J Intell Inf Syst manuscript No.
(will be inserted by the editor)

Learning to predict opinion share and detect anti-majority opinionists in social networks

Masahiro Kimura · Kazumi Saito · Kouzou Ohara · Hiroshi Motoda

Received: date / Accepted: date

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 1) each opinion has its own value and an opinion with a higher value propagates more easily/rapidly and 2) there are always people who have a tendency to disagree with any opinion expressed by the majority. We extend the basic voter model to include these two factors with the value of each opinion and the anti-majoritarian tendency of each node as new parameters, and learn these parameters from a sequence of observed opinion data over a social network. We experimentally show that it is possible to learn the opinion values correctly using a short observed opinion propagation data and to predict the opinion share in the near future correctly even in the presence of anti-majoritarians, and also show that it is possible to learn the anti-majoritarian tendency of each node if longer observation data is available. Indeed, the learned model can predict the future opinion share much more accurately than a simple polynomial extrapolation can do. Ignoring these two factors substantially degrade

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).

Masahiro Kimura
Department of Electronics and Informatics, Ryukoku University
Otsu 520-2194, Japan
E-mail: kimura@rins.ryukoku.ac.jp

Kazumi Saito
School of Administration and Informatics, University of Shizuoka
Shizuoka 422-8526, Japan
E-mail: k-saito@u-shizuoka-ken.ac.jp

Kouzou Ohara
Department of Integrated Information Technology, Aoyama Gakuin University
Kanagawa 229-8558, Japan
E-mail: ohara@it.aoyama.ac.jp

Hiroshi Motoda
Institute of Scientific and Industrial Research, Osaka University
Osaka 567-0047, Japan
E-mail: motoda@ar.sanken.osaka-u.ac.jp

the performance of share prediction. We also show theoretically that, in a situation where the local opinion share can be approximated by the average opinion share, 1) when there are no anti-majoritarians, the opinion with the highest value eventually takes over, but 2) when there are a certain fraction of anti-majoritarians, it is not necessarily the case that the opinion with the highest value prevails and wins, and further, 3) in both cases, when the opinion values are uniform, the opinion share prediction problem becomes ill-defined and any opinion can win. The simulation results support that this holds for typical real world social networks. These theoretical results help understand the long term behavior of opinion propagation.

Keywords Social networks · Opinion dynamics · Parameter learning

1 Introduction

The emergence of large scale social computing applications has made massive social network data available, and large networks formed by these services play an important role as a medium for spreading diverse information including news, ideas, opinions, and rumors (Newman et al, 2002; Newman, 2003; Gruhl et al, 2004; Domingos, 2005). Thus, investigating the spread of influence in social networks has been the focus of attention (Leskovec et al, 2007a; Crandall et al, 2008; Wu and Huberman, 2008; Romero et al, 2011; Bakshy et al, 2011; Mathioudakis et al, 2011).

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 through the network. Many new algorithms that can effectively find approximate solutions have been proposed both for estimating the expected influence and for finding good candidate nodes under different model assumptions, *e.g.*, descriptive probabilistic interaction models (Domingos and Richardson, 2001; Richardson and Domingos, 2002), and basic diffusion models such as the *independent cascade (IC) model* and the *linear threshold (LT) model* (Kempe et al, 2003; Kimura et al, 2010a; Leskovec et al, 2007b; Chen et al, 2009, 2010). This problem has good applications in sociology and "viral marketing" (Agarwal and Liu, 2008). 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, however, require a different approach where a model must handle multiple states. The model best suited for this kind of analysis would be a voter model, which is the model to analyze how different opinions spread over a social network. It is one of the most basic stochastic process model, and has the same key property with the *linear threshold (LT) model* in that a node decision is influenced by its neighbor's decision, *i.e.*, a person changes its opinion by the opinions of its neighbors. In the basic voter model which is defined on an undirected network, each node initially holds one of the two opinions, *e.g.*, yes or no, and adopts the opinion of a randomly chosen neighbor at each subsequent discrete time-step.

In this paper, we address the problem of opinion formation by using an extended voter model for which multiple states are needed. There are three extensions. As described above, the original voter model can handle only two opinions and assumes discrete time-step. We extended the basic voter model to be able to handle K opinions and to allow asynchronous opinion update. This is just to make the basic voter model to be more realistic and this extension is straightforward. Indeed, the actual opinion update is asynchronous and if we are to use the observed data, synchronous discrete time-step model does not work. The other two extensions are more fundamental. We note that when we have to make a decision from multiple choices, we consider the value of each choice, *e.g.*, quality, brand, authority, etc. because this definitely affects our choice. The same is true for opinion formation. We listen to and evaluate what our neighbors say and change our opinions. Thus, the second extension is to incorporate the *value* of each opinion as a new parameter. The extended model is referred to as the *value-weighted voter (VwV) model with multiple opinions*. Same as the basic voter model, the VwV model assumes 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, which was also addressed in Gill and Gainous (2002) and Arenson (1996). We are interested in how this affects the opinion share. Thus, the third extension is to include this anti-majority effect by linearly combining the VwV model and the anti-majority model with the *anti-majoritarian tendency* of each node as a new parameter. The extended model is referred to as the *value-weighted mixture voter (VwMV) model*. We will discuss how to learn these parameters from the observed opinion propagation data and how accurately the learned model can predict the future opinion share.

There has been a variety of work on the voter model. Liggett (1999) and Sood and Redner (2005) extensively studied dynamical properties of the basic model, including how the degree distribution and the network size affect the mean time to reach consensus, from a mathematical point of view. Castellano et al (2009) and Yang et al (2009) investigated several variants of the voter model and analyzed non-equilibrium phase transition from a physics point of view. Holme and Newman (2006) and Crandall et al (2008) extended the voter model to combine it with a network evolution model. These studies gave insights into the fundamentals of the vote model, but their focuses are different from what this paper intends to address, *i.e.*, parameter learning from the data and share prediction at a specific time T with opinion values and anti-majoritarian tendency considered. Even-Dar and Shapira (2007), whose work we think has a similar goal to ours in spirit, investigated the influence maximization problem (maximizing the spread of the opinion that supports a new technology) at a given target time T under the basic *voter model*, *i.e.*, with two opinions (one in favor of the new technology and the other against it). They showed that the most natural heuristic solution, which picks the nodes in the network with the highest degree, is indeed the optimal solution, under the condition that all nodes have the same cost. This work is close to ours in that it measures the influence at a specific time T, but is different in all others (no share prediction, no value and anti-majoritarian tendency considered, no more than two opinions, no asynchronous update and no learning). We should mention that 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 (Huber and Reinert, 2004; Donnelly and Welsh, 1984; Matloff, 1977). Each one chooses one of its neighbors randomly and decides to take the opposite opinion of the neighbor chosen. Röllin (2007) analyzed the statistical property of the anti-voter model introducing the notion of exchangeable pair couplings. Our work is different from theirs, apart from the learning mechanism and being able to handle multiple opinions, in that we consider the effect of both the voter and the anti-voter models by introducing the *anti-majoritarian tendency* of each node as a new parameter.

This paper is an extension and integration of what we have reported in Kimura et al (2010b) and Kimura et al (2011). In the former we addressed the problem of predicting the opinion share at a future time (before an consensus is reached) by learning the opinion values from a limited amount of past observed opinion diffusion data using the VwV model. In the latter we introduced the VwMV model and mainly focused on the learning performance of the *anti-majoritarian tendency* of each node. In this paper we extend our preliminary work and analyze the share prediction performance of the VwMV model when both the opinion values and the anti-majoritarian tendency are not known and have to be learned from the observed opinion propagation data, investigate how the average anti-majoritarian tendency affects the learning performance, and detect who the anti-majoritarians are. In particular, we seek for the answer for the following questions: what the opinion share will be in the near future, given only the limited amount of observed data, how easy it is to learn both opinion value and anti-majority tendency, and how much the observed data is required to learn and identify the anti-majoritarians accurately enough. It is important to learn the model quickly and predict what will happen in the near future when a new opinion appears. The model is too simple to accurately predict the far future. For this, it is more desirable to understand the asymptotic behavior by a theoretical analysis.

We conjecture that learning opinion values is easy because the number of opinion K is not many (order of tens), but learning anti-majoritarian tendency is not easy because the tendency is associated with each node and the number of nodes is huge (order of ten thousands or more). We further conjecture that predicting the opinion share is much easier than identifying the anti-majoritarians because the former is a macroscopic quantity over the whole network but the latter is defined for each node. We show that both the parameters, anti-majoritarian tendency and opinion value, can be learned by an iterative algorithm that maximizes the likelihood of the model's generating the observed data, and confirmed the above conjectures by experiments. We tested the algorithm for four real world social networks with size ranging over 4,000 to 12,000 nodes and 40,000 to 250,000 links, and showed that the parameter value update algorithm correctly identifies both the values of opinions and the anti-majoritarian tendency of each node under various situations. The opinion values can be learned in good accuracy with a small amount of data, but the anti-majoritarian tendency needs a sufficiently large amount of data to improve the accuracy. Use of the learned model can predict the opinion share in the near future very accurately despite the existence of anti-majoritarians. The theoretical analysis under the assumption in which the local opinion share can be approximated by the average opinion share shows that 1) when there are no anti-majoritarians, the opinion with the highest value eventually takes over, but 2) when there is a certain fraction of anti-majoritarians, it is not necessarily the case that the opinion with the highest value prevails and wins, and further, 3) in both cases, when the opinion values are uniform, the opinion share prediction problem becomes ill-defined and any opinion can win, and these are also supported by real world networks in which the above assumption does not hold. We want to emphasize that it is crucially important to explicitly model the anti-majority effect to obtain good results. Predicting the share by VwV model when there are anti-majoritarians does not work. There seems to be no simple way to estimate the anti-majoritarian tendency. The heuristic that simply counts the number of opinion updates in which the chosen opinion is the same as the minority opinion gives only 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-majoritarians.

The paper is organized as follows. We introduce the basic voter and anti-voter models in Section 2 and our proposed models, VwV and VwMV models in Section 3. We then perform the behavior analysis for share prediction using the mean field theory and discuss the behavior qualitatively in Section 4, and describe the parameter learning algorithm in Section 5. We detail the results of experimental evaluations in Section 6. We summarize what has been achieved and conclude the paper in Section 7.

2 Voter Models

We consider the diffusion of opinions in a social network represented by an undirected (bidirectional) graph $G = (V, E)$ 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, *i.e.*,

$$\Gamma(v) = \{u \in V;\ (u, v) \in E\}.$$

Note that $v \in \Gamma(v)$. We revisit the basic voter model that is one of the standard models of opinion dynamics, and the anti-voter model that is its variant, where the number of opinions is set to two.

2.1 Basic Voter Model

According to the work of Even-Dar and Shapira (2007), we recall the definition of the basic voter model on network G. In the 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 $t = 1, 2, 3, \cdots$ 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 $f_t : V \to \{1, 2\}$ denote the *opinion distribution* at time-step t, where $f_t(v)$ stands for the opinion of node v at time-step t. Then, $f_0 : V \to \{1, 2\}$ is the initial opinion distribution, and $f_t : V \to \{1, 2\}$ is inductively defined as follows: For any $v \in V$, node v selects its opinion according to the probability distribution,

$$P(f_t(v) = 1) = \frac{N_1(t-1, v)}{|\Gamma(v)|}$$

$$P(f_t(v) = 2) = \frac{N_2(t-1, v)}{|\Gamma(v)|}$$

where $N_k(t, v)$ is the number of v's neighbors that hold opinion k at time-step t for $k = 1, 2$.

