Physica A 449 (2016) 85–100

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Controllable uncertain opinion diffusion under confidence bound and unpredicted diffusion probability

Fuhan Yan, Zhaofeng Li, Yichuan Jiang *

School of Computer Science and Engineering, Southeast University, Jiangning District, Nanjing 211189, China

HIGHLIGHTS

- The diffusion probability is modified according to the distance between the opinions of interacted nodes.
- An optimization problem and suitable algorithm to show how to control the uncertain diffusion process.
- The performances of traditional influence maximization algorithms are often worse than that of random selection in this optimization problem.
- Reason of that the traditional influence maximization algorithms are not applicable to this problem.

ARTICLE INFO

Article history:
Received 30 July 2015
Received in revised form 17 October 2015
Available online 6 January 2016

Keywords:
Diffusion process
Opinion
Control
Strategy

ABSTRACT

The issues of modeling and analyzing diffusion in social networks have been extensively studied in the last few decades. Recently, many studies focus on uncertain diffusion process. The uncertainty of diffusion process means that the diffusion probability is unpredicted because of some complex factors. For instance, the variety of individuals' opinions is an important factor that can cause uncertainty of diffusion probability. In detail, the difference between opinions can influence the diffusion probability, and then the evolution of opinions will cause the uncertainty of diffusion probability. It is known that controlling the diffusion process is important in the context of viral marketing and political propaganda. However, previous methods are hardly feasible to control the uncertain diffusion process of individual opinion. In this paper, we present suitable strategy to control this diffusion process based on the approximate estimation of the uncertain factors. We formulate a model in which the diffusion probability is influenced by the distance between opinions, and briefly discuss the properties of the diffusion model. Then, we present an optimization problem at the background of voting to show how to control this uncertain diffusion process. In detail, it is assumed that each individual can choose one of the two candidates or abstention based on his/her opinion. Then, we present strategy to set suitable initiators and their opinions so that the advantage of one candidate will be maximized at the end of diffusion. The results show that traditional influence maximization algorithms are not applicable to this problem, and our algorithm can achieve expected performance.

1. Introduction

Previous studies have modeled and analyzed various diffusion processes in social networks [1–6]. Generally, each node in social network is assumed to associate with the binary states-active or inactive. The state of a node may be switched

* Corresponding author.
E-mail address: yjiang@seu.edu.cn (Y. Jiang).

http://dx.doi.org/10.1016/j.physa.2015.12.110

based on some rules in the diffusion process. Among the models with different rules, the independent cascade model [3,7] is one of the models that are used the most widely in diffusion analyses. In the independent cascade model, the state of a node is switched based on a probability, and each activation process is independent of others. In various cases, the diffusion probability may be constant [7] or variable [8,9].

Recently, more and more studies have focused on the uncertain diffusion process. The uncertainty of diffusion process means that the diffusion probability is unpredicted because of some complex factors. Previous study has shown that the variety of individuals' opinions is an important factor that can cause uncertainty of diffusion probability [5]. In detail, when someone receives information from others who hold similar opinions with him/her, he/she will be more likely to perform the dissemination behavior [10–12]. Besides, it has been known that opinions can evolve as individuals are reported to assimilate the viewpoints of interacted people [10,13]. Therefore, the diffusion probability becomes uncertain because of the opinion evolution.

Individuals express their opinions in online social networks frequently, such as the tastes in music or movie [10,13], the comments about innovations of product [14,15], and attitudes to the advertisements of candidate [16]. Therefore, these may be various diffusion processes of these opinions in social networks. It means that controlling these uncertain opinion diffusion processes is very important in the context of viral marketing [17] and political propaganda [16]. However, previous studies of opinion [13,18–20] just focus on how opinions evolve to the consensus or the polarization states. These studies usually discuss the influence of network structure, nodal degree or the evolution rules of opinion. As these factors are almost uncontrollable, it is costly or hardly feasible to control the diffusion process based on the conclusions obtained in these studies.

In this paper, in order to control the uncertain opinion diffusion, we discuss this diffusion process and develop feasible strategy to control this diffusion process based on the approximate estimation of the uncertain factors.

1.1. Motivation and contributions

In this paper, we raise the issue of controlling the uncertain opinion diffusion process. In order to show how to control the diffusion process, we present a diffusion model in which the diffusion probability is modified according to the distance between the opinions of interacted nodes. Moreover, each node which is exposed to the information can adjust its opinion depending on the assimilation of others' opinions.

In detail, our contributions can be divided into two aspects:

- First, based on theoretical analyses and simulations, we briefly discuss the properties of the diffusion model in which the diffusion probability is influenced by the distance between opinions. Some conclusions are obtained and helpful in developing strategy to control the diffusion process.
- Second, we present an optimization problem to show how to control the uncertain diffusion process. In detail, we model a voting mechanism in which each node can choose one of the two candidates or abstention based on its opinion. Then, we present suitable strategy to maximize the advantage of one candidate by setting suitable initiators and their opinions in the diffusion process.

This paper is organized as follows. In Section 2, we briefly introduce the related work. Section 3 presents our diffusion model, including the definition of individual opinion, the modification of diffusion probability, and the evolution rule of opinions. In Section 4, we show the properties of the diffusion model, such as how wide the initiator's opinion can spread under different initial distributions of opinions. Then, an optimization problem is provided in Section 5. Finally, we conclude this paper in Section 6.

2. Related works

2.1. Uncertain diffusion process

Many studies have focused on the diffusion process in which the diffusion probability is uncertain. For instance, structural uncertainty [3,21–23], input of multiple entities [8,9,24,25] and the variety of individuals' opinions [5] are three factors that can cause the uncertainty of diffusion probability in previous studies. Adiga et al. [3] raise the issue of diffusion process interfered by structural uncertainty. The structural uncertainty is generally modeled by rewiring edges in the networks [3,21]. Additionally, Adiga et al. [3] mathematically investigate the sensitivity of diffusion process to the structural uncertainty. They also predicted the scales of information spreading affected by the different extents of perturbation. On the other hand, the spreading of multiple entities in social networks is another interference factor to diffusion probability. For instance, two competing diseases are introduced, and being infected by one disease gives a node partial or complete immunity to another [24]. By analyzing the empirical data of multiple contagions in Twitter, Myers and Leskovec reported the great effects of interaction between cooperating and competing information on spreading probabilities [25]. In addition, the variety of individuals' opinions is also an important factor that can cause uncertainty of diffusion probability [5]. The distance between the opinions of interacted nodes can influence the diffusion probability, and the evolution of opinions makes the diffusion probability become uncertain. However, this study [5] just presents a diffusion model and analyzes its properties. In this paper, we further present an optimization problem to show how to control the uncertain opinion diffusion process.

2.2. The issue of opinion

The issue of opinion has been studied from many disciplines [13,18–20]. Three models are used widely to analyze the opinion formation. They are averaging model [18], bounded confidence model [19] and voter model [26]. In the averaging model, the nodes update their opinions by averaging their opinions with the mean of their neighboring opinions. Besides, in many studies based on this model, different weights are attached to opinions [18]. Therefore, the result of opinion evolution heavily depends on these weights. In the bounded confidence model, each node has a confidence bound. If the distance between a node's opinion and another opinion is larger than the confidence bound, the node will not care this opinion at all [19]. In the voter model, at each time step, the nodes randomly choose a neighbor and adopt its opinion [26]. The opinions in the voter model are usually represented by a binary value, and the condition that all nodes adopt the same opinion is mainly analyzed. In addition, the opinion evolution in this paper is based on the bounded confidence model.

Although the models of opinion formation are various, the previous studies always focus on how opinions evolve to the consensus or the polarization states [13,18–20,26,27]. For instance, Lorenz [19] indicates that polarization state may occur if individuals just assimilate the opinions to which they agree. On the contrary, the assimilation of all opinions usually leads to consensus state [13]. On the other hand, some studies analyze the co-evolution of opinion and network structure [26,27]. Additionally, Iyengar, Van den Bulte, and Valente [17] discussed the fundamental role of opinion leader in the diffusion of new product. However, these studies focus on the properties of the opinion evolution rather than a feasible method to control the opinion formation.

Some studies also care the feasible methods to control the opinion formation [28,29]. Afshar and Asadpour [28] discuss the influence of informed agents in opinion evolution process. The informed agent in their study is an agent that gradually and intentionally changes its neighbors' opinions towards the desired direction. Besides, the major agents are not informed agents and do not know who is informed agent. Based on various simulations, they analyze how the number of informed agents, the nodal degree and network structure influence the opinion evolution. AskariSichani and Jalili [29] further discuss how to maximize the influence of informed agents by connecting them to suitable node. In detail, the informed agents should be connected to the nodes with small in-degrees and high out-degree that are connected to high in-degree nodes. Then, the public opinion will be significantly changed towards the desired direction. The main advantage of this method is that the public opinion is very close to desired opinion after the evolution process. However, this method is based on a long process in which nodes repeatedly interact with their neighbors. If the process lasts a lot of time, there may be some noisy factors in this process, such as the change of network structure [3]. Therefore, we present a method based on a diffusion process that lasts a short time so that many noisy factors can be ignored.

