

Learning Attribute-weighted Voter Model over Social Networks

Yuki Yamagishi

*School of Management and Information
University of Shizuoka*

B08107@U-SHIZUOKA-KEN.AC.JP

Kazumi Saito

*School of Management and Information
University of Shizuoka*

K-SAITO@U-SHIZUOKA-KEN.AC.JP

Kouzou Ohara

*Department of Integrated Information Technology
Aoyama Gakuin University*

OHARA@IT.AOYAMA.AC.JP

Masahiro Kimura

*Department of Electronics and Informatics
Ryukoku University*

KIMURA@RINS.RYUKOKU.AC.JP

Hiroshi Motoda

*Institute of Scientific and Industrial Research
Osaka University*

MOTODA@AR.SANKEN.OSAKA-U.AC.JP

Editor: Chun-Nan Hsu and Wee Sun Lee

Abstract

We propose an opinion formation model, an extension of the voter model that incorporates the strength of each node, which is modeled as a function of the node attributes. Then, we address the problem of estimating parameter values for these attributes that appear in the function from the observed opinion formation data and solve this by maximizing the likelihood using an iterative parameter value updating algorithm, which is efficient and is guaranteed to converge. We show that the proposed algorithm can correctly learn the dependency in our experiments on four real world networks for which we used the assumed attribute dependency. We further show that the influence degree of each node based on the extended voter model is substantially different from that obtained assuming a uniform strength (a naive model for which the influence degree is known to be proportional to the node degree), and is more sensitive to the node strength than the node degree even for a moderate value of the node strength.

Keywords: voter model, influence degree, attribute dependency.

1. Introduction

The growth of Internet has enabled to form various kinds of large-scale social networks, and a variety of information, e.g. news, ideas, hot topics, rumors, innovations, etc. spreads in the form of “word-of-mouth” communications. It is noticeable to observe how much they affect our daily life style. The spread of information has been studied by many researchers (Newman et al., 2002; Newman, 2003; Gruhl et al., 2004; Domingos, 2005; Leskovec et al., 2006; Kimura et al., 2009, 2010a). The information diffusion models widely used 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) models. Both have been used to solve such problems as the *influence maximization problem* (Kempe et al., 2003; Kimura et al., 2007) and the *contamination minimization problem* (Kimura et al., 2009; Tong et al., 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. Thus, it can be said that the IC model emphasizes “information push” and the LT model “information pull”.

In this paper, we address a different kind of information diffusion, which is “opinion formation”, *i.e.*, spread of opinions. A well studied model for opinion dynamics is the voter model which has the same key property with the LT model that a node decision is influenced by its neighbor’s decision, *i.e.*, a person changes his or her opinion by the opinions of his or her neighbors. The basic voter model is defined on an undirected network and allows to have only two opinions. Each node adopts the opinion of a randomly chosen neighbor at each subsequent discrete time-step. There has been a variety of work on the voter model. Dynamical properties of the basic model, including how the degree distribution and the network size affect the mean time to reach consensus, have been extensively studied (Liggett, 1999; Sood and Redner, 2005) from mathematical point of view. Several variants of the voter model are also investigated (Castellano et al., 2009; Yang et al., 2009) and non equilibrium phase transition is analyzed from physics point of view. Yet another line of work extends the voter model and combines it with a network evolution model (Holme and Newman, 2006; Crandall et al., 2008).

The work which is most influential to this work is by Even-Dar and Shapria (2007) who investigated the influence maximization problem at a given target time. They showed that the most natural heuristic solution, which picks the nodes in the network with the highest degree, is indeed the optimal solution. We extended the basic voter model to be able to handle multiple opinions and asynchronous time delay and, in doing so, introduced the value for each opinion to reflect the fact that people are affected by the importance of the opinion, *e.g.*, quality, brand, authority, etc. (Kimura et al., 2010b). We called this model “Value-weighted Voter Model with Multiple Opinions (VwVM)”, and addressed the problem of predicting the expected opinion share at a target time from the observed opinion formation data. We further addressed the problem of detecting the change of the opinion values from the observed data (Saito et al., 2011)

In this work, we introduce another factor which we call *strength* of each node. This is different from the *value* of opinion that was introduced in Kimura et al. (2010b). It is based on the observation that a person is influenced not only by what each opinion is about but also by who holds/says that opinion. Some persons are more influential than others, and we consider this degree of influence by the strength value that is associated with each node. Here, we must note that the *influence* meant here can better be named as direct influence and is different from what has been used in previous studies which can better be named as indirect influence. In information diffusion the *influence degree* of a node is defined as the expected number of active nodes at the end of the random process of the information diffusion that originated from the node (Kempe et al., 2003; Kimura et al., 2007). In particular, in opinion formation, it is defined as the expected number of nodes that hold the same opinion with the starting node at the end of the random process of the opinion formation (Even-Dar and Shapria, 2007). We distinguish the strength of a node, *i.e.*, direct influence, from the influence degree of the node, *i.e.*, indirect influence. The strength we define here is assumed to be intrinsic to each node and is determined independently of the result of information diffusion or opinion formation.

The problem we want to solve in this paper is to learn this strength from the observed opinion formation data and investigate how it affects the influence degree. In principle it is possible to learn the strength of all the nodes in the network from the observed data, given the generative model of opinion formation, by maximizing the likelihood of the observed data being generated. However, the number of nodes is huge and we need prohibitively large amount of observation data to avoid the overfitting problem. We rather assume and think it more natural that the strength is determined by the attributes of each node and its attribute dependency is more or less uniform across the nodes, and try to learn the parameters that define this attribute dependency from the data. We call this model “Attribute-weighted Voter Model with Multiple Opinions (AwVM)” in contrast to the model we previously defined, “Value-weighted Voter Model with Multiple Opinions (VwVM)”.

We derived a very efficient parameter updating algorithm to maximize the likelihood function that is guaranteed to converge, and tested the performance of the algorithm on four real world networks assuming the attribute dependency of the parameters to be of a particular form. The algorithm can correctly estimate the strength of each node by way of node attributes through a learned function. We further show that the influence degree of each node based on the AwVM is substantially different from a naive AwVM that assumes a uniform strength throughout the nodes for which the influence degree is known to be proportional to the node degree, and there appears to be no simple heuristic to approximate the influence degree with good accuracy unless the network is dense. The sensitivity analysis indicates that as the degree of non-uniformity of the strength becomes greater, the influence degree becomes progressively more sensitive to the node strength than the node degree, and even for a moderate value of the non-uniformity of node strength, it is more affected by the node strength than by the node degree.

2. Opinion Formation Models

Let $G = (V, E)$ be an undirected (bidirectional) network with self-loops, where V and $E (\subset V \times V)$ are the sets of all the nodes and links in the network, respectively. For a node $v \in V$, let $\Gamma(v)$ denote the set of neighbors of v in G , that is, $\Gamma(v) = \{u \in V; (u, v) \in E\}$. Note that $v \in \Gamma(v)$.

2.1. Basic Voter Model

According to the work of [Even-Dar and Shapria \(2007\)](#), we recall the definition of the basic voter model with two opinions on networks G . In the voter model, each node of G is endowed with two states; opinions 1 and 2. The opinions are initially assigned to all the nodes in G , and the evolution process unfolds in discrete time-steps $t = 1, 2, 3, \dots$ as follows: At each time-step t , each node v picks a random neighbor u and adopts the opinion that u holds at time-step $t - 1$.

