JMLR: Workshop and Conference Proceedings 13: 193-208
2nd Asian Conference on Machine Learning (ACML2010), Tokyo, Japan, Nov. 8–10, 2010.

Generative Models of Information Diffusion with Asynchronous Time-delay

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

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

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

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

Editor: Masashi Sugiyama and Qiang Yang

Abstract

We address the problem of formalizing an information diffusion process on a social network as a generative model in the machine learning framework so that we can learn model parameters from the observation. Time delay plays an important role in formulating the likelihood function as well as for the analyses of information diffusion. We identified that there are two different types of time delay: link delay and node delay. The former corresponds to the delay associated with information propagation, and the latter corresponds to the delay due to human action. We further identified that there are two distinctions of the way the activation from the multiple parents is updated: non-override and override. The former sticks to the initial activation and the latter can decide to update the time to activate multiple times. We formulated the likelihood function of the well known diffusion models: independent cascade and linear threshold, both enhanced with asynchronous time delay distinguishing the difference in two types of delay and two types of update scheme. Simulation using four real world networks reveals that there are differences in the spread of information diffusion and they strongly depend on the choice of the parameter values and the denseness of the network.

Keywords: Information diffusion, Social network, Maximum likelihood, Asynchronous time delay

1. Introduction

There have been tremendous interests in the phenomenon of influence that members of a social network can exert on other members and how the information propagates through the network. A variety of information that includes news, innovation, hot topics, ideas, opinions and even malicious rumors, propagates in the form of so-called "word-of-mouth" communications. Social networks (both real and virtual) are now recognized as an important medium for the spread of information and a considerable number of studies have been conducted (Newman et al., 2002; Newman, 2003; Gruhl et al., 2004; Domingos, 2005; Leskovec et al., 2006).

Basic models of information diffusion which are widely used in these studies are the *independent cascade (IC)* (Goldenberg et al., 2001; Kempe et al., 2003; Kimura et al., 2009) and the *linear threshold (LT)* (Watts, 2002; Watts and Dodds, 2007). They have been used to solve such problems as the *influence maximization problem* (Kempe et al., 2003; Kimura et al., 2010) and the *contamination minimization problem* (Kimura et al., 2009). 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 (Saito et al., 2009, 2010).

This problem fits in a well defined parameter estimation problem in machine learning setting, provided that a proper model is known. Thus, having a good generative model is crucial for this approach to be successful. One important factor that needs a special care is how to treat time delay in information diffusion. Diffusion process involves time evolution. The basic models deal with time by allowing nodes to change their states in a synchronous way at each discrete time step. No time delay is considered, or one can say that every action is uniformly delayed exactly by one discrete time step. However, 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 while information propagates by various reasons. Incorporating time-delay makes the time-sequence observation data structural, which makes the analyses of diffusion process difficult because it is not self-evident from the observed sequence data which node has activated which other node. What is observed is just a sequence of time when each node has been activated. Saito et al. (2009, 2010) have extended the basic IC and LT models to incorporate asynchronous time delay and successfully solved this parameter estimation problem by maximizing the likelihood function using a variant of EM algorithm, but they have not carefully examined that there are different types of time delay and node activation scheme[1].

In this paper, we revisit the generative model and carefully analyze what kind of time delay and activation scheme is considered realistic because, in general, the way the parameters are estimated depends on how the generative model is given. We identified that there are two different types of time delay: link delay and node delay. The former corresponds to the delay associated with information propagation, and the latter corresponds to the delay due to human action. We further identified that there are two types of the way the activation from the multiple parents is updated: non-override and override. The former sticks to the initial activation and the latter can decide to update the time to activate multiple times. We rigorously formulated the likelihood function of the IC and the LT models, extending them to incorporate asynchronous time delay with the difference in two types of time delay and two types of update scheme taken into account[2]. There are a total

1. Two examples explaining the different types of time delay are given in subsection 3.1.
2. We refer to asynchronous time delay versions of the IC and the LT models as the AsIC and AsLT models, respectively.

of three different models for each of the AsIC and the AsLT models, but the theoretical analysis revealed that particular combinations of time delay and update scheme result in the same likelihood function (with a minor notational difference) and it suffices to consider two different models for each. We performed how the difference in the time delay and the update scheme affects the information diffusion results as a function of time, varying the values of diffusion parameters using four real world networks. The simulation results reveal that there are differences in the spread of information diffusion and they strongly depend on the choice of the parameter values and the denseness of the network, confirming that it is important to distinguish the different types of time delay and update scheme. The results are well interpretable.

The paper is organized as follows. We revisit the basic information diffusion models in section 2 describing the likelihood functions. In section 3 we first explain different time delay types and update schemes, and then, based on these differences, derive the rigorous likelihood function for each of the possible combinations of these types and schemes for both the AsIC and AsLT models. We show the experimental result in subsection 5.2, and summarize the main conclusions in section 6.