2.3. The issue of controlling diffusion process

Controlling the diffusion process is an important motivation of diffusion analyses. Influence maximization problem is the main aspect of controlling the diffusion process [7,14,30–33]. Kempe, Kleinberg and Tardos [7] raise the issue of influence maximization through a social network. The problem is to find a set of nodes to be active initially such that the expected number of activated nodes after the diffusion terminates is maximized. Chen et al. [14] discuss the diffusion process in which opposite information may emerge and propagate, and they present algorithm to maximize the spread of the positive influence. Chen, Lu and Zhang [30] present a study that care the influence maximization problem in a time-critical situation. Their algorithm outperforms previous algorithms that disregard the deadline constraint and delays in diffusion. Besides, Feng et al. [31] argue that past studies have not considered the impact of novelty decay on influence propagation. In their study, repeated exposures will have diminishing influence on users. However, in this paper, the traditional influence maximization algorithms are proved inapplicable to the control of the uncertain opinion diffusion.

3. Model statements

In this section, we describe our model by combining the independent cascade model [3,7] and the evolution of opinions.

3.1. Definition of individual opinion

Let $O(i, t) \in [-1, 1]$ denote node i's opinion at time t [18,19]. Here, $O(i, t) = -1$ and $O(i, t) = 1$ denote opposite extreme opinions. Moreover, let $O_0 \in [-1, 1]$ denote the opinion that is contained in the information. In other words, O_0 is the opinion of the initiator who triggers the diffusion process. In this paper, O_0 cannot be affected by others' opinions, and it is the initiator's opinion at any time.

Table 1
Main symbols and their meanings.

Symbol	Meaning
O_0	The opinion that is contained in the information (the initiator's opinion).
ε_i	The confidence bound of node i.
$f_O(x)$	The probability density function of nodes' initial opinions.
$f_t(x)$	The probability density function of the opinions of the nodes which try to activate other nodes at time t.
$\overline{p}_t$	The mean activation probability at time t.
r_a	Active rate: the rate of the active nodes in social network.
r_c	Coverage rate: the rate of the nodes that are exposed to the information.
$\|N\|$	The number of the nodes in network.

3.2. Diffusion model with individual opinion

Let $G = (N, E)$ be a social network and every node in G is considered as an individual. Let $X(i, t)$ denote the state of node i at discrete time t. Then, we have

$$X(i,t) = \begin{cases} 0, & \text{inactive;} \\ 1, & \text{active.} \end{cases} \tag{1}$$

Here, $X(i, t) = 0$ means that node i has not received the information or does not perform dissemination behavior after receiving the information. Correspondingly, $X(i, t) = 1$ means that node i has diffused the information.

According to the independent cascade model, node can try to activate its neighbors only once after it becomes active. Besides, in this paper, each active node diffuses not only the information, but also its opinion on the information.

Then, we define the influence of opinions on diffusion probability. Node i activates node j with the probability:

$$p_{i,j} = \begin{cases} p_0 \mathrm{e}^{-|O_0 - O(j,t)|}, & \text{node } i \text{ is the initiator;} \\ p_0 \mathrm{e}^{-\frac{1}{2}|(O(i,t)-O(j,t)|-\frac{1}{2}|O_0-O(j,t)|)}, & \text{otherwise.} \end{cases} \tag{2}$$

Here, p_0 is a basic probability, and $p_{i,j}$ is the modifications of p_0 [34].

Each node may change its opinion if it is exposed to the information. Let A_i denote the set of opinions which can influence $O(i, t)$, and A_i is ϕ at the beginning of each time step. If node i's neighbor node j is active and $|O(i, t) - O(j, t)| < \varepsilon_i$, $O(j, t)$ will be added to A_i. Besides, if $|O_0 - O(i, t)| < \varepsilon_i$, O_0 will also be added to A_i. Then, we have

$$O(i, t+1) = \frac{O(i,t) + \sum\limits_{O(j,t)\in A_i} O(j,t)}{1 + |A_i|}. \tag{3}$$

Here, Eq. (3) is based on the bounded confidence model, and ε_i is a confidence bound [19]. Besides, each opinion can only influence $O(i, t)$ no more than once. In other words, although a node may receive the information several times, the opinion O_0 can influence the node only once. Moreover, if node j's opinion changes after it influences $O(i, t)$, it will not influence $O(i, t)$ again.

In addition, to help clearly describe the proposed theoretical analysis framework, Table 1 provides the main symbols used in this paper and their specific meanings.

As shown in Fig. 1, each node except the initiator has four possible states. State 1 is the initial state. The possible state transitions are:

- A node in State 1 is activated at this time step. Its active neighbors' opinions and O_0 are not in its confidence interval. The node will be in State 3 at next time step.
- A node in State 1 is not activated at this time step. Its active neighbors' opinions or O_0 are in its confidence interval. The node will be in State 2 at next time step.
- A node in State 1 is activated at this time step. Its active neighbors' opinions or O_0 are in its confidence interval. The node will be in State 4 at next time step.
- A node in State 2 receives the information again and is activated at this time step. The node will be in state 4 at next time step.
- A node in State 3 receives the information again, and its active neighbors' opinions or O_0 are in its confidence interval. The node will be in State 4 at next time step.

4. Properties of the diffusion model

In this section, we briefly discuss the properties of the diffusion model based on theoretical analyses and simulations. The simulations are in two empirical networks (Enron Email network [35,36] and Facebook network [37]). In the Email network, there are 36 692 nodes while the mean degree is 10.26. Besides, there are 63 731 nodes in the Facebook network, and the mean degree is 25.3.

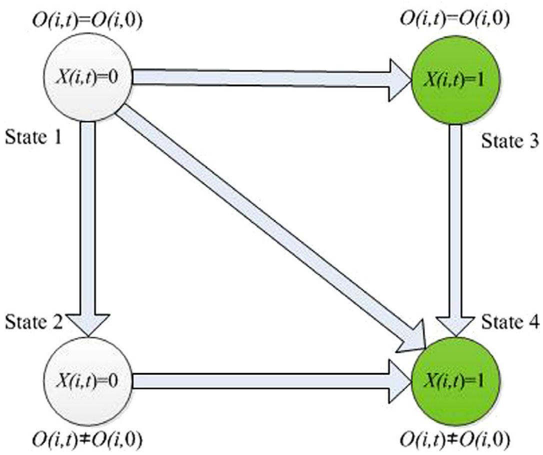


Fig. 1. The state transition of each node is shown in this figure. The green state means activation. $O(i, t) \neq O(i, 0)$ means that the node's opinion has been influenced by others' opinions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

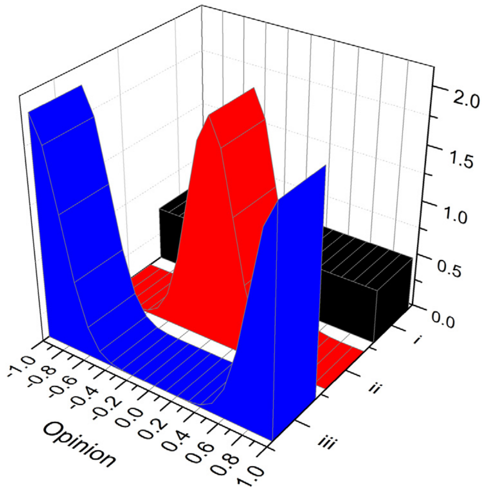


Fig. 2. The illustration of the three probability density functions: (i) $f_i(x)$; (ii) $f_{ii}(x)$; (iii) $f_{iii}(x)$.

In the simulations, the basic probability $p_0 = 0.1 * e = 0.2718$ so that $p_{i,j} \in [0.0368, 0.2718]$. Based on previous study [38], it can be known that the thresholds that limit the spread of information in the two empirical networks are between 0.0368 and 0.2718. Therefore, when $p_0 = 0.2718$, the initiator's opinion can usually widely spread but not always activate the most nodes. Then, the simulated phenomena will be more obvious. We will not test different basic probability because various initial distributions of opinions and confidence bounds can cause different diffusion probabilities. Each trial is performed with 1000 replications. The node with the largest degree will be set as the unique initiator. For the sake of simplification, it is assumed that all the nodes have the same confidence bound ε_i.

