Analyzing Mediator-Activity Effects for Trust-Network Evolution in Social Media

Keito Hatta[1], Masahito Kumano[1], Masahiro Kimura[1], Kazumi Saito[2], Kouzou Ohara[3], and Hiroshi Motoda[4,5]

[1] Department of Electronics and Informatics, Ryukoku University, Otsu, Japan
kimura@rins.ryukoku.ac.jp
[2] School of Administration and Informatics, University of Shizuoka, Shizuoka, Japan
k-saito@u-shizuoka-ken.ac.jp
[3] Department of Integrated Information Technology, Aoyama Gakuin University, Kanagawa, Japan
ohara@it.aoyama.ac.jp
[4] Institute of Scientific and Industrial Research, Osaka University, Osaka, Japan
motoda@ar.sanken.osaka-u.ac.jp
[5] School of Computing and Information Systems, University of Tasmania, Tasmania, Australia

Abstract. We analyze evolution of trust networks in social media sites from a perspective of mediators. To this end, we propose two stochastic models that simulate the dynamics of creating a trust link under the presence of mediators, *the A-ME and A-MAE models*, where the A-ME model analyzes mediator effects for trust-network evolution in terms of mediator types, and the A-MAE model, an extension of the A-ME model, analyzes mediator-activity effects for trust-network evolution. We present an efficient method of inferring the values of model parameters from an observed sequence of trust links and user activities. Using real data from Epinions, we experimentally show that the A-MAE model significantly outperforms the A-ME model for predicting trust links in the near future under the presence of mediators, and demonstrate the effectiveness of mediator-activity information for trust-network evolution. We further clarify, by using the A-ME and A-MAE models, several characteristic properties of trust-link creation probability in the Epinions data.

Keywords: Trust network evolution, Mediator activity, Trust link prediction.

1 Introduction

The emergence of Social Media has enabled us to collect large networks of trust relationships among people. The creation of a trust link from a user u to a user v implies that v is a reliable user for u, and u tends to accept and share a variety of information v presents. The structures of trust networks are important from a sociological point of view, and can also be exploited to efficiently find reliable information in an online world. Moreover, a trust network is a kind of social network, and plays an important role for information diffusion and opinion formation (Kempe et al. 2003; Kimura et al. 2010; Chen et al. 2013; Kimura et al. 2013; Saito et al. 2013). Recently, attention has been devoted to investigating the problem of predicting trust-links and social-links (Guha et

D.-N. Pham and S.-B. Park (Eds.): PRICAI 2014, LNAI 8862, pp. 297–308, 2014.

al. 2004; Liben-Nowell and Kleinberg 2007; Liu et al. 2008; Mannila and Terzi 2009; Nguyen et al. 2009; Gomez-Rodriguez et al. 2010; Tang et al. 2012; Tang et al. 2013).

Although there must be various reasons why one user u creates a trust link to another user v, the most fundamental is to explore the roles of users mediating between these two users u and v. In fact, Leskovec et al. (2010) exploited triads involving the link (u, v) as a basic feature for inferring the sign (positive or negative) of link (u, v) in a signed social network. In this paper, we investigate the evolution of a trust network from a perspective of mediators. Here, a mediator from user u to user v is defined as a user w such that w has a trust link either to or from u and also has a trust link either to or from v. Since information in online worlds can rapidly become obsolete as time passes, we only focus on mediators formed recently, and investigate their roles for creating trust links in the near future. There are four distinct types of mediators from user u to user v. For example, a mediator w of type 1 has a trust link from u to w and has a trust link from w to v while a mediator w' of type 2 has a trust link from u to w' and has a trust link from v to w' (see Fig. 1). We can naturally speculate that mediators of different types differently affect the trust link creation. We analyze the effects of mediators for creating trust links in terms of mediator-types.

Some social media sites provide a set of activities as well as the opportunity for connecting trust links, where users can select and perform one from a given set of activities. Examples include product review sites, where users can post a review and give a rating for a product in a given set of products (user activity), and can also create trust links each other (social interaction). Crandall et al. (2008) empirically proved the existence of the interplay between social interaction and behavioral similarity in online communities. This suggests a close relationship between 1) those to whom trust links are created and 2) activities performed in a social media site.

In this paper, we aim to investigate the effects of activities for trust link creation from a perspective of mediators. To this end, we propose two stochastic models that simulate the dynamics of creating trust links under the presence of mediators, *the A-ME and A-MAE models*, where the A-ME model aims at analyzing mediator effects for trust-network evolution in terms of mediator-types, and the A-MAE model enhances the A-ME model in order to analyze mediator-activity effects for trust-network evolution. For the A-ME and A-MAE models, we present an efficient method of inferring the values of parameters from an observed sequence of trust links and user activities. Using real data from Epinions[1] that is a social media site of product reviews and consumer reports, we experimentally show that the A-MAE model significantly outperforms the A-ME model for predicting trust links in the near future under the presence of mediators, and demonstrate the effectiveness of mediator-activity information for trust-network evolution. Moreover, we analyze the Epinions data based on the A-ME and A-MAE models, and clarify several characteristic properties of the probabilities of creating trust links with respect to mediator-activities and mediator-types in the Epinions data.