2. Basic Information Diffusion Models

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 Independent Cascade Model

We first recall the definition of the IC model according to Kempe et al. (2003). 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.

2.2 Linear Threshold Model

Next, we present the definition of 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.

2.3 Likelihood

As emphasized in section 1, our main focus is to formalize the information diffusion process as a generative model in machine learning problem setting. The generative model is a model of the world which can predict the future from the past, and the model must be consistent with the observation as much as possible. Thus, it is crucial to formulate the likelihood function as realistically as possible so that the model with these parameters (including $\kappa_{u,v}$ and $\omega_{u,v}$ above) that maximize the likelihood can best reflect the reality and generate the data close enough to the observation.

We denote an observed data set of M independent information diffusion results as $\{D_m;\ m = 1, \cdots, M\}$. 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 these are just sequences of $(u, t_{m,u})$ pairs and do not tell which parent node of u actually activated u. Further 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 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 $\partial C_m = \{v;\ (u, v) \in E,\ u \in C_m,\ v \notin C_m\}$.

Next we formulate the likelihood function $\mathcal{L}(\mathcal{D}_M; \mathbf{\Theta})$, where $\mathbf{\Theta}$ denotes the parameters that we want to optimize by maximizing $\mathcal{L}$. Nodes in C_m are a part of the nodes in the graph G and those not in C_m have not been activated. Since non-activated nodes, unless they get activated, never activate the other non-activated node, we only need to consider nodes in C_m and ∂C_m. Thus, the likelihood function is basically described by the product of two factors, one representing the probabilities that nodes in C_m are activated at their respective times and the other representing the probabilities that nodes in ∂C_m have not been activated during the observed time period $[t_m, T_m]$.

The likelihood functions for the IC and the LT models take slightly different forms. $\mathcal{L}$ for the IC model is given by Equation (1), and $\mathcal{L}$ for the LT model is given by Equation (2).

$$\mathcal{L}(\mathcal{D}_M; \mathbf{\Theta}) = \prod_{m=1}^{M} \prod_{v \in C_m} (h_{m,v}\ g_{m,v}), \tag{1}$$

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 $g_{m,v}$ is the probability that a node v

fails to activate its child nodes for the mth diffusion result.

$$\mathcal{L}(\mathcal{D}_M; \mathbf{\Theta}) = \prod_{m=1}^{M} \left(\prod_{v \in C_m} h_{m,v} \right) \left(\prod_{v \in \partial C_m} g_{m,v} \right), \tag{2}$$

where the definition of $h_{m,v}$ is the same as above and $g_{m,v}$ is the probability that the node v is not activated within the observed time period $[t_m, T_m]$. The specific formulae of $h_{m,v}$ and $g_{m,v}$ for the IC model are

$$h_{m,v} = 1 - \prod_{u \in B(v) \cap \tilde{C}(t_{m,v})} (1 - \kappa_{u,v}), \quad g_{m,v} = \prod_{w \in F(v) \setminus C(t_{m,v}+1)} (1 - \kappa_{v,w}), \tag{3}$$

and those for the LT model are

$$h_{m,v} = \sum_{u \in B(v) \cap \tilde{C}(t_{m,v})} \omega_{u,v}, \quad g_{m,v} = 1 - \sum_{u \in B(v) \cap C_m} \omega_{u,v}, \tag{4}$$

where $\tilde{C}(t_{m,v}) = C(t_{m,v}) \setminus C(t_{m,v} - 1)$. Note that Equations (3) have been described in Saito et al. (2008), and Equations (4) are special forms of the corresponding equations described in Saito et al. (2010).

3. Asynchronous Information Diffusion Models

In this section, we first explain notions of time-delay identifying two different types of time delay and two different types of the way the activation from the multiple parents is updated. Then, we derive the rigorous likelihood function for each of the possible combinations of these time-delay types and update schemes for asynchronous time delay versions of the IC and the LT models.

3.1 Notions of Time-delay

The basic information diffusion models briefly described in section 2 do not account for time delay. In reality it takes time for the information to diffuse by various reasons, and further, the way the delay takes place is asynchronous. Each parent u of a node v can be activated independently of the other parents in an asynchronous way and because the associated time delay from a parent to its child is different for every single pair, which parent u actually affects the node v in which order is more or less opportunistic. In case of the IC model which is sender-oriented, it may look more natural to attach the delay to the link, *i.e.*, when a node u is activated and is ready to send the information, it does not necessarily reach its child node v instantaneously but with some delay attached to the link (u, v). On the other hand, in case of the LT model which is receiver-oriented, it may look more natural to attach the delay to the node (receiver), *i.e.* when the sum of the weights from the active parents of a node v exceeds the threshold θ and the node v is ready to receive the information, it does not necessarily reach the node v instantaneously but with some delay attached to the node v. However, in both models information diffuses from a parent to its child and there is no reason to exclude other combinations than the above.