Three probability density functions [39] are used to represent the typical states [18,19] of the initial opinions. In simulations, the initial opinion of each node will be generated randomly based on the probability density functions. The three probability density functions are shown in the following and Fig. 2:

1. $f_i(x) = 0.5, x \in [-1, 1]$. This state indicates the absence of convergent opinions in social networks. For instance, individuals hold scattered opinions on some fuzzy events because nobody knows the truth.
2. $f_{ii}(x) = \frac{5}{\sqrt{2\pi}} e^{-12.5x^2}, x \in [-1, 1]$. This state represents that most social individuals hold neutral opinion. For instance, most people hold neutral opinions on some insipid topic, and they have no obvious preference.
3. $f_{iii}(x) = \begin{cases} \frac{5}{\sqrt{2\pi}} e^{-12.5(x+1)^2}, & x < 0; \\ \frac{5}{\sqrt{2\pi}} e^{-12.5(x-1)^2}, & x \geq 0. \end{cases}, x \in [-1, 1]$. Opposite to $f_{ii}(x)$, this probability density function indicates the polarized opinions which means nodes in social networks are generally distributed in two camps. For instance, individuals hold polarized opinions on some controversial issues.

Then, we discuss which value of O_0 will lead to the largest active rate.

Theorem 1. *Let $g(O_0) = r_a$. Besides, let μ denote the mathematical expectation of $f_O(x)$. It is assumed that there exist $\delta \in [-1, 1]$ and $\omega \geq 0$ satisfy the condition: $\int_y^1 \mathrm{e}^{-x+y} f_O(x)\mathrm{d}x - \int_{-1}^y \mathrm{e}^{x-y} f_O(x)\mathrm{d}x > 0$ when $y \in [\delta-\omega, \delta)$ and $\int_y^1 \mathrm{e}^{-x+y} f_O(x)\mathrm{d}x - \int_{-1}^y \mathrm{e}^{x-y} f_O(x)\mathrm{d}x < 0$ when $y \in (\delta, \delta+\omega]$. If $\varepsilon_i > 0$ and $\mu \in [\delta-\omega, \delta+\omega]$, there exists a value $\zeta \in [\min(\mu, \delta), \max(\mu, \delta)]$ so that $g(\zeta)$ is the maximum of $g(O_0)$.*

Proof. Let

$$f(y) = \int_{-1}^{1} p_0 \mathrm{e}^{-|x-y|} f_O(x)\mathrm{d}x.$$

Therefore,

$$\begin{aligned} f(y) &= \int_{-1}^{y} p_0 \mathrm{e}^{x-y} f_O(x)\mathrm{d}x + \int_{y}^{1} p_0 \mathrm{e}^{-x+y} f_O(x)\mathrm{d}x \\ &= p_0 \left[\mathrm{e}^{-y} \int_{-1}^{y} \mathrm{e}^{x} f_O(x)\mathrm{d}x + \mathrm{e}^{y} \int_{y}^{1} \mathrm{e}^{-x} f_O(x)\mathrm{d}x \right]. \end{aligned}$$

For $\forall y_0 \in [-1, 1]$,

$$\begin{aligned} \lim_{\Delta y \to 0} \frac{f(y_0+\Delta y) - f(y_0)}{\Delta y} &= \lim_{\Delta y \to 0} \left\{ \frac{p_0 \left[\mathrm{e}^{-y_0-\Delta y} \int_{-1}^{y_0+\Delta y} \mathrm{e}^{x} f_O(x)\mathrm{d}x + \mathrm{e}^{y_0+\Delta y} \int_{y_0+\Delta y}^{1} \mathrm{e}^{-x} f_O(x)\mathrm{d}x \right]}{\Delta y} \right. \\ &\quad \left. - \frac{p_0 \left[\mathrm{e}^{-y_0} \int_{-1}^{y_0} \mathrm{e}^{x} f_O(x)\mathrm{d}x + \mathrm{e}^{y_0} \int_{y_0}^{1} \mathrm{e}^{-x} f_O(x)\mathrm{d}x \right]}{\Delta y} \right\} \\ &= \lim_{\Delta y \to 0} \left\{ \frac{p_0 \mathrm{e}^{-y_0} \left[(\mathrm{e}^{\Delta y} - 1) \int_{-1}^{y_0} \mathrm{e}^{x} f_O(x)\mathrm{d}x + \mathrm{e}^{-\Delta y} \int_{y_0}^{y_0+\Delta y} \mathrm{e}^{x} f_O(x)\mathrm{d}x \right]}{\Delta y} \right. \\ &\quad \left. - \frac{p_0 \mathrm{e}^{y_0} \left[(\mathrm{e}^{\Delta y} - 1) \int_{y_0}^{1} \mathrm{e}^{-x} f_O(x)\mathrm{d}x - \mathrm{e}^{\Delta y} \int_{y_0}^{y_0+\Delta y} \mathrm{e}^{x} f_O(x)\mathrm{d}x \right]}{\Delta y} \right\} \\ &= p_0 \mathrm{e}^{-y_0} \left[\int_{-1}^{y_0} \mathrm{e}^{x} f_O(x)\mathrm{d}x \lim_{\Delta y \to 0} \frac{\mathrm{e}^{-\Delta y} - 1}{\Delta y} + \lim_{\Delta y \to 0} \frac{\mathrm{e}^{-\Delta y} \int_{y_0}^{y_0+\Delta y} \mathrm{e}^{x} f_O(x)\mathrm{d}x}{\Delta y} \right] \\ &\quad + p_0 \mathrm{e}^{y_0} \left[\int_{y_0}^{1} \mathrm{e}^{-x} f_O(x)\mathrm{d}x \lim_{\Delta y \to 0} \frac{\mathrm{e}^{\Delta y} - 1}{\Delta y} - \lim_{\Delta y \to 0} \frac{\mathrm{e}^{\Delta y} \int_{y_0}^{y_0+\Delta y} \mathrm{e}^{x} f_O(x)\mathrm{d}x}{\Delta y} \right]. \end{aligned}$$

Let $z = \mathrm{e}^{-\Delta y} - 1$ so that $\Delta y = -\ln(z+1)$. Therefore,

$$\lim_{\Delta y \to 0} \frac{\mathrm{e}^{-\Delta y} - 1}{\Delta y} = \lim_{\Delta y \to 0} \frac{z}{-\ln(z+1)} = \lim_{z \to 0} \frac{1}{-\ln(z+1)^{\frac{1}{z}}}.$$

It is easy to know that $\ln(z+1)^{\frac{1}{z}} = e$. It means that $\lim_{\Delta y \to 0} \frac{\mathrm{e}^{-\Delta y}-1}{\Delta y} = -1$. Correspondingly, $\lim_{\Delta y \to 0} \frac{\mathrm{e}^{\Delta y}-1}{\Delta y} = 1$. Based on the definition of integral, it can be known that $\int_{y_0}^{y_0+\Delta y} \mathrm{e}^{x} f_O(x)\mathrm{d}x = \lim_{n \to \infty} \frac{\Delta y}{n} \sum_{i=1}^{n} \mathrm{e}^{\eta_i} f_O(\eta_i)$ while $y_0 = \eta_0 < \eta_1 < \eta_2 < \cdots < \eta_n = y_0 + \Delta y$. Then,

$$\begin{aligned} \lim_{\Delta y \to 0} \frac{\mathrm{e}^{-\Delta y} \int_{y_0}^{y_0+\Delta y} \mathrm{e}^{x} f_O(x)\mathrm{d}x}{\Delta y} &= \lim_{\Delta y \to 0} \frac{\mathrm{e}^{-\Delta y} \lim_{n \to \infty} \frac{\Delta y}{n} \sum_{i=1}^{n} \mathrm{e}^{\eta_i} f_O(\eta_i)}{\Delta y} \\ &= \lim_{\Delta y \to 0} \left[\mathrm{e}^{-\Delta y} \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{e}^{\eta_i} f_O(\eta_i) \right]. \end{aligned}$$

Here, $\Delta y \to 0$ means $\eta_i \to y_0$. Therefore, $\lim_{\Delta y \to 0} \frac{\mathrm{e}^{-\Delta y} \int_{y_0}^{y_0+\Delta y} \mathrm{e}^{x} f_O(x)\mathrm{d}x}{\Delta y} = \mathrm{e}^{y_0} f_O(y_0)$. Correspondingly, $\lim_{\Delta y \to 0} \frac{\mathrm{e}^{\Delta y} \int_{y_0}^{y_0+\Delta y} \mathrm{e}^{x} f_O(x)\mathrm{d}x}{\Delta y} = \mathrm{e}^{y_0} f_O(y_0)$.

Therefore, $\lim_{\Delta y\to 0}\frac{f(y_0+\Delta y)-f(y_0)}{\Delta y}$ exists $\forall y_0\in[-1,1]$ so that $f(y)$ satisfies the derivable condition on $[-1, 1]$. Then,

$$f'(y)=p_0\left[\int_y^1 \mathrm{e}^{-x+y}f_O(x)\mathrm{d}x-\int_{-1}^{y}\mathrm{e}^{x-y}f_O(x)\mathrm{d}x\right].$$

Based on the condition in Theorem 1, it can be known that $f'(y)>0$ when $y\in[\delta-\omega,\delta)$ and $f'(y)<0$ when $y\in(\delta,\delta+\omega]$. Therefore, $f(\delta)$ is the local maximum of $f(y)$.