The rest of the paper is organized as follows: In Section 2, we describe the notations used in the paper. We define the A-ME and A-MAE models in Section 3, and present their learning methods in Section 4. In Section 5, we report the experimental results using real data from Epinions. We conclude the paper by summarizing the main results in Section 6.

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

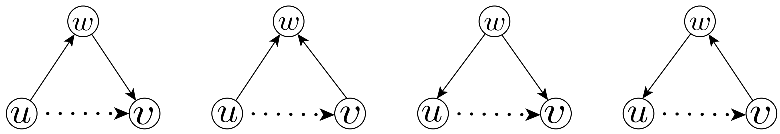


Fig. 1. Four types of mediator w from node u to node v

2 Preliminaries

We consider an online social media site offering trust links and activities. Here, we introduce the notations used throughout the paper.

For a positive integer t, let $G_t = (V, E_t)$ be the trust network created within a time-period $I_t = (t_0 + (t-1)\Delta t,\ t_0 + t\Delta t]$, where V is the set of nodes that correspond to the individual users in the site at time t_0, E_t ($\subset V \times V$) is the set of trust links created within time-period I_t, and Δt is a positive real number specified in advance. We suppose that there are no self-links and multiple-links. Note that $\bar{G}_t = (V, \bigcup_{s=1}^{t} E_s)$ is the trust network for the user set V at time $t_0 + t\Delta t$, and $E_s \cap E_{s'} = \emptyset$ if $s \neq s'$. We consider predicting the set E_{t+1} of trust links created within the next time-period I_{t+1}. We define a subset C_{t+1} of $V \times V$ by

$$C_{t+1} = (V \times V) \setminus \{(v, v) \mid v \in V\} \setminus \bigcup_{s=1}^{t} E_s.$$

Then, it is easily seen that $E_{t+1} \subset C_{t+1}$ and $\bigcup_{s=1}^{t} E_s \cap C_{t+1} = \emptyset$. Thus, we refer to C_{t+1} as the *set of candidate trust-links in time-period* I_{t+1}. For any $(u, v) \in C_{t+1}$, we investigate whether or not a trust link will be created from node u to node v in the next time-period I_{t+1}.

We assume that K activities are provided in the site. For any $u \in V$ and positive integer t, let

$$\boldsymbol{A}_t(u) = (A_{t,1}(u), \ldots, A_{t,K}(u))$$

denote the activity vector of node u within time-period I_t, where for each k, $A_{t,k}(u) = 1$ if user u selected and performed activity k within time-period I_t, and $A_{t,k}(u) = 0$ otherwise.

In this paper, we aim to investigate the roles of mediators for creating trust links. Thus, we focus on the subset C^*_{t+1} of C_{t+1} that consists of candidate trust-link $(u, v) \in C_{t+1}$ having a mediator $w \in V$ in time-period I_{t+1}, and for any $(u, v) \in C^*_{t+1}$, we consider modeling the probability $P_{t+1}(u, v)$ that a trust link is created from node u to node v in time-period I_{t+1}, i.e., $(u, v) \in E_{t+1}$. Here, node w is referred to as a *mediator* from node u to node v in time-period I_{t+1} when there exist both a trust link between u and w and a trust link between v and w that are created in time-period I_t. A mediator w from u to v is classified into four types: w is called *type 1* if $(u, w), (w, v) \in E_t$ (see Fig. 1a), w is called *type 2* if $(u, w), (v, w) \in E_t$ (see Fig. 1b), w is called *type 3* if $(w, u), (w, v) \in E_t$ (see Fig. 1c), and w is called *type 4* if $(w, u), (v, w) \in E_t$ (see Fig. 1d). In order to analyze the effects of activities for trust link creation, we aim to investigate the roles of mediators with respect to activities. Thus, we also consider incorporating activity information to model the probability $P_{t+1}(u, v)$ for any $(u, v) \in C^*_{t+1}$.

3 A-ME and A-MAE Models

For any $(u, v) \in C^*_{t+1}$, we consider modeling the probability $P_{t+1}(u, v)$ that trust link (u, v) is created in time-period I_{t+1}, i.e., $(u, v) \in E_{t+1}$. Note that by definition, there exists at least one mediator from node u to node v in I_{t+1}. In order to analyze the effect of activities for creating a trust link in terms of mediators, we propose two natural stochastic models of $P_{t+1}(u, v)$. The first model only uses mediator information, and the second model enhances the first model by adding mediator-activity effects for trust-network evolution.

3.1 A-ME Model

It is conceivable that the presence of mediators affects the creation of trust links. Moreover, we can speculate that the influence strength of a mediator depends on its type. Therefore, in order to analyze the effects of mediators for creating trust links in terms of mediator types, we propose modeling the probability $P_{t+1}(u, v)$ for any $(u, v) \in C^*_{t+1}$ by using a logistic regression model as follows:

$$P_{t+1}(u, v) = \frac{1}{1 + \exp(-\boldsymbol{\phi} \cdot \boldsymbol{y}_t(u, v))}, \tag{1}$$

where $\boldsymbol{\phi} \in \mathbf{R}^5$ is a parameter vector,

$$\boldsymbol{\phi} = (\phi_0, \phi_1, \phi_2, \phi_3, \phi_4),$$

$\boldsymbol{y}_t(u, v)$ is a feature vector of (u, v) at time $t_0 + t\Delta t$,

$$\boldsymbol{y}_t(u, v) = (1, y_{t,1}(u, v), y_{t,2}(u, v), y_{t,3}(u, v), y_{t,4}(u, v)),$$

and $\boldsymbol{\phi} \cdot \boldsymbol{y}_t(u, v)$ stands for the scalar product of vectors $\boldsymbol{\phi}$ and $\boldsymbol{y}_t(u, v)$,

$$\boldsymbol{\phi} \cdot \boldsymbol{y}_t(u, v) = \phi_0 + \sum_{i=1}^{4} \phi_i \, y_{t,i}(u, v).$$

Here, each $y_{t,i}(u, v)$ is the number of type i mediators from u to v in time-period I_{t+1}, We refer to this stochastic model to simulate the dynamics of creating a trust link as the *A-ME model*.