To explicate the information diffusion process in a more realistic setting, we think of two examples, one associated with blog posting and the other associated with electronic mailing. In case of blog posting, assume that some blogger u posts an article. Then it is natural to think that it takes some time before another blogger v comes to notice the posting. It is also natural to think that if the blogger v reads the article, he or she takes an action to respond (activated) because the act of reading the article is an active behavior. In this case, we can think that there is a delay in information diffusion from u to v but there is no delay in v taking an action. In case of electronic mailing, assume that someone u sends a mail to someone else v. It is natural to think that the mail is delivered to the receiver v instantaneously. However, this does not necessarily mean that v reads the mail as soon as it has been received because the act of receiving a mail is a passive behavior. In this case, we can think that there is no delay in information diffusion from u to v but there is a delay in v taking an action. Further, when v notices the mail, v may think to respond to it later. But before v responds, a new mail may arrive which needs a prompt response and v sends a mail immediately. We can think of this as an update of acting time[3]. These are just two examples, but it appears worth distinguishing the difference of these two kinds of time delay and update scheme (override of decision) in a more general setting.

In what follows we formulate the likelihood function distinguishing the difference of assumed time delay and override policy, and show that these distinctions indeed affect the form of the likelihood function. According to the discussion above, we define two types of delay: link delay and node delay. It is easiest to think that link delay corresponds to propagation delay and node delay corresponds to action delay. We further assume that they are mutually exclusive. This is a strong restriction as well as a strong simplification by necessity because the activation time we can observe is a sum of the two delays and we cannot distinguish between these two. Thus we have to choose either one of the two as occurring exclusively for the likelihood maximization to be feasible. In addition, we assume that there are two types of activation associated with time delay: non-override and override. The former sticks to the initial decision when to activate and the latter can decide to update (override) the time of activation multiple times each time one of the parents gets newly activated. Due to the mutual exclusiveness of link delay and node delay, override is only associated with node delay. As mentioned in section 1, we call the time delay versions of the IC and the LT models as Asynchronous Independent Cascade Model (AsIC) and Asynchronous Linear Threshold Model (AsLT), respectively.

In summary, node delay can go with either override or non-override, and link delay can only go with non-override. In the following subsections, we will derive $h_{m,v}$ and $g_{m,v}$ for each of the AsIC model and the AsLT model. We choose a delay-time δ from the exponential distribution with parameter r for the sake of convenience, but of course other distributions such as power-law and Weibull can be employed. The *time delay parameter* r is expressed explicitly as $r_{u,v}$ if it is link delay and r_v (or r_u) if it is node delay. Once the likelihood function is formalized, any optimization method can be used to find the best estimates of the parameter values. In practice, variants of EM algorithm has been shown to work satisfactorily (Saito et al., 2009, 2010).

3. Note that there are two actions here, reading and sending, but the activation time in the observed sequence data corresponds to the time v sends a mail

3.2 Asynchronous Independent Cascade Models

Link delay with non-override 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. 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. Note that this delay is due to propagation delay. Once the node v receives the information, it instantaneously gets activated and there is no action delay in v.

We order the active parent node $u \in B(v) \cap C_m(t_{m,v})$ of a node v according to the time t_u it was activated: $B(v) \cap C_m(t_{m,v}) = \{u_1, u_2, ..., u_K\}$ such that $t_{u_1} < t_{u_2} < ... < t_{u_K}$.

The probability density $h_{m,v}$ is the sum of the probability density that u_i activates v but all the other $u_j, j \neq i$ fail to activate v over all i $(i = 1, 2, ..., K)$.

$$\begin{aligned} h_{m,v} &= \sum_{k=1}^{K} \kappa_{u_k,v} r_{u_k,v} \exp(-r_{u_k,v}(t_{m,v} - t_{m,u_k})) \\ &\quad \times \prod_{i=1, i\neq k}^{K} (1 - \int_{t_{m,u_i}}^{t_{m,v}} \kappa_{u_i,v} r_{u_i,v} \exp(-r_{u_i,v}(t - t_{m,u_i}))dt) \\ &= \sum_{k=1}^{K} \kappa_{u_k,v} r_{u_k,v} \exp(-r_{u_k,v}(t_{m,v} - t_{m,u_k})) \\ &\quad \times \prod_{i=1, i\neq k}^{K} (\kappa_{u_i,v} \exp(-r_{u_i,v}(t_{m,v} - t_{m,u_i})) + (1 - \kappa_{u_i,v}))\,. \end{aligned} \tag{5}$$

The probability $g_{m,v}$ is given by

$$\begin{aligned} g_{m,v} &= \prod_{w \in F(v) \setminus C_m} \left(1 - \int_{t_{m,v}}^{T_m} \kappa_{v,w} \exp(-r_{v,w}(t - t_{m,v}))dt\right) \\ &= \prod_{w \in F(v) \setminus C_m} (\kappa_{v,w} \exp(-r_{v,w}(T_m - t_{m,v})) + (1 - \kappa_{v,w}))\,. \end{aligned} \tag{6}$$

Note that the formulation in Saito et al. (2009) corresponds to this category.