2.2 Anti-voter Model

In the basic voter model, it is assumed that people tend to follow their neighbors' majority opinion. However, since it is a common phenomenon that there are always people who do not agree with the majority and support the minority opinion, the *anti-voter model* is defined and investigated (Huber and Reinert, 2004; Röllin, 2007; Donnelly and Welsh, 1984; Matloff, 1977). In the anti-voter model, the opinion evolution process is replaced as follows: At each time-step t, each node v picks a random neighbor u and changes its opinion to the opposite of the opinion that u holds at time-step $t - 1$, *i.e.*, node v selects its opinion according to the probability distribution,

$$P(f_t(v) = 1) = \frac{N_2(t-1, v)}{|\Gamma(v)|}$$

$$P(f_t(v) = 2) = \frac{N_1(t-1, v)}{|\Gamma(v)|}$$

We note that each individual tends to adopt the minority opinion among its neighbors instead.

3 Proposed Model

3.1 Value-weighted Voter Model

We extend the basic voter model to the *value-weighted voter (VwV) model* for our purpose. In the VwV model, the total number of opinions is set to K (≥ 2), and each node of G is endowed with $(K+1)$ states; opinions $1, \cdots, K$, and *neutral* (*i.e.*, no-opinion state). We consider that a node is *active* when it holds an opinion k, and a node is *inactive* when it does not have any opinion (*i.e.*, when its state is neutral). We assume that nodes never switch their states from active to inactive. In order to discuss the competitive diffusion of K opinions, we introduce the parameter w_k (> 0) for each opinion k, which is referred to as the *opinion value* of opinion k. In the same way as the basic voter model, let $f_t : V \to \{0, 1, 2, \cdots, K\}$ denote the opinion distribution at time t, where opinion 0 denotes the neutral state. Here, f_t is defined for any non-negative real number t since the VwV model incorporates time delay in an asynchronous way, *i.e.*, t is continuous. For any $t > 0$, let $\varphi_t(v)$ denote the latest opinion of node v (before time t), and let $n_k(t, v)$ denote the number of v's neighbors that hold opinion k as the latest opinion (before time t), *i.e.*,

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

We define the evolution process of the VwV model. At the initial time $t = 0$, each opinion is assigned to only one node and all other nodes are in the neutral state. [1] Given a target time T, the evolution process unfolds in the following steps:

1. Each node v independently decides the next update time t' at its update time t according to some probability distribution such as an exponential distribution with parameter $\eta_v = 1$, [2] where the first update time is $t = 0$ for every node.
2. At update time t, the node v selects its opinion according to the probability distribution,

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

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

$$p_k(t, v, \boldsymbol{w}) = \frac{w_k\, n_k(t, v)}{\sum_{j=1}^{K} w_j\, n_j(t, v)}, \quad (k = 1, \cdots, K). \tag{2}$$

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

Note that the basic voter model with K opinions is derived from the VwV model with uniform opinion values $w_1 = \cdots = w_K$.

[1] This may look a rather unnatural assumption because it is unlikely that all the different opinions are initiated at the same time. Since each opinion is initiated by a single person and the goal is to see how it is propagated, it should be allowed that each opinion is assigned to only one node and all the remaining nodes are in neutral states, i.e. unaffected by any opinion yet. We could have changed the timing of each opinion's initial utterance, but chose the simplest case.

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

3.2 Value-weighted Mixture Voter Model

Since the anti-voter model aims to represent the phenomenon that people tend to follow their neighbors' minority opinion, the anti-voter model with K opinions can be defined by replacing Eq. (1) of the VwV model with

$$P(f_t(v) = k) \ = \ \frac{1}{K-1}\left(1 - \frac{n_k(t,v)}{\sum_{j=1}^{K} n_j(t,v)}\right), \quad (k = 1, \cdots, K).$$

Therefore, we can also extend the anti-voter model with K opinions to the *value-weighted anti-voter model* by replacing Eq. (1) with

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

For our purpose, we extend the VwV model and define the *value-weighted mixture voter (VwMV) model* by replacing Eq. (1) with

$$P(f_t(v) = k) \ = \ (1-\alpha_v)\, p_k(t,v,\boldsymbol{w}) + \alpha_v \, \frac{1 - p_k(t,v,\boldsymbol{w})}{K-1}, \quad (k = 1, \cdots, K), \tag{3}$$

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 majoritarian if the value of α_v is small, and tends to behave like an anti-majoritarian if the value of α_v is large. Therefore, we refer to α_v as the *anti-majoritarian tendency* of node v.

4 Behavior Analysis

In what follows, we first mathematically define the share prediction problem in Subsection 4.1 and explain why it is important to use a model to predict the future. Then, in Subsection 4.2 we introduce the mean field approach which is a method used in statistical physics to analyze the average behavior of a complex dynamic system. We first apply this theory to analyze the VwV model in Subsection 4.3 and discuss its asymptotic behavior and the time needed to reach consensus. We then apply this theory to analyze the VwMV model in Subsection 4.4 and discuss its asymptotic behavior in a similar way. These theoretical analysis sheds a light on the opinion formation dynamics and makes the behavior easy to understand.

4.1 Share Prediction Problem

Based on our opinion dynamics model (the VwMV model), we investigate the problem of predicting how large a share each opinion will have at a future target time T when the opinion diffusion is observed from $t = 0$ to $t = T_0$ ($< T$). Let $\mathcal{D}_{T_0}$ be the observed opinion diffusion data in time-interval $[0, T_0]$, where $\mathcal{D}_{T_0}$ consists of a sequence of (v, t, k) such that node v changed its opinion to opinion k at time t for $0 \le t \le T_0$. For any opinion k, let $h_k(t)$ denote its *population* at time t, *i.e.*,

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

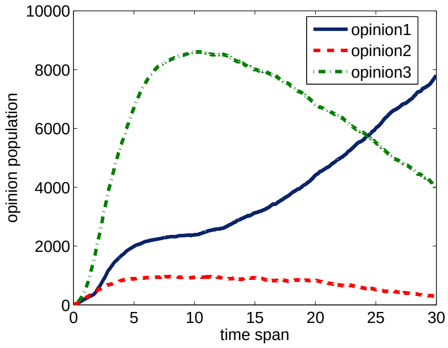


Fig. 1: An example of opinion population curves in the Blog network for $K = 3$.

Figure 1 shows an example of opinion population curves $h_1(t)$, $h_2(t)$, $h_3(t)$ for $K = 3$ in the Blog network (see Section 6 below), where the opinion values are set to $w_1 = 1.5$, $w_2 = 1.0$, $w_3 = 1.1$ and anti-majoritarian tendency α_v $(v \in V)$ is drawn from the beta distribution with shape parameters $a = 1$ and $b = 99$. Here, if we set $T_0 = 10$ and $T = 30$, we are able to observe $\mathcal{D}_{10}$ and thus $\{h_k(t);\ 0 \leq t \leq 10\}$ for $k = 1, 2, 3$ and the problem is to predict $h_1(30)$, $h_2(30)$, $h_3(30)$. Note that although the opinion dynamics is stochastic, we found that the variance of the value of $h_k(30)$ $(k = 1, 2, 3)$ is relatively small for $T_0 = 10$. We can easily see from Figure 1 that the naive time-series analysis method or a simple extrapolation method does not work well for this prediction problem. Thus, it is crucial to accurately estimate the values of the parameters of the VwMV model from the observed opinion diffusion data (more to come later on this).

Since the VwMV model gives a stochastic process, we introduce the *expected share* $g_k(t)$ of each opinion k at time t by

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

and consider the problem of predicting $g_k(t)$ $(k = 1, \cdots, K)$ from the observed data $\mathcal{D}_{T_0}$, which is referred to as the *share prediction problem*. Here, $\langle x \rangle$ denotes the expected value of a random variable x. For solving the share prediction problem, we develop a method that effectively estimates the values of the parameters w_k $(k = 1, \cdots, K)$ and α_v $(v \in V)$ from $\mathcal{D}_{T_0}$. We note that the method developed can also apply to detecting high anti-majoritarian tendency nodes (*i.e.*, anti-majoritarians) from the observed opinion diffusion data.

4.2 Mean Field Approach

Below, we theoretically investigate the asymptotic behavior of expected share $g_k(t)$ $(k = 1, \cdots, K)$ of the VwMV model for a sufficiently large t, and demonstrate that it is crucial to accurately estimate the values of the parameters, w_k, $(k = 1, \cdots, K)$ and α which is the average of α_v over all nodes $v \in V$.

According to previous work in statistical physics, (*e.g.*, Sood and Redner (2005)), we employ a mean field approach. We first consider a rate equation,

$$\frac{dg_k(t)}{dt} = (1 - g_k(t))\,P_k(t) - g_k(t)\,(1 - P_k(t)), \quad (k = 1, \cdots, K), \tag{4}$$

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. (4), $g_k(t)$ is regarded as the probability of choosing a node holding opinion k at time t. Here, we assume that the average local opinion share,

$$\left\langle \frac{n_k(t,v)}{\sum_{j=1}^{K} n_j(t,v)} \right\rangle$$

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. This assumption does not hold in general except that the network is a complete graph where every node's neighbors are all the nodes in the graph, which is not the case here. In fact, without this assumption, we cannot apply the mean field theory and analyze the average behavior of opinion dynamics. Extent to which this assumption is justified must await experimental evaluation by using the real network structure. As shown later, the assumption turned out to be acceptable. Under this assumption, we obtain the following approximation from Eq. (3):

$$P_k(t) \approx (1-\alpha)\,\tilde{p}_k(t,\boldsymbol{w}) + \alpha\,\frac{1-\tilde{p}_k(t,\boldsymbol{w})}{K-1}, \quad (k = 1,\cdots,K), \tag{5}$$

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

$$\tilde{p}_k(t,\boldsymbol{w}) = \frac{w_k\, g_k(t)}{\sum_{j=1}^{K} w_j\, g_j(t)}, \quad (k = 1,\cdots,K). \tag{6}$$

Note that Eq. (5) is exactly satisfied when G is a complete network and the anti-majoritarian tendency is node independent, *i.e.*, $\alpha_v = \alpha$, $(\forall v \in V)$.

4.3 Analysis of VwV Model

For simplicity, we begin with the analysis of the VwV model. In this case, note that $\alpha_v = 0$ $(v \in V)$, *i.e.*, $\alpha = 0$.

4.3.1 Share Analysis

We analyze the behavior of expected share $g_k(t)$ $(k = 1,\cdots,K)$ of the VwV model for a sufficiently large t according to the above mean field approach. From Eqs. (4), (5) and (6), we have

$$\begin{aligned}\frac{dg_k(t)}{dt} &= (1-g_k(t))\frac{g_k(t)w_k}{\sum_{k'=1}^{K} g_{k'}(t)w_{k'}} - g_k(t)\left(1 - \frac{g_k(t)w_k}{\sum_{k'=1}^{K} g_{k'}(t)w_{k'}}\right)\\ &= \frac{g_k(t)w_k}{\sum_{k'=1}^{K} g_{k'}(t)w_{k'}} - g_k(t).\end{aligned} \tag{7}$$

Suppose that the opinion values are non-uniform, and let k^* be the opinion with the highest value parameter such that $w_{k^*} > w_k$ for all the other opinion k $(k \neq k^*)$. Here note that $w_k/w_{k^*} < 1$ for $k \neq k^*$. Then, we can obtain the following inequality from Eq. (7) when $g_k(t) > 0$ for all k:

$$\frac{dg_{k^*}(t)}{dt} = \frac{g_{k^*}(t)w_{k^*}}{\sum_{k=1}^{K} g_k(t)w_k}\left(1-\sum_{k=1}^{K} g_k(t)\frac{w_k}{w_{k^*}}\right)$$
$$> \frac{g_{k^*}(t)w_{k^*}}{\sum_{k=1}^{K} g_k(t)w_k}\left(1-\sum_{k=1}^{K} g_k(t)\right) = 0.$$

Thus, unless $g_{k^*}(t) = 0$, the opinion k^* is expected to finally prevail the others, regardless of its current share since the function $g_{k^*}(t)$ is expected to increase as time passes until each of the other opinion shares becomes 0.

