Physica A 444 (2016) 20–34

Contents lists available at ScienceDirect

Physica A

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

Influence maximization in social networks under an independent cascade-based model

Qiyao Wang [a], Yuehui Jin [a], Zhen Lin [b,c,*], Shiduan Cheng [a], Tan Yang [d]

[a] State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

[b] Department of Technology Development, Shanghai Clearing House, Shanghai 200002, China

[c] College of Computer Science & Technology, Zhejiang University, Hangzhou 310027, China

[d] School of Software Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

HIGHLIGHTS

- A novel independent cascade model based on opinion change, called IC-OC, was proposed for information spreading.
- Two probabilities were introduced to predict opinion change when users are exposed to bilateral opinions.
- The IMIC-OC model was proposed for influence maximization.
- Experiments were conducted on three real networks to verify that the IMIC-OC model has larger influence than two baseline methods.

ARTICLE INFO

Article history:
Received 19 April 2015
Received in revised form 23 July 2015
Available online 13 October 2015

Keywords:
Influence maximization
Positive influence
Independent cascade-based model
Opinion change
Social networks

ABSTRACT

The rapid growth of online social networks is important for viral marketing. Influence maximization refers to the process of finding influential users who make the most of information or product adoption. An independent cascade-based model for influence maximization, called IMIC-OC, was proposed to calculate positive influence. We assumed that influential users spread positive opinions. At the beginning, users held positive or negative opinions as their initial opinions. When more users became involved in the discussions, users balanced their own opinions and those of their neighbors. The number of users who did not change positive opinions was used to determine positive influence. Corresponding influential users who had maximum positive influence were then obtained. Experiments were conducted on three real networks, namely, Facebook, HEP-PH and Epinions, to calculate maximum positive influence based on the IMIC-OC model and two other baseline methods. The proposed model resulted in larger positive influence, thus indicating better performance compared with the baseline methods.

1. Introduction

Online social networks (OSNs) are beneficial for researchers and companies that study user behaviors and make profits, respectively. For example, advertisers found a subset of users to maximize the adoption of products [1], which is a way of using influence maximization. The goal of influence maximization is to maximize the number of users involved in discussions

* Correspondence to: No 2 East Beijing Road, Shanghai Clearing House, Shanghai 200002, China. Tel.: +86 21 23198875.
E-mail address: linzhen@shclearing.com (Z. Lin).

http://dx.doi.org/10.1016/j.physa.2015.10.020

or adopt products in the network. The independent cascade-based model for influence maximization is proposed in this study to handle this problem.

In recent years, many researchers have focused on influence maximization. Kempe et al. [2] first used greedy algorithm to study influence maximization based on two models, namely, Independent Cascade (IC) and Linear Threshold (LT). Chen et al. [3] calculated positive influence amidst the existence of negative opinions within a network. However, the authors assumed that if users adopted one opinion, they would never change their opinions. As time passes, users are probably influenced by others to adopt opposite opinions. Thus, when users are exposed to different attitudes, probabilities are introduced to predict whether such users would change their minds until no neighbor changes opinions.

To better understand the proposed model for influence maximization, a novel IC model based on opinion change, called IC-OC, is proposed to show the process through which users build opinions. In the beginning of information spreading, the IC model is used to predict the initial opinions of the user. As several users become involved in the discussion, a transforming probability is introduced to predict whether users would change their initial opinions. The proposed model is verified by the dataset culled from Facebook.[1] Then, the IMIC-OC model is used to calculate maximum positive influence in three real networks.

We model social network as a directed influence graph. Nodes denote individuals and edges denote "who influences whom". In some social networks, such as Twitter[2] and Sina Weibo,[3] user relations are unidirectional, such that the structure of network is directly used as influence graphs. Other social networks provide bidirectional relations, such as Facebook and the co-author's network. In these networks, an undirected edge is divided into two directed edges to obtain the influence graph.