Node delay with non-override The difference of diffusion process from *Link delay with non-override* is that there is no delay in propagating the information to the node v from the node u, but there is a delay δ before the node v gets actually activated. Assume that it is the node u_i that first succeeded in activating the node v (more precisely satisfying the activation condition). Since there is no link delay, it must be the case that all the other parents that had become active before t_{u_i} must have failed in activating v (more precisely satisfying the activation condition). Since the node v decides when to actually activate itself at the time the node u_i succeeded in satisfying the activation condition and would not change its mind, other nodes which may possibly activate the node v at a later time can do nothing on the node v. Thus, the probability density $h_{m,v}$ is given by

$$h_{m,v} = \sum_{k=1}^{K} \kappa_{u_k,v} \prod_{i=1}^{k-1} (1 - \kappa_{u_i,v}) r_v \exp(-r_v(t_{m,v} - t_{m,u_k})). \quad (7)$$

The probability $g_{m,v}$ is the same as Equation (6).

Node delay with override The difference of diffusion process from *Node delay with non-override* is that here the actual activation time is allowed to be updated. For example, suppose that the node u_i first succeeded in satisfying the activation condition of the node v and the node v decided to activate itself at time $t_{u_i} + \delta_i$. At some time later but before $t_{u_i} + \delta_i$, other parent u_j also succeeded in satisfying the activation condition of the node v. Then the node v is allowed to change its actual activation time to time $t_{u_j} + \delta_j$ which may be before $t_{u_i} + \delta_i$. Thus, the probability density $h_{m,v}$ is given by

$$\begin{aligned} h_{m,v} &= \sum_{k=1}^{K} \kappa_{u_k,v} r_v \exp(-r_v(t_{m,v} - t_{m,u_k})) \\ &\quad \times \prod_{i=1, i \neq k}^{K} (1 - \int_{t_{m,u_i}}^{t_{m,v}} \kappa_{u_i,v} r_v \exp(-r_v(t - t_{m,u_i})) dt) \\ &= \sum_{k=1}^{K} \kappa_{u_k,v} r_v \exp(-r_v(t_{m,v} - t_{m,u_k})) \\ &\quad \times \prod_{i=1, i \neq k}^{K} \left(\kappa_{u_i,v} \exp(-r_v(t_{m,v} - t_{m,u_i})) + (1 - \kappa_{u_i,v})\right). \end{aligned} \quad (8)$$

The probability $g_{m,v}$ is the same as Equation (6).

3.3 Asynchronous Linear Threshold Models

Link delay with non-override The diffusion process unfolds in continuous-time t, and proceeds from a given initial active set S in the following way. When a node u_i is activated at t_{u_i}, it exerts its effect on its child node v with a delay δ_i. Suppose that the accumulated weight from the active parents of node v has become no less than θ_v at time t for the first time. The node v becomes active without any delay (no node delay) and exerts its effect on its child with a delay δ_j. Because there is no override, there is no update of the activation time of the node v. This process is repeated until no more activations are possible.

It is to be noted that because δ_i is a random variable, $t_{u_i} + \delta_i$ is not monotonic with respect to i even though u_i is ordered according to the activation time t_{u_i}. We define a new ordering of the parent node i according to the time $t_{u_i} + \delta_i$ that it exerts its effect on its child v. Suppose the node v first become activated for i of this new ordering. Then the threshold θ_v is between $\sum_{j=1}^{i-1} \omega_{u_j,v}$ and $\sum_{j=1}^{i-1} \omega_{u_j,v} + \omega_{u_i,v}$. Since θ_v is uniformly distributed, the probability that θ_v is chosen from this range is $\omega_{u_i,v}$. Thus, the probability density $h_{m,v}$

that the node v is activated at time $t_{m,v}$ can be expressed as

$$h_{m,v} = \sum_{k=1}^{K} \omega_{u_k,v} r_{u_k,v} \exp(-r_{u_k,v}(t_{m,v} - t_{m,u_k})). \tag{9}$$

The probability $g_{m,v}$ that a node v is not activated with the observed time period $[t_m, T_m]$ is given by

$$\begin{aligned} g_{m,v} &= 1 - \sum_{k=1}^{K} \omega_{u_k,v} \int_{t_{m,u_k}}^{T_m} r_{u_k,v} \exp(-r_{u_k,v}(t - t_{m,u_k}))dt \\ &= 1 - \sum_{k=1}^{K} \omega_{u_k,v}(1 - \exp(-r_{u_k,v}(T_m - t_{m,u_k}))). \end{aligned} \tag{10}$$

Node delay with non-override The difference of diffusion process from *Link delay with non-override* is that as soon as the parent node u_i is activated, its effect is immediately exerted to its child v. The delay depends on the node v's choice.