It can be obtained that

$$\overline{p}_{t=1}=\int_{-1}^{1}p_0\mathrm{e}^{-|x-O_0|}f_O(x)\mathrm{d}x.$$

Therefore, $\overline{p}_{t=1}=f(O_0)$. It means that $\overline{p}_{t=1}$ is maximum at $O_0=\delta$.

Besides,

$$\overline{p}_{t>1}=\int_{-1}^{1}f_t(y)\left[\int_{-1}^{1}p_0\mathrm{e}^{-\frac{1}{2}|x-y|-\frac{1}{2}|O_0-x|}f_O(x)\mathrm{d}x\right]\mathrm{d}y.$$

If $\varepsilon_i>0$, nodes' opinions may be influenced by O_0 and their neighbors' opinions. The co-influence of a node's neighbors' opinions leads the node's opinion to change towards the mathematical expectation of its neighbors' opinions. Therefore, nodes' opinions change towards O_0 and μ. It means that $f_t(y)$ focuses on a value between O_0 and μ. If O_0 approaches μ, the co-influence will be larger so that $f_t(y)$ will be more focused. Then, $\overline{p}_{t>1}$ will also be larger. Therefore, along with the opinion evolution, the value of O_0 that leads to the largest $\overline{p}_{t>1}$ will change from δ to μ.

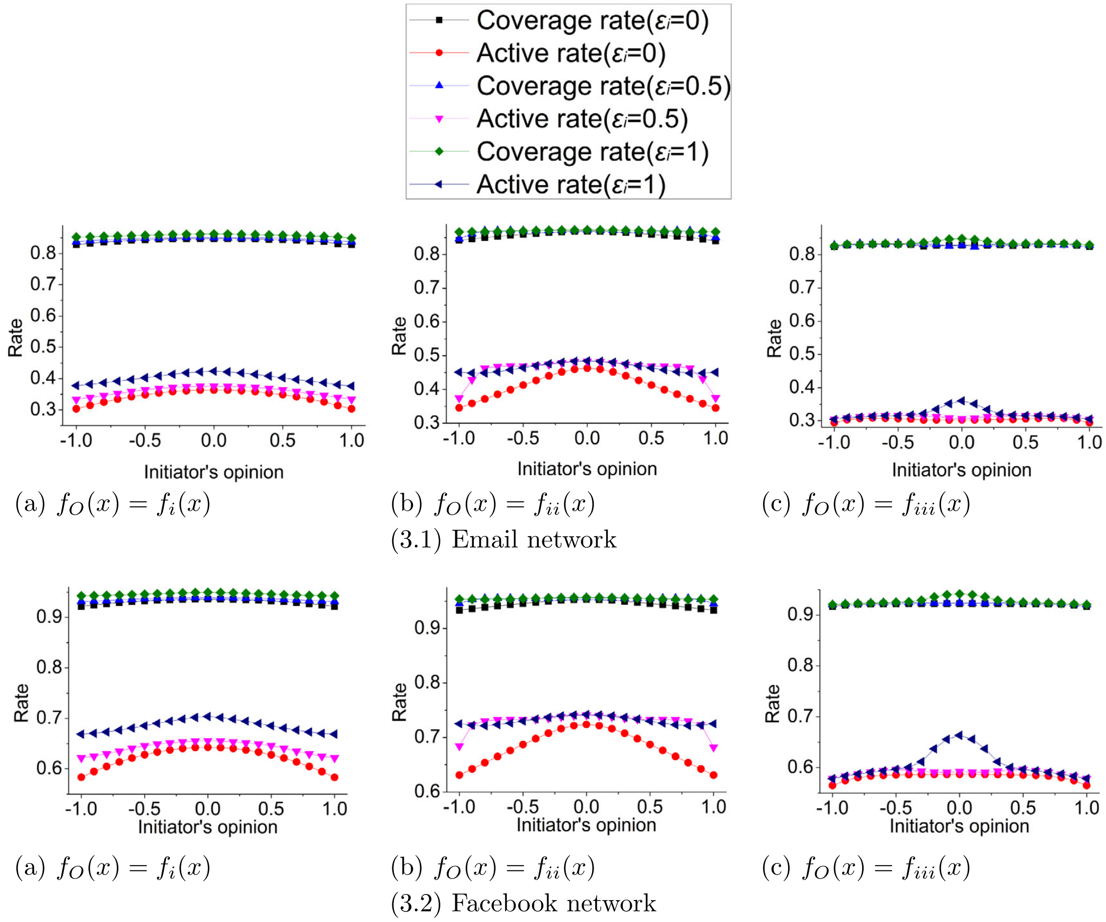
Overall, there exists a value $\zeta\in[\min(\mu,\delta),\max(\mu,\delta)]$ so that $g(\zeta)$ is the maximum of $g(O_0)$. □

Theorem 1 presents the possible interval of O_0 which will lead to the largest active rate. Then, we discuss the simulated data in Fig. 3. Considering the value δ in Theorem 1, based on Matlab, we obtain that if $f_O(x)=f_i(x)$ or $f_{ii}(x)$, $f'(y)>0$ when $y\in[-1,0)$ and $f'(y)<0$ when $y\in(0,1]$. It means that δ exists and $\delta=0$. If $f_O(x)=f_{iii}(x)$, $f'(y)>0$ when $y[-1,-0.869)\cup(0,0.869)$ and $f'(y)<0$ when $y(-0.869,0)\cup(0.869,1]$. Therefore, $\delta\approx 0.869$ or -0.869. Besides, the mathematical expectations of the three probability density functions are zero. Therefore, based on Theorem 1, it can be known that if $f_O(x)=f_i(x)$ or $f_{ii}(x)$, the active rate reaches maximum when $O_0=0$. If $f_O(x)=f_{iii}(x)$, the maximums of active rate are in the intervals $O_0\in[-0.869,0.869]$. These phenomena can be observed in Fig. 3.

There is an interesting phenomenon about the fluctuation of the active rate. In Fig. 3.1(a)(b) and 3.2(a)(b), the fluctuation of the active rate is not acute when $\varepsilon_i=1$. However, in Fig. 3.1(c) and 3.2(c), the fluctuation of the active rate is quite acute when $\varepsilon_i=1$. The reason is that large ε_i makes nodes' opinions focus on O_0 when $f_O(x)$ is not polarized. Therefore, the diffusion probability is usually large when $\varepsilon_i=1$. Then, the fluctuation of the active rate is not acute. On the contrary, if $f_O(x)$ is polarized ($f_O(x)=f_{iii}(x)$), $\varepsilon_i=0.5$ cannot make O_0 influence the two poles simultaneously so that the diffusion probability is quite small. However, when $\varepsilon_i=1$, $O_0=0$ can influence the two poles simultaneously while $O_0=-1$ or 1 cannot do. Therefore, in Fig. 3.1(c) and 3.2(c), the fluctuation of the active rate is more acute when $\varepsilon_i=1$.

In addition, the fluctuation of the active rate is acute if $f_O(x)=f_{ii}(x)$. It means that O_0 can significantly influence the active rate if most nodes hold similar opinions initially. When nodes' initial opinions are polarized, the influence of O_0 will be limited. Moreover, in Fig. 3.1(a)(c) and 3.2(a)(c), larger ε_i leads to larger active rate. The reason is that large ε_i makes each node's opinion be able to influence most other nodes. Then, the opinions will evolve to the consensus state because of the interaction. Similar conclusion has been presented in Refs. [13,19]. With the convergence of opinions, the average diffusion probability will increase. On the contrary, if ε_i is small, the opinions will not necessarily be convergent. The final distribution of opinions heavily depends on the initial opinions. For instance, if $f_O(x)=f_i(x)$, the opinions may become several opinion clusters, and then these opinion clusters cannot influence other clusters. Therefore, the diffusion probability is usually small if ε_i is small. The exception occurs in Fig. 3.1(b) and 3.2(b), r_a with $\varepsilon_i=0.5$ may be larger than r_a with $\varepsilon_i=1$ when O_0 is near 0.75 or -0.75. It is because that the opinions are convergent initially, and these opinions can be considered as an opinion cluster. Larger ε_i will not make the opinions be more convergent. In this case, $O_0=0.75$ or -0.75 can influence the opinion cluster when $\varepsilon_i=1$ but it cannot when $\varepsilon_i=0.5$. The influence of $O_0=0.75$ makes the opinion cluster become more scattered in Fig. 3.1(b) and 3.2(b). Therefore, the active rate with $\varepsilon_i=0.5$ is larger in this special case. To sum up, larger ε_i leads to larger active rate in major cases. In real world, it means that the opinion can spread widely among the people who will assimilate the opinion that is quite different from their opinions.