On the other hand, suppose that the opinion values are uniform (*i.e.*, $w_1 = \cdots = w_K$). Then, we obtain from Eq. (7) that

$$\frac{dg_k(t)}{dt} = 0, \quad (k = 1, \cdots, K).$$

Thus, if there exists some $t_0 > 0$ such that $g_1(t_0) = \cdots = g_K(t_0) = 1/K$, then $g_k(t) = 1/K$, $(t \geq t_0)$ for every opinion k. This implies that any opinion can in general become the majority.

Hence, we have the following results:

1. When the opinion values are uniform (*i.e.*, $w_1 = \cdots = w_K$), any opinion can become a winner.
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, $\lim_{t\to\infty} g_{k^*}(t) = 1$.

These results suggest that it is crucially important to accurately estimate the opinion values of the VwV model from the observed data $\mathcal{D}_{T_0}$, [3] and imply that the share prediction problem can be well-defined only when the opinion values are non-uniform. We experimentally confirmed the results for several realistic networks, although the above analysis is valid only when the approximation (see Eq. (5)) holds.

4.3.2 Consensus Time Analysis

We further analyze the consensus time of the VwV model by using the above mean field approach when opinion values are non-uniform. For simplicity, we assume that $w_k = w$ if $k \neq k^*$, *i.e.*, the opinion values of the other opinions are the same. [4] Let r be the ratio of the value parameters defined by $r = w/w_{k^*}$. Then, we obtain the following differential equation for $g_{k^*}(t)$ from Eq. (7):

$$\frac{dg_{k^*}(t)}{dt} = \frac{g_{k^*}(t)}{r(1-g_{k^*}(t))+g_{k^*}(t)} - g_{k^*}(t)$$
$$= \frac{(1-r)g_{k^*}(t)(1-g_{k^*}(t))}{r+(1-r)g_{k^*}(t)}.$$

From this differential equation, we can easily derive the following solution:

$$\frac{r}{1-r}\log(g_{k^*}(t)) - \frac{1}{1-r}\log(1-g_{k^*}(t)) = t + C,$$

[3] If the goal is to predict which opinion wins eventually, it is sufficient to identify which opinion has the highest value, but if we want to estimate the share of each opinion, we need to estimate the values accurately.

[4] This makes the analysis drastically simpler, but the results remains valid qualitatively.

where C stands for a constant of integration. Figure 2 shows examples of expected share curves based on the above solution with different ratios of the opinion values, where the ratio r is set to $r = 1-2^{-d}$ $(d = 1,2,3,4,5)$, and each curve is plotted from $t = 0$ by assuming $g_{k^*}(0) = 0.01$ until $t = T$ that satisfies $g_{k^*}(T) = 0.99$. From Figure 2, we can see that the consensus time is quite short when the ratio r is small, while it takes somewhat longer when the ratio r approaches to 1. More importantly, this result indicates that the consensus time of the VwV model is extremely short even when the ratio r is close to 1, compared with the basic voter model studied in previous work (*e.g.*, Even-Dar and Shapira (2007)). [5] Therefore, we consider that voter model can become more practical by introducing the opinion values.

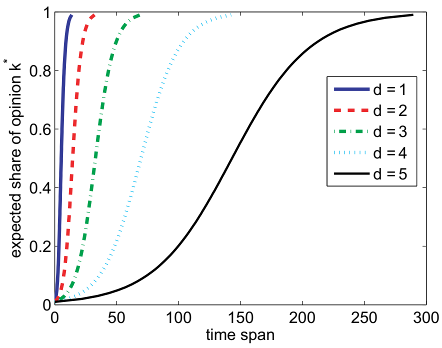


Fig. 2: Examples of expected share curves.

4.4 Analysis of VwMV model

Next, we analyze the behavior of expected share $g_k(t)$ $(k = 1,\cdots,K)$ of the VwMV model for a sufficiently large t according to the above mean field approach.

4.4.1 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. (6), we obtain

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

Thus, we can easily derive from Eqs. (4) and (5) 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

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

[5] Their results is that the basic voter model converges after $O(n^3 \log n)$ steps with probability 1- $o(1)$ where n is the number of nodes.

4.4.2 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 as before for simplicity that

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

where w is a positive constant. We also assume that there exists some $t_0 > 0$ such that

$$g_1(t_0) = \cdots = g_K(t_0) = 1/K.$$

We can see from the symmetry of the setting that $g_k(t) = g_\ell(t)$, $(t \geq t_0)$ if $k, \ell \neq k^*$. This implies that opinion k^* is the winner at time t if and only if $g_{k^*}(t) > 1/K$. Then, from Eqs. (4) and (6), 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. (5) that

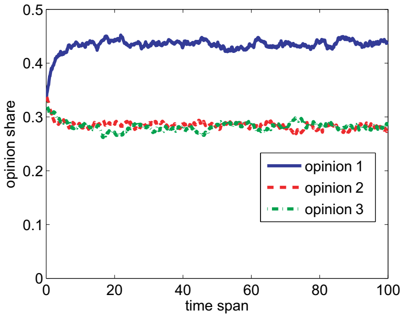
$$\left. \frac{dg_{k^*}(t)}{dt} \right|_{t=t_0} = \frac{s_{k^*} - 1}{K - 1} \left(1 - \frac{1}{K} - \alpha \right).$$

Therefore, we obtain the following results:

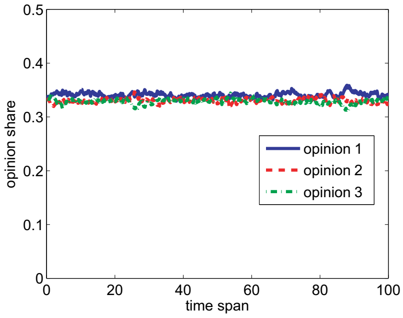
1. When $\alpha < 1 - 1/K$,
$$g_{k^*}(t) > 1/K, \quad (t > t_0),$$
that is, opinion k^* is expected to spread most widely and become the majority.
2. When $\alpha = 1 - 1/K$,
$$g_k(t) = 1/K, \quad (t \geq t_0),$$
for any opinion k, that is, any opinion can become a winner.
3. When $\alpha > 1 - 1/K$,
$$g_{k^*}(t) < 1/K, \quad (t > t_0),$$
that is, opinion k^* is expected to spread least widely and become the minority.

4.4.3 Experiments:

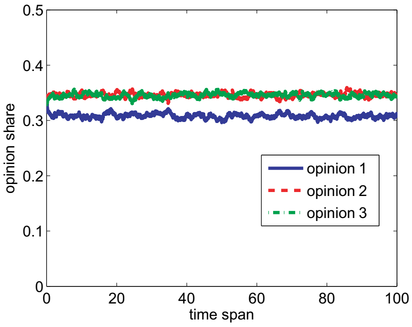
The above theoretical results are justified only when the approximation (see Eq. (5)) holds, which is always true in the case of complete networks. Real social networks are much more sparse and thus, we need to verify the extent to which the above results are true for real networks. We experimentally 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 6), 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 certain combinations of shape parameters a and b. Figure 3 shows the results of opinion share curves, $t \mapsto h_k(t) \Big/ \sum_{j=1}^{K} h_j(t)$, $(k = 1, 2, 3)$,



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

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

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

We obtained similar results to those in Figure 3 also for many other trials. These results support the validity of our theoretical analysis.

5 Learning Method

In this section we describe a method for estimating parameter values of the VwMV model from a given observed opinion spreading data $\mathcal{D}_{T_0}$. Based on the evolution process of our model (see Eq. (3)), we can obtain the likelihood function,

$$\mathcal{L}(\mathcal{D}_{T_0}; \boldsymbol{w}, \alpha) = \log\left(\prod_{(v,t,k)\in\mathcal{D}_{T_0}} P(f_t(v) = k) \right), \tag{8}$$

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 $\mathcal{L}(\mathcal{D}_{T_0}; \boldsymbol{w}, \boldsymbol{\alpha})$ with respect to $\boldsymbol{w}$ and $\boldsymbol{\alpha}$. Note from Eqs. (2), (3) and (8) that $\mathcal{L}(\mathcal{D}_{T_0}; c\boldsymbol{w}, \boldsymbol{\alpha}) = c\mathcal{L}(\mathcal{D}_{T_0}; \boldsymbol{w}, \boldsymbol{\alpha})$ for any $c > 0$. Note also that each opinion value w_k is positive. Thus, we transform the parameter vector $\boldsymbol{w}$ by $\boldsymbol{w} = \boldsymbol{w}(\boldsymbol{x})$, where

$$\boldsymbol{w}(\boldsymbol{x}) = (e^{x_1}, \cdots, e^{x_{K-1}}, 1), \quad \left(\boldsymbol{x} = (x_1, \cdots, x_{K-1}) \in \mathbf{R}^{K-1}\right). \tag{9}$$

Namely, our problem is to estimate the values of $\boldsymbol{x}$ and $\boldsymbol{\alpha}$ that maximize $\mathcal{L}(\mathcal{D}_{T_0}; \boldsymbol{w}(\boldsymbol{x}), \boldsymbol{\alpha})$.

We derive an 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 $k, j \in \{1, \cdots, K\}$,

$$\beta_{v,k,j}(\boldsymbol{\alpha}) = \begin{cases} 1 - \alpha_v & \text{if} \quad j = k, \\ \alpha_v/(K-1) & \text{if} \quad j \neq k. \end{cases} \tag{10}$$

Then, from the definition of $P(f_t(v) = k)$ (see Eq. (3)), by noting $1 - p_k(t, v, \boldsymbol{w}) = \sum_{j \neq k} p_j(t, v, \boldsymbol{w})$, we can express Eq. (8) as follows:

$$\mathcal{L}(\mathcal{D}_{T_0}; \boldsymbol{w}(\boldsymbol{x}), \boldsymbol{\alpha}) = \sum_{(v,t,k) \in \mathcal{D}_{T_0}} \log \left(\sum_{j=1}^{K} \beta_{v,k,j}(\boldsymbol{\alpha}) \; p_j(t, v, \boldsymbol{w}(\boldsymbol{x})) \right).$$

Now, let $\bar{z}$ and $\bar{\boldsymbol{\alpha}}$ be the current estimates of $\boldsymbol{x}$ and $\boldsymbol{\alpha}$, respectively. Then, we define $q_{v,t,k,j}(\boldsymbol{x}, \boldsymbol{\alpha})$ by

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

($v \in V$, $0 \leq t \leq T_0$, $k, j = 1, \cdots, K$), and transform our objective function as follows:

$$\mathcal{L}(\mathcal{D}_{T_0}; \boldsymbol{w}(\boldsymbol{x}), \boldsymbol{\alpha}) = \mathcal{Q}(\boldsymbol{x}, \boldsymbol{\alpha}; \bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}}) - \mathcal{H}(\boldsymbol{x}, \boldsymbol{\alpha}; \bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}}), \tag{11}$$

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

$$\mathcal{Q}(\boldsymbol{x}, \boldsymbol{\alpha}; \bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}}) = \mathcal{Q}_1(\boldsymbol{x}; \bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}}) + \mathcal{Q}_2(\boldsymbol{\alpha}; \bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}}), \tag{12}$$

$$\mathcal{Q}_1(\boldsymbol{x}; \bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}}) = \sum_{(v,t,k) \in \mathcal{D}_{T_0}} \sum_{j=1}^{K} q_{v,t,k,j}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}}) \log p_j(t, v, \boldsymbol{w}(\boldsymbol{x})), \tag{13}$$

$$\mathcal{Q}_2(\boldsymbol{\alpha}; \bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}}) = \sum_{(v,t,k) \in \mathcal{D}_{T_0}} \sum_{j=1}^{K} q_{v,t,k,j}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}}) \log \beta_{v,k,j}(\boldsymbol{\alpha}), \tag{14}$$

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

$$\mathcal{H}(\boldsymbol{x}, \boldsymbol{\alpha}; \bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}}) = \sum_{(v,t,k) \in \mathcal{D}_{T_0}} \sum_{j=1}^{K} q_{v,t,k,j}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}}) \log q_{v,t,k,j}(\boldsymbol{x}, \boldsymbol{\alpha}).$$

Since $\mathcal{H}(\boldsymbol{x}, \boldsymbol{\alpha}; \bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}})$ is maximized at $\boldsymbol{x} = \bar{\boldsymbol{x}}$ and $\boldsymbol{\alpha} = \bar{\boldsymbol{\alpha}}$, we can increase the value of $\mathcal{L}(\mathcal{D}_{T_0}; \boldsymbol{w}(\boldsymbol{x}), \boldsymbol{\alpha})$ by maximizing $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{\alpha}; \bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}})$ with respect to $\boldsymbol{x}$ and $\boldsymbol{\alpha}$ (see Eq. (11)). From Eq. (12), we can maximize $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{\alpha}; \bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}})$ by independently maximizing $\mathcal{Q}_1(\boldsymbol{x}; \bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}})$ and $\mathcal{Q}_2(\boldsymbol{\alpha}; \bar{\boldsymbol{x}}, \bar{\boldsymbol{\alpha}})$ with respect to $\boldsymbol{x}$ and $\boldsymbol{\alpha}$, respectively.

