega_{u,v} \exp(-r_v(T_m - t_{m,u}))).$$

It is easy to see that $Q(\boldsymbol{r}, \boldsymbol{\omega}; \bar{\boldsymbol{r}}, \bar{\boldsymbol{\omega}})$ is convex with respect to $\boldsymbol{r}$ and $\boldsymbol{\omega}$, and $H(\boldsymbol{r}, \boldsymbol{\kappa}; \bar{\boldsymbol{r}}, \bar{\boldsymbol{\omega}})$ is defined by

$$\begin{aligned} H(\boldsymbol{r}, \boldsymbol{\omega}; \bar{\boldsymbol{r}}, \bar{\boldsymbol{\omega}}) &= \sum_{m=1}^{M} \left(\sum_{v \in C_m} H^{(1)}_{m,v} + \sum_{v \in \partial C_m} H^{(2)}_{m,v} \right), \\ H^{(1)}_{m,v} &= \sum_{u \in \mathcal{B}_{m,v}} \bar{\phi}_{m,u,v} \log(\phi_{m,u,v}), \\ H^{(2)}_{m,v} &= \sum_{u \in \{v\} \cup B(v) \setminus C_m} \bar{\varphi}_{m,u,v} \log(\varphi_{m,u,v}) + \sum_{u \in \mathcal{B}_{m,v}} \bar{\psi}_{m,u,v} \log(\psi_{m,u,v}). \end{aligned} \tag{7}$$

Since $H(\boldsymbol{r}, \boldsymbol{\omega}; \bar{\boldsymbol{r}}, \bar{\boldsymbol{\omega}})$ is maximized at $\boldsymbol{r} = \bar{\boldsymbol{r}}$ and $\boldsymbol{\omega} = \bar{\boldsymbol{\omega}}$ from equation (7), we can increase the value of $\mathcal{L}(\boldsymbol{r}, \boldsymbol{\omega}; \mathcal{D}_M)$ by maximizing $Q(\boldsymbol{r}, \boldsymbol{\omega}; \bar{\boldsymbol{r}}, \bar{\boldsymbol{\omega}})$ (see equation (5)).

Thus, we obtain the following update formulas of our estimation method as the solution which maximizes $Q(\boldsymbol{r}, \boldsymbol{\omega}; \bar{\boldsymbol{r}}, \bar{\boldsymbol{\omega}})$ with respect to $\boldsymbol{r}$:

$$r_v = \left(\sum_{m \in \mathcal{M}^{(1)}_v} \sum_{u \in \mathcal{B}_{m,v}} \bar{\phi}_{m,u,v} \right) \times \left(\sum_{m \in \mathcal{M}^{(1)}_v} \sum_{u \in \mathcal{B}_{m,v}} \bar{\phi}_{m,u,v}(t_{m,v} - t_{m,u}) + \sum_{m \in \mathcal{M}^{(2)}_v} \sum_{u \in \mathcal{B}_{m,v}} \bar{\psi}_{m,u,v}(T_m - t_{m,u}) \right)^{-1}$$

where $\mathcal{M}^{(1)}_v$ and $\mathcal{M}^{(2)}_v$ are defined by

$$\mathcal{M}^{(1)}_v = \{m \in \{1, \cdots, M\};\ v \in C_m\}, \quad \mathcal{M}^{(2)}_v = \{m \in \{1, \cdots, M\};\ v \in \partial C_m\}.$$

As for $\boldsymbol{\omega}$, we have to take the constraints $\omega_{v,v} + \sum_{u \in B(v)} \omega_{u,v} = 1$ into account for each v, which can easily be made using the Lagrange multipliers method, and we obtain the following update formulas of our estimation method:

$$\omega_{u,v} \propto \sum_{m \in \mathcal{M}^{(1)}_{u,v}} \bar{\phi}_{m,u,v} + \sum_{m \in \mathcal{M}^{(2)}_{u,v}} \bar{\varphi}_{m,u,v} + \sum_{m \in \mathcal{M}^{(3)}_{u,v}} \bar{\psi}_{m,u,v}, \quad \omega_{v,v} \propto \sum_{m \in \mathcal{M}^{(2)}_v} \bar{\varphi}_{m,v,v}$$