Empirically, the knowledge on an objective will influence people's opinions and the spread of a certain opinion. In detail, when people have little knowledge or data about an objective, they do not have clear opinions. Because of conformist mentality, many people just join the mainstream faction and express the opinion that is the most popular. Therefore, the diffusion probability is usually large so that a rumor will spread fast, just like when $f_O(x)=f_{ii}(x)$. On the contrary, when people have enough knowledge or data, they will hold clear opinions. The opinion divergence may be more significant in this case. Therefore, a certain opinion will spread slowly.



(a) $f_O(x) = f_i(x)$ (b) $f_O(x) = f_{ii}(x)$ (c) $f_O(x) = f_{iii}(x)$

(3.1) Email network

(a) $f_O(x) = f_i(x)$ (b) $f_O(x) = f_{ii}(x)$ (c) $f_O(x) = f_{iii}(x)$

(3.2) Facebook network

Fig. 3. The active rates and coverage rates are shown in this figure. The values of ε_i are 0, 0.5 and 1. The initiator's opinion changes from −1 to 1.

5. The algorithm to control the uncertain opinion diffusion process

In this section, we discuss how to control the uncertain opinion diffusion process.

Nowadays, online social networks have been greatly popular through which millions of users are connected. It has been known that the interactions in online social networks can change people's opinions and their choices [10,13]. For instance, the information in Twitter has proved to be very powerful in the 2008 US presidential elections [40,41]. Therefore, we present an optimization problem at the background of voting. It is assumed that each node can choose one of the two candidates or abstention based on its opinion. As shown in the model statements, the nodes' opinions may change in the diffusion process. Therefore, the voting result will be influenced. Then, we formulate the problem of which the goal is to set suitable initiators and their opinions to maximize the advantage of one candidate.

5.1. Voting mechanism

When a node is asked to vote on an issue, it has three choices. Let V_i denote the choice of node i:

$$V_i = \begin{cases} -1; \\ 0; \\ 1. \end{cases} \tag{4}$$

Here, $V_i = -1$ can represent the support to one of the two candidates while $V_i = 1$ represents the opposite choice. Meanwhile, $V_i = 0$ represents abstention.

Each node makes a choice based on its opinion. Then, we define the mapping from opinion to the probabilities of making choices:

$$p(V_i = -1) = \begin{cases} -O(i,t), & O(i,t) < 0; \\ 0, & \text{otherwise}. \end{cases} \tag{5}$$

$$p(V_i = 1) = \begin{cases} O(i,t), & O(i,t) > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

$$p(V_i = 0) = \begin{cases} 1 - p(V_i = 1), & O(i,t) > 0; \\ 1 - p(V_i = -1), & O(i,t) < 0; \\ 1, & O(i,t) = 0. \end{cases} \tag{7}$$

In this section, we will develop strategy to maximize the advantage of one candidate at the end of diffusion. Therefore, $O(i, t)$ in the three equations is node *i*' opinion at the end of diffusion by default.

5.2. Problem formulation and properties

Then, we formulate an optimization problem, called Voting Victory Problem (VVP). The voting victory problem is formally defined as follows.

Voting Victory Problem (VVP). *Let two sets of nodes* $L = \{i|V_i = -1\}$ *and* $R = \{i|V_i = 1\}$*. Given* $k > 0$*, the goal is to select a set* S *of* k *initiators and set their opinion* O_0 *so that* $|R| - |L|$ *is maximized at the end of diffusion.*

The confidence bound ε_i and the initial distributions of opinions are uncontrollable in VVP. We can just develop strategy based on the estimations of them. Therefore, ε_i and the initial distributions of opinions are assumed to be known in the problem. For the sake of simplification, it is assumed that all the nodes have the same confidence bound ε_i.

Overall, we have two sub-tasks for VVP:

- Set the parameter O_0;
- Select k initiators to send information.

The selection of initiators will be influenced by O_0. Therefore, we will estimate O_0 first. Theorem 2 shows the property of the problem.

Theorem 2. *VVP is NP-hard.*

Proof. It is assumed that the clustering coefficient of the network is small, and the initial distribution of opinion is scattered. If ε_i is small enough, for each node i and its neighbor node j, the inequality $|O(i, t) - O(j, t)| > \varepsilon_i$ can be satisfied. It means that any nodes cannot influence their neighbors' opinions, and only O_0 is able to influence nodes' opinions. In this case, VVP is equivalent to traditional influence maximization problem because larger scope of diffusion will make O_0 influence more nodes. In other words, traditional influence maximization problem can be considered as a special case of VVP. Traditional influence maximization problem has been proved as NP-hard [7]. Therefore, VVP is also NP-hard. □

5.3. Theoretical analyses and algorithm design

5.3.1. The estimation of the most suitable o_0

Let O_d denotes the most suitable opinion for initiator in VVP. In other words, $|R| - |L|$ will be maximized if $O_0 = O_d$. Then, we estimate the value of O_d.

The confidence bound ε_i decides the interval of opinions that O_0 can influence directly. It means the actual interval that O_0 can influence is $[O_0 - \varepsilon_i, O_0 + \varepsilon_i]$. Besides, when a node become active, it diffuses not only the information that contains O_0, but also its own opinion. O_0 may have influence on the active node's opinion. Therefore, active node's opinion should also be considered. Then, we estimate the average opinion that is expressed by the active nodes. Let

$$w_1 = \int_{O_0 - \varepsilon_i}^{O_0} f_O(x)\mathrm{d}x \tag{8}$$

and

$$w_2 = \int_{O_0}^{O_0 + \varepsilon_i} f_O(x)\mathrm{d}x. \tag{9}$$

Here, $w_1 + w_2$ denotes the rate of the opinions that can be influenced by O_0. Then, we have

$$l_1 = \frac{1}{w_1} \int_{O_0 - \varepsilon_i}^{O_0} x f_O(x)\mathrm{d}x \tag{10}$$

and

$$l_2 = \frac{1}{w_2}\int_{O_0}^{O_0+\varepsilon_i} x f_O(x)\mathrm{d}x. \tag{11}$$

Here, l_1 denotes the mathematical expectations of opinions that are in the confidence interval of O_0 and smaller than O_0. l_2 denotes the mathematical expectations of opinions that are in the confidence interval of O_0 and larger than O_0. Based on l_1 and l_2, we can estimate the active nodes' opinions which have been influenced by O_0.

$$l_1' = \frac{1}{2}(O_0 + l_1) \tag{12}$$

and

$$l_2' = \frac{1}{2}(O_0 + l_2). \tag{13}$$

Here, l_1' denotes average opinion that is expressed by the active nodes whose initial opinions are smaller than O_0. Correspondingly, l_2' denotes average opinion that is expressed by the active nodes whose initial opinions are larger than O_0.

Based on O_0, l_1' and l_2', we can estimate the co-influence of them. Therefore, we have

$$O' = \frac{O_0 + w_1 l_1' + w_2 l_2'}{1 + w_1 + w_2}. \tag{14}$$

Here, O' represents the composite opinion of O_0, l_1' and l_2'. The opinions in $[O_0 - \varepsilon_i, O_0]$ change towards positive direction so that the $p(V_i = 1)$ may increase and $p(V_i = -1)$ may decrease. It means that $|R| - |L|$ will increase. However, the opinions in $[O_0, O_0 + \varepsilon_i]$ change towards negative direction so that $|R| - |L|$ decreases. Therefore, the initiator should try to increase $\int_{O_0-\varepsilon_i}^{O_0} f_O(x)\mathrm{d}x$ and decrease $\int_{O_0}^{O_0+\varepsilon_i} f_O(x)\mathrm{d}x$. Besides, for the opinions in $[O_0 - \varepsilon_i, O_0 + \varepsilon_i]$, the degrees of the influence of O_0 are not the same. Therefore, $|O_0 - x|$ should also be considered. Then, we have

$$\rho_1 = \int_{O_0-\varepsilon_i}^{O_0} |O' - x| f_O(x)\mathrm{d}x - \int_{O_0}^{O_0+\varepsilon_i} |O' - x| f_O(x)\mathrm{d}x. \tag{15}$$

Here, ρ_1 denotes the influence of O'. Moreover, l_1' and l_2' may influence the opinions that are not in the confidence interval of O_0. Therefore, it can be obtained that

$$\rho_2 = \int_{l_1'-\varepsilon_i}^{O_0-\varepsilon_i} (l_1' - x) f_O(x)\mathrm{d}x \tag{16}$$

and

$$\rho_3 = \int_{O_0+\varepsilon_i}^{l_2'+\varepsilon_i} (l_2' - x) f_O(x)\mathrm{d}x. \tag{17}$$

Overall, we can obtain that

$$\sigma(O_0) = \rho_1 + w_1\rho_2 + w_2\rho_3. \tag{18}$$

Then, $\sigma(O_d)$ should be the maximum of $\sigma(O_0)$.

5.3.2. The selection of the initiators

In this section, we present algorithm to select the initiators.