Suppose the node v first became activated for i of the parent ordering according to the time t_{u_i}. Then by the same reasoning as before, the threshold θ_v is between $\sum_{j=1}^{i-1} \omega_{u_j,v}$ and $\sum_{j=1}^{i-1} \omega_{u_j,v} + \omega_{u_i,v}$, and the probability density $h_{m,v}$ can be expressed as

$$h_{m,v} = \sum_{k=1}^{K} \omega_{u_k,v} r_v \exp(-r_v(t_{m,v} - t_{m,u_k})). \tag{11}$$

The probability $g_{m,v}$ is the same as Equation (10). Note that the formulation in Saito et al. (2010) corresponds to this category.

Node delay with override The difference of diffusion process from *Node delay with non-override* is that multiple updates of the activation time of the node v is allowed. Suppose that the node v first became activated by receiving the effect of the parent u_i. All the parents that have become activated after that can still influence the updates. Considering the probability that the node u_i's effect eventually leads the node v's activation at a time later than $t_{m,v}$, the probability density that the node v is activated by the node u_k at time $t_{m,v}$ is

$$\begin{aligned} h_{m,u_k,v} &= \omega_{u_k,v} \sum_{j=k}^{K} r_v \exp(-r_v(t_{m,v} - t_{m,u_k})) \prod_{i=k, i\neq j}^{K} \int_{t_{m,v}}^{\infty} r_v \exp(-r_v(t - t_{m,u_i}))dt \\ &= \omega_{u_k,v}(K - k + 1) r_v \prod_{i=k}^{K} \exp(-r_v(t_{m,v} - t_{m,u_i})). \end{aligned} \tag{12}$$

Thus, finally we obtain

$$h_{m,v} = \sum_{k=1}^{K} h_{m,u_k,v}. \tag{13}$$

The probability $g_{m,v}$ is the same as Equation (10).

4. Properties of Asynchronous Time-delay models

In this section, we describe some properties of asynchronous time-delay models in terms of the expected influence degree and behavioral analysis.

4.1 Expected Influence Degree

The expected influence degree of each node v, which is defined by the expected length of information diffusion sequence starting from the node v, plays a crucial role to solve the several important problems such as influence maximization and contamination minimization. Here we can easily see that for the same diffusion parameters $\kappa_{u,v}$ (or $\omega_{u,v}$), the expected influence degree of each node obtained by the basic IC (or LT) model is equal to the one obtained by any variant of AsIC (or AsLT) model after a substantially large time has passed. This is because what the asynchronous time models are doing is simply controlling the activation time of each node in relative to the basic models, but the asymptotic values of the expected influence degree remain the same. Thus, for the purpose of obtaining the expected influence degree, it suffices to use the basic models and we can apply any kind of efficient methods such as the bond percolation method (Kimura et al., 2010).

However, for some applications, such as the maximization of information spread to promote sales during a certain period of time, estimating the expected influence degree at a specific time or at a specific time interval may become very important and essential, *i.e.* a transient phenomenon becomes important. In particular, we can naturally conceive that each variant of the asynchronous time-delay models shows a different effect of time-delay on information diffusion. Since it is quite difficult to obtain analytical results, we attempt to clarify such effect by performing the experimental evaluation shown in the next section.

4.2 Behavioral Analysis

It has been shown in Saito et al. (2009, 2010) that behavioral analysis can reveal intrinsic characteristics of a given information diffusion sequence, under the assumption that people behave quite similarly for the same topic of information diffusion. Thus far, we have assumed that $\mathbf{\Theta}$ can vary with respect to nodes and links but is independent of the topic of information diffused. However, as predicted, they may be sensitive to the topic. If we place a constraint that $\mathbf{\Theta}$ depends only on topics but not on nodes and links of the network G, we can assign a different m to a different topic. Under this setting, we can set $r_{m,u,v} = r_m$ or $r_{m,v} = r_m$, $\kappa_{m,u,v} = \kappa_m$ and $\omega_{m,u,v} = \omega_m = q_m|B(v)|^{-1}$ for any link $(u, v) \in E$ and for any node $v \in V$. Here note that the newly introduced parameter $q_m (0 < q_m < 1)$ is the one which corresponds to κ in the AsIC model and $\omega_{v,v} = 1 - q_m$. Using each pair of the estimated parameters, (r_m, κ_m) for the AsIC model and (r_m, q_m) for the AsLT 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 the two-dimensional space. The validity of the above assumption has been confirmed using a real diffusion dataset in blogsphare as exemplified in Saito et al. (2009, 2010).

Looking through the results in the previous subsections, we note that in case of the AsIC model $h_{m,v}$ takes the same form for *Link delay with non-override* and *Node delay with override*, and in case of the AsLT model $h_{m,v}$ takes the same form for *Node delay with* *non-override* and *Link delay with non-override*. This means that in terms of the behavioral analysis as is explained above, interestingly these respective two different time delay models give the same results.

5. Evaluation of Effect of Time-delay on Information Diffusion

As mentioned earlier, we empirically evaluate the effect of the difference in the time-delay models on information diffusion using four real world networks. To this end, we introduce the following unified measure to quantify the average speed of propagation for networks of different sizes, as well as with various parameter settings in the information diffusion models.