3.2 A-MAE Model

It is also conceivable that the influence degree of a mediator depends on activity. For $(u, v) \in C^*_{t+1}$, let us consider mediators w_k and w_ℓ from node u to node v in time-period I_{t+1} such that in time-period I_t, u, v and w_k performed the same activity k, and u, v and

w_ℓ did the same activity ℓ, that is, $A_{t,k}(u) = A_{t,k}(v) = A_{t,k}(w_k) = 1$, $A_{t,\ell}(u) = A_{t,\ell}(v) = A_{t,\ell}(w_\ell) = 1$, where $k \neq \ell$. Then, for creating a trust link from u to v, the influence that w_k and w_ℓ exert can be different. In order to analyze the effects of activities in terms of mediators, we propose modeling the probability $P_{t+1}(u, v)$ for any $(u, v) \in C^*_{t+1}$ by combining co-occurrence information with respect to activities with the A-ME model as follows:

$$P_{t+1}(u, v) = \sum_{k=1}^{K} \lambda_k \frac{1}{1 + \exp(-\boldsymbol{\theta}_k \cdot \boldsymbol{x}_{t,k}(u, v))}, \tag{2}$$

where $\boldsymbol{\lambda}$ is a parameter vector,

$$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K); \quad \sum_{k=1}^{K} \lambda_k = 1, \quad \lambda_k > 0 \;\; (k = 1, \ldots, K),$$

each $\boldsymbol{\theta}_k \in \mathbf{R}^5$ is a parameter vector with respect to activity k,

$$\boldsymbol{\theta}_k = (\theta_{k,0}, \theta_{k,1}, \theta_{k,2}, \theta_{k,3}, \theta_{k,4}),$$

and each $\boldsymbol{x}_{t,k}(u, v)$ is a feature vector of (u, v) with respect to activity k at time $t_0 + t\Delta t$,

$$\boldsymbol{x}_{t,k}(u, v) = (1, x_{t,k,1}(u, v), x_{t,k,2}(u, v), x_{t,k,3}(u, v), x_{t,k,4}(u, v)).$$

Here, each $x_{t,k,i}(u, v)$ is the number of type i mediators w from u to v in time-period I_{t+1} such that u, v and w performed activity k in I_t, that is, $x_{t,k,i}(u, v) \geq 0$. In particular, we assume that for any $(u, v) \in C^*_{t+1}$, there exist a mediator w' in time-period I_{t+1} and an activity k' such that nodes u, v and w' performed activity k' in I_t, that is, $x_{t,k',i'}(u, v) > 0$, where w' is of type i'. We refer to this stochastic model to simulate the dynamics of creating a trust link as the *A-MAE model*.

4 Inference Method

Suppose we are given the observed data in time-period I_t, consisting of a set of trust links E_t and a set of user activities $\mathcal{A}_t = \{\boldsymbol{A}_t(u) \mid u \in V\}$, for $t = 1, \ldots, T$, where T is a positive integer with $T \geq 2$. Let $\mathcal{D}_T$ denote this observed sequence data for trust links and user activities, i.e.,

$$\mathcal{D}_T = \{(E_t, \mathcal{A}_t) \mid t = 1, \ldots, T\}.$$

We consider inferring the values of parameters of the A-ME and A-MAE models from $\mathcal{D}_T$. Notice that the A-ME model can be regarded as a special case of the A-MAE model by setting $K = 1$ (see Eqs. (1) and (2)). Thus, we only describe the inference method for the A-MAE model.

We conform to the framework of MAP estimation. We assume a Dirichlet prior for $\boldsymbol{\lambda}$ and Gaussian priors for $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K$. Thus, we estimate the values of $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K$ by maximizing the objective function

$$\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K) = \sum_{t=1}^{T-1} \left(\sum_{(u,v) \in C^*_{t+1} \cap E_{t+1}} \log \left(\sum_{k=1}^{K} \frac{\lambda_k}{1 + \exp(-\boldsymbol{\theta}_k \cdot \boldsymbol{x}_{t,k}(u,v))} \right) \right.$$
$$\left. + \sum_{(u,v) \in C^*_{t+1} \setminus E_{t+1}} \log \left(\sum_{k=1}^{K} \frac{\lambda_k \exp(-\boldsymbol{\theta}_k \cdot \boldsymbol{x}_{t,k}(u,v))}{1 + \exp(-\boldsymbol{\theta}_k \cdot \boldsymbol{x}_{t,k}(u,v))} \right) \right)$$
$$+ \gamma \sum_{k=1}^{K} \log \lambda_k - \sum_{k=1}^{K} \frac{1}{2\sigma_k^2} \sum_{i=0}^{4} \theta_{k,i}^2, \tag{3}$$