Traditional influence maximization algorithms usually select the node that has the most neighbors. When the number of initiators is more than one, the set of nodes that have the most neighbors totally will be selected. For instance, Greedy algorithm [30] is a classic algorithm in influence maximization problem. In Greedy algorithm, the node that maximizes the increment of neighbors will be selected as a new initiator. Let n_i denote the number of the neighbors of node i. Besides, let $F(i) = n_i$, and $F(S)$ be the number of the total neighbors of the nodes in set S. Then, the Greedy algorithm is shown in Algorithm 1.

Algorithm 1. Greedy algorithm

1) initialize $S = \phi$
2) **for** $i = 1$ to k **do**
3) select $u = argmax_{w \in N \setminus S}(F(S \cup \{w\}) - F(S))$
4) $S = S \cup \{u\}$
5) **end for**
6) output S

The performance of Greedy algorithm is quite good. However, the runtime of Greedy algorithm is very high when the number of initiators or the number of nodes increases. In order to reduce the runtime of algorithm, Chen, Wang and Yang present the DegreeDiscount algorithm [42]. The runtime of DegreeDiscount algorithm is much lower than that of Greedy algorithm. The DegreeDiscount algorithm is shown in Algorithm 2.

Algorithm 2. DegreeDiscount algorithm

initialize $S = \phi$
for $i = 1$ to $|N|$ **do**
    $dd_i = F(i)$
    initialize t_i to 0
end for
for $i = 1$ to k **do**
    select $u = argmax_w\{dd_w | w \in N \setminus S\}$
    $S = S \cup \{u\}$
    for each neighbor v of u and $v \in N \setminus S$ **do**
        $t_v = t_v + 1$
        $dd_v = d_v - 2t_v - (d_v - t_v)t_v p$
    end for
end for
output S

In DegreeDiscount algorithm, p is the diffusion probability. However, the diffusion probability in our model is uncertain. Therefore, in this section, p is replaced by the average probability $\overline{p}_t$.

The goal of the two traditional algorithms is to select a set of nodes to maximize the active rate r_a. However, the maximum r_a is not the request in VVP. If r_a is large, nodes may receive the information more than once. As mentioned in the model statements, each node can be influenced by O_0 at most once. Therefore, larger r_a cannot increase the influence of O_0. On the contrary, a node is influenced by various opinions when it receives the information more than once. Then, the influence of O_0 will be diluted. On the other hand, larger r_c can make $|R| - |L|$ become larger if O_0 is suitable. The reason is that larger r_c means more nodes can be exposed to the information and may be influenced by O_0. Therefore, a larger r_c with a smaller r_a is needed in VVP. In order to cause large r_c with small r_a, we present a heuristic algorithm to select the initiators. Then, we will compare the heuristic algorithm with the Greedy algorithm and DegreeDiscount algorithm in the simulations.

In order to decrease r_a, the degrees of initiators should not be very large. However, a too small degree will lead to small r_c. Therefore, if node i is chosen as initiator, n_i should satisfy the inequality

$$n_i * \overline{p}_t > \frac{1}{\overline{p}_t}. \tag{19}$$

When k increases, the limitation of the degrees of initiators can be weakened. Then, the inequality becomes

$$n_i * \overline{p}_t * k > \frac{1}{\overline{p}_t}. \tag{20}$$

If the inequality above is satisfied, r_c will be an acceptable value. Besides, the interval of opinions that O_0 can influence should also be considered. Therefore,

$$n_i \geq \frac{1}{\int_{O_0 - \varepsilon_i}^{O_0} f_O(x)\mathrm{d}x}. \tag{21}$$

In the heuristic algorithm, the nodes that cannot satisfy the inequality (20) and (21) will be excluded. Besides, in some unconnected networks, a node with small degree may be isolated. In order to make the information be able to spread from the initiator to most nodes, the isolated nodes will also be excluded. Then, among the nodes that are not excluded, the k nodes with the smallest degrees will be selected. In addition, if the number of nodes that are not excluded is less than k, the k nodes with the largest degrees in the network will be selected. Overall, Algorithm 3 is obtained.

Then, we discuss the time complexity of Algorithm 3. As mentioned above, the nodes with small degree will be selected in Algorithm 3. Therefore, the probability that the initiators are the neighbor of each other is very small. It means that the connection between initiators can be ignored. It is noteworthy that considering the connection between the initiators leads to the large complexity of Greedy algorithm. Therefore, the complexity of Algorithm 3 is much smaller than that of Greedy algorithm. If the number of nodes is $|N|$ and the number of initiators is k, the time complexity of Algorithm 3 will be $O(|N| * \sqrt{|N|} + k * |N|)$.

Algorithm 3. Heuristic algorithm for VVP

initialize $S = \phi$
initialize $V = \phi$
initialize $w = 0$
for $i = 1$ to $|N|$ **do**
    initialize $b_i = 0$
end for
select $v = argmax_{w \in N}(F(w))$
$b_v = 1$
for each edge $(i, j) \in \{(i, j)|(i, j) \in E, b_i = 0, b_j = 1\}$ **do**
    $b_i = 1$
end for
for $i = 1$ to $|N|$ **do**
    if $b_i = 1$ and $F(i) * \overline{p}_t * k \geq \frac{1}{\overline{p}_t}$ and $F(i) \geq \frac{1}{\int_{O_0-\varepsilon_i}^{O_0} f_O(x)dx}$ **do**
        $V = V \cup \{i\}$
    end if
end for
if $|V| \geq k$ **do**
    for $i = 1$ to k **do**
        select $u = argmin_{w \in (V \setminus S)}(F(w))$
        $S = S \cup \{u\}$
    end for
end if
else do
    for $i = 1$ to k **do**
        select $u = argmax_{w \in (N \setminus S)}(F(w))$
        $S = S \cup \{u\}$
    end for
end else
output S

5.4. Experiments

In this section, we test our theoretical estimation and algorithm by simulations. The basic probability $p_0 = 0.1 * e = 0.2718$. Each trial is performed with 1000 replications.

5.4.1. The most suitable O_0

In each trial, k initiators will be selected randomly, and $k = 1, 11, \ldots, 101$. In the simulations, the initiator's opinion $O_0 = \{-1, -0.9, -0.8, \ldots, 0.8, 0.9, 1\}$, and the value which can cause the largest simulated $|R| - |L|$ is the most suitable O_0.

As shown in Tables 2 and 3, the most suitable O_0 in Email network and Facebook network are the same. Besides, in the two networks, the theoretical values are equal to the simulated values in most cases. The mistake just occurs when $f_O(x) = f_{iii}(x)$ and $\varepsilon_i = 1$, and the distance between the theoretical value and the simulated value is 0.1. In Tables 2 and 3, the mean distance between the theoretical values and the simulated values is 0.0167. In addition, the most suitable O_0 is constant when k is varied. The reason is that the fluctuation of r_a and r_c is not acute enough to change the most suitable O_0.

5.4.2. The initiators

In this section, the performance of the heuristic algorithm is tested. The initiators in various simulations are selected based on the three algorithms or selected randomly. Besides, O_0 is set as the most suitable value obtained in above simulations.

Figs. 4 and 5 show the performance of different algorithms. The curves are the simulated values of $(|R| - |L|)/|N|$ with different initiators. Besides, $\varepsilon_i = 0.5$ in Fig. 4 while $\varepsilon_i = 1$ in Fig. 5. It can be known from the figures that Heuristic algorithm can lead to expected results. When k is small, none of the four curves is always above others. However, when k increases, Heuristic algorithm usually leads to the largest $|R| - |L|$, especially when $k > 50$.

As shown in Figs. 4 and 5, the initial distribution of opinions and the value of ε_i can influence the performances of algorithms. In Fig. 4, if $f_O(x) = f_i(x)$, the simulated data based on different algorithms are approximately overlapping, and the distances between them are quite small. When $f_O(x) = f_{ii}(x)$ or $f_{iii}(x)$, the advantage of the Heuristic algorithm is more obvious. The reason is that opinions are scattered when $f_O(x) = f_i(x)$ and ε_i is small in Fig. 4. The fluctuation of active rate

Table 2
Initiator's most suitable opinion in Email network.

k	$f_O(x)$	ε_i	The most suitable O_0	
1–101	$f_i(x)$	0.5	Theoretical	1
			Simulated	1
		1	Theoretical	1
			Simulated	1
	$f_{ii}(x)$	0.5	Theoretical	0.4
			Simulated	0.4
		1	Theoretical	0.8
			Simulated	0.8
	$f_{iii}(x)$	0.5	Theoretical	−0.5
			Simulated	−0.5
		1	Theoretical	−0.3
			Simulated	−0.4

Table 3
Initiator's most suitable opinion in Facebook network.

k	$f_O(x)$	ε_i	The most suitable O_0	
1–101	$f_i(x)$	0.5	Theoretical	1
			Simulated	1
		1	Theoretical	1
			Simulated	1
	$f_{ii}(x)$	0.5	Theoretical	0.4
			Simulated	0.4
		1	Theoretical	0.8
			Simulated	0.8
	$f_{iii}(x)$	0.5	Theoretical	−0.5
			Simulated	−0.5
		1	Theoretical	−0.3
			Simulated	−0.4

does not influence the $(|R| - |L|)/|N|$ obviously in this case. In Fig. 5, if $f_O(x) = f_i(x)$, the heuristic algorithm is significantly better than other algorithms. The reason is that ε_i is large in Fig. 5 so that larger active rate increases the distance between the performances of algorithms. In addition, if the initial distribution of opinions, network structure and algorithm are fixed, the performance of algorithm in Fig. 5 is always better than that in Fig. 4. The reason is that larger ε_i makes opinions become convergent, and O_0 is able to influence more opinions.