More formally, let $f_t : V \rightarrow \{1, 2\}$ denote the opinion distribution at time-step t , where $f_t(v)$ stands for the opinion of node v at time-step t . Then, $f_0 : V \rightarrow \{1, 2\}$ is the initial opinion distribution, and $f_t : V \rightarrow \{1, 2\}$ is inductively defined as follows: For any $v \in V$,

$$\begin{cases} f_t(v) = 1, & \text{with probability } |n_1(t, v)|/|\Gamma(v)|, \\ f_t(v) = 2, & \text{with probability } |n_2(t, v)|/|\Gamma(v)|, \end{cases}$$

where $n_k(t, v) = \{u \in \Gamma(v); f_{t-1}(u) = k\}$ is the set of v 's neighbors that hold opinion k at time-step $t - 1$ for $k = 1, 2$.

Again, according to the work of [Even-Dar and Shapria \(2007\)](#), we define the expected influence degree of each node v , denoted by $\sigma_b(v)$. Consider an initial opinion distribution that $f_0(v) = 1$ and $f_0(w) = 2$ if $w \neq v$. Namely, only v 's opinion is 1 and that of all the other nodes is 2. Then $\sigma_b(v)$

is defined as the expected number of nodes that hold the opinion 1 (node v 's initial opinion) after enough time has passed. More formally, for a given network $G = (V, E)$, we identify each node with a unique integer from 1 to $|V|$. Then we can define the adjacency matrix $\mathbf{A} \in \{0, 1\}^{|V| \times |V|}$ by setting $a_{u,v} = 1$ if $(u, v) \in E$; otherwise $a_{u,v} = 0$. We also define the diagonal matrix \mathbf{D} , each element of which, $d_{v,v} = d(v)$, is the degree of node v , *i.e.*, $d(v) = |\Gamma(v)| = \sum_{u \in V} a_{u,v}$. Here note that $d(v) \geq 1$ due to self-loops. Let $\mathbf{h}_v \in \{0, 1\}^{|V|}$ be a vector whose v -th element is 1 and all the other elements are 0, and $\mathbf{h} \in \{1\}^{|V|}$ be a vector whose elements are all 1. Now, we can calculate the expected number of nodes that hold opinion 1 at time $t = 1$ starting from a node v by $\mathbf{h}_v^T \mathbf{A} \mathbf{D}^{-1} \mathbf{h}$. Thus, the vector \mathbf{b} of the expected influence degree, each element of which is $b_v = \sigma_b(v)$, is defined as a limiting solution of the following iterative process with the initial setting $\mathbf{b}^{(0)} = \mathbf{h}$.

$$\mathbf{b}^{(t)} = \mathbf{A} \mathbf{D}^{-1} \mathbf{b}^{(t-1)}. \quad (1)$$

Especially, in case of an undirected network, the analytical solution can be derived, *i.e.*, $b_v = \sigma_b(v) = |V| |\Gamma(v)| / \sum_{u \in V} |\Gamma(u)|$. The result clearly states that the influence degree b_v of a node v is proportional to its node degree $|\Gamma(v)|$.

2.2. Attribute-weighted Voter Model

We extend the basic voter model by allowing to hold K opinions ($K \geq 2$). Further as explained in Section 1, we introduce the strength for each node v , denoted by s_v ¹. Then, we can define the following probability of opinion adoption for the new voter model with the node strength.

$$P(f_t(v) = k) = \frac{\sum_{u \in n_k(t,v)} s_u}{\sum_{u \in \Gamma(v)} s_u}, \quad (k = 1, \dots, K), \quad (2)$$

where $n_k(t, v) = \{u \in \Gamma(v); f_{t-1}(u) = k\}$ is the set of v 's neighbors that hold opinion k at time-step $t - 1$ for $k = 1, 2, \dots, K$.

Similarly to the basic voter model, we can define the expected influence degree $\sigma_a(v)$ for the new voter model, which is the expected number of nodes that hold the opinion k after enough time has passed when only the node v has the opinion k and all the other nodes have different opinions at $t = 0$. To this end, we define the diagonal matrix \mathbf{W} , each diagonal element of which, $w_{v,v} = w(v)$, represents the total strength of v 's neighbors, $w(v) = \sum_{u \in \Gamma(v)} s_u$. According to the arguments of the basic voter model, the vector \mathbf{a} of the expected influence degree, each element of which is $a_v = \sigma_a(v)$, is defined as a limiting solution of the following iterative process with the initial setting $\mathbf{a}^{(0)} = \mathbf{h}$.

$$\mathbf{a}^{(t)} = \mathbf{S} \mathbf{W}^{-1} \mathbf{a}^{(t-1)}, \quad (3)$$

where \mathbf{S} is the strength matrix which is obtained by replacing $a_{u,v}$ of \mathbf{A} with $a_{u,v} s_u$. Note that unlike Eq. (1), no analytical solution is known for Eq. (3), and how the strength s_v affects the influence degree a_v is not clear. Thus, we solve it numerically by iteratively calculating Eq. (3) until the difference $\|\mathbf{a}^{(t)} - \mathbf{a}^{(t-1)}\|$ becomes a reasonably small value. The convergence is guaranteed by the Perron-Frobenius theorem because the network is connected. Further note that Eq. (3) is defined independently of the number of opinions K .

In general, we don't know the adequate value for the strength of each node apriori. As one possible approach, we consider estimating each of them from a set of observed opinion formation

1. One of the anonymous reviewers pointed that our formulation is similar to the policy gradient learning in the reinforcement learning, which we were not aware of, *i.e.*, Eq. (2) corresponds to a Boltzman distribution policy, Eq. (3) to the transition model of a Markov decision problem (MDP), and estimating the strength of a node to learning the value function of the MDP.

results. However, as each node has its own strength, the number of parameters to learn is so huge that we need prohibitively large amount of training data to learn them all individually. As described in Section 1, we note that it is more natural to think that the strength is determined by the attributes of each node (person), *e.g.*, occupation, physical appearance, income, social status, etc., and its attribute dependency is more or less uniform across the nodes. We, thus, propose an Attribute-weighted Voter Model with Multiple Opinions (AwVM) that explicitly considers the dependency of node strength on its attributes. We assume that each node can have multiple attributes. Let $x_{v,j}$ be a value that node v takes for the j -th attribute, and J the total number of the attributes. We denote the J -dimensional vector of the attribute values for each node v by \mathbf{x}_v . Then we propose to model the strength s_v of node v by the following formula²:

$$s_v = s(\mathbf{x}_v, \boldsymbol{\theta}) = \exp(\boldsymbol{\theta}^T \mathbf{x}_v), \quad (4)$$

where $\boldsymbol{\theta}^T = (\theta_1, \dots, \theta_J)$ is the J -dimensional parameter vector for the attributes to determine the strength value of each node.

