Detecting Anti-majority Opinionists Using Value-weighted Mixture Voter Model

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

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

Abstract. We address the problem of detecting anti-majority opinionists using the value-weighted mixture voter (VwMV) model. This problem is motivated by the fact that some people have a tendency to disagree with any opinion expressed by the majority. We extend the value-weighted voter model to include this phenomenon with the anti-majoritarian tendency of each node as a new parameter, and learn this parameter as well as the value of each opinion from a sequence of observed opinion data over a social network. We experimentally show that it is possible to learn the anti-majoritarian tendency of each node correctly as well as the opinion values, whereas a naive approach which is based on a simple counting heuristic fails. We also show theoretically that, in a situation where the local opinion share can be approximated by the average opinion share, it is not necessarily the case that the opinion with the highest value prevails and wins when the opinion values are non-uniform, whereas the opinion share prediction problem becomes ill-defined and any opinion can win when the opinion values are uniform. The simulation results support that this holds for typical real world social networks.

1 Introduction

The emergence of large scale social computing applications has made massive social network data available as well as our daily life much depends on these networks through which news, ideas, opinions and rumors can spread [17, 16, 7, 5]. Thus, investigating the spread of influence in social networks has been the focus of attention [14, 4, 20]. The most well studied problem would be the *influence maximization problem*, that is, the problem of finding a limited number of influential nodes that are effective for spreading information. Many new algorithms that can effectively find approximate solutions have been proposed both for estimating the expected influence and for finding good candidate nodes [9, 11, 15, 2, 3]. However, the models used above allow a node in the network to take only one of the two states, *i.e.*, either active or inactive, because the focus is on *influence*.

Applications such as an on-line competitive service in which a user can choose one from multiple choices and decisions require a different approach where a model must handle multiple states. Also important is to consider the value of each choice, *e.g.*, quality, brand, authority, etc., because this affects others' choices. Opinion formation and its spread fit in the same class of problems. The model best suited for this kind of analysis would be a voter model [19, 8, 6, 4, 1, 21], which is one of the most basic stochastic process models and has the same key property with the *linear threshold model* used in information diffusion that a node decision is influenced by its neighbor's decision, *i.e.*, a person changes his/her opinion by the opinions of his/her neighbors. In [12], we extended the voter model to include opinion values, and addressed the problem of predicting the opinion share at a future time by learning the opinion values from a limited amount of past observed opinion diffusion data. Interestingly, theoretical analysis 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), revealed that the expected share prediction problem is well-defined only when the opinion values are non-uniform, in which case the final consensus is winner-take-all, *i.e.*, the opinion with the highest value wins and all the others die, and when they are uniform, any opinion can be a winner.

The problem we address in this paper tackles the same problem, but from a different angle. In the voter model including its variants, it is assumed that people naturally tend to follow their neighbors' majority opinion. However, we note that there are always people who do not agree with the majority and support the minority opinion. We are interested in how this affects the opinion share, and have extended the value-weighted voter model with multiple opinions to include this anti-majority effect with the *anti-majoritarian tendency* of each node as a new parameter. We are not the first to introduce the notion of anti-majority. There is a model called anti-voter model where only two opinions are considered. Each one chooses one of its neighbors randomly and decides to take the opposite opinion of the neighbor chosen. Röllin [18] analyzed the statistical property of the anti-voter model introducing the notion of exchangeable pair couplings. We have extended the simple anti-voter model to value-weighted anti-voter model with multiple opinions, and combined it linearly with the value-weighted voter model with multiple opinions. The model now has a new parameter at each node which is a measure for the the anti-majoritarian tendency (weight for the value-weighted anti-voter model) in addition to the original parameter (opinion value), and we call the combined model the *value-weighted mixture voter (VwMV) model*.

Both the parameters, anti-majoritarian tendency and opinion value, can be efficiently learned by an iterative algorithm (EM algorithm) that maximizes the likelihood of the model's generating the observed data. We tested the algorithm for three real world social networks with size ranging over 4,000 to 10,000 nodes and 40,000 to 250,000 links, and experimentally showed that the parameter value update algorithm correctly identifies the anti-majoritarian tendency of each node under various situations provided that there are enough data. The anti-majoritarian tendency estimated by using a heuristic that simply counts the number of opinion updates in which the chosen opinion is the same as the minority opinion turns out to be a very poor approximation. These results show that the model learned by the proposed algorithm can be used to predict the future opinion share and provides a way to analyze such problems as influence maximization or minimization for opinion diffusion under the presence of anti-majority opinionists. A similar analysis as in [12] revealed interesting results for the average behavior that the opinion share crucially depends on the anti-majoritarian tendency and that the opinion with the highest value does not necessarily prevail when the values are non-uniform, which is in contrast to the result of the value-weighted voter model, whereas the share prediction problem becomes ill-defined when the opinion values are uniform, *i.e.*, any opinion can win, which is the same as in the value-weighted voter model. The simulation results also support that this holds for typical real world social networks.

2 Opinion Dynamics Models

We define the VwMV model. 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 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 $v \in \Gamma(v)$. Given an integer K with $K \geq 2$, we consider the spread of K opinions (opinion $1, \cdots$, opinion K) on G, where each node holds exactly one of the K opinions at any time t (≥ 0). We assume that each node of G initially holds one of the K opinions with equal probability at time $t = 0$. Let $f_t : V \to \{1, \cdots, K\}$ denote the *opinion distribution* at time t, where $f_t(v)$ stands for the opinion of node v at time t. Note that f_0 stands for the initial opinion distribution. For any $v \in V$ and $k \in \{1, 2, \cdots, K\}$, let $n_k(t, v)$ be the number of v's neighbors that hold opinion k as 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).

2.1 Voter and Anti-voter Models

We revisit the voter model, which is one of the standard models of opinion dynamics, where K is usually set to 2. The evolution process of the 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 $r_v = 1$.[1] The successive update time is determined similarly at each update time t.
2. At update time t, the node v adopts the opinion of a randomly chosen neighbor u, *i.e.*,

$$f_t(v) = \varphi_t(u).$$

[1] This assumes that the average delay time is 1.