In addition, it is unexpected that $|R| - |L|$ based on Greedy algorithm and DegreeDiscount algorithm are smaller than that based on random selection in many cases. It means that traditional influence maximization algorithms are not applicable to VVP. In other words, influencing people's opinions and their choices in voting may be significantly different from traditional influence maximization problem. The reason is that the goal of the traditional influence maximization algorithms is to maximize the number of people who are influenced, not to maximize the total influence on the people. However, in this problem, the essential goal is to maximize the change of opinions instead of making most people be influenced.

Fig. 6 shows the runtime of different algorithms. The confidence bound ε_i and the initial distribution of opinions do not influence the runtime of the algorithms. Therefore, we just show the runtime in different networks. The runtime of Greedy algorithm increases fast when k increases. The runtime of heuristic algorithm is large although k is small, nonetheless it increases very slowly when k increases. The reason is that the cost of excluding some nodes is high. DegreeDiscount algorithm is a famous algorithm with low cost. Although the runtime of heuristic algorithm is always larger than that of DegreeDiscount algorithm, the distance between the runtime of the two algorithms is not very big if $k > 50$. Therefore, when k is larger than 50, the runtime of the heuristic algorithm is acceptable.

6. Conclusions

In this paper, we study the issue of controlling the uncertain diffusion process in which the diffusion probability is influenced by evolutionary opinions. We present a model by extending the independent cascade model. In our model, each individual in social network has own opinion, and the probability with which he/she forwards the information is modified by the distance between his/her opinion and others'. Besides, individual may change his/her opinion by assimilating others' opinions.

Based on theoretical analyses and simulations, we briefly discuss the properties of diffusion model and obtain some conclusions. For instance, if most nodes hold similar opinions initially, the initiator's opinion can significantly influence

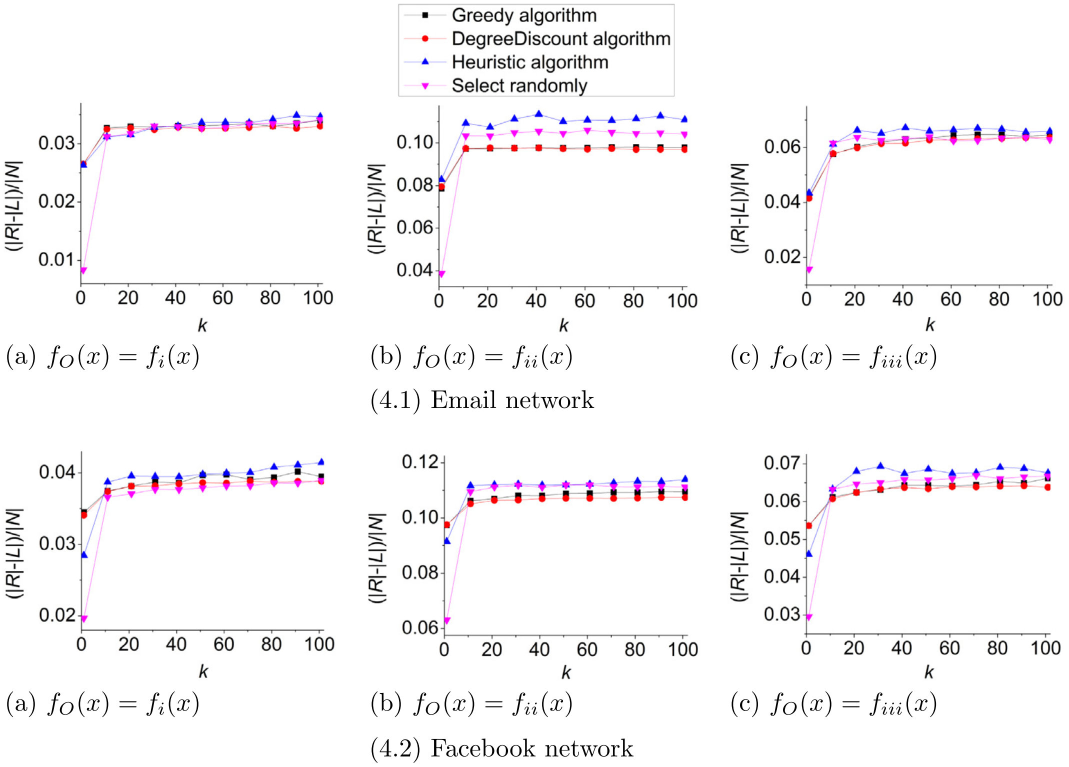


Fig. 4. $(|R| - |L|)/|N|$ with different selection of initiators are shown in this figure. The confidence level $\varepsilon_i = 0.5$ in these simulations. The initiator's opinion is the most suitable value obtained in simulations.

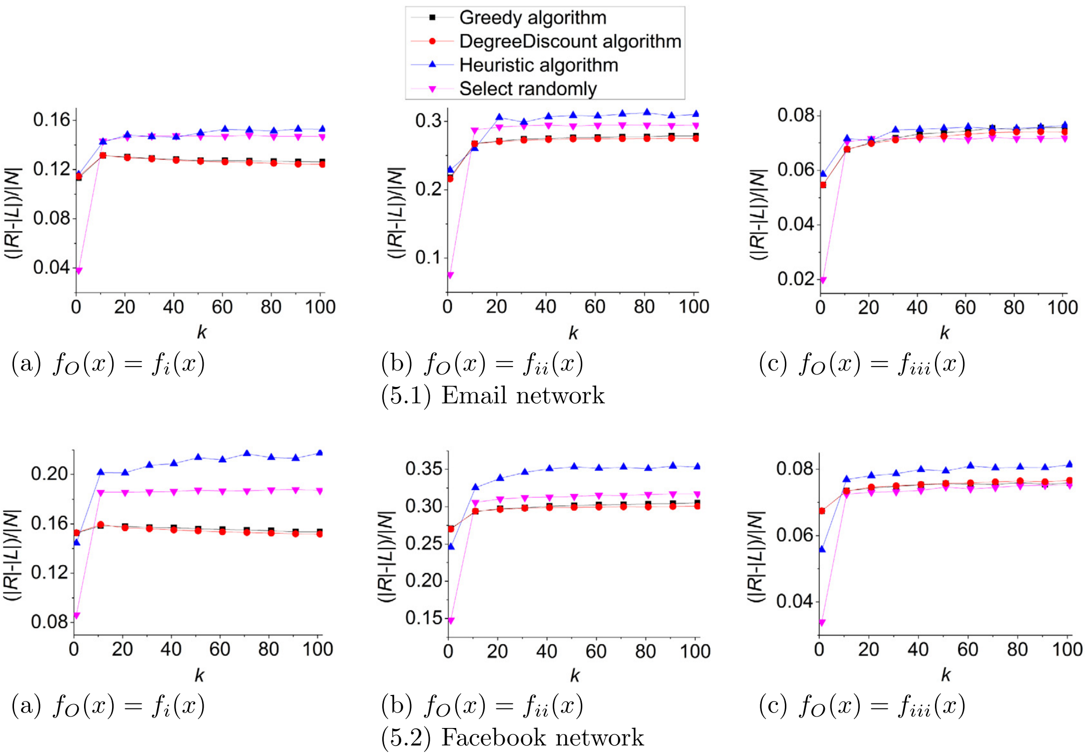
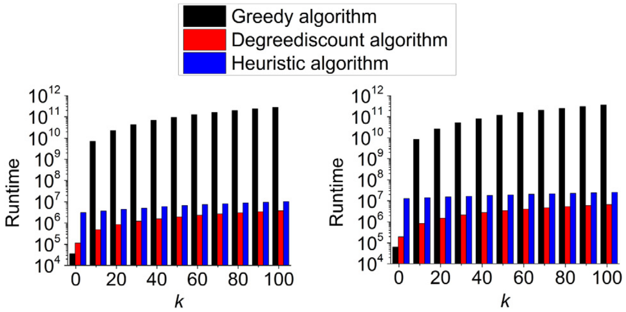


Fig. 5. $(|R| - |L|)/|N|$ with different selection of initiators are shown in this figure. The confidence level $\varepsilon_i = 1$ in these simulations. The initiator's opinion is the most suitable value obtained in simulations.