where $\mathcal{M}^{(1)}_{u,v}$, $\mathcal{M}^{(2)}_{u,v}$ and $\mathcal{M}^{(3)}_{u,v}$ are defined by

$$\begin{aligned} \mathcal{M}^{(1)}_{u,v} &= \{m \in \{1, \cdots, M\};\ v \in C_m,\ u \in \mathcal{B}_{m,v}\}, \\ \mathcal{M}^{(2)}_{u,v} &= \{m \in \{1, \cdots, M\};\ v \in \partial C_m,\ u \in B(v) \setminus \mathcal{B}_{m,v}\}, \\ \mathcal{M}^{(3)}_{u,v} &= \{m \in \{1, \cdots, M\};\ v \in \partial C_m,\ u \in \mathcal{B}_{m,v}\}. \end{aligned}$$

The actual values are obtained after normalization. Recall that we can regard our estimation method as a kind of the EM algorithm. It should be noted that each time the iteration proceeds the value of the likelihood function never decreases and the iterative algorithm is guaranteed to converge.

3.3 Model Selection

Next, we describe our model selection method. We select the model based on predictive accuracy. Here, note that we cannot use an information theoretic criterion such as AIC (Akaike Information Criterion) or MDL (Minimum Description Length) because we need to select one from models with completely different probability distributions. Moreover, for both models, it is quite difficult to efficiently calculate the exact activation probability of each node more than two information diffusion cascading steps ahead. In oder to avoid these difficulties, we propose a method based on a hold-out strategy, which attempts to predict the activation probabilities at one step later.

For simplicity, we assume that for each D_m, the initial observation time t_m is zero, i.e., $t_m = 0$ for $m = 1, \cdots, M$. Then, we introduce a set of observation periods

$$\mathcal{I} = \{[0, \tau_n);\ n = 1, \cdots, N\},$$

where N is the number of observation data we want to predict sequentially and each τ_n has the following property: There exists some $(v, t_{m,v}) \in D_m$ such that $0 < \tau_n < t_{m,v}$. Let $D_{m;\tau_n}$ denote the observation data in the period $[0, \tau_n)$ for the mth diffusion result, i.e.,

$$D_{m;\tau_n} = \{(v, t_{m,v}) \in D_m;\ t_{m,v} < \tau_n\}.$$

We also set $\mathcal{D}_{M;\tau_n} = \{(D_{m;\tau_n}, \tau_n);\ m = 1, \cdots, M\}$. Let $\boldsymbol{\Theta}$ denote the set of parameters for either the AsIC or the AsLT models, i.e., $\boldsymbol{\Theta} = (\boldsymbol{r}, \boldsymbol{\kappa})$ or $\boldsymbol{\Theta} = (\boldsymbol{r}, \boldsymbol{\omega})$. We can estimate the values of $\boldsymbol{\Theta}$ from the observation data $\mathcal{D}_{M;\tau_n}$ by using the learning algorithms in Sections 3.1 and 3.2. Let $\widehat{\boldsymbol{\Theta}}_{\tau_n}$ denote the estimated values of $\boldsymbol{\Theta}$. Then, we can calculate the activation probability $q_{\tau_n}(v, t)$ of node v at time t ($\geq \tau_n$) using $\widehat{\boldsymbol{\Theta}}_{\tau_n}$.