3. The process is repeated from the initial time $t = 0$ until the next update-time passes a given final-time T.

We note that in the voter model each individual tends to adopt the majority opinion among its neighbors. Thus, we can extend the original voter model with 2 opinions to a voter model with K opinions by replacing Step 2 with: At update time t, the node v selects one of the K opinions according to the probability distribution,

$$P(f_t(v) = k) = \frac{n_k(t, v)}{|\Gamma(v)|}, \quad (k = 1, \cdots, K). \tag{1}$$

The anti-voter model is defined in a similar way. In this model K is set to 2 and Step 2 is replaced with: At update time t, the node v adopts the opposite opinion of a randomly chosen neighbor u, *i.e.*,

$$f_t(v) = 3 - \varphi_t(u).$$

We note that each individual tends to adopt the minority opinion among its neighbors instead. The anti-voter model with K opinions is obtained by replacing Eq. (1) with

$$P(f_t(v) = k) = \frac{1}{K-1}\left(1 - \frac{n_k(t, v)}{|\Gamma(v)|}\right), \quad (k = 1, \cdots, K). \tag{2}$$

2.2 Value-weighted Mixture Voter Model

In order to investigate the competitive spread of K opinions, it is important to consider each opinion's value because this affects others' choices. In [12], we extended the voter model with K opinions to the *value-weighted voter model* by introducing the parameter (*opinion value* of opinion k) w_k (> 0). In this model, Eq. (1) was replaced with

$$P(f_t(v) = k) = p_k(t, v, \boldsymbol{w}), \quad (k = 1, \cdots, K),$$

where $\boldsymbol{w} = (w_1, \cdots, w_K)$ and

$$p_k(t, v, \boldsymbol{w}) = \frac{w_k\, n_k(t, v)}{\sum_{j=1}^{K} w_j\, n_j(t, v)}, \quad (k = 1, \cdots, K). \tag{3}$$

We can also extend the anti-voter model with K opinions to the *value-weighted anti-voter model* by replacing Eq. (2) with

$$P(f_t(v) = k) = \frac{1 - p_k(t, v, \boldsymbol{w})}{K-1}, \quad (k = 1, \cdots, K). \tag{4}$$

Further, we can define the *value-weighted mixture voter (VwMV) model* by replacing Eq. (4) with

$$P(f_t(v) = k) = (1 - \alpha_v)\, p_k(t, v, \boldsymbol{w}) + \alpha_v \frac{1 - p_k(t, v, \boldsymbol{w})}{K-1}, \quad (k = 1, \cdots, K), \tag{5}$$

where α_v is a parameter with $0 \le \alpha_v \le 1$. Note that each individual located at node v tends to behave like a majority opinionist if the value of α_v is small, and tends to behave like an anti-majority opinionist if the value of α_v is large. Therefore, we refer to α_v as the *anti-majoritarian tendency* of node v.

3 Learning Problem and Behavior Analysis

We consider the problem of identifying the VwMV model on network G 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 changed its opinion to opinion k at time t for $0 \leq t \leq T$. The identified model can be used to predict how much of the share each opinion will have at a future time T' $(> T)$, and to identify both high anti-majoritarian tendency nodes (*i.e.*, anti-majority opinionists) and low anti-majoritarian tendency nodes (*i.e.*, majority opinionists). Below, we theoretically investigate some basic properties of the VwMV model, and demonstrate that it is crucial to accurately estimate the values of the parameters, w_k, $(k = 1, \cdots, K)$ and α_v, $(v \in V)$.

For any opinion k, let $h_k(t)$ denote its *population* at time t, *i.e.*,

$$h_k(t) = |\{v \in V;\ \ f_t(v) = k\}|,$$

and let $g_k(t)$ denote its expected *share* at time t, *i.e.*,

$$g_k(t) = \left\langle \frac{h_k(t)}{\sum_{j=1}^{K} h_j(t)} \right\rangle.$$

We investigate the behavior of expected share $g_k(t)$ for a sufficiently large t. According to previous work in statistical physics (*e.g.*, [19]), we employ a mean field approach. We first consider a rate equation,

$$\frac{dg_k(t)}{dt} = (1 - g_k(t))\, P_k(t) - g_k(t)\, (1 - P_k(t)), \quad (k = 1, \cdots, K), \tag{6}$$

where $P_k(t)$ denotes the probability that a node adopts opinion k at time t. Note that in the right-hand side of Eq. (6), $g_k(t)$ is regarded as the probability of choosing a node holding opinion k at time t. Here, we assume that the average local opinion share $\langle n_k(t, v)/ \sum_{j=1}^{K} n_j(t, v)\rangle$ in the neighborhood of a node v can be approximated by the expected opinion share $g_k(t)$ of the whole network for each opinion k. Then, we obtain the following approximation from Eq. (5):

$$P_k(t) \approx (1 - \alpha)\, \tilde{p}_k(t, \boldsymbol{w}) + \alpha\, \frac{1 - \tilde{p}_k(t, \boldsymbol{w})}{K - 1}, \quad (k = 1, \cdots, K), \tag{7}$$

where α is the average value of anti-majoritarian tendency α_v, $(v \in V)$, and

$$\tilde{p}_k(t, \boldsymbol{w}) = \frac{w_k\, g_k(t)}{\sum_{j=1}^{K} w_j\, g_j(t)}, \quad (k = 1, \cdots, K). \tag{8}$$

Note that Eq. (7) is exactly satisfied when G is a complete network and the anti-majoritarian tendency is node independent, *i.e.*, $\alpha_v = \alpha$, $(\forall v \in V)$.

For the value-weighted voter model (*i.e.*, $\alpha = 0$), we theoretically showed the following results in [12]:

1. When the opinion values are uniform (*i.e.*, $w_1 = \cdots = w_K$), any opinion can become a winner, that is, if $g_1(0) = \cdots = g_K(0) = 1/K$, then $g_k(t) = 1/K$, $(t \geq 0)$ for each opinion k.
2. When the opinion values are non-uniform, the opinion k^* with highest opinion value is expected to finally prevail over the others, that is, $\lim_{t\to\infty} g_{k^*}(t) = 1$.

We extend these results to the VwMV model below.