where $\gamma > 0$ and $\sigma_k > 0$ $(k = 1, \ldots, K)$ are regularization constants. We find optimal values of parameters by an EM algorithm. Let $\bar{\lambda} = (\bar{\lambda}_1, \ldots, \bar{\lambda}_K)$ be the current estimate of $\boldsymbol{\lambda}$, and let $\bar{\boldsymbol{\theta}}_k = (\bar{\theta}_{k,0}, \ldots, \bar{\theta}_{k,4})$ be the current estimate of $\boldsymbol{\theta}_k$ for $k = 1, \ldots, K$. By Jensen's inequality from Eq. (3), we have

$$\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K) - \mathcal{L}(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\theta}}_1, \ldots, \bar{\boldsymbol{\theta}}_K) \geq Q(\boldsymbol{\lambda}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K \mid \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\theta}}_1, \ldots, \bar{\boldsymbol{\theta}}_K), \tag{4}$$

where

$$Q(\boldsymbol{\lambda}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K \mid \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\theta}}_1, \ldots, \bar{\boldsymbol{\theta}}_K)$$
$$= \sum_{t=1}^{T-1} \sum_{k=1}^{K} \left(\sum_{(u,v) \in C^*_{t+1} \cap E_{t+1}} \bar{c}_{t,k}(u,v) \left\{ \log \lambda_k - \log\left(1 + \exp(-\boldsymbol{\theta}_k \cdot \boldsymbol{x}_{t,k}(u,v))\right) \right\} \right.$$
$$\left. + \sum_{(u,v) \in C^*_{t+1} \setminus E_{t+1}} \bar{d}_{t,k}(u,v) \left\{ \log \lambda_k - \boldsymbol{\theta}_k \cdot \boldsymbol{x}_{t,k}(u,v) - \log\left(1 + \exp(-\boldsymbol{\theta}_k \cdot \boldsymbol{x}_{t,k}(u,v))\right) \right\} \right)$$
$$+ \gamma \sum_{k=1}^{K} \log \lambda_k - \sum_{k=1}^{K} \frac{1}{2\sigma_k^2} \sum_{i=0}^{4} \theta_{k,i}^2 + \bar{Q}_0(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\theta}}_1, \ldots, \bar{\boldsymbol{\theta}}_K), \tag{5}$$

and for each (t, k, u, v),

$$\bar{c}_{t,k}(u,v) = \frac{\bar{\lambda}_k}{1 + \exp\left(-\bar{\boldsymbol{\theta}}_k \cdot \boldsymbol{x}_{t,k}(u,v)\right)} \Bigg/ \sum_{\ell=1}^{K} \frac{\bar{\lambda}_\ell}{1 + \exp\left(-\bar{\boldsymbol{\theta}}_\ell \cdot \boldsymbol{x}_{t,\ell}(u,v)\right)} > 0, \tag{6}$$

$$\bar{d}_{t,k}(u,v) = \frac{\bar{\lambda}_k \exp\left(-\bar{\boldsymbol{\theta}}_k \cdot \boldsymbol{x}_{t,k}(u,v)\right)}{1 + \exp\left(-\bar{\boldsymbol{\theta}}_k \cdot \boldsymbol{x}_{t,k}(u,v)\right)} \Bigg/ \sum_{\ell=1}^{K} \frac{\bar{\lambda}_\ell \exp\left(-\bar{\boldsymbol{\theta}}_\ell \cdot \boldsymbol{x}_{t,\ell}(u,v)\right)}{1 + \exp\left(-\bar{\boldsymbol{\theta}}_\ell \cdot \boldsymbol{x}_{t,\ell}(u,v)\right)} > 0. \tag{7}$$

Here, $\bar{Q}_0(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\theta}}_1, \ldots, \bar{\boldsymbol{\theta}}_K)$ is a function of $\bar{\boldsymbol{\lambda}}$ and $\bar{\boldsymbol{\theta}}_1, \ldots, \bar{\boldsymbol{\theta}}_K$ such that $Q(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\theta}}_1, \ldots, \bar{\boldsymbol{\theta}}_K \mid \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\theta}}_1, \ldots, \bar{\boldsymbol{\theta}}_K) = 0$. From Eqs. (4) and (5), the update formulas of $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K$ can be derived by maximizing $Q(\boldsymbol{\lambda}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K \mid \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\theta}}_1, \ldots, \bar{\boldsymbol{\theta}}_K)$ with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K$. Thus, we can first obtain the following update formula for $\boldsymbol{\lambda}$ from Eqs. (5), (6) and (7):

$$\lambda_k = \frac{\sum_{t=1}^{T-1} \left(\sum_{(u,v) \in C^*_{t+1} \cap E_{t+1}} \bar{c}_{t,k}(u,v) + \sum_{(u,v) \in C^*_{t+1} \setminus E_{t+1}} \bar{d}_{t,k}(u,v) \right) + \gamma}{\sum_{\ell=1}^{K} \sum_{t=1}^{T-1} \left(\sum_{(u,v) \in C^*_{t+1} \cap E_{t+1}} \bar{c}_{t,\ell}(u,v) + \sum_{(u,v) \in C^*_{t+1} \setminus E_{t+1}} \bar{d}_{t,\ell}(u,v) \right) + \gamma K}$$