For each τ_n, we select the node $v(\tau_n)$ and the time $t_{m(\tau_n),v(\tau_n)}$ by

$$t_{m(\tau_n),v(\tau_n)} = \min\left\{ t_{m,v};\ (v, t_{m,v}) \in \bigcup_{m=1}^{M} (D_m \setminus D_{m;\tau_n}) \right\}.$$

Note that $v(\tau_n)$ is the first active node in $t \geq \tau_n$. We evaluate the predictive performance for the node $v(\tau_n)$ at time $t_{m(\tau_n),v(\tau_n)}$. Approximating the empirical distribution by

$$p_{\tau_n}(v, t) \ = \ \delta_{v,v(\tau_n)}\, \delta(t - t_{m(\tau_n),v(\tau_n)})$$

with respect to $(v(\tau_n), t_{m(\tau_n),v(\tau_n)})$, we employ the Kullback-Leibler (KL) divergence

$$KL(p_{\tau_n} \| q_{\tau_n}) \ = \ -\sum_{v \in V} \int_{\tau_n}^{\infty} p_{\tau_n}(v, t) \log \frac{q_{\tau_n}(v, t)}{p_{\tau_n}(v, t)}\, dt,$$

where $\delta_{v,w}$ and $\delta(t)$ stand for Kronecker's delta and Dirac's delta function, respectively. Then, we can easily show

$$KL(p_{\tau_n} \| q_{\tau_n}) \ = \ -\log h_{m(\tau_n),v(\tau_n)}. \tag{8}$$

By averaging the above KL divergence with respect to $\mathcal{I}$, we propose the following model selection criterion $\mathcal{E}$ (see Equation (8)):

$$\mathcal{E}(\mathcal{X}; D_1 \cup \cdots \cup D_M) = -\frac{1}{N} \sum_{n=1}^{N} \log h_{m(\tau_n),v(\tau_n)}, \tag{9}$$

where $\mathcal{X}$ expresses the information diffusion model (i.e., the AsIC or the AsLT models). In our experiments, we adopted

$$\mathcal{I} = \{[0, t_{m,v}); (v, t_{m,v}) \in D_1 \cup \cdots \cup D_M, t_{m,v} \geq \tau_0\},$$

where τ_0 is the median time of all the observed activation time points.

3.4 Behavioral Analysis

Thus far, we assumed that $\boldsymbol{\Theta}$ can vary with respect to nodes and links but is independent of the topic of information diffused. However, they may be sensitive to the topic. We follow [13] and place a constraint that $\boldsymbol{\Theta}$ depends only on topics but not on nodes and links of the network G, and assign a different m to a different topic. Therefore, we set $r_{m,u,v} = r_m$ and $\kappa_{m,u,v} = \kappa_m$ for any link $(u, v) \in E$ in case of the AsIC model and $r_{m,v} = r_m$ and $\omega_{m,u,v} = q_m |B(v)|^{-1}$ for any node $v \in V$ and link $(u, v) \in E$ in case of the AsLT model. Here note that $0 < q_m < 1$ and $\omega_{v,v} = 1 - q_m$. Without this constraint, we only have one piece of observation for each (m, u, v) and there is no way to learn $\boldsymbol{\Theta}$.

Using each pair of the estimated parameters, (r_m, q_m) for the AsLT model and (r_m, κ_m) for the AsIC model, we can discuss which model is more appropriate for each topic, and analyze the behavior of people with respect to the topics of information by simply plotting them as a point in 2-dimensional space.

4 Performance Evaluation by Artificial Data

Our goal here is to evaluate the parameter learning and model selection methods to see how accurately it can detect the true model that generated the data, using topological structure of four large real networks. Here, we assumed the true model by which the data are generated to be either AsLT or AsIC.

4.1 Data Sets

We employed four datasets of large real networks (all bidirectionally connected). The first one is a trackback network of Japanese blogs used in [8] and has 12,047 nodes and 79,920 directed links (the blog network). The second one is a network of people derived from the "list of people" within Japanese Wikipedia, also used in [8], and has 9,481 nodes and 245,044 directed links (the Wikipedia network). The third one is a network derived from the Enron Email Dataset [16] 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 coauthorship network used in [17] and has 12,357 nodes and 38,896 directed links (the coauthorship network).