(a) Email network (b) Facebook network

Fig. 6. The runtime of different algorithms is shown in this figure. The number of initiators changes from 1 to 101.

the number of people who forward the information. If nodes' initial opinions are polarized initially, the influence of the initiator's opinion will be limited. What is more, larger confidence bound usually lead to larger active rate.

Then, we present an optimization problem at the background of voting to show how to control the diffusion process. In detail, we model a voting mechanism in which individuals can choose one of the two candidates or abstention, and they vote with probabilities that depend on their opinions. In order to maximize the advantage of one candidate, we present theoretical analyses for the problem and develop strategy to influence individuals' opinions based on the diffusion process. The results show that traditional influence maximization algorithms are not applicable to the problem, and our strategy can achieve the expected performance.

Our work and conclusions can present some inspirations to the real world applications. For instance, when the distribution of opinions is symmetrical, it is usually thought that the initiator's most suitable opinion should support our faction. However, in the simulations, the most suitable value is less than zero when opinions are initially polarized. In this case, although consolidating the supporters of our faction is useful, reducing the supporters of hostile faction will be more effective. In addition, at the background of marketing, polarized opinion can be consider as mightily commending a product while slightly biased opinion can represent the euphemistic commendation. The conclusions in Section 5.4.1 show that when most people's opinions are neutral and their confidence bound is small, slightly biased opinion is better than polarized opinion for the initiators. It means that when the consumers are not biased towards the product but suspicious, euphemistically commending the product may be better.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 61170164, 61472079 and 71201077), the Funds for Distinguished Young Scholars of the Natural Science Foundation of Jiangsu Province (No. BK2012020), and the Program for Distinguished Talents of Six Domains in Jiangsu Province (No. 2011-DZ023).

References

[1] L.-L. Xia, G.-P. Jiang, B. Song, Y.-R. Song, Rumor spreading model considering hesitating mechanism in complex social networks, Physica A 437 (2015) 295–303.
[2] Y. Jiang, J. Jiang, Diffusion in social networks: A multiagent perspective, IEEE Trans. Syst. Man Cybern. 45 (2) (2015) 198–213.
[3] A. Adiga, C. Kuhlman, H. Mortveit, A.K.S. Vullikanti, Sensitivity of diffusion dynamics to network uncertainty, in: Twenty-Seventh AAAI Conference on Artificial Intelligence, 2013.
[4] Y. Jiang, J. Jiang, Understanding social networks from a multiagent perspective, IEEE Trans. Parallel Distrib. Syst. 25 (10) (2014) 2743–2759.
[5] F. Yan, Z. Li, Y. Jiang, Noised diffusion dynamics with individual biased opinion, in: ECAI, 2014, pp. 1129–1130.
[6] Z. Li, Y. Jiang, Cross-layers cascade in multiplex networks, in: Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems, International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 269–276.
[7] D. Kempe, J. Kleinberg, É Tardos, Maximizing the spread of influence through a social network, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2003, pp. 137–146.
[8] A. Beutel, B.A. Prakash, R. Rosenfeld, C. Faloutsos, Interacting viruses in networks: can both survive? in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2012, pp. 426–434.
[9] S. Goyal, M. Kearns, Competitive contagion in networks, in: Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing, ACM, 2012, pp. 759–774.
[10] K. Lewis, M. Gonzalez, J. Kaufman, Social selection and peer influence in an online social network, Proc. Natl. Acad. Sci. 109 (1) (2012) 68–72.
[11] S. Aral, L. Muchnik, A. Sundararajan, Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks, Proc. Natl. Acad. Sci. 106 (51) (2009) 21544–21549.
[12] C.R. Shalizi, A.C. Thomas, Homophily and contagion are generically confounded in observational social network studies, Sociol. Methods Res. 40 (2) (2011) 211–239.
[13] P. Dandekar, A. Goel, D.T. Lee, Biased assimilation, homophily, and the dynamics of polarization, Proc. Natl. Acad. Sci. 110 (15) (2013) 5791–5796.
[14] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, Y. Yuan, Influence maximization in social networks when negative opinions may emerge and propagate, in: SDM, Vol. 11, SIAM, 2011, pp. 379–390.
[15] D. Cosley, S.K. Lam, I. Albert, J.A. Konstan, J. Riedl, Is seeing believing?: how recommender system interfaces affect users' opinions, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2003, pp. 585–592.

[16] R. Huckfeldt, M.T. Pietryka, J. Reilly, Noise, bias, and expertise in political communication networks, Social Networks 36 (2014) 110–121.
[17] R. Iyengar, C. Vanden Bulte, T.W. Valente, Opinion leadership and social contagion in new product diffusion, Mark. Sci. 30 (2) (2011) 195–212.
[18] A. Tsang, K. Larson, Opinion dynamics of skeptical agents, in: Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems, International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 277–284.
[19] J. Lorenz, Heterogeneous bounds of confidence: Meet, discuss and find consensus!, Complexity 15 (4) (2010) 43–52.
[20] K. Fan, W. Pedrycz, Emergence and spread of extremist opinions, Physica A 436 (2015) 87–97.
[21] K. Kaneko, Spatiotemporal intermittency in coupled map lattices, Progr. Theoret. Phys. 74 (5) (1985) 1033–1044.
[22] J. Chen, F. Huang, M. Khan, M. Marathe, P. Stretz, H. Xia, The effect of demographic and spatial variability on epidemics: A comparison between beijing, delhi, and los angeles, in: Critical Infrastructure (CRIS), 2010 5th International Conference on, IEEE, 2010, pp. 1–8.
[23] S. Eubank, C. Barrett, R. Beckman, K. Bisset, L. Durbeck, C. Kuhlman, B. Lewis, A. Marathe, M. Marathe, P. Stretz, Detail in network models of epidemiology: are we there yet ? J. Biol. Dyn. 4 (5) (2010) 446–455.
[24] B. Karrer, M. Newman, Competing epidemics on complex networks, Phys. Rev. E 84 (3) (2011) 036106.
[25] S.A. Myers, J. Leskovec, Clash of the contagions: Cooperation and competition in information diffusion, in: Data Mining (ICDM), 2012 IEEE 12th International Conference on, IEEE, 2012, pp. 539–548.
[26] P. Holme, M.E. Newman, Nonequilibrium phase transition in the coevolution of networks and opinions, Phys. Rev. E 74 (5) (2006) 056108.
[27] S. Chatterjee, F.M. Hafizoglu, S. Sen, Predicting migration and opinion adoption patterns in agent communities, in: Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems, International Foundation for Autonomous Agents and Multiagent Systems, 2013, pp. 1287–1288.
[28] M. Afshar, M. Asadpour, Opinion formation by informed agents, J. Artif. Soc. Social Simul. 13 (4) (2010) 5.
[29] O. AskariSichani, M. Jalili, Influence maximization of informed agents in social networks, Appl. Math. Comput. 254 (2015) 229–239.
[30] W. Chen, W. Lu, N. Zhang, Time-critical influence maximization in social networks with time-delayed diffusion process, ArXiv Preprint arXiv:1204.3074.
[31] S. Feng, X. Chen, G. Cong, Y. Zeng, Y.M. Chee, Y. Xiang, Influence maximization with novelty decay in social networks, in: Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014.
[32] Y. Wu, Y. Yang, F. Jiang, S. Jin, J. Xu, Coritivity-based influence maximization in social networks, Physica A 416 (2014) 467–480.
[33] X. He, D. Kempe, Stability of influence maximization, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2014, pp. 1256–1265.
[34] E. Bonabeau, M. Dorigo, G. Theraulaz, Swarm Intelligence: From Natural to Artificial Systems, Vol. 1, Oxford university press, 1999.
[35] B. Klimt, Y. Yang, Introducing the enron corpus., in: CEAS, 2004.
[36] J. Leskovec, K.J. Lang, A. Dasgupta, M.W. Mahoney, Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters, Internet Math. 6 (1) (2009) 29–123.
[37] B. Viswanath, A. Mislove, M. Cha, K.P. Gummadi, On the evolution of user interaction in facebook, in: Proceedings of the 2nd ACM Workshop on Online Social Networks, ACM, 2009, pp. 37–42.
[38] Y. Moreno, R. Pastor-Satorras, A. Vespignani, Epidemic outbreaks in complex heterogeneous networks, Eur. Phys. J. B 26 (4) (2002) 521–529.
[39] W.J. Dixon, F.J. Massey, et al., Introduction to Statistical Analysis, Vol. 344, McGraw-Hill, New York, 1969.
[40] A.L. Hughes, L. Palen, Twitter adoption and use in mass convergence and emergency events, Int. J. Emerg. Manag. 6 (3–4) (2009) 248–260.
[41] A. Guille, H. Hacid, C. Favre, D.A. Zighed, Information diffusion in online social networks: A survey, ACM SIGMOD Rec. 42 (2) (2013) 17–28.
[42] W. Chen, Y. Wang, S. Yang, Efficient influence maximization in social networks, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2009, pp. 199–208.