First, we estimate the value of $\boldsymbol{x}$ that maximizes $Q_1(\boldsymbol{x};\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}})$. Here, note from Eqs.(2) and (9) that for $j = 1,\cdots,K$ and $\lambda = 1,\cdots,K-1$,

$$\frac{\partial p_j(t,v,\boldsymbol{w}(\boldsymbol{x}))}{\partial x_\lambda} = \delta_{j,\lambda}\, p_j(t,v,\boldsymbol{w}(\boldsymbol{x})) - p_j(t,v,\boldsymbol{w}(\boldsymbol{x}))\, p_\lambda(t,v,\boldsymbol{w}(\boldsymbol{x})), \tag{15}$$

where $\delta_{j,\lambda}$ is Kronecker's delta. From Eqs. (13) and (15), we have

$$\frac{\partial Q_1(\boldsymbol{x};\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}})}{\partial x_\lambda} = \sum_{(v,t,k)\in\mathcal{D}_{T_0}} \sum_{j=1}^{K} q_{v,t,k,j}(\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}}) \left(\delta_{j,\lambda} - p_\lambda(t,v,\boldsymbol{w}(\boldsymbol{x}))\right), \tag{16}$$

for $\lambda = 1,\cdots,K-1$. Moreover, from Eqs. (15) and (16), we have

$$\frac{\partial^2 Q_1(\boldsymbol{x};\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}})}{\partial x_\lambda\,\partial x_\mu} = \sum_{(v,t,k)\in\mathcal{D}_{T_0}} \sum_{j=1}^{K} q_{v,t,k,j}(\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}}) \Big(p_\lambda(t,v,\boldsymbol{w}(\boldsymbol{x}))\, p_\mu(t,v,\boldsymbol{w}(\boldsymbol{x})) - \delta_{\lambda,\mu}\, p_\lambda(t,v,\boldsymbol{w}(\boldsymbol{x}))\Big),$$

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

$$\begin{aligned}
&\sum_{\lambda,\mu=1}^{K-1} \frac{\partial^2 Q_1(\boldsymbol{x};\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}})}{\partial x_\lambda\,\partial x_\mu} y_\lambda y_\mu \\
&= \sum_{(v,t,k)\in\mathcal{D}_{T_0}} \sum_{j=1}^{K} q_{v,t,k,j}(\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}}) \left[\left(\sum_{\lambda=1}^{K-1} p_\lambda(t,v,\boldsymbol{w}(\boldsymbol{x}))\, y_\lambda\right)^2 - \sum_{\lambda=1}^{K-1} p_\lambda(t,v,\boldsymbol{w}(\boldsymbol{x}))\, {y_\lambda}^2\right] \\
&= -\sum_{(v,t,k)\in\mathcal{D}_{T_0}} \sum_{j=1}^{K} q_{v,t,k,j}(\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}}) \left[\sum_{\lambda=1}^{K-1} p_\lambda(t,v,\boldsymbol{w}(\boldsymbol{x})) \left(y_\lambda - \sum_{\mu=1}^{K-1} p_\mu(t,v,\boldsymbol{w}(\boldsymbol{x}))\, y_\mu\right)^2 \right. \\
&\qquad\qquad \left. + \left(1 - \sum_{\lambda=1}^{K-1} p_\lambda(t,v,\boldsymbol{w}(\boldsymbol{x}))\right) \left(\sum_{\mu=1}^{K-1} p_\mu(t,v,\boldsymbol{w}(\boldsymbol{x}))\, y_\mu\right)^2\right] \\
&\leq 0,
\end{aligned} \tag{17}$$

for any $(y_1,\cdots,y_{K-1}) \in \mathbf{R}^{K-1}$. Hence, by solving the equations $\partial Q_1(\boldsymbol{x};\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}})/\partial x_\lambda = 0$, $(\lambda = 1,\cdots,K-1)$ (see Eq. (16)), we can find the value of $\boldsymbol{x}$ that maximizes $Q_1(\boldsymbol{x};\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}})$. We employed a standard Newton Method in our experiments.

Next, we estimate the value of $\boldsymbol{\alpha}$ that maximizes $Q_2(\boldsymbol{\alpha};\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}})$. From Eqs. (10) and (14), we have

$$Q_2(\boldsymbol{\alpha};\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}}) = \sum_{(v,t,k)\in\mathcal{D}_{T_0}} \left(q_{v,t,k,k}(\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}}) \log(1-\alpha_v) + (1 - q_{v,t,k,k}(\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}})) \log\left(\frac{\alpha_v}{K-1}\right)\right).$$

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

$$\alpha_v = \frac{1}{|\mathcal{D}_{T_0}(v)|} \sum_{(t,k)\in\mathcal{D}_{T_0}(v)} (1 - q_{v,t,k,k}(\bar{\boldsymbol{x}},\bar{\boldsymbol{\alpha}})),$$

for each $v \in V$, where $\mathcal{D}_{T_0}(v) = \{(t,k);\, (v,t,k) \in \mathcal{D}_{T_0}\}$.

When $\alpha_v = 0$ for any $v \in V$, the VwMV model is reduced to the VwV model. Thus, a straightforward application of the above learning algorithm for the VwMV model gives the learning algorithm for the VwV model. Note here that for the VwV model, the objective function becomes

$$\mathcal{L}(\mathcal{D}_{T_0}; \boldsymbol{w}(\boldsymbol{x}), \boldsymbol{0}) = \sum_{(v,t,k)\in\mathcal{D}_{T_0}} \log p_k(t, v, \boldsymbol{w}(\boldsymbol{x})),$$

(see Eqs. (1) and (8)), and its second derivatives become

$$\frac{\partial^2 \mathcal{L}(\mathcal{D}_{T_0}; \boldsymbol{w}(\boldsymbol{x}), \boldsymbol{0})}{\partial x_\lambda \, \partial x_\mu} = \sum_{(v,t,k)\in\mathcal{D}_{T_0}} \Big(p_\lambda(t, v, \boldsymbol{w}(\boldsymbol{x}))\, p_\mu(t, v, \boldsymbol{w}(\boldsymbol{x})) - \delta_{\lambda,\mu}\, p_\lambda(t, v, \boldsymbol{w}(\boldsymbol{x})) \Big),$$

$(\lambda, \mu = 1, \cdots, K-1)$. In a similar way to Eq. (17), we can easily prove that the Hessian matrix $(\partial^2 \mathcal{L}(\mathcal{D}_{T_0}; \boldsymbol{w}(\boldsymbol{x}), \boldsymbol{0}) / \partial x_\lambda \, \partial x_\mu)$ is negative semi-definite. Therefore, we can guarantee that the optimal solution of the objective function is global optimal for the VwV model. Here, we mention that although it is not guaranteed that the optimal solution of the objective function of the VwMV model is global optimal, their estimated parameter values converged very closely to their true values in our experiments when there is an enough amount of training data.

6 Experimental Evaluation

Using large real networks, we experimentally investigate the capability of the proposed model and the performance of the proposed learning method. We first show the results of the accuracies of predicting future opinion shares. We then 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-majoritarians).

6.1 Experimental Settings

We employed four datasets of large real networks, which are all bidirectionally connected networks [6] and exhibit many of the key features of social networks. [7] The first 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 second one is a Coauthor network (Palla et al, 2005) and has 12,357 nodes and 38,896 directed links (the Coauthor network). The third one is a network derived from the Enron Email Dataset (Klimt and Yang, 2004) by extracting the senders and the recipients and linking those that had bidirectional communications. It has 4,254 nodes and 44,314 directed links (the Enron network). The last one is a network of people that was derived from the "list of people" within Japanese Wikipedia (Kimura et al, 2009), which has 9,481 nodes and 245,044 directed links (the Wikipedia network). Just to provide a sense of how fast the opinion can propagate, the average shortest path of each network is given here: 8.175 for the Blog network, 8.160 for the Coauthor network, 3.726 for the Enron network and 4.700 for the Wikipedia network.

[6] Opinion propagation is directional. Choosing bidirectional networks means that opinion can propagate in both directions.

[7] It would be the best if we can use the real opinion propagation data. However, as we are not able to find such data, the next best is to use the network structures constructed from the real world social media data (not synthetic networks).

To do experiments, we have to first determine the values of parameters: the number of opinions K, the true value of each opinion w_k^*, the true value of each anti-majoritarian tendency α_v^*, $(v \in V)$. We varied $K = 2, 3, \cdots, 10$, and chose w_k^* from the interval $[0.5, 1.5]$ uniformly at random and α_v^* by drawing it from the beta distribution with the shape parameters a and b. We chose the beta distribution simply because of the easiness of controlling the average and variance of the distribution. As implied in Subsection 3.1, we used the exponential distribution with $\eta_v = 1$ to determine the opinion update time. Which nodes to start from is another problem. As explained also in Subsection 3.1 we assigned each opinion to only one node initially and all other nodes were set in the neutral state. Those initially assigned K nodes are taken from the top K nodes with respect to the node degree ranking. We start simulating the opinion propagation process from these K nodes using the parameter values which are assumed true, and generated $\mathcal{D}_{T_0}$. As for our learning settings, we set the initial value of each value parameter to $w_k = 1$, and the initial value of each anti-majoritarian tendency to $\alpha_v = 0.5$, $(v \in V)$. We terminated the learning iteration when the increase of our objective function becomes sufficiently small, *i.e.*,

$$\frac{\mathcal{L}(\mathcal{D}_{T_0}; \boldsymbol{w}, \boldsymbol{\alpha}) - \mathcal{L}(\mathcal{D}_{T_0}; \bar{\boldsymbol{w}}, \bar{\boldsymbol{\alpha}})}{\mathcal{L}(\mathcal{D}_{T_0}; \bar{\boldsymbol{w}}, \bar{\boldsymbol{\alpha}})} < 10^{-8},$$

where $\boldsymbol{w}$ and $\boldsymbol{\alpha}$ mean the parameter vectors updated from $\bar{\boldsymbol{w}}$ and $\bar{\boldsymbol{\alpha}}$. Note that our learning algorithm always increases our objective function as described in the previous section.