Here, according to [13], we assumed the simplest case where the parameter values are uniform across all links and nodes, i.e., $\omega_{u,v} = q|B(v)|^{-1}$, $r_v = r$ for AsLT, and $r_{u,v} = r$, $\kappa_{u,v} = \kappa$ for AsIC. Under this assumption there is no need for the observation sequence data to pass through every link or node at least once. This drastically reduces the amount of data necessary to learn the parameters. Then, our task is to estimate the values of these parameters from data. The true value of q was set to 0.9 for every network to achieve reasonably long diffusion results, and the true value of r was set to 1.0.

Table 1. Parameter estimation error of the learning method for four networks

Network		Blog	Wiki	Enron	Coauthor
$\mathcal{D}_M(AsLT)$	r	0.248	0.253	0.200	0.244
	q	0.080	0.078	0.077	0.089
$\mathcal{D}_M(AsIC)$	r	0.114	0.026	0.029	0.167
	κ	0.020	0.013	0.002	0.054

Table 2. Accuracy of the model selection method for four networks

Network	Blog	Wiki	Enron	Coauthor
$\mathcal{D}_M(AsLT)$	79	86	99	76
	(28.2)	(54.0)	(47.7)	(19.0)
$\mathcal{D}_M(AsIC)$	92	100	100	93
	(370.2)	(920.8)	(1500.6)	(383.5)

According to [7], we set κ to a value smaller than $1/\bar{d}$, where $\bar{d}$ is the mean out-degree of a network. Thus, the true value of κ was set to 0.2 for the coauthorship network, 0.1 for the blog and Enron networks, and 0.02 for the Wikipedia network. Using these values, two sets of data were generated for each network, one for the true AsLT model and the other for the true AsIC model, denoted by $\mathcal{D}_M(AsLT)$ and $\mathcal{D}_M(AsIC)$, respectively. For each of these, sequences of data were generated, each starting from a randomly selected initial active node and having at least 10 nodes. In our experiments, we set $M = 100$ and evaluated our model selection method in the framework of behavioral analysis. Parameter updating is terminated when either the iteration number reaches its maximum (set to 100) or the following condition is first satisfied: $|r(s+1) - r(s)| + |q(s+1) - q(s)| \leq 10^{-6}$ for AsLT, $|r(s+1) - r(s)| + |\kappa(s+1) - \kappa(s)| \leq 10^{-6}$ for AsIC. In most of the cases, the latter inequality is satisfied in less than 100 iterations. The converged values are rather insensitive to the initial values, and we confirmed that the parameter updating algorithm stably converges to the correct values. In actual computation, the learned values for τ_n is used as the initial values for τ_{n+1} for efficiency purpose.

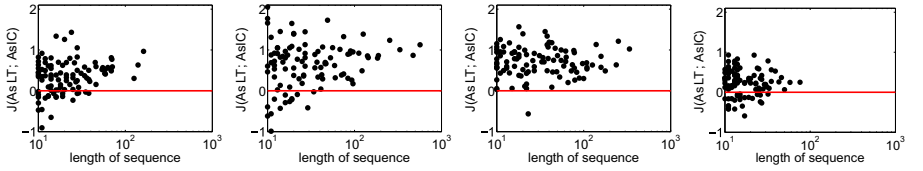
4.2 Learning Results

Table 1 shows the error in the estimated parameters for four networks by the proposed learning method. In this evaluation we treated each sequence as a separate observation and learned the parameters from each, repeated this M (=100) times and took the average. More specifically, the parameters of AsLT were estimated from $\mathcal{D}_M(AsLT)$, and those of AsIC from $\mathcal{D}_M(AsIC)$. Even though each pair of the parameters for individual models was estimated by using only one sequence data, we can see that the estimated values were reasonably close to the true one. This confirms that our proposed learning methods work well. The results indicate that the estimation performance on AsIC is substantially better than that on AsLT. We consider that this performance difference is attributed to the average sequence length, as discussed later.