Case of uniform opinion values: We suppose that $w_1 = \cdots = w_K$. Then, since $\sum_{k=1}^{K} g_k(t) = 1$, from Eq. (8), we obtain

$$\tilde{p}_k(t, \boldsymbol{w}) = g_k(t), \quad (k = 1, \cdots, K).$$

Thus, we can easily derive from Eqs. (6) and (7) that

$$\frac{dg_k(t)}{dt} = -\frac{\alpha}{1 - 1/K}\left(g_k(t) - \frac{1}{K}\right), \quad (k = 1, \cdots, K).$$

Hence, we have

$$\lim_{t \to \infty} g_k(t) = 1/K, \quad (k = 1, \cdots, K).$$

Case of non-uniform opinion values: We assume that the opinion values are non-uniform. We parameterize the non-uniformity by the ratio,

$$s_k = \frac{w_k}{\sum_{j=1}^{K} w_j / K}, \quad (k = 1, \cdots, K).$$

Let k^* be the opinion with the highest opinion value. Note that $s_{k^*} > 1$. We assume for simplicity that

$$w_k = w' \; (< w_{k^*}) \quad \text{if } k \neq k^*,$$

where w' is a positive constant. We also assume that

$$g_1(0) = \cdots = g_K(0) = 1/K.$$

We can see from the symmetry of the setting that $g_k(t) = g_\ell(t)$, $(t \geq 0)$ if $k, \ell \neq k^*$. This implies that opinion k^* is the winner at time t if and only if $g_{k^*}(t) > 1/K$. Here, suppose that there exists some time $t_0 > 0$ such that

$$g_{k^*}(t_0) = 1/K.$$

Then, from Eqs. (6) and (8), we obtain

$$\left.\frac{dg_{k^*}(t)}{dt}\right|_{t=t_0} = P_{k*}(t_0) - \frac{1}{K}, \qquad \tilde{p}_{k^*}(t_0, \boldsymbol{w}) = \frac{s_{k^*}}{K}.$$

Thus we have from Eq. (7) that

$$\left.\frac{dg_{k^*}(t)}{dt}\right|_{t=t_0} = \frac{s_{k^*} - 1}{K - 1}\left(1 - \frac{1}{K} - \alpha\right).$$

Therefore, we obtain the following results:

1. When $\alpha < 1 - 1/K$,

   $$g_{k^*}(t) > 1/K, \quad (t > 0),$$

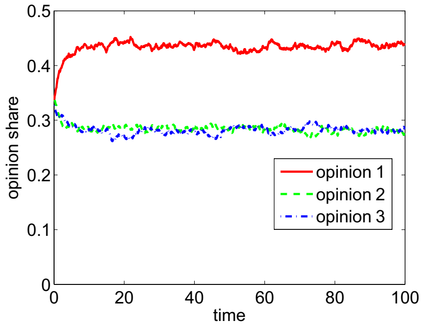
   that is, opinion k^* is expected to spread most widely and become the majority.
2. When $\alpha = 1 - 1/K$,

   $$g_k(t) = 1/K, \quad (t \geq 0),$$

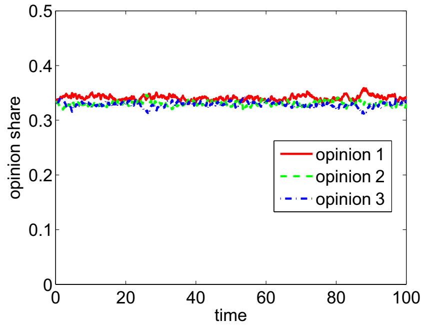
   for any opinion k, that is, any opinion can become a winner.
3. When $\alpha > 1 - 1/K$,

   $$g_{k^*}(t) < 1/K, \quad (t > 0),$$

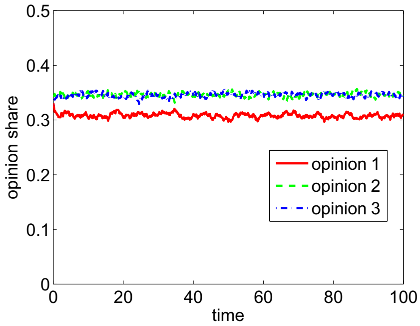
   that is, opinion k^* is expected to spread least widely and become the minority.



(a) $a = 2$, $b = 4$, ($\alpha < 1 - 1/3$)



(b) $a = 4$, $b = 2$, ($\alpha = 1 - 1/3$)



(c) $a = 18$, $b = 2$, ($\alpha > 1 - 1/3$)

Fig. 1: Results of the opinion share curves for different distributions of anti-majoritarian tendency in the Blog network.

Experiments: The above theoretical results are justified only when the approximation (see Eq. (7)) holds, which is always true in the case of complete networks. Real social networks are much more sparse and thus, we need to verify the extent to which the above results are true for real networks. We experimentally confirmed the above theoretical results for several real-world networks. Here, we present the experimental results for $K = 3$ in the Blog network (see Section 5), where the opinion values are $w_1 = 2$, $w_2 = w_3 = 1$, and anti-majoritarian tendency α_v, ($v \in V$) is drawn from the beta distribution with shape parameters a and b. Figure 1 shows the results of opinion share curves, $t \mapsto h_k(t) \big/ \sum_{j=1}^{K} h_j(t)$, ($k = 1, 2, 3$), when the distribution of anti-majoritarian tendency changes, where each node adopted one of three opinions with equal probability at time $t = 0$. Note that

$$\begin{aligned} &\alpha = 0.33\ (< 1 - 1/3), && \text{if } a = 2, b = 4, \\ &\alpha = 1 - 1/3, && \text{if } a = 4, b = 2, \\ &\alpha = 0.9\ (> 1 - 1/3), && \text{if } a = 18, b = 2. \end{aligned}$$

We obtained similar results to those in Figures 1a, 1b and 1c also for many other trials. These results support the validity of our theoretical analysis.