6.2 Share Prediction

For each number of opinions $(k = 2, 3, \cdots, K)$ we predicted the expected share $g_k(T)$ for the observed data $\mathcal{D}_{T_0}$, where we set $T = 30$, and investigated the cases $T_0 = 10, 15$ and $\alpha = 0.5, 0.1, 0.01$ by generating α_v with $(a, b) = (2, 2), (1, 9), (1, 99)$, respectively. As we mentioned in Section 1 we think it is important to learn the model using a small amount of data and predict the near future. Since the average shortest path of each network is less than 10, $T_0 = 10$ is the minimum training time required to learn the parameters for all the nodes. Note that α means the average anti-majoritarian tendency, which is given by $a/(a+b)$. Namely, after we have estimated the values of each w_k and each α_v, we predicted the value of $g_k(T)$ by simulating the model M times from $\mathcal{D}_{T_0}$ and taking their average, where we used $M = 100$. In fact, our preliminary experiments indicate that the results for $M = 100$ are not much different from those for $M = 1,000$ and $10,000$ in the networks we used.

In order to investigate the importance of introducing the anti-majoritarian tendency of each node, we compared the proposed method with the VwV model which has no anti-majoritarian component. Moreover, in order to investigate the importance of introducing the opinion values, we also 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 this method as the *uniform value method*. Furthermore, given the observed data $\mathcal{D}_{T_0}$, we can simply apply a polynomial extrapolation for predicting the expected share of opinion k at a target time T, since we can naively speculate that the recent trend for each opinion captured by the polynomial function approximation continues. Thus, we consider predicting the values of $g_1(T)$, $\cdots$, $g_K(T)$, by estimating the value of the population $h_k(T)$ of opinion k at time T based on the polynomial function of degree L that interpolates the $L+1$ data points $\{(T_0 - \Delta + \ell\Delta/L, h_k(T_0 - \Delta + \ell\Delta/L)); \ell = 0, 1, \cdots, L\}$, where Δ is the parameter with $0 < \Delta \leq T_0$. We refer to this prediction method as the *polynomial extrapolation method*. In our experiments,

Table 1: Results of opinion share prediction for the Blog network ($T_0 = 10$, $\alpha = 0.1$, $K = 10$). Note that the two-side 0.05 point of the t-distribution with 9 degrees of freedom is $t^*_{9,0.05} = 2.262$.

Method	Average of error $\mathcal{E}_g$	t-value $\mathcal{T}_g^{PC}$
proposed	0.0396	—
VwV	0.4520	15.4645
uniform value	0.5172	13.4893
linear ($\varDelta = 1$)	0.4996	12.5024
linear ($\varDelta = 3$)	0.4243	12.0177
linear ($\varDelta = 5$)	0.3247	14.4648
quadratic ($\varDelta = 1$)	1.0845	13.2210
quadratic ($\varDelta = 3$)	1.2795	26.9768
quadratic ($\varDelta = 5$)	1.3296	15.6869
cubic ($\varDelta = 1$)	1.3710	18.3478
cubic ($\varDelta = 3$)	1.1799	12.5790
cubic ($\varDelta = 5$)	1.1219	16.7674
quartic ($\varDelta = 1$)	1.1963	19.0506
quartic ($\varDelta = 3$)	1.1079	16.2728
quartic ($\varDelta = 5$)	1.0956	11.9049

Table 2: Results of opinion share prediction for the Coauthor network ($T_0 = 10$, $\alpha = 0.1$, $K = 10$). Note that the two-side 0.05 point of the t-distribution with 9 degrees of freedom is $t^*_{9,0.05} = 2.262$.

Method	Average of error $\mathcal{E}_g$	t-value $\mathcal{T}_g^{PC}$
proposed	0.0590	—
VwV	0.4634	15.9587
uniform value	0.4422	13.5598
linear ($\varDelta = 1$)	0.4193	14.1568
linear ($\varDelta = 3$)	0.2814	10.2062
linear ($\varDelta = 5$)	0.2097	9.5952
quadratic ($\varDelta = 1$)	1.0794	17.6792
quadratic ($\varDelta = 3$)	1.2158	12.9942
quadratic ($\varDelta = 5$)	1.6140	22.1016
cubic ($\varDelta = 1$)	1.1616	16.4268
cubic ($\varDelta = 3$)	1.1615	17.8509
cubic ($\varDelta = 5$)	0.9575	18.6748
quartic ($\varDelta = 1$)	1.1852	14.3082
quartic ($\varDelta = 3$)	1.0971	14.1797
quartic ($\varDelta = 5$)	1.1889	17.3193

we adopted $L = 1, 2, 3, 4$, *i.e.*, the linear, quadratic, cubic, and quartic polynomial functions, and examined $\varDelta = 1$, $\varDelta = 3$, and $\varDelta = 5$. We evaluated the effectiveness of the proposed share prediction method by comparing it with the above six methods (VwV, uniform and four polynomial).

Let $\widehat{g}_k(T)$ be the estimate of $g_k(T)$ by a share prediction method. We measured the performance of the share prediction method by the prediction error $\mathcal{E}_g$ defined by [8]

$$\mathcal{E}_g = \sum_{k=1}^{K} |\widehat{g}_k(T) - g_k(T)|.$$

[8] It may sound more reasonable to weight each difference by the share itself, but we decided not to do so. We rather considered the prediction problem as the classification problem.

Table 3: Results of opinion share prediction for the Enron network ($T_0 = 10$, $\alpha = 0.1$, $K = 10$). Note that the two-side 0.05 point of the t-distribution with 9 degrees of freedom is $t^*_{9,0.05} = 2.262$.

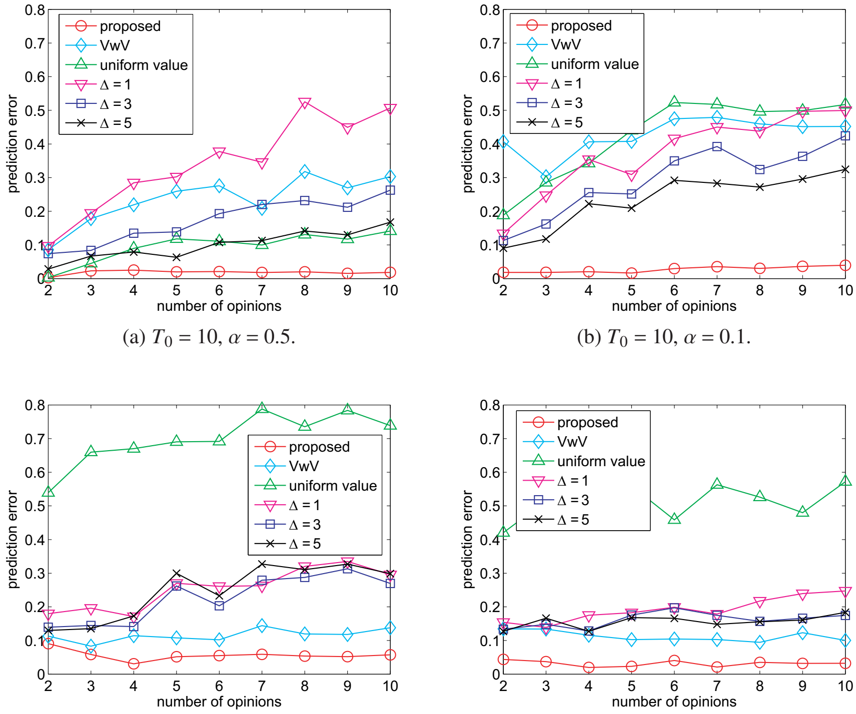
Method	Average of error $\mathcal{E}_g$	t-value $\mathcal{T}_g^{PC}$
proposed	0.0731	—
VwV	0.6030	12.7367
uniform value	0.6088	20.4684
linear ($\Delta = 1$)	0.6909	8.4882
linear ($\Delta = 3$)	0.6511	11.2945
linear ($\Delta = 5$)	0.5577	15.1556
quadratic ($\Delta = 1$)	1.1341	11.4784
quadratic ($\Delta = 3$)	1.0765	13.9631
quadratic ($\Delta = 5$)	1.1763	14.7290
cubic ($\Delta = 1$)	1.2378	15.1644
cubic ($\Delta = 3$)	1.1378	16.2330
cubic ($\Delta = 5$)	1.2605	16.6612
quartic ($\Delta = 1$)	1.1699	11.4147
quartic ($\Delta = 3$)	1.3411	32.1862
quartic ($\Delta = 5$)	1.1910	15.0670

Table 4: Results of opinion share prediction for the Wikipedia network ($T_0 = 10$, $\alpha = 0.1$, $K = 10$). Note that the two-side 0.05 point of the t-distribution with 9 degrees of freedom is $t^*_{9,0.05} = 2.262$.

Method	Average of error $\mathcal{E}_g$	t-value $\mathcal{T}_g^{PC}$
proposed	0.0390	—
VwV	0.4429	12.8927
uniform value	0.6000	11.6327
linear ($\Delta = 1$)	0.5151	13.2910
linear ($\Delta = 3$)	0.4073	12.2377
linear ($\Delta = 5$)	0.3968	14.8808
quadratic ($\Delta = 1$)	1.1122	12.8117
quadratic ($\Delta = 3$)	1.1521	15.0864
quadratic ($\Delta = 5$)	1.1674	16.0370
cubic ($\Delta = 1$)	1.2193	13.7714
cubic ($\Delta = 3$)	1.1950	16.4728
cubic ($\Delta = 5$)	1.0156	16.7386
quartic ($\Delta = 1$)	1.0679	12.1467
quartic ($\Delta = 3$)	1.2045	18.3987
quartic ($\Delta = 5$)	1.3886	27.4023

We first examined the case of $T_0 = 10$, $\alpha = 0.1$ and $K = 10$. Tables 1, 2, 3 and 4 are the results of opinion share prediction for the Blog, the Coauthor, the Enron and the Wikipedia networks, respectively. We conducted 10 trials varying the true values of value parameters for each K, and the second column in Tables 1, 2, 3 and 4 indicates the average of $\mathcal{E}_g$ over the 10 trials. In order to investigate whether the difference of the prediction error $\mathcal{E}_g$ between the proposed method and each of the other methods used for comparison is statistically significant or not, we performed a t-test. Let $\mathcal{E}_g^P$ and $\mathcal{E}_g^C$ denote the values of $\mathcal{E}_g$ for the proposed method and the compared method, respectively. We calculated t-value

$$\mathcal{T}_g^{PC} = \frac{\sqrt{10}\,\mathrm{mean}\left(\mathcal{E}_g^P - \mathcal{E}_g^C\right)}{\mathrm{std}\left(\mathcal{E}_g^P - \mathcal{E}_g^C\right)},$$



(a) $T_0 = 10$, $\alpha = 0.5$. (b) $T_0 = 10$, $\alpha = 0.1$.

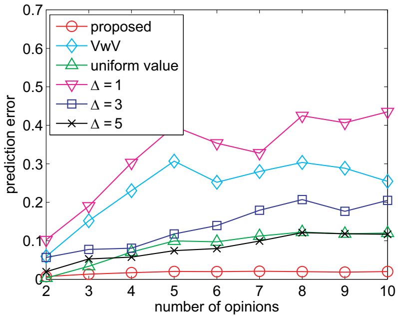
(c) $T_0 = 10$, $\alpha = 0.01$. (d) $T_0 = 15$, $\alpha = 0.01$.

Fig. 4: Results of opinion share prediction for the Blog network.

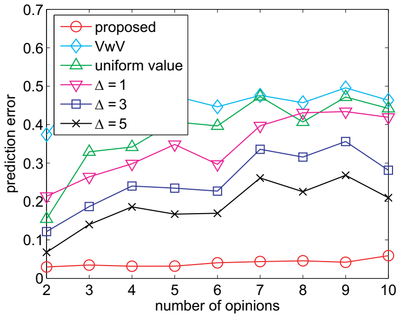
where mean(x) and std(x) denote the standard average and the sample standard deviation of sample x, respectively. In Tables 1, 2, 3 and 4, the third column indicates the t-value $\mathcal{T}_g^{PC}$. Here, note that the two-side 0.05 point of the t-distribution with 9 degrees of freedom is $t^*_{9,0.05} = 2.262$. Thus, we see that in the case of $T_0 = 10$, $\alpha = 0.1$ and $K = 10$, the difference between the proposed method and each of the compared methods in prediction error $\mathcal{E}_g$ is statistically significant by the t-test at significance level 0.05. Moreover, from Tables 1, 2, 3 and 4, we see that the linear extrapolation method performed best among the polynomial extrapolation methods in the case of $T_0 = 10$, $\alpha = 0.1$ and $K = 10$. We obtained the same results for the other cases with different combinations of T_0, α and K. Thus, we show only the results of the linear extrapolation method for the polynomial extrapolation method.