So far, we assumed a discrete time step. However, the actual opinion formation takes place in an asynchronous way along the continuous time axis, and the time stamps of the observed data are not equally spaced. Thus, there is a need to extend the model to make the state changes asynchronous. In order to describe the asynchronous voter model, we need to extend the definition of $n_k(t, v)$, the set of v 's neighbors that hold opinion k at $t - 1$, to be the one with the latest opinion before time t , *i.e.*,

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

where $\varphi_t(u)$ is the latest opinion of u before time t .³ Then, the evolution process of the asynchronous voter model is defined as follows:

1. At time 0, each node v independently decides its update time t according to some probability distribution such as an exponential distribution with parameter 1.⁴ The successive update time is determined similarly at each update time t .
2. At update time t , the node v adopts a new opinion according to Eq. (2).
3. The process is repeated from the initial time $t = 0$ until the next update-time exceeds a given final-time T .

3. Learning Problem and Method

We consider the problem of estimating the parameters for the attributes from observed data \mathcal{D}_T in time-span $[0, T]$, where \mathcal{D}_T consists of a sequence of (v, t, k) such that node v updated its opinion to opinion k at time t for $0 \leq t \leq T$.⁵ By estimating parameters, we can identify the influential nodes, *i.e.*, those with large influence degree, as well as the relevant attributes for determining the strength of nodes.

We formulate our problem for estimating the parameter values of the AwVM from a given observed opinion formation data \mathcal{D}_T . Based on the evolution process of our model (see Eq. (2)), we

2. This is a simple smooth function with respect to $\boldsymbol{\theta}$ that guarantees $s_v > 0$.

3. Note that the opinion update takes place at time t and we need the distribution before the time t .

4. This assumes that the average delay time is 1.

5. The data come in sequence each time the update takes place, but in the formulation we treat them as a set for easiness of the mathematical treatment.

can obtain the log likelihood function,

$$\begin{aligned}\mathcal{L}(\mathcal{D}_T; \boldsymbol{\theta}) &= \log \left(\prod_{(v,t,k) \in \mathcal{D}_T} P(f_t(v) = k) \right) \\ &= \sum_{(v,t,k) \in \mathcal{D}_T} \left(\log \left(\sum_{u \in n_k(t,v)} \exp(\boldsymbol{\theta}^T \mathbf{x}_u) \right) - \log \left(\sum_{u \in \Gamma(v)} \exp(\boldsymbol{\theta}^T \mathbf{x}_u) \right) \right).\end{aligned}\quad (5)$$

Thus our estimation problem is formulated as a maximization problem of the objective function $\mathcal{L}(\mathcal{D}_T; \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

We derive an EM like iterative algorithm for obtaining the maximum likelihood estimators. Now, let $\bar{\boldsymbol{\theta}}$ be the current estimates of $\boldsymbol{\theta}$. Then, by considering the posterior probabilities,

$$q_{v,t,k,u}(\boldsymbol{\theta}) = \frac{\exp(\boldsymbol{\theta}^T \mathbf{x}_u)}{\sum_{w \in n_k(t,v)} \exp(\boldsymbol{\theta}^T \mathbf{x}_w)},$$

($v \in V, 0 \leq t \leq T, k = 1, \dots, K, u \in n_k(t,v)$), we can transform our objective function as follows:

$$\mathcal{L}(\mathcal{D}_T; \boldsymbol{\theta}) = Q(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}}) - \mathcal{H}(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}}), \quad (6)$$

where $Q(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}})$ is defined by

$$Q(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}}) = \sum_{(v,t,k) \in \mathcal{D}_T} \left(\left(\sum_{u \in n_k(t,v)} q_{v,t,k,u}(\bar{\boldsymbol{\theta}}) \boldsymbol{\theta}^T \mathbf{x}_u \right) - \log \left(\sum_{u \in \Gamma(v)} \exp(\boldsymbol{\theta}^T \mathbf{x}_u) \right) \right), \quad (7)$$

and $\mathcal{H}(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}})$ is defined by

$$\mathcal{H}(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}}) = \sum_{(v,t,k) \in \mathcal{D}_T} \sum_{u \in n_k(t,v)} q_{v,t,k,u}(\bar{\boldsymbol{\theta}}) \log q_{v,t,k,u}(\boldsymbol{\theta}). \quad (8)$$

It is self-evident that $\mathcal{H}(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}})$ is maximized when $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$, which is easily proved by noting the constraint $\sum_u q_{v,t,k,u}(\bar{\boldsymbol{\theta}}) = 1$. Then, it follows $\mathcal{L}(\mathcal{D}_T; \boldsymbol{\theta}) - \mathcal{L}(\mathcal{D}_T; \bar{\boldsymbol{\theta}}) \geq Q(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}}) - Q(\bar{\boldsymbol{\theta}}; \bar{\boldsymbol{\theta}})$. Thus, $\mathcal{L}(\mathcal{D}_T; \boldsymbol{\theta})$ always increases by repeatedly maximizing $Q(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\theta}$ and updating $\bar{\boldsymbol{\theta}}$. When $Q(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}})$ reaches a fixed point, *i.e.*, Q is the maximum at $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$ for $\bar{\boldsymbol{\theta}}$, it holds that $(\partial Q(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}}) / \partial \boldsymbol{\theta})_{\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}} = (\partial \mathcal{L}(\mathcal{D}_T; \boldsymbol{\theta}) / \partial \boldsymbol{\theta})_{\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}} = 0$, *i.e.*, when Q reaches a fixed point, $\mathcal{L}(\mathcal{D}_T; \boldsymbol{\theta})$ also reaches a maximum, not necessarily a global maximum.

In order to derive our maximization algorithm for $Q(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}})$, we further define the following probability that the node v adopts the opinion of node u at time t whatever opinion u has:

$$r_{v,t,u}(\boldsymbol{\theta}) = \frac{\exp(\boldsymbol{\theta}^T \mathbf{x}_u)}{\sum_{w \in \Gamma(v)} \exp(\boldsymbol{\theta}^T \mathbf{x}_w)}.$$

Then, we can obtain the gradient vector of $Q(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\theta}$ as follows:

$$\frac{\partial Q(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \sum_{(v,t,k) \in \mathcal{D}_T} \left(\left(\sum_{u \in n_k(t,v)} q_{v,t,k,u}(\bar{\boldsymbol{\theta}}) \mathbf{x}_u \right) - \left(\sum_{u \in \Gamma(v)} r_{v,t,u}(\boldsymbol{\theta}) \mathbf{x}_u \right) \right).$$

Similarly, we can obtain the Hessian matrix with respect to $\boldsymbol{\theta}$ as follows:

$$\frac{\partial^2 Q(\theta; \bar{\theta})}{\partial \theta \partial \theta^T} = - \sum_{(v,t,k) \in \mathcal{D}_T} \left(\left(\sum_{u \in \Gamma(v)} r_{v,t,u}(\theta) \mathbf{x}_u \mathbf{x}_u^T \right) - \left(\sum_{u \in \Gamma(v)} r_{v,t,u}(\theta) \mathbf{x}_u \right) \left(\sum_{u \in \Gamma(v)} r_{v,t,u}(\theta) \mathbf{x}_u \right)^T \right).$$