In the current study, the IC-OC model is proposed to explain how users build opinions during the process of information spreading. The IMIC-OC model is proposed to find influential users who have maximum positive influence. The following items summarize our contributions to the literature.

- Transforming probability is introduced to predict whether users change their initial opinions when bilateral opinions exist in the network.
- The IC-OC model for information spreading is proposed to gain a better explanation for opinion formation.
- The IMIC-OC model is proposed to tackle the influence maximization problem; the properties of the proposed model are also analyzed.

Experiments were conducted on a real network. First, the IC-OC model was verified for information spreading. Experimental results indicated that the proposed model had better performance than the baseline methods. Maximum positive influence was calculated by applying the IMIC-OC model on three real datasets. The findings showed that the proposed model had better performance compared with the baseline methods. Further explanations were given.

The reminder of this study is organized into sections. Prior works on influence maximization and opinion formation are discussed in Section 2. Section 3 presents the problem statement. Section 4 introduces the IC-OC model for information spreading and the IMIC-OC model for influence maximization. Section 5 shows the experimental results. Finally, Section 6 presents the conclusion.

2. Related works

2.1. Influence maximization

Many researchers have attempted to tackle the influence maximization problem. Kempe et al. [2] proposed two basic information diffusion models, namely, IC and LT. An algorithm was used to obtain an approximate solution for this NP-hard problem. Chen et al. [4] categorized the solutions to the problem of influence maximization into two classes: reducing running time of algorithms and using new methods to calculate influence.

Many studies have focused on improving the efficiency of algorithms on influence maximization [5–7]. Zhang et al. [8], for example, reduced running time and memory consumption by mapping several networks into one network. Goyal et al. [9] proposed CELF++, a novel CELF algorithm, to reduce running time. Heidari et al. [10] used Monte-Carlo simulation to improve time complexity, whereas Li et al. [11] considered wide influence spread to tackle location-aware influence maximization.

Other studies proposed new algorithms to calculate influence [12,13]. Zhou et al. [14] employed the voter model to handle the situation wherein different users activate the same user. Wu et al. [15] used core users in the forefront of a network to calculate influence, and Li et al. [16] proposed a novel conformity-aware cascade model, which used conformity theory to calculate influence. Meanwhile, Zhu et al. [17] combined continuous-time Markov chain and IC model for influence maximization. Guo et al. [18] proposed personal influence maximization, which aimed to find users who can activate targeted users. Lee et al. [19] maximized influence on specific users by query processing. In sum, the problem of influence maximization is based on information spreading. Researchers have proposed new models to exploit the structure of network

[1] http://www.facebook.com.
[2] http://www.twitter.com.
[3] http://www.weibo.com.

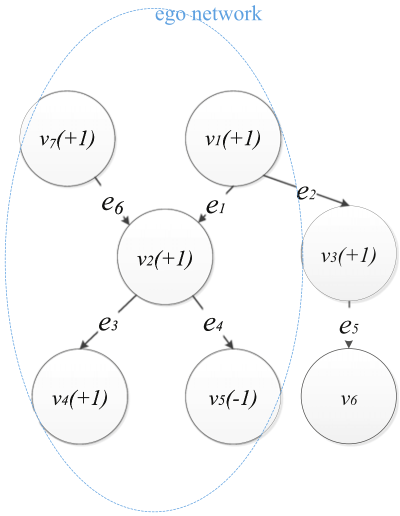

Fig. 1. Structure of the influence graph.

or user relations so as to improve the performance of information spreading. Based on these models, the authors proposed corresponding models for influence maximization to find influential users.

Some researchers have also attempted to calculate influence when different opinions exist in a network. For instance, Chen et al. [3] proposed the IC-N model to calculate influence when negative opinions were present in a given network. They assumed that users never change opinions once activated. However, in reality, users could change their opinions, that is, the accumulation of opposite opinions in the network could possibly make users change their minds. Thus, we introduce transforming probability to predict whether users would change opinions when they are exposed to different views. Li et al. [20] introduced broadcast and current attitude in studying user attitude modification. Based on the above method, the authors proposed the IMLT-IOA model for influence maximization.