4.3 Model Selection Results

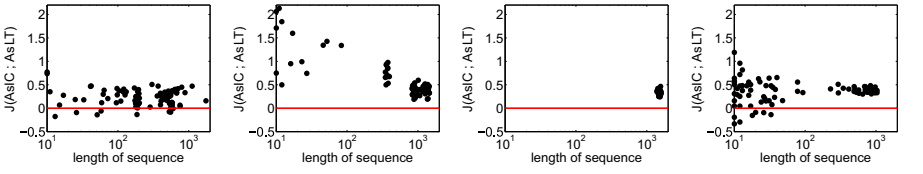
The average KL divergence given by equation (9) is the measure for the goodness of the model $\mathcal{X}$, given the data D_m. The smaller its value is, the better the model explains the data in terms of predictability. Thus, we can estimate the true model from which D_m is generated to be AsLT if $\mathcal{E}(AsLT; D_m) < \mathcal{E}(AsIC; D_m)$, and vice versa.

Table 2 summarizes the number of sequences for which the model selection method correctly identified the true model. The number within the parentheses is the average length of the sequences in each dataset. From these results, we can say that the proposed



(a) Blog (b) Wikipedia (c) Enron (d) Coauthorship

Fig. 1. Relation between the length of sequence and the the accuracy of model selection for $\mathcal{D}_M(AsLT)$



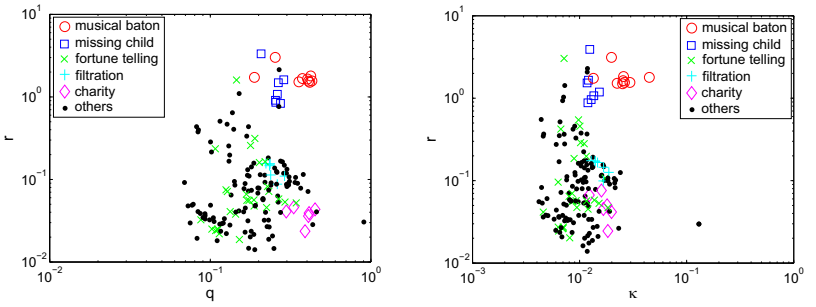
(a) Blog (b) Wikipedia (c) Enron (d) Coauthorship

Fig. 2. Relation between the length of sequence and the the accuracy of model selection for $\mathcal{D}_M(AsIC)$

method achieved a good accuracy, 90.6% on average. Especially, for the Enron network, its estimation was almost perfect. To analyze the performance of the proposed method more deeply, we investigated the relation between the length of sequence and the model selection result. It is shown in Fig. 1 for $\mathcal{D}_M(AsLT)$, where the horizontal axis denotes the length of sequence in each dataset and the vertical axis is the difference of the average KL divergence defined by $J(AsLT;AsIC) = \mathcal{E}(AsIC;D_m) - \mathcal{E}(AsLT;D_m)$. Thus, $J(AsLT;AsIC) > 0$ means that the proposed method correctly estimated the true model for the dataset $D_m(AsLT)$ because it means $\mathcal{E}(AsLT;D_m)$ is smaller than $\mathcal{E}(AsIC;D_m)$. From these figures, we can see that there is a correlation between the length of sequence and the estimation accuracy, and that the misselection occurs only in short sequences for every network. We notice that the overall accuracy becomes 95.5% when considering only the sequences that contain no less than 20 nodes. This means that the proposed model selection method is highly reliable for a long sequence and its accuracy could asymptotically approach to 100% as the sequence gets longer. Figure 2 is the results for $\mathcal{D}_M(AsIC)$, where $J(AsIC;AsLT) = \mathcal{E}(AsLT;D_m) - \mathcal{E}(AsIC;D_m)$. The results are better than for $\mathcal{D}_M(AsLT)$. In particular, Wikipedia and Blog networks have no misselection. We note that the plots are shifted to the right for all networks, meaning that the data sequences are longer for $\mathcal{D}_M(AsIC)$ than for $\mathcal{D}_M(AsLT)$. The better accuracy is attributed to this.