for $k = 1, \ldots, K$. Next, we derive the update formulas for $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K$. From Eqs. (5), (6) and (7), we have

$$\frac{\partial Q}{\partial \theta_{k,i}} = \sum_{t=1}^{T-1} \Bigg(\sum_{(u,v) \in C^*_{t+1} \cap E_{t+1}} \frac{\bar{c}_{t,k}(u,v)\, x_{t,k,i}(u,v)\, \exp(-\boldsymbol{\theta}_k \cdot \boldsymbol{x}_{t,k}(u,v))}{1 + \exp(-\boldsymbol{\theta}_k \cdot \boldsymbol{x}_{t,k}(u,v))} - \sum_{(u,v) \in C^*_{t+1} \setminus E_{t+1}} \frac{\bar{d}_{t,k}(u,v)\, x_{t,k,i}(u,v)}{1 + \exp(-\boldsymbol{\theta}_k \cdot \boldsymbol{x}_{t,k}(u,v))} \Bigg) - \frac{1}{\sigma_k^2} \theta_{k,i} \tag{8}$$

for $k = 1, \ldots, K$ and $i = 0, 1, 2, 3, 4$. Moreover, we have

$$\frac{\partial^2 Q}{\partial \theta_{k,i}\, \partial \theta_{k,j}} = -\sum_{t=1}^{T-1} \Bigg(\sum_{(u,v) \in C^*_{t+1} \cap E_{t+1}} \frac{\bar{c}_{t,k}(u,v)\, x_{t,k,i}(u,v)\, x_{t,k,j}(u,v)\, \exp(-\boldsymbol{\theta}_k \cdot \boldsymbol{x}_{t,k}(u,v))}{\{1 + \exp(-\boldsymbol{\theta}_k \cdot \boldsymbol{x}_{t,k}(u,v))\}^2} + \sum_{(u,v) \in C^*_{t+1} \setminus E_{t+1}} \frac{\bar{d}_{t,k}(u,v)\, x_{t,k,i}(u,v)\, x_{t,k,j}(u,v)\, \exp(-\boldsymbol{\theta}_k \cdot \boldsymbol{x}_{t,k}(u,v))}{\{1 + \exp(-\boldsymbol{\theta}_k \cdot \boldsymbol{x}_{t,k}(u,v))\}^2} \Bigg) - \frac{1}{\sigma_k^2} \delta_{i,j} \tag{9}$$

for $k = 1, \ldots, K$ and $i, j = 0, 1, 2, 3, 4$, where $\delta_{i,j}$ is the Kronecker delta. From Eq. (9), we can guarantee that the matrix $(\partial^2 Q / \partial \theta_{k,i} \partial \theta_{k,j})$ is negative definite. By solving the equations $\partial Q / \partial \theta_{k,i} = 0$ ($k = 1, \ldots, K$, $i = 0, 1, 2, 3, 4$) (see Eq. (8)), we can find the values of $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K$ that maximize $Q(\boldsymbol{\lambda}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K \mid \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\theta}}_1, \ldots, \bar{\boldsymbol{\theta}}_K)$, that is, obtain the update formulas for $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K$. We employed a standard Newton method in our experiments (see Eqs. (8) and (9)).

5 Experiments

Using real data from Epinions, we first evaluate the A-ME and A-MAE models in terms of trust-link prediction. Next, we try to analyze the properties of trust-link creation behavior in the Epinions data on the basis of the A-ME and A-MAE models.

5.1 Social Media Data

We collected real data for a trust network and a set of user activities from Epinions, which is a social media site of product reviews and consumer reports. In Epinions, a user u can create a trust link to another user v by registering v as a trust user. We examined the evolution of the trust network constructed from trust links among users. Also, in Epinions, a user can post a review and give a rating for a product in a given set of products, where those products are classified into K categories. We say that *user u performed activity k* when u posted a review or gave a rating for some product of category k.