Thus we can obtain the following modification vector $\Delta \theta$ for updating θ :

$$\Delta \theta = - \left(\frac{\partial^2 Q(\theta; \bar{\theta})}{\partial \theta \partial \theta^T} \right)^{-1} \frac{\partial Q(\theta; \bar{\theta})}{\partial \theta}. \quad (9)$$

Here note that the following quadratic form of the Hessian matrix is non-positive for an arbitrary J -dimensional non-zero vector $\mathbf{z} = (z_1, \dots, z_J)$,

$$\mathbf{z}^T \frac{\partial^2 Q(\theta; \bar{\theta})}{\partial \theta \partial \theta^T} \mathbf{z} = - \sum_{(v,t,k) \in \mathcal{D}_T} \left(\left(\sum_{u \in \Gamma(v)} r_{v,t,u}(\theta) (\mathbf{x}_u^T \mathbf{z})^2 \right) - \left(\sum_{u \in \Gamma(v)} r_{v,t,u}(\theta) \mathbf{x}_u^T \mathbf{z} \right)^2 \right) \leq 0$$

The Hessian matrix of Q is non-positive definite, and thus, the optimal solution of Q is obtained by using the Newton method. We can regard our estimation method as a variant of the EM algorithm. Namely, calculating Q by Eq. (7) and updating θ by Eq. (9) correspond to Expectation and Maximization steps, respectively. Which one in the $|n_k(t, v)|$ active parents actually activated its child is not known when $|n_k(t, v)| \geq 2$, which corresponds to the existence of the latent variables although they are not explicit in our formulation. We want to emphasize here that each time iteration proceeds the value of the likelihood function never decreases and the iterative algorithm is guaranteed to converge due to the convexity of Q .

Below we summarize the algorithm of the proposed method.

1. Initialize parameter vector θ as $\theta_j = 0$ for $j = 1, \dots, J$.
2. Calculate the gradient vector at the current parameter vector θ .
3. If the gradient vector is sufficiently small, *i.e.*, $\|\partial Q(\theta; \bar{\theta}) / \partial \theta\| < \eta$, output the parameter vector $\bar{\theta}$, and then terminate. Otherwise, go to 4.
4. Update the parameter vector θ by Eq. (9), and return to 2.

Here η is a parameter for the termination condition. In our experiments, η is set to a sufficiently small number, *i.e.*, $\eta = 10^{-12}$.

Finally, we briefly discuss the computational complexity of the learning algorithm. Evidently the most computationally expensive part is the Hessian matrix, *i.e.*, for each $(v, t, k) \in \mathcal{D}_T$ and its corresponding neighbor $u \in \Gamma(v)$, the required computational complexity is the square of the parameter size J . The average number of neighbors $\Gamma(v)$ is $|E|/|V|$, and the expected number of elements in \mathcal{D}_T is $T|V|$ (this is true for the exponential distribution with parameter 1). Thus, let M be the number of iterations required to obtain the solution; then the computational complexity of the algorithm is given by $O(MT|E|J^2)$.

4. Experiments

First, we evaluated experimentally our learning algorithm using synthetic opinion formation data that were generated from four large real world networks for which we assumed a specific relation between the node attributes and the node strength (Eq. (4)). Second, we evaluated how the node strength affects the influence degree (the expected number of nodes that hold the same opinion with the starting node at the end of the random process of the opinion formation).

4.1. Network Datasets

We used four large real networks, which are all bidirectionally connected. The first one is a reader network of “Ameba”⁶ that is a Japanese blog service site. We crawled the reader lists of 117,374 blogs of the Ameba blog service site in June 2006, and collected a large connected network, which has 56,604 nodes and 1,071,080 directed links (the Ameblo network). The second one is a trackback network of Japanese blogs (Kimura et al., 2009) that has 12,047 nodes and 79,920 directed links (the blog network). The third one is a Coauthorship network (Palla et al., 2005) and has 12,357 nodes and 38,896 directed links (the coauthorship network). The last one is a network of people that was derived from the “list of people” within Japanese Wikipedia (Kimura et al., 2008), which has 9,481 nodes and 245,044 directed links (the Wikipedia network).

4.2. Experimental Setting

For each network we generated synthetic opinion formation data \mathcal{D}_T of time span $[0, T]$ in the following way: 1) artificially generate node attributes and determine their values in a random manner; 2) determine a parameter vector θ which is assumed to be true; and then 3) generate \mathcal{D}_T multiple times for the given K and T by running the true AwVM that uses θ , each of which starts from the state in which the initial opinion of each node is selected uniformly at random from the K opinions. The values of K and T are varied as needed. We generated a total of 10 attributes for every node in each network, each with a real value of $[-1, 1]$. The true parameter vector θ was determined so that the distribution of s_v becomes uniform, that is, the expected value of $\theta^T \mathbf{x}_v$, *i.e.*, $\langle \theta^T \mathbf{x}_v \rangle$ becomes 0. As one such instance, we chose the parameter vector $\theta = (2.0, -1.0, 1.0, -2.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0)$. Note that this is different from limiting the number of attributes to 4.⁷

We conducted two kinds of experiments to evaluate the proposed method, one to evaluate the performance of the learning algorithm, and the other to evaluate the effect of introducing the strength on the influence degree. The learning performance was evaluated in terms of the accuracy of the parameter values $\hat{\theta}$ that are estimated from \mathcal{D}_T by our learning algorithm with different values for T and K , and its computation time. The effect of the strength was evaluated by comparing the influence degree (Eq. (3)) and the node ranking (with respect to the influence degree) of the AwVM that uses the learned parameters with the naive AwVM in which uniform strength *i.e.*, $s_v = 1$, is used, which is equivalent to the basic voter model with K opinions. Further, the node ranking of AwVM is compared with the ranking based on the node degree and other heuristic. We denote the true influence degree of AwVM as $\sigma_a(v)$ (Section 2.2), the estimated influence degree of AwVM as $\hat{\sigma}_a(v)$ and the influence degree of the naive AwVM as $\sigma_b(v)$ (Section 2.1). We terminated the iterative calculation of Eq. (3) if either of the following conditions is satisfied: the number of iteration exceeds 1,000, or $\|\mathbf{a}^{(t)} - \mathbf{a}^{(t-1)}\| < 10^{-8}$.

4.3. Evaluation of Parameter Estimation

Figure 1 shows the mean absolute error which is defined as $\sum_{j=1}^J |\theta_j - \hat{\theta}_j|/J$. It is the average over 10 trials that were conducted on the 10 distinct opinion formation data \mathcal{D}_T , each generated independently for each combination of the corresponding T and K for each network. Figures 1(a) to 1(c) is to show how the time span $[0, T]$ affects the results, and Figs. 1(d) to 1(f) is to show how