4 Learning Method

We describe a method for estimating parameter values of the VwMV model from given observed opinion spreading data $\mathcal{D}_T$. Based on the evolution process of our model (see Eq. (5)), we can obtain the likelihood function,

$$\mathcal{L}(\mathcal{D}_T; \boldsymbol{w}, \boldsymbol{\alpha}) = \log \left(\prod_{(v,t,k) \in \mathcal{D}_T} P(f_t(v) = k) \right), \tag{9}$$

where $\boldsymbol{w}$ stands for the K-dimensional vector of opinion values, *i.e.*, $\boldsymbol{w} = (w_1, \cdots, w_K)$, and $\boldsymbol{\alpha}$ is the $|V|$-dimensional vector with each element α_v being the anti-majoritarian tendency of node v. Thus our estimation problem is formulated as a maximization problem of the objective function $\mathcal{L}(\mathcal{D}_T; \boldsymbol{w}, \boldsymbol{\alpha})$ with respect to $\boldsymbol{w}$ and $\boldsymbol{\alpha}$. Note from Eqs. (3), (5) and (9) that $\mathcal{L}(\mathcal{D}_T; c\boldsymbol{w}, \boldsymbol{\alpha}) = c\mathcal{L}(\mathcal{D}_T; \boldsymbol{w}, \boldsymbol{\alpha})$ for any $c > 0$. Note also that each opinion value w_k is positive. Thus, we transform the parameter vector $\boldsymbol{w}$ by $\boldsymbol{w} = \boldsymbol{w}(\boldsymbol{z})$, where

$$\boldsymbol{w}(\boldsymbol{z}) = (e^{z_1}, \cdots, e^{z_{K-1}}, 1), \quad \left(\boldsymbol{z} = (z_1, \cdots, z_{K-1}) \in \mathbf{R}^{K-1} \right). \tag{10}$$

Namely, our problem is to estimate the values of $\boldsymbol{z}$ and $\boldsymbol{\alpha}$ that maximize $\mathcal{L}(\mathcal{D}_T; \boldsymbol{w}(\boldsymbol{z}), \boldsymbol{\alpha})$.

We derive an EM like iterative algorithm for obtaining the maximum likelihood estimators. To this purpose, we introduce the following parameters that depend on $\boldsymbol{\alpha}$: For any $v \in V$ and $k, j \in \{1, \cdots, K\}$,

$$\beta_{v,k,j}(\boldsymbol{\alpha}) = \begin{cases} 1 - \alpha_v & \text{if } \; j = k, \\ \alpha_v/(K-1) & \text{if } \; j \neq k. \end{cases} \tag{11}$$

Then, from the definition of $P(f_t(v) = k)$ (see Eq. (5)), by noting $1 - p_k(t, v, \boldsymbol{w}) = \sum_{j \neq k} p_j(t, v, \boldsymbol{w})$, we can express Eq. (9) as follows:

$$\mathcal{L}(\mathcal{D}_T; \boldsymbol{w}(\boldsymbol{z}), \boldsymbol{\alpha}) = \sum_{(v,t,k) \in \mathcal{D}_T} \log \left(\sum_{j=1}^{K} \beta_{v,k,j}(\boldsymbol{\alpha}) \; p_j(t, v, \boldsymbol{w}(\boldsymbol{z})) \right).$$

Now, let $\bar{\boldsymbol{z}}$ and $\bar{\boldsymbol{\alpha}}$ be the current estimates of $\boldsymbol{z}$ and $\boldsymbol{\alpha}$, respectively. Then, by considering the posterior probabilities,

$$q_{v,t,k,j}(\boldsymbol{z}, \boldsymbol{\alpha}) = \frac{\beta_{v,k,j}(\boldsymbol{\alpha}) \; p_j(t, v, \boldsymbol{w}(\boldsymbol{z}))}{\sum_{i=1}^{K} \beta_{v,k,i}(\boldsymbol{\alpha}) \; p_i(t, v, \boldsymbol{w}(\boldsymbol{z}))},$$

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

$$\mathcal{L}(\mathcal{D}_T; \boldsymbol{w}(\boldsymbol{z}), \boldsymbol{\alpha}) = \mathcal{Q}(\boldsymbol{z}, \boldsymbol{\alpha}; \bar{\boldsymbol{z}}, \bar{\boldsymbol{\alpha}}) - \mathcal{H}(\boldsymbol{z}, \boldsymbol{\alpha}; \bar{\boldsymbol{z}}, \bar{\boldsymbol{\alpha}}), \tag{12}$$

where $\mathcal{Q}(\boldsymbol{z}, \boldsymbol{\alpha}; \bar{\boldsymbol{z}}, \bar{\boldsymbol{\alpha}})$ is defined by

$$\mathcal{Q}(\boldsymbol{z}, \boldsymbol{\alpha}; \bar{\boldsymbol{z}}, \bar{\boldsymbol{\alpha}}) = \mathcal{Q}_1(\boldsymbol{z}; \bar{\boldsymbol{z}}, \bar{\boldsymbol{\alpha}}) + \mathcal{Q}_2(\boldsymbol{\alpha}; \bar{\boldsymbol{z}}, \bar{\boldsymbol{\alpha}}), \tag{13}$$

$$\mathcal{Q}_1(\boldsymbol{z}; \bar{\boldsymbol{z}}, \bar{\boldsymbol{\alpha}}) = \sum_{(v,t,k) \in \mathcal{D}_T} \sum_{j=1}^{K} q_{v,t,k,j}(\bar{\boldsymbol{z}}, \bar{\boldsymbol{\alpha}}) \; \log p_j(t, v, \boldsymbol{w}(\boldsymbol{z})), \tag{14}$$

$$\mathcal{Q}_2(\boldsymbol{\alpha}; \bar{\boldsymbol{z}}, \bar{\boldsymbol{\alpha}}) = \sum_{(v,t,k) \in \mathcal{D}_T} \sum_{j=1}^{K} q_{v,t,k,j}(\bar{\boldsymbol{z}}, \bar{\boldsymbol{\alpha}}) \; \log \beta_{v,k,j}(\boldsymbol{\alpha}), \tag{15}$$

and $\mathcal{H}(\boldsymbol{z}, \alpha; \bar{\boldsymbol{z}}, \bar{\alpha})$ is defined by

$$\mathcal{H}(\boldsymbol{z}, \alpha; \bar{\boldsymbol{z}}, \bar{\alpha}) = \sum_{(v,t,k)\in\mathcal{D}_T} \sum_{j=1}^{K} q_{v,t,k,j}(\bar{\boldsymbol{z}}, \bar{\alpha}) \log q_{v,t,k,j}(\boldsymbol{z}, \alpha).$$