Figure 4 is the results for the Blog network, where circles, diamonds and upward triangles indicate the prediction errors of the proposed method, the VwV method, and the uniform value method, respectively, and downward triangles, squares, and crosses indicate those of the linear extrapolation method adopting $\varDelta = 1$, $\varDelta = 3$, and $\varDelta = 5$, respectively. Figure 4 (a), (b), (c) and (d) are the results for $(T_0, \alpha) = (10, 0.5)$, $(10, 0.1)$, $(10, 0.01)$, and $(15, 0.01)$, respectively. Figures 5, 6, and 7 are the results for the other three networks, *i.e.*, the Coauthor network, the Enron network, and the Wikipedia network, respectively.

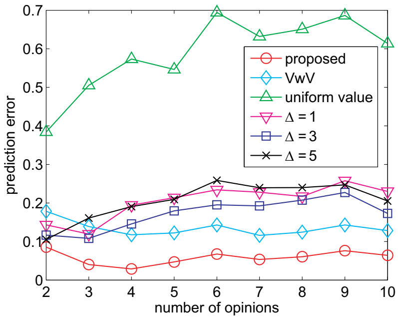
From these figures, we see that the proposed method worked substantially better than the other methods. More specifically, the VwV method worked poorly when values for the anti-majoritarian tendency were relatively large. Conversely, the uniform value method



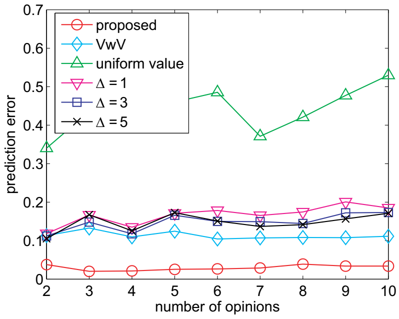
(a) $T_0 = 10$, $\alpha = 0.5$.



(b) $T_0 = 10$, $\alpha = 0.1$.



(c) $T_0 = 10$, $\alpha = 0.01$.

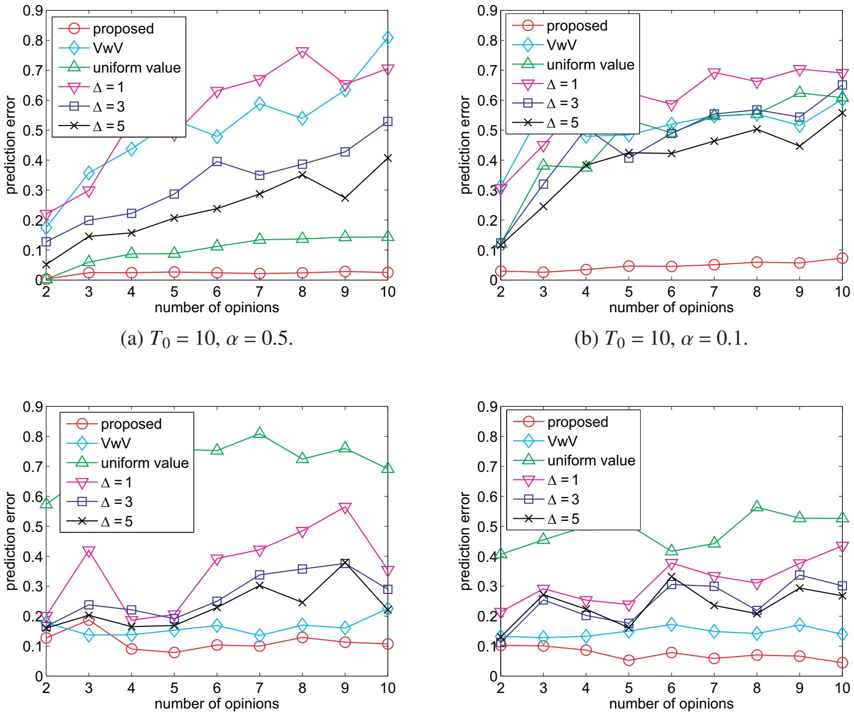


(d) $T_0 = 15$, $\alpha = 0.01$.

Fig. 5: Results of opinion share prediction for the Coauthor network

worked poorly when they were relatively small. These results are predictable because the VwV method cannot cope with the effect of the anti-majoritarian tendency and the uniform value method cannot cope with the effect of opinion value. We further see that the proposed method significantly outperformed the *polynomial extrapolation method* in every case. Especially, we observed that the proposed method accurately predicted the share at T even in the case that the share ranking at T_0 got reversed at the target time T as shown in Figure 1. This is attributed to the use of the estimated value parameters which take different values for different opinions, and is consistent with the results of the mean field analysis. We also observe that compared with cases of $\alpha = 0.5$ and $\alpha = 0.1$, the performance of the proposed method in case of $\alpha = 0.01$ becomes worse for $T_0 = 10$. This is because the opinion change driven by the anti-majoritarians is smaller when α is smaller, thereby providing less effective training data for learning α. Larger error for α negatively affects the results of share prediction despite the effect of anti-majoritarians is less. However, it becomes better and comparable to the other cases for $T_0 = 15$ as expected since the amount of training data increases.

During the experiments we noticed that the time needed to reach the consensus gets longer when the difference between the largest and the second largest values of the opinion value parameters is small. This can also be predicted by the consensus time analysis, *i.e.*, considering the case where the highest two values are the same and the rest are also the same.



(a) $T_0 = 10$, $\alpha = 0.5$.

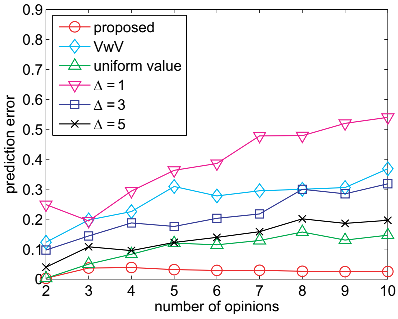
(b) $T_0 = 10$, $\alpha = 0.1$.

(c) $T_0 = 10$, $\alpha = 0.01$.

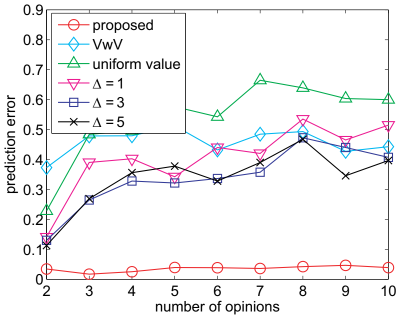
(d) $T_0 = 15$, $\alpha = 0.01$.

Fig. 6: Results of opinion share prediction for the Enron network

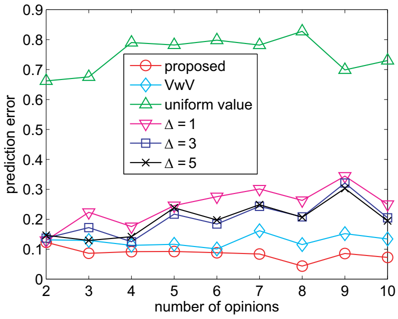
In this subsection, we focused only on the accuracy of share prediction and did not discuss the accuracy of parameter learning. As conjectured in Section 1, learning the opinion values is easy and learning the anti-majoritarian tendency is hard. Indeed, all the opinion values can be estimated in good accuracy. The average error was 6% even using a training data for such a short period of time. However, as predicted, the average error of the estimated anti-majoritarian tendency is large. For example, in the case of $T_0 = 10$, $\alpha = 0.5$ and $K = 10$, the average value of error $\mathcal{E}_\alpha$ was more than 0.17 for all the four networks. Namely, the estimation error of anti-majoritarian tendency for each node was more than $(0.17/0.5) * 100 = 34\%$ on the average. This is because the number of parameters is the same as the number of nodes which is very large. Nevertheless, the accuracy of share prediction is very good. For example, in the case of $T_0 = 10$, $\alpha = 0.5$ and $K = 10$, the average value of error $\mathcal{E}_g$ was less than 0.026 for all the four networks. Namely, the share prediction error for each opinion was less than $((0.026/10)/(1/10)) * 100 = 2.6\%$ on the average. This looks strange at a glance, but we can explain the reason as follows. We started with the K distinct initial nodes and all the other nodes were neutral in the beginning. Recall that we set the average time delay to 1.0, which means that on the average each node updates its opinion every single time unit. Thus when $T_0 = 10$ the opinion updates can propagate 10 steps on the average. As explained in Subsection 6.2, considering that the average shortest path of the network is less than 10 for all the networks, opinion update takes place barely almost all the nodes. For some nodes the number of updates is 10 and for other nodes it is 1. The accuracy



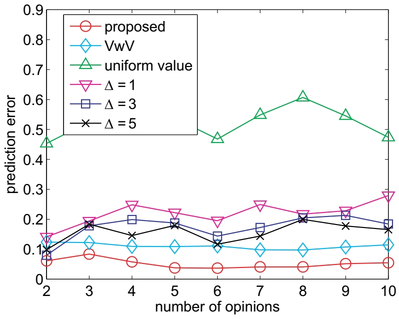
(a) $T_0 = 10$, $\alpha = 0.5$.



(b) $T_0 = 10$, $\alpha = 0.1$.



(c) $T_0 = 10$, $\alpha = 0.01$.



(d) $T_0 = 15$, $\alpha = 0.01$.

Fig. 7: Results of opinion share prediction for the Wikipedia network

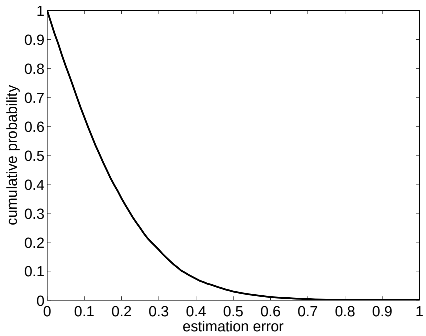


Fig. 8: Distribution of estimation error for anti-majority tendency of each node in the Blog network ($T_0 = 10$, $\alpha = 0.5$, $K = 10$).

of the anti-majoritarian tendency for these nodes where the opinion updates are very few is indeed very bad (no valid learning took place), but the accuracy for the nodes that undergo several opinion updates is good. The variance of the node-wise accuracy is large. Figure 8 which is the cumulative error probability $P(|\hat{\alpha}_v - \alpha_v^*| \geq x)$ in case of $T_0 = 10$, $\alpha = 0.5$ and $K = 10$ for the Blog network clearly indicates this, where each α_v^* and $\hat{\alpha}_v$ denote the true and the estimated anti-majoritarian tendencies of node v, respectively. The average error is indeed large and about 30% of nodes have errors greater than 50%. However, as the mean field analysis implies, it is the average of the anti-majoritarian tendency that matters, as the first approximation, as far as the opinion share is concerned. In this case, we can verify that $\sum_{v \in V} \left| \hat{\alpha}_v - \alpha^*_v \right| / |V| = 0.1811$ and $\left| \sum_{v \in V} \hat{\alpha}_v / |V| - \sum_{v \in V} \alpha^*_v / |V| \right| = 0.0015$. The latter is three orders of magnitude less. This explains the good accuracy of the opinion share despite the bad accuracy of the anti-majoritarian tendency. In the next subsection we will describe the accuracy of the anti-majoritarian tendency using more training data.