6. <http://www.ameba.jp/>

7. The algorithm should be able to identify the useless attributes for which the parameter values are 0.

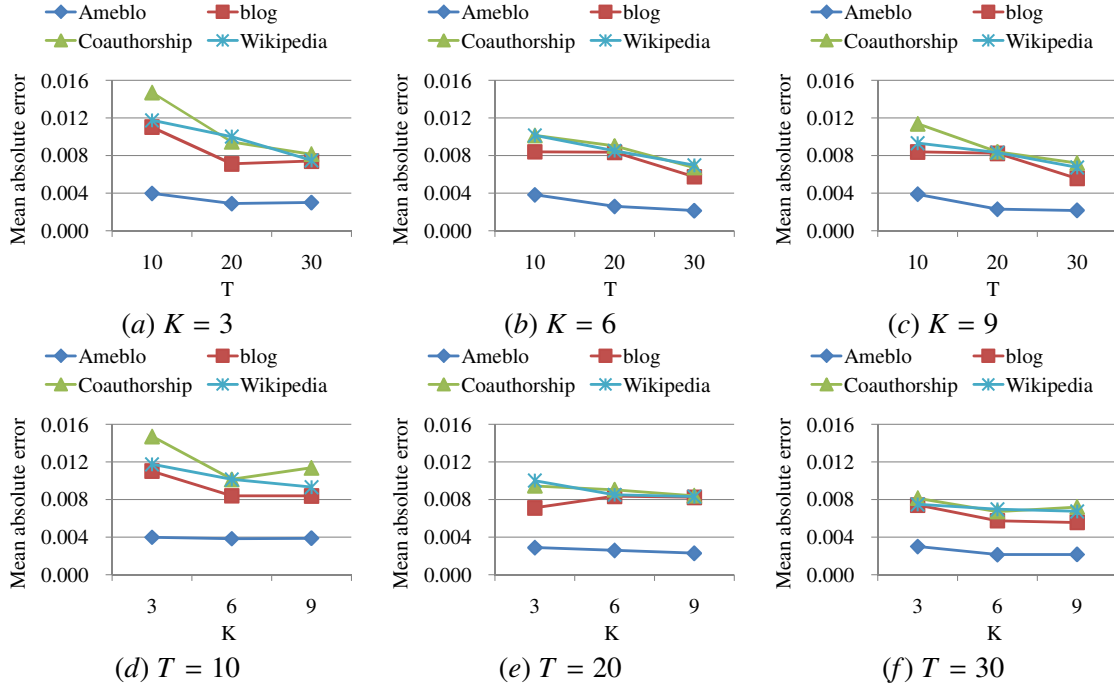


Figure 1: Mean absolute errors of the estimated parameter values for each network.

Table 1: Computation time (sec.) and the number of iterations of the proposed learning algorithm (in parentheses).

| | | $K = 3$ | $K = 6$ | $K = 9$ |
|--------------|----------|---------------|--------------|--------------|
| Ameblo | $T = 10$ | 36.49 (24.9) | 22.70 (16.0) | 19.49 (13.5) |
| | $T = 20$ | 71.58 (25.0) | 45.70 (16.2) | 40.70 (14.0) |
| | $T = 30$ | 103.51 (25.7) | 70.45 (17.0) | 62.85 (14.1) |
| blog | $T = 10$ | 2.24 (23.0) | 1.76 (17.0) | 1.58 (15.0) |
| | $T = 20$ | 4.55 (24.8) | 3.61 (18.5) | 3.36 (16.9) |
| | $T = 30$ | 6.77 (25.6) | 5.62 (19.9) | 5.26 (18.1) |
| Coauthorship | $T = 10$ | 1.06 (19.0) | 0.97 (15.4) | 0.90 (14.0) |
| | $T = 20$ | 2.10 (21.0) | 1.91 (17.0) | 1.83 (16.0) |
| | $T = 30$ | 2.97 (21.0) | 2.79 (17.5) | 2.71 (16.1) |
| Wikipedia | $T = 10$ | 8.07 (34.3) | 5.69 (23.9) | 4.95 (20.2) |
| | $T = 20$ | 17.14 (37.6) | 12.75 (27.1) | 11.08 (23.9) |
| | $T = 30$ | 27.84 (41.3) | 19.96 (29.2) | 18.55 (26.9) |

the number of opinions K affects the results. The mean absolute error is extremely small and in general decreases as both T and K increase. A larger T implies a larger training sample size of \mathcal{D}_T which naturally contributes to reducing the error. A larger K implies a larger chance of updating to a different opinion at each update time which contributes to increasing the diversity of the data, again contributing to reducing the error. Said differently, when K is smaller, the time to reach local

consensus gets earlier and the amount of data to be effective for learning gets smaller. The error for Ameblo network being the smallest is explained by the fact that this network has by far the largest number of nodes, indicating the largest expected number of samples which is $T|V|$ as explained in the last paragraph in Section 3.

Table 1 shows the computation time. The machine used is Intel(R) Xeon(R) CPU W5590 @3.33GHz with 32GB memory. The computation time roughly follows the results of computational complexity analysis. Under the situation where the number of iteration M is about the same for the same K , which is the case here, the computation time is proportional to the number of links for the same T . Ameblo network takes 4 times longer than Wikipedia network, Wikipedia network takes 3 times longer than Blog network and Blog network takes twice longer than Coauthorship network. The number of opinions K affects favorably the computation time. This is because the larger diversity of the data as explained above accelerates the convergence of iterative procedure.

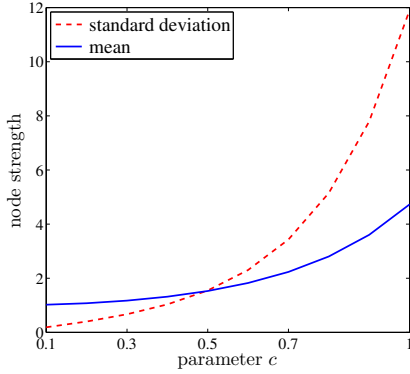
Overall, we can say that our learning method can accurately estimate the parameter values of the AwVM, and its performance with respect to the time span and the number of opinions are interpretable.

4.4. Comparative Study on Influence Degree

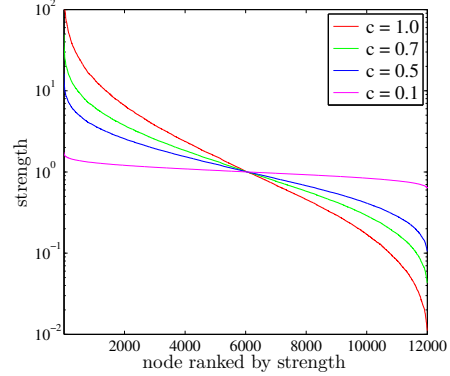
We evaluated the effect of the node strength in terms of the influence degree. To this end, we introduced a new parameter c to control the true parameter by $\theta_c = c\theta$. The parameter c is a measure to indicate the non-uniformity of the node strength, and is called the non-uniformity parameter. It affects the value of the strength directly, too, but as Eq.(2) shows, what matters is the relative strength. We use $c = 0.1, 0.5$, and 1.0 in the following experiments. Figure 2(a) illustrates how the mean of the node strength and the standard deviation (non-uniformity) change with c . Both increase exponentially as c increases. The distribution of the node strength for $c = 1.0$ is much more non-uniform than that for $c = 0.5$ as shown in Fig. 2(b). The strength is almost uniform across the nodes for $c = 0.1$, but some nodes have the strength which is 100 times as high as the average for $c = 1.0$. This can be expected from the form of Eq.(4). We think such a deviation for $c = 1.0$ is not rare even in our real human relations.