Since $\mathcal{H}(\boldsymbol{z}, \alpha; \bar{\boldsymbol{z}}, \bar{\alpha})$ is maximized at $\boldsymbol{z} = \bar{\boldsymbol{z}}$ and $\alpha = \bar{\alpha}$, we can increase the value of $\mathcal{L}(\mathcal{D}_T; \boldsymbol{w}(\boldsymbol{z}), \alpha)$ by maximizing $\mathcal{Q}(\boldsymbol{z}, \alpha; \bar{\boldsymbol{z}}, \bar{\alpha})$ with respect to $\boldsymbol{z}$ and α (see Eq. (12)). From Eq. (13), we can maximize $\mathcal{Q}(\boldsymbol{z}, \alpha; \bar{\boldsymbol{z}}, \bar{\alpha})$ by independently maximizing $\mathcal{Q}_1(\boldsymbol{z}; \bar{\boldsymbol{z}}, \bar{\alpha})$ and $\mathcal{Q}_2(\alpha; \bar{\boldsymbol{z}}, \bar{\alpha})$ with respect to $\boldsymbol{z}$ and α, respectively.

First, we estimate the value of $\boldsymbol{z}$ that maximizes $\mathcal{Q}_1(\boldsymbol{z}; \bar{\boldsymbol{z}}, \bar{\alpha})$. Here, note from Eqs.(3) and (10) that for $j = 1, \cdots, K$ and $\lambda = 1, \cdots, K-1$,

$$\frac{\partial p_j(t, v, \boldsymbol{w}(\boldsymbol{z}))}{\partial z_\lambda} = \delta_{j,\lambda}\, p_j(t, v, \boldsymbol{w}(\boldsymbol{z})) - p_j(t, v, \boldsymbol{w}(\boldsymbol{z}))\, p_\lambda(t, v, \boldsymbol{w}(\boldsymbol{z})), \tag{16}$$

where $\delta_{j,\lambda}$ is Kronecker's delta. From Eqs. (14) and (16), we have

$$\frac{\partial Q_1(\boldsymbol{z}; \bar{\boldsymbol{z}}, \bar{\alpha})}{\partial z_\lambda} = \sum_{(v,t,k)\in\mathcal{D}_T} \sum_{j=1}^{K} q_{v,t,k,j}(\bar{\boldsymbol{z}}, \bar{\alpha}) \left(\delta_{j,\lambda} - p_\lambda(t, v, \boldsymbol{w}(\boldsymbol{z}))\right), \tag{17}$$

for $\lambda = 1, \cdots, K-1$. Moreover, from Eqs. (16) and (17), we have

$$\frac{\partial^2 Q_1(\boldsymbol{z}; \bar{\boldsymbol{z}}, \bar{\alpha})}{\partial z_\lambda\, \partial z_\mu} = \sum_{(v,t,k)\in\mathcal{D}_T} \sum_{j=1}^{K} q_{v,t,k,j}(\bar{\boldsymbol{z}}, \bar{\alpha}) \left(p_\lambda(t, v, \boldsymbol{w}(\boldsymbol{z}))\, p_\mu(t, v, \boldsymbol{w}(\boldsymbol{z})) - \delta_{\lambda,\mu}\, p_\lambda(t, v, \boldsymbol{w}(\boldsymbol{z}))\right),$$

for $\lambda, \mu = 1, \cdots, K-1$. Thus, the Hessian matrix $(\partial^2 Q_1(\boldsymbol{z}; \bar{\boldsymbol{z}}, \bar{\alpha})/\partial z_\lambda \partial z_\mu)$ is negative semi-definite since

$$\begin{aligned}
&\sum_{\lambda,\mu=1}^{K-1} \frac{\partial^2 Q_1(\boldsymbol{z}; \bar{\boldsymbol{z}}, \bar{\alpha})}{\partial z_\lambda\, \partial z_\mu}\, x_\lambda x_\mu \\
&= \sum_{(v,t,k)\in\mathcal{D}_T} \sum_{j=1}^{K} q_{v,t,k,j}(\bar{\boldsymbol{z}}, \bar{\alpha}) \left(\left(\sum_{\lambda=1}^{K-1} p_\lambda(t, v, \boldsymbol{w}(\boldsymbol{z}))\, x_\lambda\right)^2 - \sum_{\lambda=1}^{K-1} p_\lambda(t, v, \boldsymbol{w}(\boldsymbol{z}))\, {x_\lambda}^2 \right) \\
&= -\sum_{(v,t,k)\in\mathcal{D}_T} \sum_{j=1}^{K} q_{v,t,k,j}(\bar{\boldsymbol{z}}, \bar{\alpha}) \left(\sum_{\lambda=1}^{K-1} p_\lambda(t, v, \boldsymbol{w}(\boldsymbol{z})) \left(x_\lambda - \sum_{\mu=1}^{K-1} p_\mu(t, v, \boldsymbol{w}(\boldsymbol{z}))\, x_\mu \right)^2 \right. \\
&\qquad\qquad \left. + \left(1 - \sum_{\lambda=1}^{K-1} p_\lambda(t, v, \boldsymbol{w}(\boldsymbol{z}))\right) \left(\sum_{\mu=1}^{K-1} p_\mu(t, v, \boldsymbol{w}(\boldsymbol{z}))\, x_\mu\right)^2 \right) \\
&\leq 0,
\end{aligned}$$

for any $(x_1, \cdots, x_{K-1}) \in \mathbf{R}^{K-1}$. Hence, by solving the equations $\partial Q_1(\boldsymbol{z}; \bar{\boldsymbol{z}}, \bar{\alpha})/\partial z_\lambda = 0$, $(\lambda = 1, \cdots, K-1)$ (see Eq. (17)), we can find the value of $\boldsymbol{z}$ that maximizes $\mathcal{Q}_1(\boldsymbol{z}; \bar{\boldsymbol{z}}, \bar{\alpha})$. We employed a standard Newton Method in our experiments.