2.2. *Opinion formation*

In real life, people may actually change their initial opinions when their neighbors, who have a profound impact on them, hold opposite opinions. Some researchers have studied the problem of opinion formation [21,22]. Das et al. [23] categorized models for opinion formation into three models, namely, averaging, flocking and voter. In averaging models, users update their opinions by using the average of neighbors. In flocking models [24], conformity bias exists. In voter models [25,26], users randomly pick up the opinions of their neighbors at each time step. Meanwhile, Wang et al. [27] introduced a freezing period in the voter model. During the freezing period, users have a high probability to change their minds. After that time, users could easily reach consensus in the network.

3. Problem statement

Various definitions of social networks are available. First, the influence graph is introduced, which is a directed graph expressed as $G = (V, E)$.

Definition 1. An influence graph $G = (V, E)$ is a directed graph that shows the relation of "who influences whom". Here, $V = \{v_1, \ldots, v_n\}$ denotes a set of users, while $E = \{e_1, \ldots, e_m\}$ denotes a set of edges.

Definition 2. Neighbors $N(v)$ denote entire neighbors of user v. Here, $N_{in}(v)$ and $N_{out}(v)$ denote in- and out-neighbors, respectively. If user v_j is an out-neighbor of user v_i, $v_j \in N_{out}(v_i)$, user v_i has an impact on user v_j. Meanwhile, if user v_j is an in-neighbor of user v_i, $v_j \in N_{in}(v_i)$, user v_j influences user v_i.

For example, Fig. 1 shows that edge e_1 connects users v_1 and v_2. Edge e_1 denotes that user v_1 influences user v_2. User v_1 is an in-neighbor of user v_2, that is, $v_1 \in N_{in}(v_2)$, and user v_2 is an out-neighbor of user v_1, that is, $v_2 \in N_{out}(v_1)$.

Definition 3. Users are categorized into three classes, namely, inactive (I), positive (P), and negative (N). Here, I users are not activated in the network. P users hold positive opinions in the network, and N users hold negative opinions in the network.

In Fig. 1, only user v_6 is inactive, $I = \{v_6\}$. Five users are positive, $P = \{v_1, v_2, v_3, v_4, v_7\}$, whereas one user is negative, $N = \{v_5\}$.

Three classes of users can be expressed as

$$|V| = |I| + |P| + |N|, \tag{1}$$

where $|*|$ denotes the cardinality of $*$.

In Eq. (2), $N^i_{out}(v)$, $N^p_{out}(v)$, and $N^n_{out}(v)$ denote inactive, positive and negative out-neighbors, respectively. The term is expressed as

$$|N_{out}(v)| = |N^i_{out}(v)| + |N^p_{out}(v)| + |N^n_{out}(v)|. \tag{2}$$

In Eq. (3), $N^i_{in}(v)$, $N^p_{in}(v)$, and $N^n_{in}(v)$ denote inactive, positive and negative in-neighbors, respectively. We then obtain

$$|N_{in}(v)| = |N^i_{in}(v)| + |N^p_{in}(v)| + |N^n_{in}(v)|. \tag{3}$$

Definition 4. Ego network $G_e = (V_e, E_e)$ is a network that indicates relations between a specific user v_i and neighbors $N(v_i)$, $V_e = v_i \cup N(v_i)$. Edges connecting users are included, $E_e \subset E$.

In Fig. 1, ego network $G_e(v_2)$ of user v_2 shows relations between user v_2 and neighbors, $N(v_2) = \{v_1, v_4, v_5, v_7\}$. The ego network of user v_2 includes five users, $V_e = \{v_1, v_2, v_4, v_5, v_7\}$, and four edges, $E_e = \{e_1, e_3, e_4, e_6\}$.