5 Behavioral Analysis of Real World Blog Data

We analyzed the behavior of topics in a real world blog data. Here, again, we assumed the true model behind the data to be either AsLT or AsIC. Then, we first applied our



(a) the AsLT model based method (b) the AsIC model based method

Fig. 3. Results for the Doblog database

learning method to behavioral analysis based on the method described in Section 3.4, assuming two possibilities, i.e. the true model being either AsLT or AsIC for all the topics, and investigated how each topic spreads throughout the network by comparing the learned parameter values. Next, we estimated the true model of each data sequence for each topic by applying the model selection method described in Section 3.3.

5.1 Data Sets

We used the real blogroll network used in [13], which was generated from the database of a blog-hosting service in Japan called *Doblog* [3]. In the network, bloggers are connected to each other and we assume that topics propagate from blogger x to another blogger y when there is a blogroll link from y to x. In addition, according to [18], it is assumed that a topic is represented as a URL which can be tracked down from blog to blog. We used the propagation sequences of 172 URLs for this analysis, each of which has at least 10 time steps. Please refer to [13] for more details.

5.2 Behavioral Analysis

We ran the experiments for each identified URL and obtained the parameters q and r for the AsLT model based method and κ and r for the AsIC model based method. Figures 3a and 3b are the plots of the results for the major URLs (topics) by the AsLT and AsIC methods, respectively. The horizontal axis is the diffusion parameter q for the AsLT method and κ for the AsIC method, while the vertical axis is the delay parameter r for both. The latter axis is normalized such that $r = 1$ corresponds to a delay of one day, meaning $r = 0.1$ corresponds to a delay of 10 days. In these figures, we used five kinds of markers other than dots, to represent five different typical URLs: the circle ($\circ$) stands for a URL that corresponds to the musical baton which is a kind of telephone game on the Internet (the musical baton), the square ($\square$) for a URL that corresponds to articles about a missing child (the missing child), the cross ($\times$) for a URL that corresponds to articles about fortune telling (the fortune telling), the diamond ($\diamond$) for a URL of a certain charity site (the charity), and the plus ($+$) for a URL of a site for flirtatious

[3] Doblog(`http://www.doblog.com/`), provided by NTT Data Corp. and Hotto Link, Inc.

Table 3. Results of model selection for the Doblog dataset

Topic	Total	AsLT	AsIC
Musical baton	9	5	4
Missing child	7	0	7
Fortune telling	28	4	24
Charity	6	5	1
Flirtation	7	7	0
Others	115	11	104

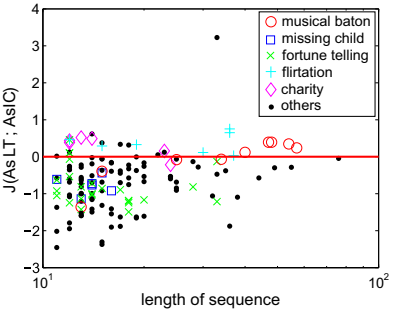


Fig. 4. The relation between the KL difference and sequence length for the Doblog database

tendency test (the flirtation). All the other topics are denoted by dots (·), which means they are a mixture of many topics.

The results indicates that in general both the AsLT and AsIC models capture reasonably well the characteristic properties of topics in a similar way. For example, it captures the urgency of the missing child, which propagates quickly. Musical baton which actually became the latest craze on the Internet also propagates quickly. In contrast non-emergency topics such as the flirtation and the charity propagate very slowly. Unfortunately, this highlights the people's low interest level of the charity activity in the real world. We further note that the dependency of topics on the parameter r is almost the same for both AsLT and AsIC, but that on the parameters q and κ is slightly different, e.g., relative difference of musical baton, missing child and charity. Although q and κ are different parameters but both are the measures that represent how easily the diffusion takes place. We showed in [13] that the influential nodes are very sensitive to the model used and this can be attributed to the differences of these parameter values.

5.3 Model Selection

In the analysis of previous subsection, we assumed that each topic follows the same diffusion model. However, in reality this is not true and each topic should propagate following more closely to either one of the AsLT and AsIC models. Thus, in this subsection, we attempt to estimate the underlying behavior model of each topic by applying the model selection method to individual sequence as described in section 4. Namely, we regard that each observation consists of only one observed data sequence, i.e., $\mathcal{D}_1$, and calculate its KL divergences by equation (9) for the both models, and compare the goodness.