Now for each c , we define the mean absolute error of the AwVM by $\epsilon_a = \frac{1}{|V|} \sum_{v \in V} |\sigma_a(v) - \hat{\sigma}_a(v)|$, and that of the naive model by $\epsilon_b = \frac{1}{|V|} \sum_{v \in V} |\sigma_a(v) - \sigma_b(v)|$. Figure 3 shows the results of the mean absolute errors ϵ_a and ϵ_b . Here, we used the 10 distinct opinion formation data, each generated with $T = 30$ and $K = 3$ for each network, and $\hat{\theta}$ was estimated by our learning method for each trial. All of the results shown in Fig. 3 are the average over the 10 trials. The mean absolute error of the AwVM is reduced to 0.5 to 2% of the naive model for $c = 1.0$, and 10 to 20% even for $c = 0.1$ where the node strength has only a small effect (Fig. 3(a)). Fig. 3(b) shows the results of the standard deviation of the mean absolute errors. Both the mean and the standard deviation are of the same order, but the standard deviation is more sensitive to c , which can be expected by the increase in the standard deviation (non-uniformity) shown in Fig. 2(a).

To do a more detailed analysis, we illustrate in Fig. 4 the actual influence degree of each node in the blog network for one particular run, randomly chosen from the 10 independent trials. The nodes in the horizontal axis are ranked in descending order of the true influence degree σ_a . Also from this result, we can find that the difference between σ_a (black solid line) and $\hat{\sigma}_a$ (red cross “x”) is quite small for any value of c , while the difference between σ_a and σ_b (blue plus “+”) becomes

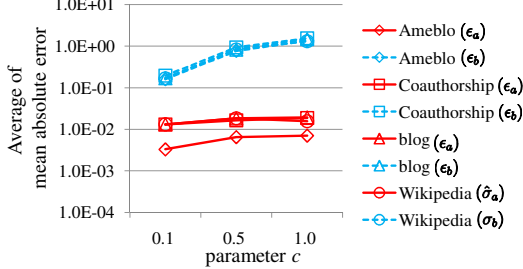


(a) Change of the mean and the standard deviation of node strength with the non-uniformity parameter c

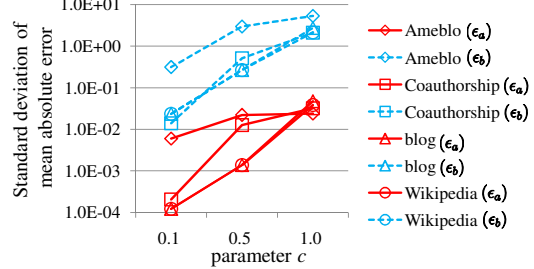


(b) The distribution of node strength with the non-uniformity parameter $c = 0.1, 0.5, 0.7, 1.0$.

Figure 2: The relation between the node strength and the non-uniformity parameter c .



(a) Average of the mean absolute errors of $\hat{\sigma}_a(v)$ and $\sigma_b(v)$



(b) Standard deviation of the mean absolute errors of $\hat{\sigma}_a(v)$ and $\sigma_b(v)$

Figure 3: Average and standard deviation of the mean absolute errors of $\hat{\sigma}_a(v)$ and $\sigma_b(v)$ for each network.

larger as c increases⁸. Especially, the difference between σ_a and σ_b is quite large for the top 100 nodes for $c = 0.5$ and $c = 1.0$. Note that the horizontal axis is logarithmic scale. This implies that for a majority of nodes the difference between $\sigma_a(v)$ and $\sigma_b(v)$ is very small. This explains why the mean absolute error is low as shown in Fig.3. We observed the same tendency for the other three networks, and here show only the results for the Coauthorship network in Fig. 5 for a reference. In summary, these results indicate that it is important to obtain the parameter values accurately in order to estimate the influence degree of each node in good accuracy.

In case of the basic model, it has been proven that the influence degree $\sigma_b(v)$ of a node is proportional to the degree of the node (Section 2.1). We explored to find a measure that correlates well to the influence degree in case of AwVM. Figure 6 shows the relation between the node degree and the influence degree (nodes ranked according to the degree) and Fig. 7 the relation between the node strength and the influence degree (nodes ranked according to the strength), both for the same trial for the blog network as in Fig. 4. The black solid line $\bar{d}(v)$ in Fig. 6 represents the value of $d(v)$

8. Note that the scale of the vertical axis of Fig. 4(c) is much larger than the other two figures.

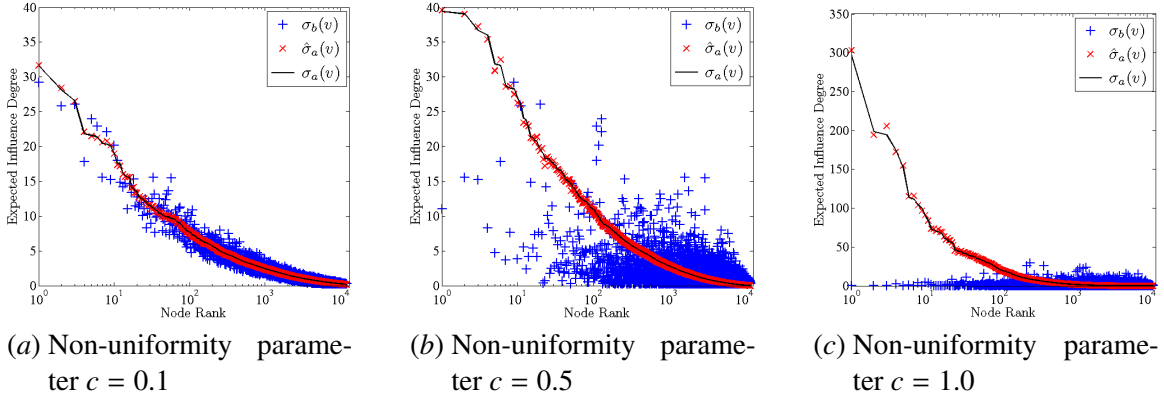


Figure 4: Comparison of influence degree, $\sigma_a(v)$, $\hat{\sigma}_a(v)$, and $\sigma_b(v)$ for the blog network for one particular run, randomly selected from the 10 independent trials.

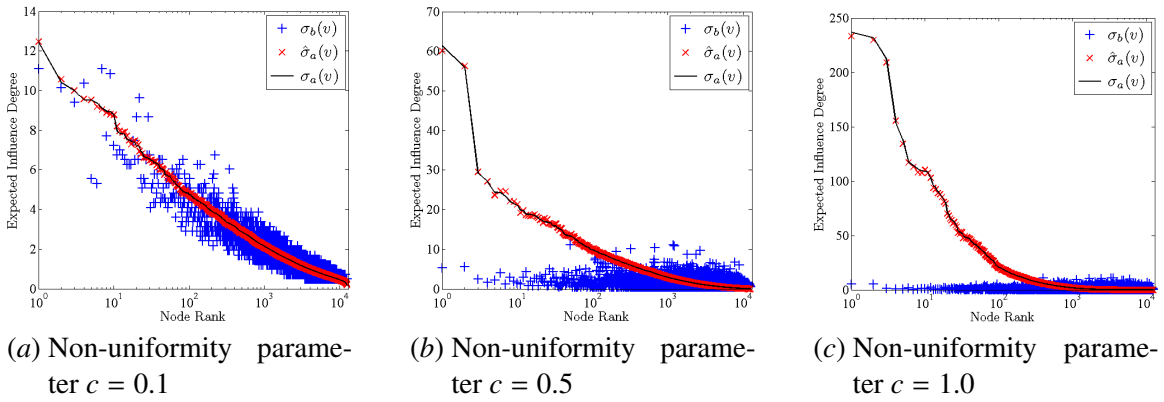


Figure 5: Comparison of influence degree, $\sigma_a(v)$, $\hat{\sigma}_a(v)$, and $\sigma_b(v)$ for the Coauthorship network for one particular run, randomly selected from the 10 independent trials.