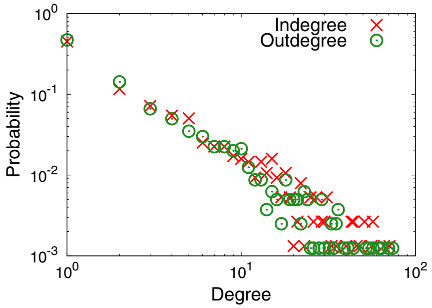


Fig. 2. Degree distribution of the Epinions data

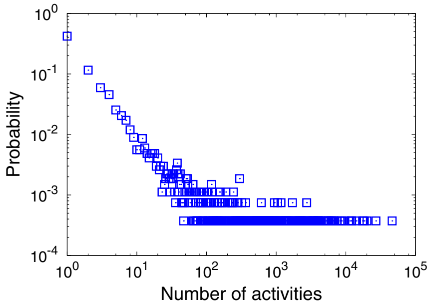


Fig. 3. Activity distribution of the Epinions data

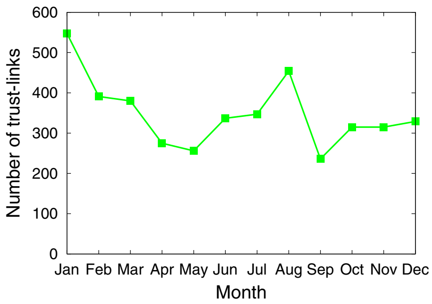


Fig. 4. Fluctuation in the number of trust-links created in the Epinions data

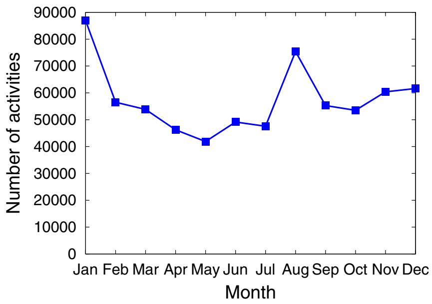


Fig. 5. Fluctuation in the number of activities performed in the Epinions data

By the breadth-first search, we traced in the trust links from a user who was featured as the most popular user in October 2012 until no new users appeared, and collected both a set of trust links and a set of product reviews and ratings. The collected data contains 27,873 users, 218,686 trust links, and 809,521 reviews and 14,105,311 ratings for 268,897 products, where the number of categories was 19, i.e., $K = 19$. The 19 categories (i.e., activities) were as follows: 1) Hotel & Travel, 2) Web Sites & Internet Services, 3) Business & Technology, 4) Kids & Family, 5) Sports & Outdoors, 6) Computers & Internet, 7) Electronics, 8) Games, 9) Restaurants & Gourmet, 10) Media, 11) Cars & Motorsports, 12) Wellness & Beauty, 13) Gifts, 14) Home & Garden, 15) Personal Finance, 16) Education, 17) Miscellaneous, 18) Archived Reviews, 19) Local Services. On the basis of stability consideration, we exploited only the data generated in 2010, and constructed a dataset from those users that had trust-links and produced activities in 2010. We refer to this dataset as the Epinions data, where the number of users was 749.

We first examined basic statistical properties of the Epinions data. Figures 2 show the in-degree and out-degree distributions (see (Newman 2003)). Figure 3 displays the activity distribution, that is, the fraction of the number of activities a user performed. We observe that the in-degree, out-degree and activity distributions follow power-laws (see (Newman 2003)). These results imply that the Epinions data satisfies the typical

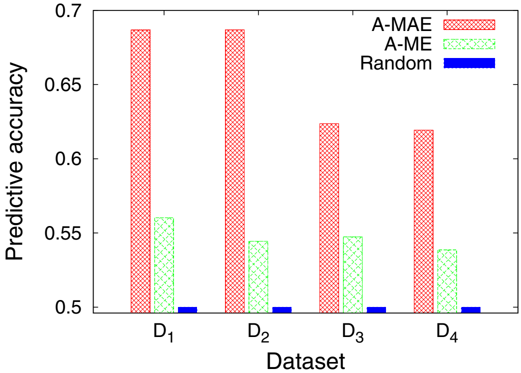


Fig. 6. Results of trust-link prediction for the Epinions data

properties of social networks and user activities. Figure 4 shows the fluctuation in the number of trust links created within a month. Also, Figure 5 displays the fluctuation in the number of activities performed within a month. We observe that trust links were almost constantly created as well as user activities were almost constantly performed.

5.2 Trust Link Prediction

In order to investigate the effects of mediator-types and mediator-activities, we evaluate the A-ME and A-MAE models for predicting trust links in the near future under the presence of mediators. We estimate the values of parameters for the A-ME and A-MAE models from an observed sequence $\mathcal{D}_T$, and consider predicting the trust links belonging to the subset C^*_{T+1} by using these models. For any $(u, v) \in C^*_{T+1}$, let $\hat{P}_{T+1}(u, v)$ denote the estimate of probability $P_{T+1}(u, v)$ by the A-ME and A-MAE models. For any $(u, v) \in C^*_{T+1}$, we make a prediction as follows: (u, v) will become a trust-link in time-period I_{T+1} if $\hat{P}_{T+1}(u, v) \geq 1/2$, and (u, v) will not become a trust-link in time-period I_{T+1} otherwise. We measure the predictive accuracy according to Leskovec et al. (2010) in the following way: We randomly select a subset F_{T+1} of C^*_{T+1} satisfying $E_{T+1} \cap F_{T+1} = \emptyset$ and $|E_{T+1} \cap C^*_{T+1}| = |F_{T+1}|$. For each $(u, v) \in (E_{T+1} \cap C^*_{T+1}) \cup F_{T+1}$, we make a prediction based on the trust-link prediction method. Let $\hat{E}_{T+1}$ denote the set of links $(u, v) \in (E_{T+1} \cap C^*_{T+1}) \cup F_{T+1}$ such that the method predicts $(u, v) \in E_{T+1} \cap C^*_{T+1}$. Let $\hat{F}_{T+1}$ denote the set of links $(u, v) \in (E_{T+1} \cap C^*_{T+1}) \cup F_{T+1}$ such that the method predicts $(u, v) \in F_{T+1}$. Then, we calculate

$$PA = \left(\left|\hat{E}_{T+1} \cap (E_{T+1} \cap C^*_{T+1})\right| + \left|\hat{F}_{T+1} \cap F_{T+1}\right|\right) \Big/ \left(\left|E_{T+1} \cap C^*_{T+1}\right| + |F_{T+1}|\right).$$

In our experiments, we conducted five trials for selecting F_{T+1}, and evaluated the predictive accuracy by the average of PA over the five trials. Note that the predictive accuracy of random guessing is 0.5.