Table 3 and Fig. 4 summarize the results. From these results, we can see that most of the diffusion behaviors on this blog network follows the AsIC model. It is interesting to note that the model estimated for the musical baton is not identical to that for the missing child although their diffusion patterns are very similar in the previous analysis. The missing child strictly follows the AsIC model. This is attributed to its greater urgency. On the other, musical baton seems to follow more closely to AsLT. This is because the longer sequence results in a better accuracy and the models selected in longer sequences

are all AsLT in Fig. 4 although the numbers are almost tie (4 vs. 5) in Table 3. This can be interpreted that people follow their friends in this game. Likewise, it is easy to imagine that one would align oneself with the opinions of those around when requested to raise funds. This explains that charity follows AsLT. The flirtation clearly follows AsLT. This is probably because the information of this kind of play site easily diffuses within close friends. Note that there exists one dot at near the top center in Fig. 4, showing the greatest tendency to follow AsLT. This dot represents a typical circle site that distributes one's original news article on personal events.

6 Discussion

We now have ways to compare the diffusion process with respect to two models (the AsLT model and the AsIC model) for the same observed dataset. Being able to learn the parameters of these models enable us to analyze the diffusion process more precisely. Comparing the results bring us deeper insights into the relation between models and information diffusion processes.

We note that the formulation in Sections 2 and 3 allows the parameters to depend on links and nodes, but the analysis we showed in Section 4 is for the simplest case where the parameters are uniform across the whole network. Actually, if all the parameters are node and link dependent, the number of the parameters becomes so huge and it is not practical (almost impossible) to estimate them accurately because the amount of observation data needed is prohibitively huge and there is always a problem of overfitting. However, this can be alleviated. In a more realistic setting we can divide E into subsets $E_1, E_2, ..., E_L$ and assign the same value for each parameter within each subset. For example, we may divide the nodes into two groups: those that strongly influence others and those not, or we may divide the nodes into another two groups: those that are easily influenced by others and those not. If there is some background knowledge about the node grouping, our method can make the best use of it. Obtaining such background knowledge is also an important research topic in the knowledge discovery from social networks.

The discussion above is also related to the use of the data for model selection in Section 5 in which we used each sequence separately to learn the model parameter values and select the model rather than using them altogether for the same topic and obtaining a single set of parameter values. The results in Section 5 show that the model parameters thus obtained for each sequence are very similar to each other for the same topic. This in turn justifies the use of the same parameter values for multiple sequence observation data (the way we formulated in Section 3.3).

As we mentioned in Section 5.2 but did not show in this paper due to the space limitation, the ranking results that involve detailed probabilistic simulation is very sensitive to the underlying model which is assumed to generate the observed data. In other words, it is very important to select an appropriate model for the analysis of information diffusion from which the data has been generated if the node characteristics are the main objective of analysis, e.g. such problems as the influence maximization problem [7, 11], a problem at a more detailed level. However, it is also true that the parameters for the topics that actually propagated quickly/slowly in observation converged to the values that enable them to propagate quickly/slowly on the model, regardless of the model chosen. Namely, we can say that the difference of models does not have much influence on the

relative difference of topic propagation which indeed strongly depends on topic itself. Both models are well defined and can explain this property at this level of abstraction. Nevertheless, the model selection is very important if we want to characterize how each topic propagates through the network.

Finally, the proposed learning method is efficient and the runtime is not an issue. The convergence is fast and it can handle networks of millions of nodes because the complexity depends directly on the data size, not the number of nodes. In particular, the complexity of learning from a single sequence is proportional to the number of active nodes, their average degree, and the EM iteration number.