Next, we estimate the value of $\boldsymbol{\alpha}$ that maximizes $Q_2(\boldsymbol{\alpha}; \bar{\mathbf{z}}, \bar{\boldsymbol{\alpha}})$. From Eqs. (11) and (15), we have

$$Q_2(\boldsymbol{\alpha}; \bar{\mathbf{z}}, \bar{\boldsymbol{\alpha}}) = \sum_{(v,t,k) \in \mathcal{D}_T} \left(q_{v,t,k,k}(\bar{\mathbf{z}}, \bar{\boldsymbol{\alpha}}) \log(1 - \alpha_v) + (1 - q_{v,t,k,k}(\bar{\mathbf{z}}, \bar{\boldsymbol{\alpha}})) \log \left(\frac{\alpha_v}{K-1} \right) \right).$$

Note that $Q_2(\boldsymbol{\alpha}; \bar{\mathbf{z}}, \bar{\boldsymbol{\alpha}})$ is also a convex function of $\boldsymbol{\alpha}$. Therefore, we obtain the unique solution $\boldsymbol{\alpha}$ that maximizes $Q(\mathbf{z}, \boldsymbol{\alpha}; \bar{\mathbf{z}}, \bar{\boldsymbol{\alpha}})$ as follows:

$$\alpha_v = \frac{1}{|\mathcal{D}_T(v)|} \sum_{(t,k) \in \mathcal{D}_T(v)} (1 - q_{v,t,k,k}(\bar{\mathbf{z}}, \bar{\boldsymbol{\alpha}})),$$

for each $v \in V$, where $\mathcal{D}_T(v) = \{(t, k); (v, t, k) \in \mathcal{D}_T\}$.

5 Experimental Evaluation

Using large real networks, we experimentally investigate the performance of the proposed learning method. We show the results of the estimation error of anti-majoritarian tendency, and the accuracies of detecting nodes with high anti-majoritarian tendency (*i.e.*, anti-majority opinionists) and nodes with low anti-majoritarian tendency (*i.e.*, majority opinionists), respectively.

5.1 Experimental Settings

We used three datasets of large real networks, which are all bidirectional connected networks and exhibit many of the key features of social networks. The first one is a trackback network of Japanese blogs used by [10] and has $12,047$ nodes and $79,920$ directed links (the Blog network). The second one is a network derived from the Enron Email Dataset [13] 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 third one is a network of people derived from the “list of people” within Japanese Wikipedia, also used by [10] and has $9,481$ nodes and $245,044$ directed links (the Wikipedia network).

We drew the true anti-majoritarian tendency α_v^* of each node $v \in V$ from the beta distribution of $a = b = 2$, and set the true opinion values as follows:

$$w_k^* = 5 - \frac{4(k-1)}{K-1}, \quad (k = 1, \cdots, K).$$

Note that the average value of α_v is expected to be 0.5, *i.e.*, $\alpha = 0.5$, and

$$w_1^* = 5,\ w_2^* = 5 - \frac{4}{K-1},\ \cdots,\ w_K^* = 1.$$

For each of three networks, we selected the initial opinion of each node uniformly at random, and generated the opinion diffusion data $\mathcal{D}_T$ of time span $[0, T]$ based on the true VwMV model. Then, we investigated the problem of estimating the anti-majoritarian tendency from the observed data $\mathcal{D}_T$.

We measured the error in estimating the anti-majoritarian tendency by estimation error $\mathcal{E}$,

$$\mathcal{E} = \frac{1}{|V|} \sum_{v \in V} |\hat{\alpha}_v - \alpha_v^*|,$$

where each $\hat{\alpha}_v$ denotes the estimated anti-majoritarian tendency of node v. We also measured the accuracies of detecting the high and the low anti-majoritarian tendency nodes by F-measures $\mathcal{F}_A$ and $\mathcal{F}_N$, respectively. Here, $\mathcal{F}_A$ and $\mathcal{F}_N$ are defined as follows:

$$\mathcal{F}_A = \frac{2|\hat{A} \cap A^*|}{|\hat{A}| + |A^*|}, \qquad \mathcal{F}_N = \frac{2|\hat{N} \cap N^*|}{|\hat{N}| + |N^*|},$$

where A^* and $\hat{A}$ are the sets of the true and the estimated top 15% nodes of high anti-majoritarian tendency, respectively, and N^* and $\hat{N}$ are the sets of the true and the estimated top 15% nodes of low anti-majoritarian tendency, respectively.

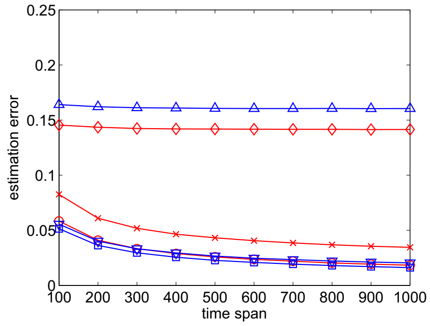
5.2 Comparison Methods

In order to investigate the importance of introducing the opinion values, we first compared the proposed method with the same VwMV model in which the opinion values are constrained to take a uniform value and the anti-majoritarian tendency of each node is the only parameter to be estimated. We refer to the method as the *uniform value method*. We also compared the proposed method with the naive approach in which the anti-majoritarian tendency of a node is estimated by simply counting the number of opinion updates in which the opinion chosen by the node is the minority opinion in its neighborhood. We refer to the method as the *naive method*.

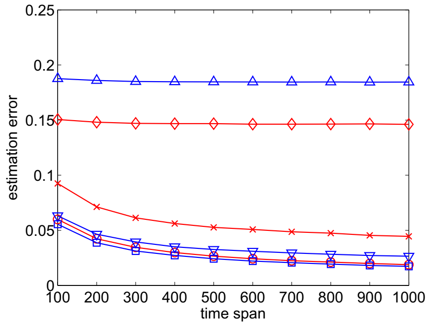
5.3 Experimental Results

We examined the results for both a small ($K = 3$) and a large ($K = 10$) K. Figures 2a, 2b and 2c show the estimation error $\mathcal{E}$ of each method as a function of time span T. Figures 3a, 3b and 3c show the F-measure $\mathcal{F}_A$ of each method as a function of time span T. Figures 4a, 4b and 4c show the F-measure $\mathcal{F}_N$ of each method as a function of time span T. Here, we repeated the same experiment five times independently, and plotted the average over the five results.