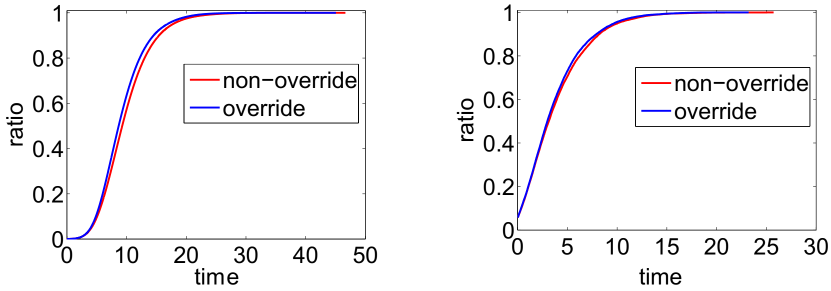
$$p(t) \quad = \quad \frac{\sum_{m=1}^{M} |\{(v_m, t_{m,v}) \in \mathcal{D}_m; t_{m,v} \leq t\}|}{\sum_{m=1}^{M} |\mathcal{D}_m|}. \tag{14}$$

For a given set of information results $\{D_m;\ m = 1, \cdots, M\}$ and a specified time t, this measure gives the expected ratio of the number of activated nodes until t to that of the total activated nodes. In our experiments, the initial and final times were set to $t_m = 0$ and $T_m = \infty$, respectively, for each information diffusion sequence m.

5.1 Network Data

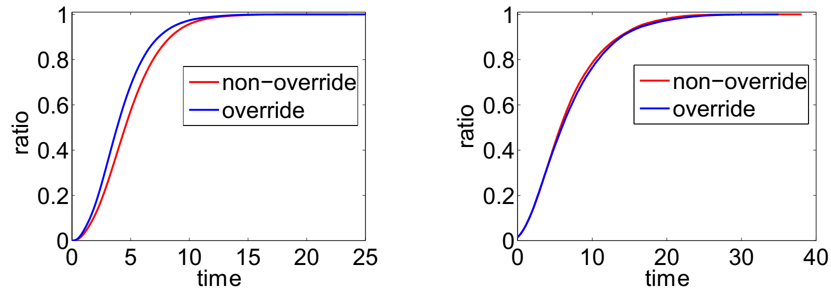
We employed four datasets of large real networks (all bidirectionally connected) and used their structures to generate diffusion data. The first one is a coauthorship network used in Palla et al. (2005) and has $12,357$ nodes and $38,896$ directed links (the coauthorship network). The second one is a trackback network of Japanese blogs used in Kimura et al. (2009) and has $12,047$ nodes and $79,920$ directed links (the blog network). The third one is a network derived from the Enron email dataset (Klimt and Yang, 2004) by extracting the senders and the recipients and linking those that had bidirectional communications. It has $4,254$ nodes and $44,314$ directed links (the Enron network). The fourth one is a network of people derived from the "list of people" within Japanese Wikipedia, also used in Kimura et al. (2009), and has $9,481$ nodes and $245,044$ directed links (the Wikipedia network).

As a practical situation, we evaluated the information diffusion models in the framework of behavioral analyses. Then, as explained in the previous section, link delay with non-override and node delay with override are indistinguishable for the AsIC model, while link delay and node delay both with non-override are indistinguishable for the AsLT model. Thus, we focused on node delay and evaluated the effect of override and non-override for both the AsIC and AsLT models, *i.e.*, $\kappa_{u,v} = \kappa$, $r_v = r$ for AsIC, and $\omega_{u,v} = q|B(v)|^{-1}$, $r_v = r$ for AsLT. In our preliminary experiments, changing the parameter r worked only for scaling the time axis of the diffusion results. Thus, we fixed its value at 1 ($r = 1$) for all cases and evaluated the effects of other diffusion parameters (κ for the AsIC model and q for the AsLT model). We prepared two different values (small and big) for both κ and q for each network. The values for κ were chosen to be the double and the half of the baseline value which is defined by $1/\bar{d}$, where $\bar{d}$ is the mean out-degree of a network. Each baseline value of κ becomes 0.2 for the coauthorship network, 0.1 for the blog and Enron networks, and 0.04 for the Wikipedia network. The values for q were set to 1 and 0.5, respectively



(a) $\kappa = 0.4$ (b) $\kappa = 0.1$

Figure 1: Results for the AsIC models in the coauthor network.