that is normalized so that the total sum of $d(v)$ over all nodes becomes equal to the total sum of the influence degree, and the black solid line $\bar{s}(v)$ in Fig. 7 represents the value of s_v that is normalized in the same way. From Fig. 6, we can see that the node degree $d(v)$ can be a good estimator of $\hat{\sigma}_a(v)$ for $c = 0.1$, but it does not work well for a larger value of c . This is because the AwVM becomes closer to the basic voter model as the parameter c becomes closer to 0. Indeed, in Fig. 6, σ_b overlaps with the curve that represents $\bar{d}(v)$ because the naive model with $s_v = 1.0$, *i.e.*, $c = 0$, is identical to the basic voter model. Conversely, Fig. 7 shows that s_v can be a good estimator of $\hat{\sigma}_a(v)$ in case of $c = 1.0$, but it does not work well for a smaller c . Also in this case, the node degree $\bar{d}(v) = \sigma_b(v)$ can approximate $\hat{\sigma}_a(v)$ in good accuracy only for $c = 0.1$ ⁹.

These results suggest that although either the node strength s_v or the node degree $d(v)$ alone cannot be a good estimator of the influence degree of the AwVM, their combination could be a good estimator of the influence degree. To confirm this hypothesis, we further investigated the relation between their product, *i.e.*, $g(v) = s_v d(v)$, which is referred to as the *strength-weight degree*, and the

9. It does not look so at first glance, but note that the number of plots are the same for both $\sigma_b(v)$ and $\hat{\sigma}_a(v)$ and there are many overlaps.

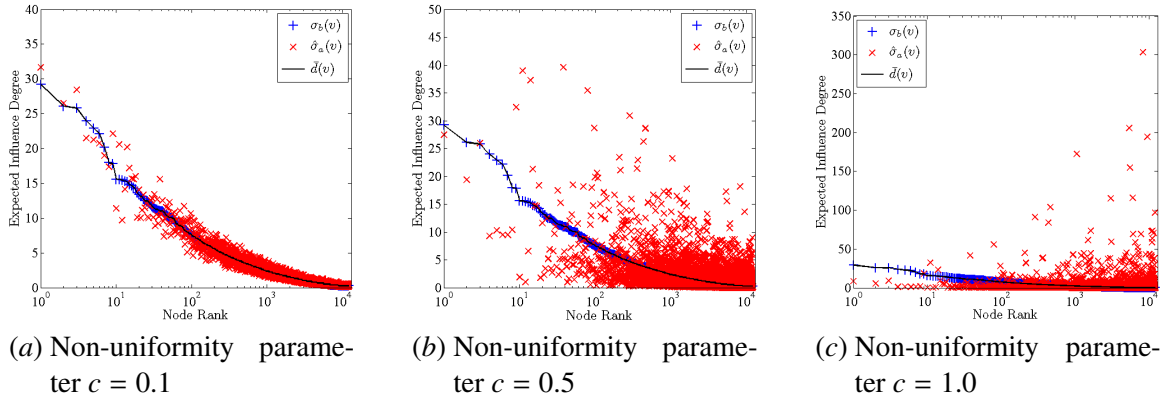


Figure 6: Relation between the node degree $d(v)$ and the influence degree, $(\hat{\sigma}_a(v)$ and $\sigma_b(v)$) for the blog network for one particular run, randomly selected from the 10 independent trials.

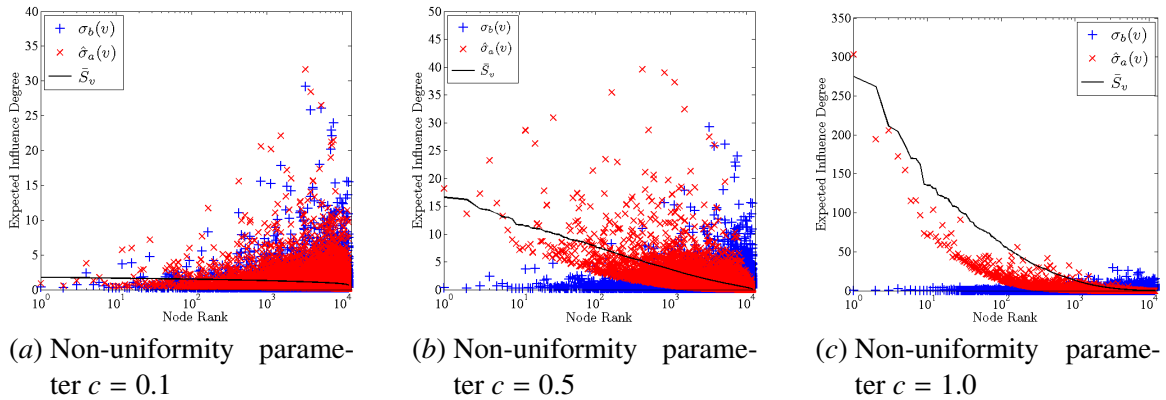


Figure 7: Relation between the node strength s_v and the influence degree $(\hat{\sigma}_a(v)$ and $\sigma_b(v)$) for the blog network for one particular run, randomly selected from the 10 independent trials.

influence degree for the AwVM. The result is shown in Fig. 8, where the black solid line $\bar{g}(v)$ represents the value of $g(v)$ that is normalized in the same way as $\bar{d}(v)$ in Fig. 6 (nodes ranked according to $g(v)$). Again, we used the results that are obtained from the same trial for the blog network as in Fig. 4. Unfortunately, this result refutes the aforementioned hypothesis. It can approximate the influence degree well only for $c = 0.1$ as $d(v)$ does, but not for a larger c . We examined the other three networks and found that the results strongly depend on the sparseness of the network. Wikipedia and Ameblo networks are much denser than Blog network. They showed better results for a larger c . We then randomly deleted links from these two networks and confirmed that the results become worse as more links are deleted. The reason behind this needs further investigation.

From these results, we can conclude that the opinion formation process for the AwVM becomes quite complex due to the introduction of the node strength, which is not the case for the basic voter model, and it seems that there is no simple heuristic measure that is able to estimate the influence degree in good accuracy for a wide range of networks. The *strength-weight degree* can be a good measure in some cases but not in all.