Definition 5. Sentiment $\boldsymbol{s_i(t)}$ denotes the sentiment of user v_i at time t. Here, $s_i(t) \in \{-1, 1\}$ represents negative and positive sentiments, respectively.

Definition 6. Influential users $\boldsymbol{S}$ refers to a set of users that has a profound impact on activating other users. The number of influential users is given as $|S| = k$.

Definition 7. Positive influence $\boldsymbol{\phi(S)}$ denotes a set of users with positive opinions, which are the same as influential users S. Positive influence is defined as

$$\phi(S) = \{v_j | s_j * s_i > 0, v_i \in S\}, \tag{4}$$

where user $v_i \in S$ is an influential user, s_i is the sentiment of user v_i, and user v_j is activated in the network and holds positive opinions.

Definition 8. Negative influence denotes a set of users with negative opinions when influential users S spreads negative opinions. Owing to the asymmetry between positive and negative influence, we mainly study positive influence. Further explanations are given in Section 4.

As shown in Fig. 1, the network has seven users $V = \{v_1, \ldots, v_7\}$, and six edges $E = \{e_1, \ldots, e_6\}$. If user v_1 is an influential user, $S_1 = \{v_1\}$, we set user v_1 to spread a positive message. Users v_2, v_3, v_4, v_5 adopt this information and only user v_5 holds a negative opinion. Thus, positive influence is $\phi(S_1) = \{v_2, v_3, v_4\}$ and $|\phi(S_1)| = 3$. If user v_2 is an influential user, $S_2 = \{v_2\}$, positive influence is $\phi(S_2) = \{v_4\}$ and $|\phi(S_2)| = 1$. If user v_3 is an influential user, $S_3 = \{v_3\}$, positive influence is $\phi(S_3) = \emptyset$ and $|\phi(S_3)| = 0$.

Definition 9. Maximum positive influence $\boldsymbol{\phi_m(S)}$ denotes the maximum number of users who are activated by influential users S when the number of influential users is given as $|S| = k$. Maximum positive influence is defined as

$$\phi_m(S) = max\{|\phi(S)|\}. \tag{5}$$

In Fig. 1, if the cardinality of influential users is 1, $|S| = 1$, many users can be influential users, $S_1 = \{v_1\}$, $S_2 = \{v_2\}$, *and* $S_3 = \{v_3\}$. In addition, the maximization positive influence is $\phi_m(S) = \{v_2, v_3, v_4\}$ and $S = S_1$.

Problem Definition: Given number k, we aim to find a set of influential users S^* who activate the maximum positive influence $\phi(S)$. Hence,

$$S^* = argmax\ |\phi(S)|, \tag{6}$$

where the cardinality of influential users S is k.

4. Model

4.1. IC-OC model

We introduce the IC-OC model in this section. Information spreading is initiated when some users spread information within the network. Out-neighbors are activated by adoption probability and hold the same opinion by conforming probability. Explicit definitions of probabilities are given in Section 4.2.1. As time passes, bilateral opinions emerge in the network. When in-neighbors hold different opinions, users change opinions by transforming probability while keeping their opinions by stubborn probability. The explicit definitions of these probabilities are provided in Section 4.2.1.

4.2. IMIC-OC model

The IMIC-OC model aims to find a set of users who have maximum positive influence. This section analyzes properties of the IMIC-OC model.

According to Chen et al. [3], negative and positive influences are not symmetric. If the initial opinion of user v_i is positive, users easily change to negative opinions. If user v_i initially holds a negative opinion, user v_i hardly changes to a positive one. Therefore, we mainly analyze positive influence in the following sub-sections.

4.2.1. Introduction of IMIC-OC model

In the IMIC-OC model, we first propose how influential users utilize paths to reach a user. Second, some probabilities are introduced to explain how users build opinions in the process of information spreading. Third, probabilities are introduced to calculate positive influence in the IMIC-OC model.