(a) $\kappa = 0.2$ (b) $\kappa = 0.05$

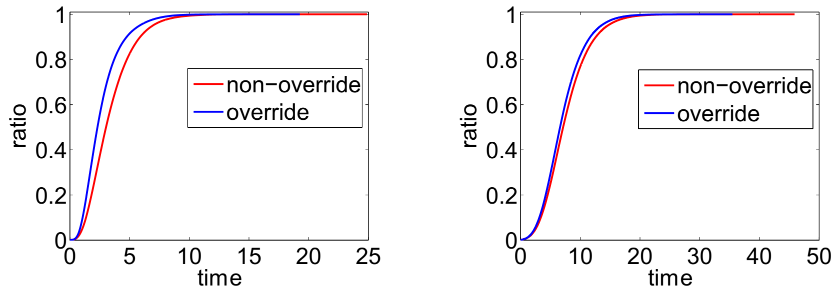
Figure 2: Results for the AsIC models in the blog network.

and used for all networks. Eventually, $M = 1,000$ information diffusion results with the sequence length of at least 10 were generated for each of these parameter values for each network, randomly selecting an initial active node for each diffusion result.

5.2 Experimental Results

In Figures 1, 2, 3, and 4, we show experimental results for the AsIC models by using the respective networks: coauthorship, blog, Enron, and Wikipedia. We note that it takes longer for the ratio to converge to 1.0 in Figure 1a than in Figure 1b although the diffusion probability κ is larger in Figure 1a than in Figure 1b. This does not necessarily mean that the diffusion is slower for the case where the diffusion probability is larger. The main reason is due to the difference of the number of active nodes. A larger diffusion probability generates a longer diffusion sequence which, in turn, takes a longer time. This tendency is not clear for the other figures because the diffusion probability is at most $\kappa = 0.2$. The same is true for the AsLT model.

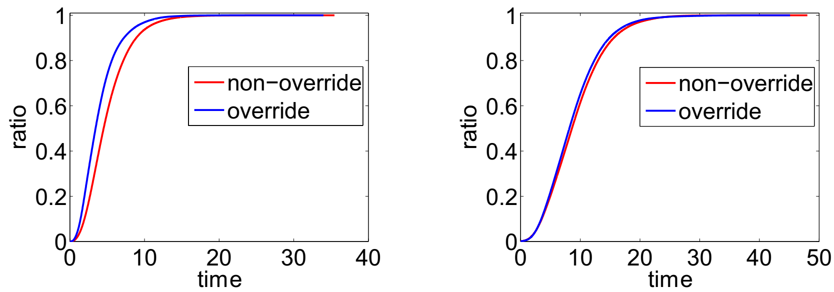
We further see that there is very little difference between non-override and override schemes when the diffusion parameter is small (half of the baseline value), but the difference becomes larger and the speed of information diffusion becomes faster for override scheme when the diffusion parameter is large (double of the baseline value). Here note that we chose each diffusion parameter according to the ratio of the numbers of nodes to links. This means that the value for κ is set to be reversely proportional to the network denseness and



(a) $\kappa = 0.2$

(b) $\kappa = 0.05$

Figure 3: Results for the AsIC models in the enron network.



(a) $\kappa = 0.08$

(b) $\kappa = 0.02$

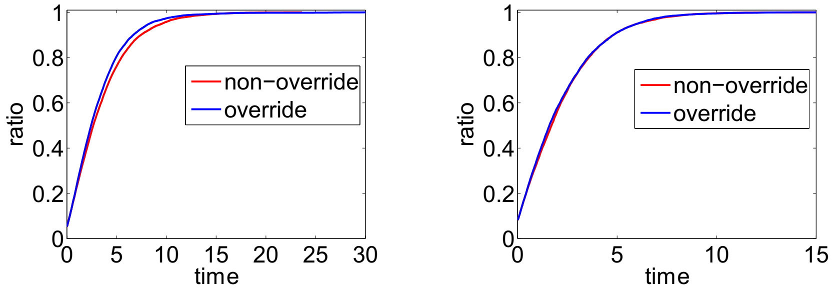
Figure 4: Results for the AsIC models in the wiki network.

the diffusion properties are supposed to be similar to each other. Thus, the results indicate that the effects of time-delay on the information diffusion models become much larger for a denser network when the diffusion parameter is large although no such difference is observed among the networks of different denseness when the diffusion parameter is small.

In Figures 5, 6, 7, and 8, we show experimental results for the AsLT models by using the respective networks: coauthorship, blog, Enron, and Wikipedia. Similarly to the AsIC model, here again we see hardly the difference between non-override and override schemes when the diffusion parameters are small ($q = 0.5$), but we do see that there is the difference between the two schemes and the speed of information diffusion becomes faster for override scheme when the diffusion parameters are large ($q = 1$). The effect of the difference of the network denseness is similar to the results for the AsIC model. In particular, we observe this difference is larger in the order of the Wikipedia, Enron, blog, and coauthorship networks. Here note that this order coincides with the descending order of their average degrees, *i.e.*, denseness of the network. This suggests that the effects of time-delay on the information diffusion models become much larger for a denser network when the diffusion parameter is large.