For the trust-link prediction task, we employed the Epinions data. By letting $T = 2$ for the observed data $\mathcal{D}_T$, and setting Δt as three months, we constructed four datasets

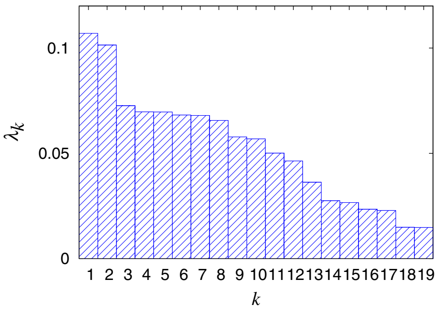


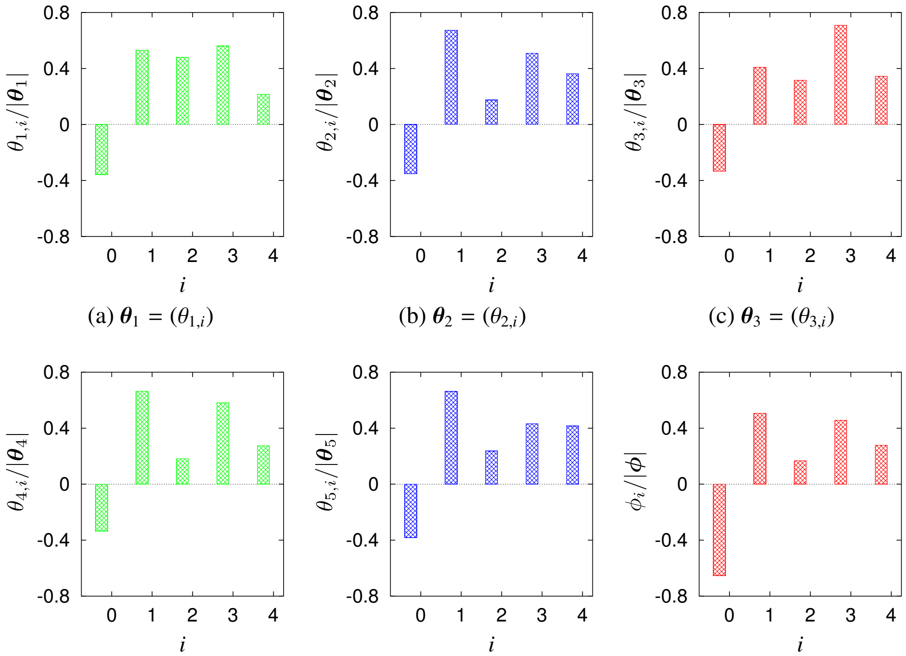
Fig. 7. Estimated values of $\boldsymbol{\lambda} = (\lambda_k)$ for dataset D_4

D_1, D_2, D_3 and D_4, where each D_j contains the data in time-periods I_1 and I_2 as training data and the data in time-period I_3 as test data. Here, the initial time t_0 is set as January 1 for D_1, February 1 for D_2, March 1 for D_3 and April 1 for D_4, respectively. For example, for dataset D_1, time-period I_1 is January 1 to March 31, time-period I_2 is April 1 to June 30, and time-period I_3 is July 1 to September 30. Using datasets D_1, D_2, D_3 and D_4, we compared the A-MAE model, A-ME model and random guessing, where the random guessing is applied as a kind of statistical test confirming whether the mediator effect for trust-link creation is significant or not. Figure 6 show the results. Compared to the random guessing, the A-ME model improved the predictive accuracy to about 0.55. This shows that incorporating mediator-type information has a positive effect for predicting trust-links. Moreover, the A-MAE model significantly outperformed the A-ME model. These results demonstrate the effectiveness of mediator-activity information for modeling the dynamics of creating a trust link in trust-network evolution.

5.3 Behavior Analysis

Based on the A-ME and A-MAE models, we analyzed the probabilities of creating trust links in terms of mediator-activities and mediator-types, and tried to clarify the properties of user behaviors in the Epinions data. Here, we only report the analysis results for the dataset D_4.

Figure 7 shows the estimated values of $\boldsymbol{\lambda} = (\lambda_k)$. We see that different mediator-activities can differently affect trust-link creation. The top-five influential activities in this period (i.e., April to September in 2010) were activities $k = 1$ (Hotel & Travel), $k = 2$ (Web Sites & Internet Services), $k = 3$ (Business & Technology), $k = 4$ (Kids & Family) and $k = 5$ (Sports & Outdoors). We observe that activities $k = 1$ and $k = 2$ especially played important roles for trust-link creation in this period. Next, we investigated the $\boldsymbol{\theta}_k$ for activity k ($k = 1, 2, 3, 4, 5$) and the $\boldsymbol{\phi}$ for the A-ME model that does not make use of mediator-activity information. Figure 8 shows the estimated values of $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{\theta}_3$, $\boldsymbol{\theta}_4$, $\boldsymbol{\theta}_5$ and $\boldsymbol{\phi}$, where they are normalized such that their Euclidean norms are equal to one. We see that different mediator-types can differently affect trust-link creation. The results of $\boldsymbol{\phi}$ implies that this period had the following tendencies: A mediator of type 1 was the most influential, that of type 3 followed, and that of type 2 was the least. We ob-