$Path(S, v_i) = \{path_1, \dots, path_n\}$ denotes a set of paths from influential users S to user v_i, whereas $L = \{l_1, \dots, l_n\}$ denotes corresponding lengths of paths, $Path(S, v_i) = \{path_1, \dots, path_n\}$. Many influential users can reach user v_i. Thus, the shortest path, $path_{min}(S, v_i)$, from influential users S to user v_i, is calculated to obtain the initial opinion of user v_i. In addition, $l_{min}(S, v_i)$ is the corresponding length, $path_i = \langle v_1, v_2, \dots, v_i \rangle$ denotes entire users in $path_i$, and $path_{min}(S, v_i) = \langle v_1, v_2, \dots, v_{l_{min}}, v_i \rangle$ shows the entire users in $path_{min}(S, v_i)$.

Probabilities are introduced to describe the spread of information from one user to another as well as the changing of opinions during the process of information spreading. The explicit explanations are given below.

Definition 10. Adoption probability p_a is the probability of adopting information. If user v_i has an influence on user v_j, adoption probability p_a can be expressed as

$$p_a = \frac{n}{N_1} * \frac{n}{N_2}, \tag{7}$$

where N_1 is the number of posts that are published by user v_i, N_2 is the number of posts that are adopted by user v_j, and n is the number of posts adopted by v_j from user v_i.

Definition 11. Conforming probability p_c is the probability that represents how user v_j conforms to user v_i when user v_j adopts a piece of information from user v_i. Conforming probability is defined as

$$p_c = \frac{n_c}{n}, \tag{8}$$

where n_c is the number of posts whose sentiment is consistent with the sentiment expressed by user v_i, and n is the number of posts adopted by v_j from user v_i.

Definition 12. Positive propagation probability pp is the probability that user v_j adopts information from user v_i and user v_j holds the same opinion as user v_i. Positive propagation probability is defined as

$$pp_{ij} = p_a * p_c = \frac{n_c * n}{N_1 * N_2}, \tag{9}$$

where pp_{ij} denotes the positive propagation probability between users v_i and user v_j, p_a denotes adoption probability, p_c denotes conforming probability, N_1 is the number of posts that are published by user v_i, N_2 is the number of posts adopted by user v_j, n is the number of posts adopted by v_j from user v_i, and n_c is the number of posts whose sentiment is consistent with the sentiment expressed by user v_i.

Definition 13. Negative propagation probability np is the probability that user v_j adopts the information from user v_i, and user v_j holds the opposite opinion against user v_i. Negative propagation probability is defined as

$$np_{ij} = p_a * (1 - p_c) = \left(1 - \frac{n_c}{n}\right) * \frac{n * n}{N_1 * N_2} = \frac{(n - n_c) * n}{N_1 * N_2}, \tag{10}$$

where np_{ij} denotes negative propagation probability between user v_i and user v_j, p_a denotes the adoption probability, p_c denotes the conforming probability, N_1 is the number of posts published by user v_i, N_2 is the number of posts adopted by user v_j, n is the number of posts adopted by v_j from user v_i, and n_c is the number of posts whose sentiment is consistent with the sentiment expressed by user v_i.

When more users become involved in discussions, other users may begin to hold different opinions in the network. Some users could change their minds when some in-neighbors hold opposite opinions. Transforming and stubborn probabilities are introduced to predict whether users change their minds. Prior to the introduction of probabilities, two weights, stubborn w_{ii} and influential w_{ji}, are introduced in the ego network to calculate these two probabilities. The definitions of stubborn weight w_{ii} and influential weight w_{ji} are expressed below.

Definition 14. Stubborn weight w_{ii} represents how user v_i adheres to his opinion.

Definition 15. Influential weight w_{ji} represents how user $v_j \in N_{in}(v_i)$ influences user v_i.