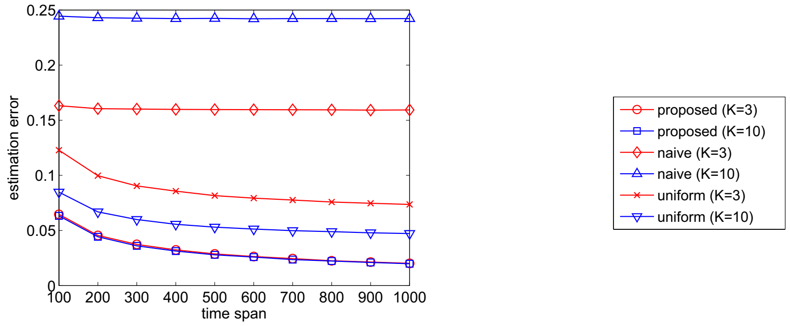
As expected, $\mathcal{E}$ decreases, and $\mathcal{F}_A$ and $\mathcal{F}_N$ increase as T increases (*i.e.*, the amount of training data $\mathcal{D}_T$ increases). We observe that the proposed method performs the best, the uniform value method follows, and the naive method behaves very poorly for all the networks. The proposed method can detect both the anti-majority and the majority opinionists with the accuracy greater than 90% at $T = 1000$ for all cases. We can also see that the proposed method is not sensitive to both K and the network structure, but the other two methods are so. For example, although the uniform value method of $K = 10$ performs well in $\mathcal{F}_A$ for the Blog and Enron networks, it does not so in $\mathcal{F}_A$ for the Wikipedia network, and in $\mathcal{F}_N$ for all the networks. Moreover, the uniform value method of $K = 3$ does not work well for all the cases. These results clearly demonstrate the advantage of the proposed method, and it does not seem feasible to detect even roughly



(a) Blog network



(b) Enron network



(c) Wikipedia network

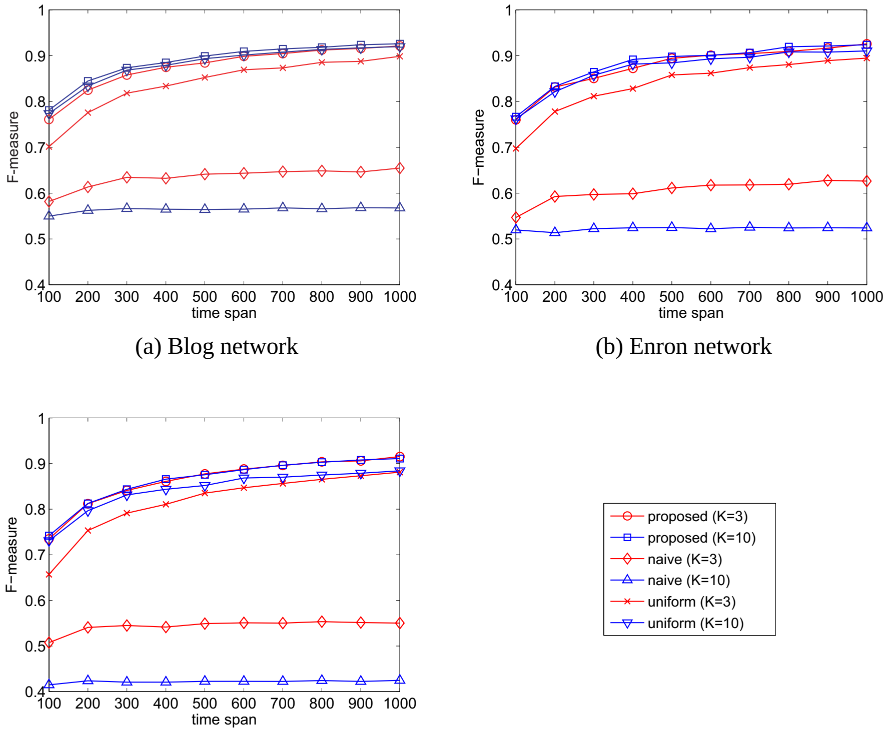
Fig. 2: Results for estimation errors of anti-majoritarian tendency.

the high anti-majoritarian tendency nodes and the low anti-majoritarian tendency nodes without using the explicit model and solving the optimization problem.

Here, we also note that the proposed method accurately estimated the opinion values. In fact, the average estimation errors of opinion value were less than 1% at $T = 1000$ for all cases. Moreover, we note that the processing times of the proposed method at $T = 1000$ for $K = 3$ and $K = 10$ were less than 3 min. and 4 min., respectively. All our experiments were undertaken on a single PC with an Intel Core 2 Duo 3GHz processor, with 2GB of memory, running under Linux.

6 Conclusion

We addressed the problem of how different opinions with different values spread over a social network under the presence of anti-majority opinionists by Value-weighted Mixture Voter Model which combines the value-weighted voter and the anti-voter models both with multiple opinions. The degree of anti-majority (anti-majoritarian tendency) is quantified by the weight of the two models, and is treated as a parameter. We formulated the model in the machine learning framework, and learned the anti-majoritarian