7 Conclusion

We considered the problem of analyzing information diffusion process in a social network using two kinds of information diffusion models, incorporating asynchronous time delay, the AsLT model and the AsIC model, and investigated how the results differ according to the model used. To this end, we proposed novel methods of 1) learning the parameters of the AsLT model from the observed data (the method for learning the parameters of the AsIC model has already been reported), and 2) selecting models that better explains the observation. We experimentally confirmed that the learning method converges to the correct values very stably and the model selection method can correctly identifies the diffusion models by which the observed data is generated based on extensive simulations on four real world datasets. We further applied the methods to the real blog data and analyzed the behavior of topic propagation. The relative propagation speed of topics, i.e. how far/near and how fast/slow each topic propagates, that are derived from the learned parameter values is rather insensitive to the model selected, but the model selection algorithm clearly identifies the difference of model goodness for each topic. We found that many of the topics follow the AsIC models in general, but some specific topics have clear interpretations for them being better modeled by either one of the two, and these interpretations are consistent with the model selection results. There are numerous factors that affect the information diffusion process, and there can be a number of different models. Model selection is a big challenge in social network analysis and this work is the first step towards this goal.

Acknowledgments

This work was partly supported by Asian Office of Aerospace Research and Development, Air Force Office of Scientific Research under Grant No. AOARD-10-4053, and JSPS Grant-in-Aid for Scientific Research (C) (No. 20500147).

References

1. Newman, M.E.J., Forrest, S., Balthrop, J.: Email networks and the spread of computer viruses. Physical Review E 66, 035101 (2002)
2. Newman, M.E.J.: The structure and function of complex networks. SIAM Review 45, 167–256 (2003)

3. Gruhl, D., Guha, R., Liben-Nowell, D., Tomkins, A.: Information diffusion through blogspace. SIGKDD Explorations 6, 43–52 (2004)
4. Domingos, P.: Mining social networks for viral marketing. IEEE Intelligent Systems 20, 80–82 (2005)
5. Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. In: Proceedings of the 7th ACM Conference on Electronic Commerce (EC'06), pp. 228–237 (2006)
6. Goldenberg, J., Libai, B., Muller, E.: Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters 12, 211–223 (2001)
7. Kempe, D., Kleinberg, J., Tardos, E.: 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), pp. 137–146 (2003)
8. Kimura, M., Saito, K., Motoda, H.: Blocking links to minimize contamination spread in a social network. ACM Transactions on Knowledge Discovery from Data 3, 9:1–9:23 (2009)
9. Watts, D.J.: A simple model of global cascades on random networks. Proceedings of National Academy of Science, USA 99, 5766–5771 (2002)
10. Watts, D.J., Dodds, P.S.: Influence, networks, and public opinion formation. Journal of Consumer Research 34, 441–458 (2007)
11. Kimura, M., Saito, K., Nakano, R.: Extracting influential nodes for information diffusion on a social network. In: Proceedings of the 22nd AAAI Conference on Artificial Intelligence (AAAI-07), pp. 1371–1376 (2007)
12. Saito, K., Kimura, M., Nakano, R., Motoda, H.: Finding influential nodes in a social network from information diffusion data. In: Proceedings of the International Workshop on Social Computing and Behavioral Modeling (SBP'09), pp. 138–145 (2009)
13. Saito, K., Kimura, M., Ohara, K., Motoda, H.: Learning continuous-time information diffusion model for social behavioral data analysis. In: Zhou, Z.-H., Washio, T. (eds.) ACML 2009. LNCS, vol. 5828, pp. 322–337. Springer, Heidelberg (2009)
14. Goyal, A., Bonchi, F., Lakshhmanan, L.V.S.: Learning influence probabilities in social networks. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, pp. 241–250 (2010)
15. Bakshy, E., Karrer, B., Adamic, L.A.: Social influence and the diffusion of user-created content. In: Proceedings of the Tenth ACM Conference on Electronic Commerce, pp. 325–334 (2009)
16. Klimt, B., Yang, Y.: The enron corpus: A new dataset for email classification research. In: Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (eds.) ECML 2004. LNCS (LNAI), vol. 3201, pp. 217–226. Springer, Heidelberg (2004)
17. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. Nature 435, 814–818 (2005)
18. Adar, E., Adamic, L.A.: Tracking information epidemics in blogspace. In: Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05), pp. 207–214 (2005)