Given user v_i, the sum of stubborn weight w_{ii} and influential weight w_{ji} is equal to 1, and is expressed as

$$w_{ii} + \sum_{v_j \in N_{in}(v_i)} w_{ji} = 1. \tag{11}$$

Here, $w_{ii} = 0$ indicates that users are not insistent on their initial opinions and are easily influenced by neighbors. In this case, the proposed model is close to the IC model. In addition, $w_{ii} = 1$ indicates that users are insistent and are not about to change their initial opinions. In this case, the proposed model is close to the IC-N model [3].

According to Das et al. [23], stubborn weight is $w_{ii} \geq c * \sum_{v_j \in N_{in}(v_i)} w_{ji}$ where parameter $c \geq 0.25$. We discuss the case

$$w_{ii} = c * \sum_{v_j \in N_{in}(v_i)} w_{ji} \quad \text{where } c \geq 0.25. \tag{12}$$

w'_{ii} and w'_{ji} denote temporary stubborn and temporary influential weights, respectively, to calculate w_{ii} and w_{ji}. Temporary influential weight w'_{ji} is equal to conforming probability $p_{c_{ji}}$, as shown in Eq. (13)

$$w'_{ji} = p_{c_{ji}}. \tag{13}$$

Temporary stubborn weight and temporary influential weight are satisfied in Eq. (12). Thus, temporary stubborn weight w'_{ii} is calculated by

$$w'_{ii} = c * \sum_{v_j \in N_{in}(v_i)} w'_{ji}. \tag{14}$$

Stubborn weight w_{ii} and influential weight w_{ji} are the normalizations of w'_{ii} and w'_{ji}, respectively, and are expressed as

$$w_{ii} = \frac{w'_{ii}}{w'_{ii} + \sum\limits_{v_j \in N_{in}(v_i)} w'_{ji}} = \frac{c}{c+1}, \tag{15}$$

$$\begin{aligned} w_{ji} &= \frac{w'_{ji}}{(c+1) * \sum\limits_{v_j \in N_{in}(v_i)} w'_{ji}} \\ &= \frac{p_{c_{ji}}}{(c+1) * \sum\limits_{v_j \in N_{in}(v_i)} p_{c_{ji}}}. \end{aligned} \tag{16}$$

Owing to the asymmetry between positive and negative influences, stubborn weight w^p_{ii} and negative stubborn weight w^n_{ii} are expressed as

$$w^p_{ii} = c_p * \sum_{v_j \in N_{in}(v_i)} w_{ji} \quad \text{and} \quad w^n_{ii} = c_n * \sum_{v_j \in N_{in}(v_i)} w_{ji}, \tag{17}$$

where c_p denotes parameter c for positive influence, and c_n denotes parameter c for negative influence.

Based on stubborn weight w_{ii} and influential weight w_{ji}, the sentiment of user v_i at time $t+1$ balances neighbors $v_j \in N_{in}(v_i)$ and the sentiment of user v_i at time t. We multiply stubborn weight w_{ii} and the sentiment $s_i(t)$ of user v_i at time t, influential weight w_{ji} and neighbor sentiment $s_j(t)$. The equation is expressed as

$$\begin{aligned} s_i(t+1) &= f(s_i(t), s_j(t)) \\ &= sgn\left(w_{ii} * s_i(t) + \sum_{v_j \in N_{in}(v_i)} w_{ji} * s_j(t)\right), \end{aligned} \tag{18}$$

where $sgn(*)$ is a sign function, w_{ii} denotes the stubborn weight of user v_i, and w_{ji} denotes the influential weight when user $v_j \in N_{in}(v_i)$ has an impact on user v_i.

Definition 16. Transforming probability p_t represents how users change their opinions from one sentiment to another. Here, p_{t1} denotes the transforming probability from positive to negative, whereas p_{t2} denotes the transforming probability from negative to positive.