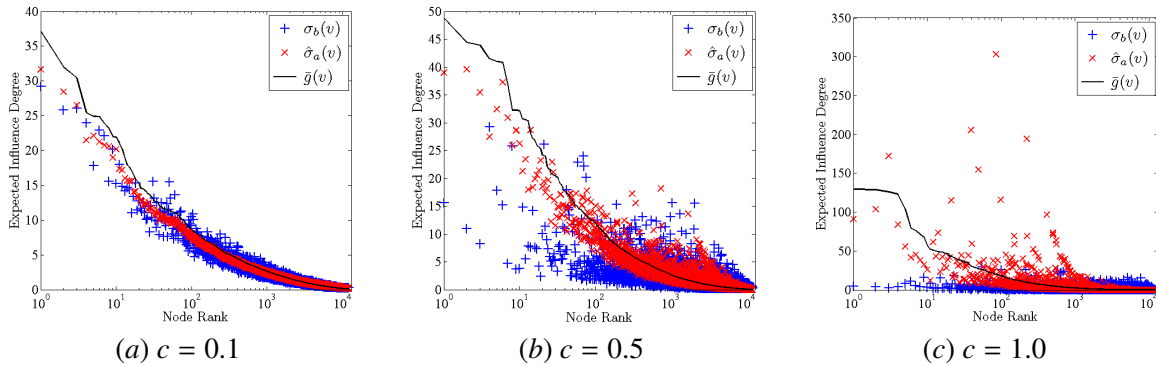


Figure 8: Relation between the product of the node strength and the node degree ($g(v) = s_v d(v)$) and the influence degree ($\hat{\sigma}_a$ and σ_b) for the blog network for one particular run, randomly selected from the 10 independent trials.

4.5. Discussion

We discuss the relation of the AwVM to the existing two models mentioned earlier. Based on the basic voter model, [Even-Dar and Shapria \(2007\)](#) investigated the influence maximization problem. They assumed that there need some initial costs to make each node accept an initial opinion. This cost is completely different from the node strength argued in this paper in that the initial cost of each node does not affect its influence degree at all. In contrast, the node strength directly affects the influence degree of each node, as shown in our experiments. Clearly, it is straightforward to incorporate initial costs to the AwVM. The theoretical analysis for the VwVM revealed that the opinion with the highest value wins and all the others die (winner-take-all process) for a situation where the local opinion share can be approximated by the average opinion share over the whole network, (e.g., the case of a complete network) ([Kimura et al., 2010b](#)). There does not exist an effective formula of calculating the influence degree for the VwVM, which corresponds to Eq. (3) for the AwVM, and obtaining the influence degree for the VwVM is computationally expensive. Clearly, node strength and opinion value are different concept, and it is possible to introduce the opinion values to the AwVM. However, how the node strength affects the winner-take-all process remains as an open question.

In this paper, we assumed that the network is static. Dynamic social network is easily handled by the current method by updating the neighbor nodes right before time t , incorporating the newly added/deleted nodes and updating the neighbors opinion distribution $n_k(t, v)$ accordingly. The same algorithm runs without any other changes.

5. Conclusion

Opinion formation over a social network was analyzed by modeling the cascade of interactions of neighboring nodes as probabilistic process of state changes. We modeled this process as a variant of the well known voter model with emphasis given on the strength of each node, called Attribute-weighted Voter Model with Multiple Opinions (AwVM). The strength reflects the degree of direct influence of the node, and we addressed the problem of estimating this strength from the observed opinion formation data. As each node has its own strength, the number of variables we want to

estimate is as large as the number of nodes in the network, which requires a prohibitively large amount of training data. We avoided this problem by assuming a functional dependency of the node strength on the small number of selected node attributes, which we believe to reflect the reality, and learn the parameter values that specify the functional dependency without a need for such a large amount of data. The task was formulated as the maximum likelihood estimation problem, and an efficient parameter value update algorithm that guarantees the convergence was derived. We evaluated the performance of the learning algorithm on four real world networks assuming a particular class of attribute dependency, and confirmed that the dependency can be correctly learned. We further showed that the influence degree of each node (expected number of the nodes at the end of opinion formation process that have the same opinion as the starting node considered) based on our AwVM is substantially different from that obtained assuming a model with uniform strength, *i.e.* without the strength, and the sensitivity analysis indicated that the influence degree is more sensitive to the node strength than the node degree even for a moderate value of the node strength. The results, at first glance, suggested to use the strength-weight degree as a rough measure of approximating influence degree, but quantitative evaluation refuted this hypothesis. Introduction of node strength, which is quite natural and reflects the reality, brings complexity to the opinion formation process, and there appears to be no simple heuristic measure that can predict the influence degree in good accuracy for a wide range of networks.

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. 23500194).

References

- C. Castellano, M. A. Munoz, and R. Pastor-Satorras. Nonlinear q -voter model. *Physical Review E*, 80:041129, 2009.
- D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri. Feedback effects between similarity and social influence in online communities. In *Proceedings of KDD 2008*, pages 160–168, 2008.
- P. Domingos. Mining social networks for viral marketing. *IEEE Intell. Syst.*, 20:80–82, 2005.
- E. Even-Dar and A. Shapria. A note on maximizing the spread of influence in social networks. In *Proceedings of WINE 2007*, pages 281–286, 2007.
- 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.
- P. Holme and M. E. J. Newman. Nonequilibrium phase transition in the coevolution of networks and opinions. *Physical Review E*, 74:056108, 2006.

- D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD-2003*, pages 137–146, 2003.
- M. Kimura, K. Saito, and R. Nakano. Extracting influential nodes for information diffusion on a social network. In *AAAI-07*, pages 1371–1376, 2007.
- M. Kimura, K. Saito, and H. Motoda. Minimizing the spread of contamination by blocking links in a network. In *AAAI-08*, pages 1175–1180, 2008.
- M. Kimura, K. Saito, and H. Motoda. Blocking links to minimize contamination spread in a social network. *ACM Trans. Knowl. Discov. 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 Min. and Knowl. Disc.*, 20:70–97, 2010a.
- M. Kimura, K. Saito, K. Ohara, and H. Motoda. Learning to predict opinion share in social networks. In *AAAI-10*, pages 1364–1370, 2010b.
- J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. In *EC'06*, pages 228–237, 2006.
- T. M. Liggett. *Stochastic interacting systems: contact, voter, and exclusion processes*. Springer, New York, 1999.
- M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, 45:167–256, 2003.
- M. E. J. Newman, S. Forrest, and J. Balthrop. Email networks and the spread of computer viruses. *Phys. Rev. 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. Detecting changes in opinion value distribution for voter model. In *Proceedings of International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction (SBP11)*, pages 89–96. Springer, LNAI 6589, 2011.
- V. Sood and S. Redner. Voter model on heterogeneous graphs. *Physical Review Letters*, 94:178701, 2005.
- H. Tong, B. A. Prakash, C. Tsoourakakis, T. Eliassi-Rad, C. Faloutsos, and D. H. Chau. On the vulnerability of large graphs. In *ICDM 2010*, pages 1091–1096, 2010.
- D. J. Watts. A simple model of global cascades on random networks. *PNAS*, 99:5766–5771, 2002.
- D. J. Watts and P. S. Dodds. Influence, networks, and public opinion formation. *J. Cons. Res.*, 34:441–458, 2007.
- H. Yang, Z.. Wu, C. Zhou, T. Zhou, and B. Wang. Effects of social diversity on the emergence of global consensus in opinion dynamics. *Physical Review E*, 80:046108, 2009.