(a) $\boldsymbol{\theta}_1 = (\theta_{1,i})$ (b) $\boldsymbol{\theta}_2 = (\theta_{2,i})$ (c) $\boldsymbol{\theta}_3 = (\theta_{3,i})$

(d) $\boldsymbol{\theta}_4 = (\theta_{4,i})$ (e) $\boldsymbol{\theta}_5 = (\theta_{5,i})$ (f) $\boldsymbol{\phi} = (\phi_i)$

Fig. 8. Estimated values of $\boldsymbol{\theta}_k$ ($k = 1, 2, 3, 4, 5$) (A-MAE model) and $\boldsymbol{\phi}$ (A-ME model) for dataset D_4

serve that $\boldsymbol{\theta}_2$, $\boldsymbol{\theta}_4$, $\boldsymbol{\theta}_5$ and $\boldsymbol{\phi}$ were qualitatively similar. This suggests that activities $k = 2$, $k = 4$ and $k = 5$ were typical activities in this period in terms of user behaviors for trust-link creation. Moreover, we can find the following characteristic properties of the trust-link creation behaviors in this period: type 2 and type 3 mediators had especially strong influence for activities $k = 1$ and $k = 3$, respectively.

6 Conclusion

We addressed the problem of modeling the evolution of a trust network in a social media site. In particular, we focused on investigating the roles of mediators for trust-link creation. To this end, we proposed two stochastic models for simulating the dynamics of creating a trust link under the presence of mediators, the A-ME and A-MAE models, where the A-ME model aims to examine mediator effects for trust-network evolution in terms of mediator types, and the A-MAE model enhances the A-ME model to analyze mediator-activity effects for trust-network evolution. For these proposed models, we presented an efficient method of estimating the values of parameters from an observed sequence of trust links and user activities. Using real data from Epinions, we experimentally evaluated the A-ME and A-MAE models for predicting trust links in the near future under the presence of mediators. First, by comparing the A-ME model and random guessing, we demonstrated that incorporating mediator-type information has a positive effect for predicting trust-links. Next, we showed that the A-MAE model significantly outperforms the A-ME model, and demonstrated the effectiveness of mediator-activity information for trust-network evolution. We also showed that different mediator-activities differently affect trust-link creation, and different mediator-types differently affect trust-link creation. Moreover, by using the A-ME and A-MAE models, we found several characteristic properties of trust-link creation probability in the Epinions data in terms of mediator-activities and mediator-types.

Acknowledgments. This work was partly supported by Asian Office of Aerospace Research and Development, Air Force Office of Scientific Research under Grant No. AOARD-13-4042, and JSPS Grant-in-Aid for Scientific Research (C) (No. 26330352).

References

Chen, W., Lakshmanan, L., Castillo, C.: Information and influence propagation in social networks. Synthesis Lectures on Data Management 5, 1–177 (2013)

Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., Suri, S.: Feedback effects between similarity and social influence in online communities. In: Proceedings of KDD 2008, pp. 160–168 (2008)

Gomez-Rodriguez, M., Leskovec, J., Krause, A.: Inferring networks of diffusion and influence. In: Proceedings of KDD 2010, pp. 1019–1028 (2010)

Guha, R., Kumar, R., Raghavan, P., Tomkins, A.: Propagation of trust and distrust. In: Proceedings of WWW 2004, pp. 403–412 (2004)

Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of KDD 2003, pp. 137–146 (2003)

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

Kimura, M., Saito, K., Ohara, K., Motoda, H.: Learning to predict opinion share and detect anti-majority opinionists in social networks. J. Intell. Inf. Syst. 41, 5–37 (2013)

Leskovec, J., Huttenlocher, D., Kleinberg, J.: Predicting positive and negative links in online social networks. In: Proceedings of WWW 2010, pp. 641–650 (2010)

Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. J. Am. Soc. Inf. Sci. Technol. 58, 1019–1031 (2007)

Liu, H., Lim, E., Lauw, H., Le, M., Sun, A., Srivastava, J., Kim, Y.: Predicting trusts among users of online communities: an epinion case study. In: Proceedings of EC 2008, pp. 310–319 (2008)

Mannila, H., Terzi, E.: Finding links and initiators: A graph-reconstruction problem. In: Proceedings of SDM 2009, pp. 1207–1217 (2009)

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

Nguyen, V., Lim, E., Jiang, J., Sun, A.: To trust or not to trust? predicting online trusts using trust antecedent framework. In: Proceedings of ICDM 2009, pp. 896–901 (2009)

Saito, K., Kimura, M., Ohara, K., Motoda, H.: Learning asynchronous-time information diffusion models and its application to behavioral data analysis over social networks. Journal of Computer Engineering and Informatics 1, 30–57 (2013)

Tang, J., Gao, H., Hu, X., Liu, H.: Exploiting homophily effect for trust prediction. In: Proceedings of WSDM 2013, pp. 53–62 (2013)

Tang, J., Gao, H., Liu, H., Sarma, A.D.: etrust: Understanding trust evolution in an online world. In: Proceedings of KDD 2012, pp. 253–261 (2012)