To sum up, we confirmed that the results of our theoretical analyses hold in these real networks and that the proposed method outperforms the *polynomial extrapolation method*. On the average, the prediction error of the proposed method was about four times less for a given T_0. Besides, it achieved a comparable prediction accuracy with the observation time three times less compared with the *polynomial extrapolation method*.

6.3 Discovery of Anti-majority Opinionists

We examined the accuracy of discovering anti-majoritarian opinionists (and majoritarian opinionists) for both a small ($K = 3$) and a large ($K = 10$) K, by varying $T_0 = 100, 200, \cdots,$ 1000. The error is measured by $\mathcal{E}_\alpha$,

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

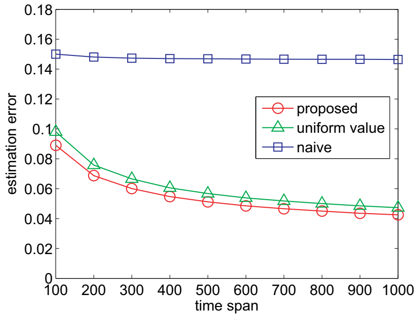
We also measured the accuracies of detecting the high and the low anti-majoritarian tendency nodes by F-measures $\mathcal{F}_A$ and $\mathcal{F}_N$, respectively. Here, $\mathcal{F}_A$ and $\mathcal{F}_N$ are defined as follows:

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

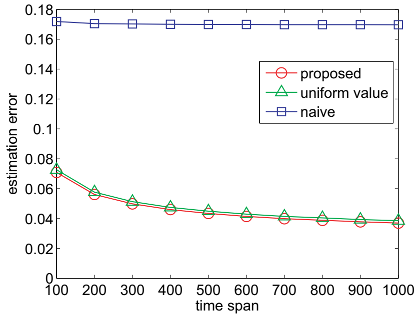
where A^* and $\hat{A}$ are the sets of the true and the estimated top 15% nodes of high anti-majoritarian tendency, respectively, and N^* and $\hat{N}$ are the sets of the true and the estimated top 15% nodes of low anti-majoritarian tendency, respectively.

We 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's opinion in its neighborhood. We refer to the method as the *naive counting method*. We also compared the proposed method with the uniform value method mentioned in the previous subsection.

Figures 9 and 10 are the results for the Blog network, where circles, upward triangles, and squares indicate the prediction errors and the F-measure performance of the proposed method, the uniform value method, and *naive method*, respectively. Figures 9 (a), and (b) show the estimation error $\mathcal{E}_\alpha$ of each method as a function of time span T_0 with $K = 3$ and $K = 10$, respectively, while Figures 10 (a) and (b) the F-measure $\mathcal{F}_A$ of each method as a function of time span T_0 with $K = 3$ and $K = 10$, respectively. Here, we repeated the same experiment 10 times independently, and plotted the average over the 10 results. Figures 11, 12, 13, 14, 15 and 16 are the results for the other three networks, *i.e.*, the Coauthor network, the Enron network, and the Wikipedia network, respectively. Note that we only showed the results for $\alpha = 0.5$, *i.e.*, $a = b = 2$, because we obtained quite similar results for the other anti-majoritarian tendency α.

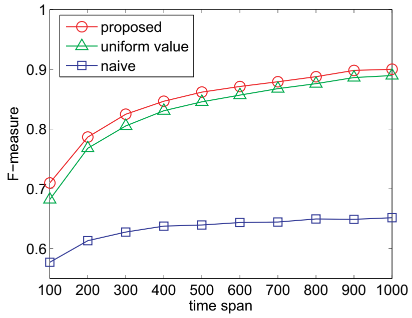


(a) Results for $K = 3$.

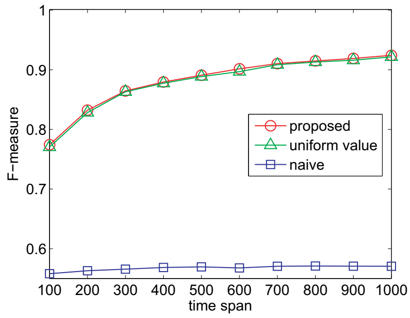


(b) Results for $K = 10$.

Fig. 9: Estimation errors of anti-majoritarian tendency for the Blog network.



(a) Results for $K = 3$.



(b) Results for $K = 10$.

Fig. 10: Accuracies of extracting nodes with high anti-majoritarian tendency for the Blog network.

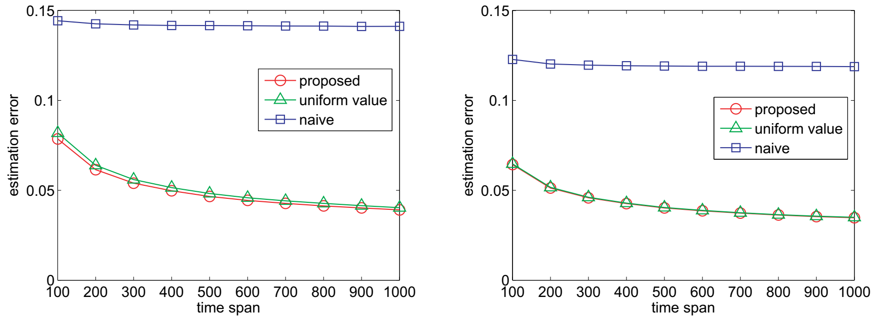
Table 5: Results for estimation errors of anti-majoritarian tendency for the Blog network ($T_0 = 1000$). Note that the two-side 0.05 point of the t-distribution with 9 degrees of freedom is $t^*_{9,0.05} = 2.262$.

Method	$K = 3$ Average of error $\mathcal{E}_\alpha$	$K = 3$ t-value $\mathcal{T}_\alpha^{PC}$	$K = 10$ Average of error $\mathcal{E}_\alpha$	$K = 10$ t-value $\mathcal{T}_\alpha^{PC}$
proposed	0.0229	—	0.0169	—
uniform value	0.0280	5.6016	0.0186	5.0033
naive	0.1403	229.4537	0.1607	577.3649

In order to investigate whether the difference between the proposed method and each of the other methods is statistically significant or not, we in particular performed a t-test for estimation error $\mathcal{E}_\alpha$. Let $\mathcal{E}_\alpha^P$ and $\mathcal{E}_\alpha^C$ denote the values of $\mathcal{E}_\alpha$ for the proposed method and a compared method, respectively. We calculated t-value

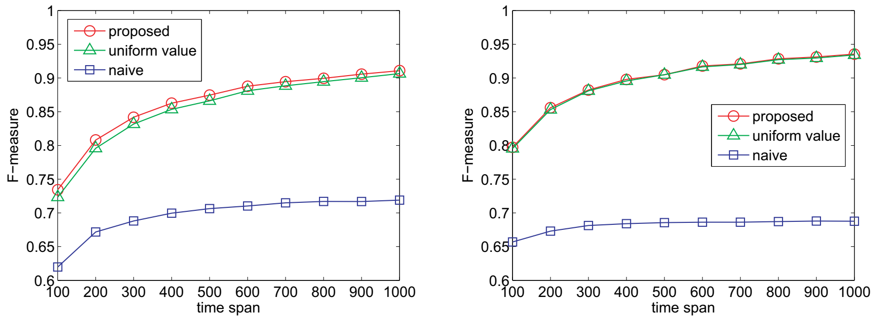
$$\mathcal{T}_\alpha^{PC} = \frac{\sqrt{10}\,\mathrm{mean}\left(\mathcal{E}_\alpha^P - \mathcal{E}_\alpha^C\right)}{\mathrm{std}\left(\mathcal{E}_\alpha^P - \mathcal{E}_\alpha^C\right)},$$

where mean(x) and std(x) are defined in the previous section. Tables 5, 6, 7, and 8 show the results for estimation errors of anti-majoritarian tendency in the case of $T_0 = 1000$ for the



(a) Results for $K = 3$. (b) Results for $K = 10$.

Fig. 11: Estimation errors of anti-majoritarian tendency for the Coauthor network.



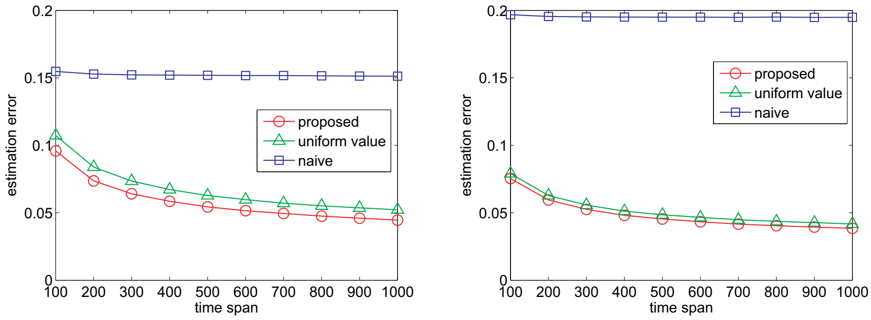
(a) Results for $K = 3$. (b) Results for $K = 10$.

Fig. 12: Accuracies of extracting nodes with high anti-majoritarian tendency for the Coauthor network.

Table 6: Results for estimation errors of anti-majoritarian tendency for the Coauthor network ($T_0 = 1000$). Note that the two-side 0.05 point of the t-distribution with 9 degrees of freedom is $t^*_{9,0.05} = 2.262$.

Method	$K = 3$ Average of error $\mathcal{E}_\alpha$	$K = 3$ t-value $\mathcal{T}^{PC}_\alpha$	$K = 10$ Average of error $\mathcal{E}_\alpha$	$K = 10$ t-value $\mathcal{T}^{PC}_\alpha$
proposed	0.0195	—	0.0147	—
uniform value	0.0208	4.0840	0.0150	9.9920
naive	0.1350	404.8052	0.1074	526.8500

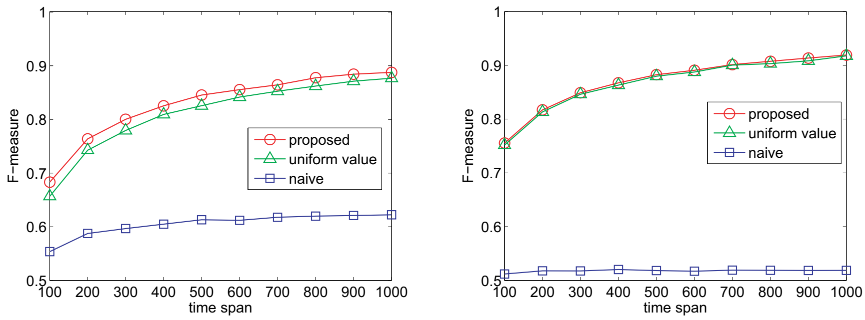
Blog, the Coauthor, the Enron, and the Wikipedia networks, respectively. Here, the second and the fourth columns indicate the average of $\mathcal{E}_\alpha$ over the 10 trials for the cases of $K = 3$ and $K = 10$, respectively. Also, the third and the fifth columns indicate t-value $\mathcal{T}^{PC}_\alpha$ for the cases of $K = 3$ and $K = 10$, respectively. Note that the two-side 0.05 point of the t-distribution with 9 degrees of freedom is $t^*_{9,0.05} = 2.262$. Thus, from Tables 5, 6, 7, and 8, we see that in the case of $T_0 = 1000$, the difference between the proposed and each comparison methods in prediction error $\mathcal{E}_\alpha$ is statistically significant by the t-test at significance level 0.05. Note that we only showed the results for $T_0 = 1000$, because we obtained quite similar results for other values of $T_0 \geq 100$. As explained in Subsection 6.2, $T_0 = 10$ is too short for learning anti-majoritarians.



(a) Results for $K = 3$.

(b) Results for $K = 10$.

Fig. 13: Estimation errors of anti-majoritarian tendency for the Enron network.



(a) Results for $K = 3$.