(a) Blog network

(b) Enron network

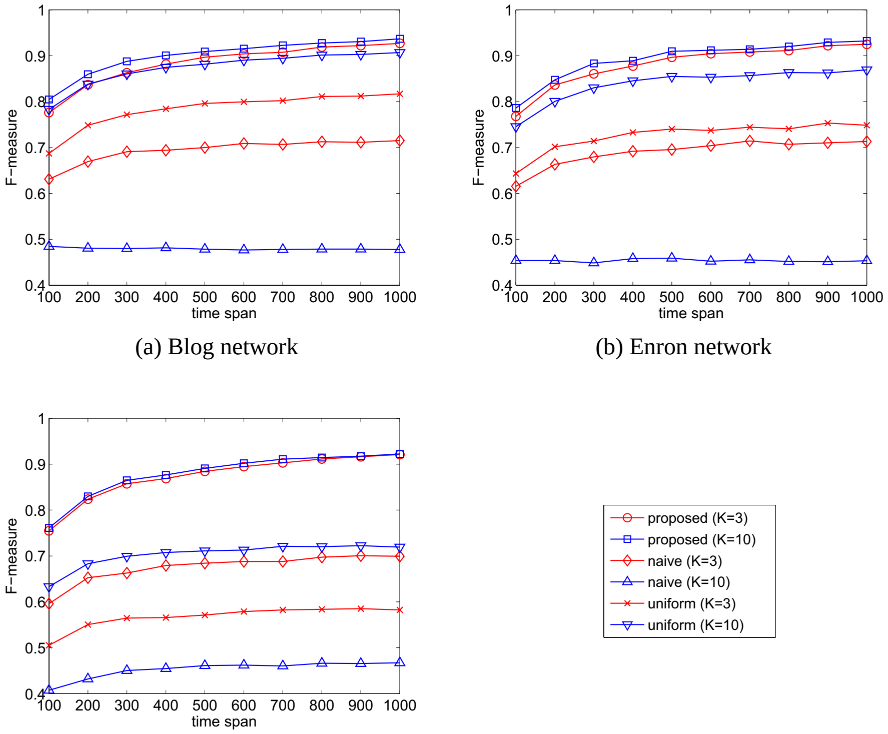
(c) Wikipedia network

Fig. 3: Results for accuracies of extracting nodes with high anti-majoritarian tendency.

tendency of each node and the value of each opinion from a sequence of observed opinion diffusion such that the likelihood of the model's generating the data is maximized.

The iterative parameter update algorithm is efficient and correctly identifies both the anti-majoritarian tendency and the opinion value if there are enough data. We confirmed this by applying the algorithm to three real world social networks (Blog, Enron and Wikipedia) under various situations. We compared the results with the naive approach in which the anti-majoritarian tendency is estimated by simply counting the number of opinion updates such that the chosen opinion is the same as the minority opinion. The naive approach behaves very poorly and our algorithm far outperformed it.

The opinion share crucially depends on the anti-majoritarian tendency and it is important to be able to accurately estimate it. The model learned by the proposed algorithm can be used to predict future opinion share and provides a useful tool to do various analyses. The theoretical analysis showed that in a situation where the local opinion share can be approximated by the average opinion share over the whole network, the opinion with the highest value does not necessarily prevails when the values are non-uniform, which is in contrast to the result of the value-weighted voter model (winner-take-all), whereas the opinion share prediction problem becomes ill-defined when the opinion



(a) Blog network

(b) Enron network

(c) Wikipedia network

Fig. 4: Results for accuracies of extracting nodes with low anti-majoritarian tendency.

values are uniform, *i.e.*, any opinion can win, which is the same as in the value-weighted voter model. The simulation results support that this holds for typical real world social networks. Our immediate future work is to apply the model to an interesting problem of influential maximization for opinion diffusion under the presence of anti-majority opinionists.

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

1. Castellano, C., Munoz, M.A., Pastor-Satorras, R.: Nonlinear *q*-voter model. Physical Review E 80, Article 041129 (2009)

2. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09). pp. 199–208 (2009)
3. Chen, W., Yuan, Y., Zhang, L.: Scalable influence maximization in social networks under the linear threshold model. In: Proceedings of the 10th IEEE International Conference on Data Mining (ICDM'10). pp. 88–97 (2010)
4. Crandall, D., Cosley, D., Huttenlocner, D., Kleinberg, J., Suri, S.: Feedback effects between similarity and social influence in online communities. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08). pp. 160–168 (2008)
5. Domingos, P.: Mining social networks for viral marketing. IEEE Intelligent Systems 20, 80–82 (2005)
6. Even-Dar, E., Shapira, A.: A note on maximizing the spread of influence in social networks. In: Proceedings of the 3rd International Workshop on Internet and Network Economics (WINE'07). pp. 281–286 (2007)
7. Gruhl, D., Guha, R., Liben-Nowell, D., Tomkins, A.: Information diffusion through blogspace. SIGKDD Explorations 6, 43–52 (2004)
8. Holme, P., Newman, M.E.J.: Nonequilibrium phase transition in the coevolution of networks and opinions. Physical Review E 74, Article 056108 (2006)
9. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03). pp. 137–146 (2003)
10. Kimura, M., Saito, K., Motoda, H.: Blocking links to minimize contamination spread in a social network. ACM Transactions on Knowledge Discovery from Data 3, Article 9 (2009)
11. Kimura, M., Saito, K., Nakano, R., Motoda, H.: Extracting influential nodes on a social network for information diffusion. Data Mining and Knowledge Discovery 20, 70–97 (2010)
12. Kimura, M., Saito, K., Ohara, K., Motoda, H.: Learning to predict opinion share in social networks. In: Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI'10). pp. 1364–1370 (2010)
13. Klimt, B., Yang, Y.: The enron corpus: A new dataset for email classification research. In: Proceedings of the 15th European Conference on Machine Learning (ECML'04). pp. 217–226 (2004)
14. Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. ACM Transactions on the Web 1, Article 5 (2007)
15. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N.: Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07). pp. 420–429 (2007)
16. Newman, M.E.J.: The structure and function of complex networks. SIAM Review 45, 167–256 (2003)
17. Newman, M.E.J., Forrest, S., Balthrop, J.: Email networks and the spread of computer viruses. Physical Review E 66, Article 035101 (2002)
18. Röllin, A.: Translated poisson approximation using exchangeable pair couplings. Annals of Applied Probablity 17, 1596–1614 (2007)
19. Sood, V., Redner, S.: Voter model on heterogeneous graphs. Physical Review Letters 94, Article 178701 (2005)
20. Wu, F., Huberman, B.A.: How public opinion forms. In: Proceedings of the 4th International Workshop on Internet and Network Economics (WINE'08). pp. 334–341 (2008)
21. Yang, H., Wu, Z., Zhou, C., Zhou, T., Wang, B.: Effects of social diversity on the emergence of global consensus in opinion dynamics. Physical Review E 80, Article 046108 (2009)