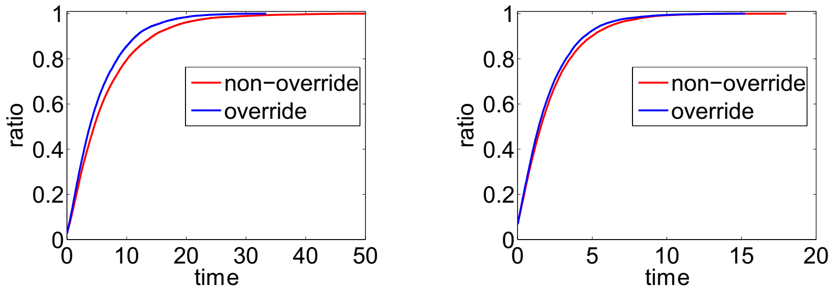
6. Conclusion

We formalized an information diffusion process as a generative model in the machine learning framework. In particular, we emphasized that the treatment of the time delay is important



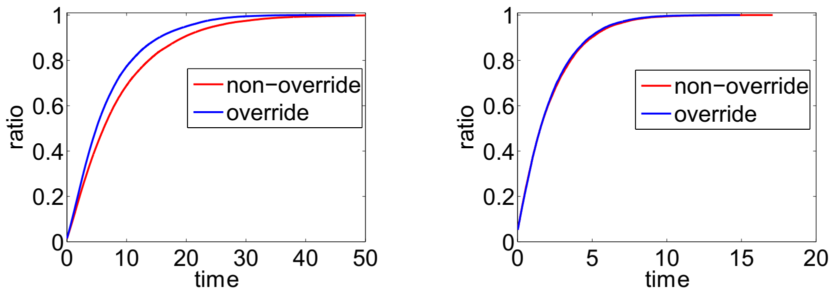
(a) $q = 1$ (b) $q = 0.5$

Figure 5: Results for the AsLT models in the coauthor network.



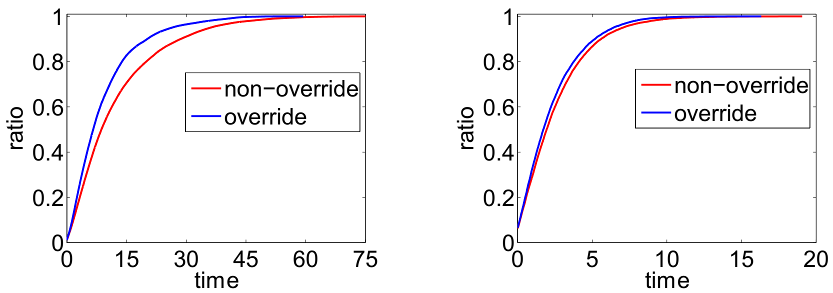
(a) $q = 1$ (b) $q = 0.5$

Figure 6: Results for the AsLT models in the blog network.



(a) $q = 1$ (b) $q = 0.5$

Figure 7: Results for the AsLT models in the enron network.



(a) $q = 1$ (b) $q = 0.5$

Figure 8: Results for the AsLT models in the wiki network.

in deriving the likelihood function. Diffusion comes with the notion of time and the probabilistic nature of the diffusion model hides the time delay structure from the surface of the observed sequence data, and makes the analysis difficult. We identified that there are two different types of time delay which we named link delay and node delay. The former corresponds to the delay associated with information propagation, and the latter corresponds to the delay associated with human action. We further identified that there are two different schemes of the way the activation from the multiple parents is updated which we named non-override and override. The former sticks to the initial activation and the latter can decide to update the time to activate multiple times. We applied these different notions of time delay to the well known basic information diffusion models: independent cascade (IC) and linear threshold (LT), and formalized asynchronous time delay versions of the IC and the LT models (AsIC and AsLT). We then derived a rigorous likelihood function for the feasible combinations of the time delay and update scheme for each of the AsIC and the AsLT models. There are a total of three different models for each diffusion models (AsIC and AsLT), but the theoretical analysis revealed that particular combinations of time delay and update scheme result in the same likelihood function (with a minor notational difference) and it is sufficient to consider two different models for each. We performed experiments to see how the difference in the time delay and the update scheme affects the information diffusion results as a function of time, varying the values of diffusion parameters using four real world networks. The simulation results reveal that there are differences in the spread of information diffusion and they strongly depend on the choice of the parameter values and the denseness of the network. We confirmed that it is important to distinguish the different types of time delay and update scheme in particular for a dense network that has a large information diffusion parameter value.

References

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.

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. Blocking links to minimize contamination spread in a social network. *ACM Transactions on Knowledge Discovery from Data*, 3:9:1–9:23, 2009.

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

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.

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.

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, 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, 2009.

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 International Workshop on Social Computing and Behavioral Modeling (SBP10)*, pages 149–158, 2010.

K. Saito, R. Nakano, and M. Kimura. Prediction of information diffusion probabilities for independent cascade model. In *Proceedings of the 12th International Conference, Knowledge-based Intelligent Information and Engineering Systems (KES2008)*, pages 41–49. LNAI 5179, 2008.

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.