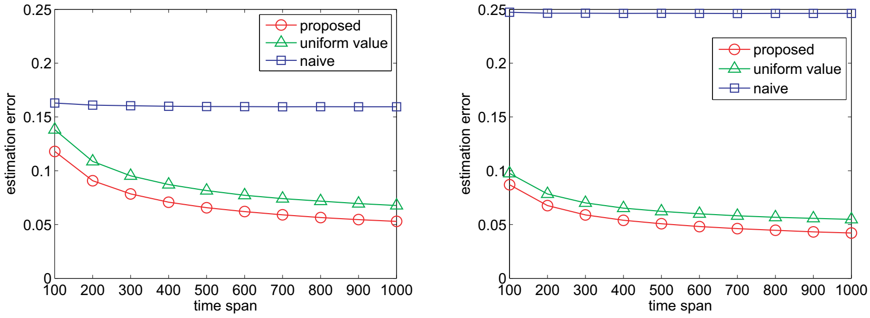
(b) Results for $K = 10$.

Fig. 14: Accuracies of extracting nodes with high anti-majoritarian tendency for the Enron network.

Table 7: Results for estimation errors of anti-majoritarian tendency for the Enron network ($T_0 = 1000$). Note that the two-side 0.05 point of the t-distribution with 9 degrees of freedom is $t^*_{9,0.05} = 2.262$.

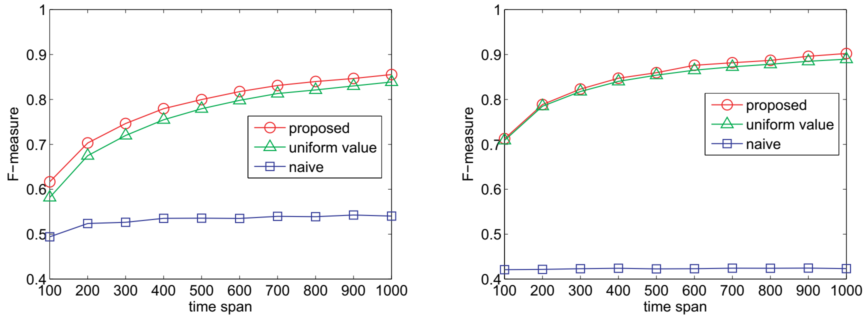
Method	$K = 3$ Average of error $\mathcal{E}_\alpha$	$K = 3$ t-value $\mathcal{T}^{PC}_\alpha$	$K = 10$ Average of error $\mathcal{E}_\alpha$	$K = 10$ t-value $\mathcal{T}^{PC}_\alpha$
proposed	0.0254	—	0.0186	—
uniform value	0.0331	3.8280	0.0220	8.5671
naive	0.1453	101.3125	0.1863	306.6563

As expected, $\mathcal{E}_\alpha$ decreases, and $\mathcal{F}_A$ increases as T_0 increases (*i.e.*, the amount of training data $\mathcal{D}_{T_0}$ 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. Here, we note that quite similar results were also observed for $\mathcal{F}_N$, *i.e.*, extracting nodes with low anti-majoritarian tendency although those results are not reported in this paper. The proposed method can detect both the anti-majoritarians and the majoritarians 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 because of the explicit use of the model, 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, Coauthor and Enron networks, it does not so in $\mathcal{F}_A$ for the Wikipedia



(a) Results for $K = 3$. (b) Results for $K = 10$.

Fig. 15: Estimation errors of anti-majoritarian tendency for the Wikipedia network.



(a) Results for $K = 3$. (b) Results for $K = 10$.

Fig. 16: Accuracies of extracting nodes with high anti-majoritarian tendency for the Wikipedia network.

Table 8: Results for estimation errors of anti-majoritarian tendency for the Wikipedia network ($T_0 = 1000$). Note that the two-side 0.05 point of the t-distribution with 9 degrees of freedom is $t^*_{9,0.05} = 2.262$.

Method	$K = 3$ Average of error $\mathcal{E}_\alpha$	$K = 3$ t-value $\mathcal{T}^{PC}_\alpha$	$K = 10$ Average of error $\mathcal{E}_\alpha$	$K = 10$ t-value $\mathcal{T}^{PC}_\alpha$
proposed	0.0336	—	0.0224	—
uniform value	0.0489	3.2202	0.0360	9.0308
naive	0.1550	51.3607	0.2409	392.8008

network. These results clearly demonstrate the advantage of the proposed method, and it does not seem feasible to detect even roughly the high 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 $T_0 = 1000$ for all cases. Moreover, we note that the processing times of the proposed method at $T_0 = 1000$ for $K = 3$ and $K = 10$ were less than 3 min. and 4 min., respectively. All our experiments were executed on a single PC with an Intel Core 2 Duo 3GHz processor, with 2GB of memory, running under Linux.

7 Conclusion

Unlike the popular probabilistic model such as Independent Cascade and Linear Threshold models for information diffusion where the node in the network takes only one of the two states (active or inactive), applications such as on-line competitive service in which a user can choose one from multiple choices and opinion formation in which a person listens to his/her neighbors" different opinions and decides whether to change his/her opinion require a model that can handle multiple states.

We extended a voter model, a model of opinion formation dynamics where the basic assumption adopted is that people change their opinions following their neighbors' majority opinion, and proposed a new opinion formation model called Value-weighted Mixture Voter (VwMV) Model to analyze how the multiple opinions spread over a large social network and predict future opinion share. The model has two new features. One is that each opinion can have a value, a measure of opinion's importance, and the other is that each node can have an anti-majoritarian tendency, a measure of deviation from the ordinary behavior. In particular, the latter reflects the fact that there are always people who do not agree with the majority and support the minority opinion. Both are parameters in the model, and their values are not known in general.

Our goal was to 1) learn the parameters from a limited amount of observed opinion propagation data and predict the opinion share in the near future, 2) identify the anti-majoritarians from the learned results, and 3) analyze asymptotic behavior of average opinion dynamics to uncover its intrinsic characteristics.

For the first and the second goals we showed that these parameters are learnable from a sequence of observed opinion data by iteratively maximizing the likelihood function. We further showed that it is enough to learn the opinion values and the average anti-majoritarian tendency in good accuracy if the target is to predict the future opinion share, which can be done easily using a limited amount of observed data, but identifying the anti-majoritarians in good accuracy requires much longer observation data because the anti-majoritarian tendency of each node has to be learned. The learning algorithm is guaranteed to find the global optimal solution when there are no anti-majoritarians but may be trapped to a local optimal solution when there are anti-majoritarians. However, the numerical experiment shows that the algorithm converges to a global optimal if there is enough amount of data. We emphasize that use of the learned model can predict the future opinion share much more accurately than a simple polynomial extrapolation can do, and a model ignoring these parameters (opinion values and the anti-majoritarian tendencies) substantially degrades the performance of share prediction. We tried to find a simpler way to estimate the anti-majoritarian tendency of each node, but there seems to be no way. The heuristic that simply counts the number of opinion updates in which the chosen opinion is the same as the minority opinion gives only a very poor approximation. Thus, it is important to explicitly model the anti-majoritarian tendency to predict the correct future opinion share. For the third goal we applied the mean field theory and uncovered the following features. In a situation where the local opinion share can be approximated by the average opinion share, 1) when there are no anti-majoritarians, the opinion with the highest value eventually takes over, but 2) when there is a certain fraction of anti-majoritarians, it is not necessarily the case that the opinion with the highest value prevails and wins, and further, 3) in both cases, when the opinion values are uniform, the opinion share prediction problem becomes ill-defined and any opinion can win. Although the mean field approximation does not hold in real networks, the simulation that uses the real world network structure supports that this holds for real world social networks that we used in this study. We believe that these findings are useful in deepening our understanding the behavior of opinion dynamics.

References

Agarwal N, Liu H (2008) Blogosphere: Research issues, tools, and applications. SIGKDD Explorations 10:18–31

Arenson AJ (1996) Rejection of the power of judicial review in britain. Deakin Law Review 3:37–53

Bakshy E, Hofman J, Mason W, Watts D (2011) Everyone's an influencer: Quantifying influences on twitter. In: Proceedings of the 4th International Conference on Web Search and Data Mining (WSDM'11), pp 65–74

Castellano C, Munoz MA, Pastor-Satorras R (2009) Nonlinear q-voter model. Physical Review E 80:041129:1–041129:8

Chen W, Wang Y, Yang S (2009) 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

Chen W, Yuan Y, Zhang L (2010) 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

Crandall D, Cosley D, Huttenlocner D, Kleinberg J, Suri S (2008) 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

Domingos P (2005) Mining social networks for viral marketing. IEEE Intelligent Systems 20:80–82

Domingos P, Richardson M (2001) Mining the network value of customers. In: Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'01), pp 57–66

Donnelly P, Welsh D (1984) The antivoter problem: random 2-colourings of graphs. Graph Theory and Combinatorics pp 133–144

Even-Dar E, Shapira A (2007) 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

Gill J, Gainous J (2002) Why does voting get so complicated? a review of theories for analyzing democratic participation. Statistical Science 17:383–404

Gruhl D, Guha R, Liben-Nowell D, Tomkins A (2004) Information diffusion through blogspace. SIGKDD Explorations 6:43–52

Holme P, Newman MEJ (2006) Nonequilibrium phase transition in the coevolution of networks and opinions. Physical Review E 74:056108:1–056108:5

Huber M, Reinert G (2004) The stationary distribution in the antivoter model: exact sampling and approximations. In: Stein's Method: Expository Lectures and Applications, IMS Lecture Notes, vol 46, pp 75–92

Kempe D, Kleinberg J, Tardos E (2003) 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

Kimura M, Saito K, Motoda H (2009) Blocking links to minimize contamination spread in a social network. ACM Transactions on Knowledge Discovery from Data 3:9:1–9:23

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

Kimura M, Saito K, Ohara K, Motoda H (2010b) Learning to predict opinion share in social networks. In: Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI'10), pp 1364–1370

Kimura M, Saito K, Ohara K, Motoda H (2011) Detecting anti-majority opinionists using value-weighted mixture voter model. In: Proceedings of the 14th International Conference on Discovery Science (DS'11), LNAI 6926, pp 150–164

Klimt B, Yang Y (2004) 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

Leskovec J, Adamic LA, Huberman BA (2007a) The dynamics of viral marketing. ACM Transactions on the Web 1:5:1–5:39

Leskovec J, Krause A, Guestrin C, Faloutsos C, VanBriesen J, Glance N (2007b) 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

Liggett TM (1999) Stochastic interacting systems: contact, voter, and exclusion processes. Spriger, New York

Mathioudakis M, Bonch F, Castillo C, Gionis A, Ukkonen A (2011) Sparsification of influence networks. In: Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'11), pp 529–537

Matloff N (1977) Ergodicity conditions for a dissonant voting model. Annals of Probability 5:371–386

Newman MEJ (2003) The structure and function of complex networks. SIAM Review 45:167–256

Newman MEJ, Forrest S, Balthrop J (2002) Email networks and the spread of computer viruses. Physical Review E 66:035101:1–035101:4

Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435:814–818

Richardson M, Domingos P (2002) Mining knowledge-sharing sites for viral marketing. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'02), pp 61–70

Röllin A (2007) Translated poisson approximation using exchangeable pair couplings. Annals of Applied Probablity 17:1596–1614

Romero D, Meeder B, Kleinberg J (2011) Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In: Proceedings of the 20th International World Wide Web Conference (WWW'11), pp 695–704

Sood V, Redner S (2005) Voter model on heterogeneous graphs. Physical Review Letters 94:178701:1–178701:4

Wu F, Huberman BA (2008) How public opinion forms. In: Proceedings of the 4th International Workshop on Internet and Network Economics (WINE'08), pp 334–341

Yang H, Wu Z, Zhou C, Zhou T, Wang B (2009) Effects of social diversity on the emergence of global consensus in opinion dynamics. Physical Review E 80:046108:1–046108:5