Definition 17. Stubborn probability p_s indicates that users keep their opinions even when neighbors hold different opinions. Similarly, we obtain two corresponding stubborn probabilities: p_{s1} denotes the probability that users keep positive opinions, and p_{s2} denotes the probability that users keep negative opinions.

Given user v_i, the sum of transforming probability p_t and stubborn probability p_s is equal to 1, and is expressed as

$$p_t + p_s = 1. \tag{19}$$

Eq. (19) is rewritten for the positive and the negative situations, which are expressed as

$$p_{t1} + p_{s1} = 1 \quad \text{or} \quad p_{t2} + p_{s2} = 1. \tag{20}$$

Transforming probability p_{t1} from positive to negative is the sum of influential weights from negative in-neighbors, and is expressed as

$$p_{t1} = \sum_{v_j \in N_{in}^{n}(v_i)} w_{ji}, \tag{21}$$

where w_{ji} is the influential weight from user v_j to user v_i, and $v_j \in N_{in}^{n}(v_i)$ denotes that user v_j is a negative in-neighbor of user v_i.

According to Eq. (19), stubborn probability p_{s1} is the sum of positive stubborn weight and influential weights from positive in-neighbors, and is expressed as

$$p_{s1} = w_{ii}^{p} + \sum_{v_j \in N_{in}^{p}(v_i)} w_{ji}, \tag{22}$$

where w_{ii}^{p} is the stubborn weight for the positive, w_{ji} is the influential weight from user v_j to user v_i, and $v_j \in N_{in}^{p}(v_i)$ denotes that user v_j is a positive in-neighbor of user v_i.

Similarly, transforming probability p_{t2} from the negative to the positive requires the sum of influential weights of positive in-neighbors $N_{in}^{p}(v_i)$. Stubborn probability p_{s2} for the negative is the sum of negative stubborn weight and influential weights from negative in-neighbors. These relations are respectively expressed as

$$p_{t2} = \sum_{v_j \in N_{in}^{p}(v_i)} w_{ji}, \tag{23}$$

$$p_{s2} = w_{ii}^{n} + \sum_{v_j \in N_{in}^{n}(v_i)} w_{ji}. \tag{24}$$

In conclusion, the process of information spreading follows a certain procedure discussed here. First, within a network, users are categorized into three classes, namely, inactive (I), positive (P), and negative (N). Users transfer from one class to another based on probabilities. Fig. 2 shows entire probabilities used in the process of information spreading. At time t, user v_i publishes a positive post, which is visible to out-neighbors $N_{out}(v_i)$ of user v_i. User $v_j \in N_{out}(v_i)$ adopts this information with adoption probability p_a. Then, user v_j holds positive opinion with conforming probability p_c after user v_i adopts this information. Thus, user $v_j \in N_{out}^{p}(v_i)$ adopts a positive opinion with positive propagation probability pp. Meanwhile, user $v_j \in N_{out}^{n}(v_i)$ holds a negative opinion with negative propagation probability np. User $v_j \in N_{out}^{i}(v_i)$ remains inactive with probability $(1 - p_a)$. As time passes, more users become involved in discussions, and some users present different opinions. When the in-neighbors $N_{in}(v_j)$ of user v_j hold opposite opinions, and they have a profound impact on user v_j, user v_j could change his initial opinion. If user v_j is positive at time t, he keeps a positive opinion at time $t+1$ with stubborn probability p_{s1} and changes to negative opinion with transforming probability p_{t1}. However, if user v_j is negative at time t, he keeps a negative opinion with stubborn probability p_{s2} and holds a positive opinion at time $t+1$ with transforming probability p_{t2}.

Probabilities are introduced to calculate positive influence. The shortest path from $path_{min}(S, v_i)$ is introduced to calculate initial opinions. The probability $P(s_0 = 1)$ that the initial opinion of user v_i is positive depends on probabilities among users $path_{min}(v_i) = \langle v_1, v_2, \ldots, v_{lmin}, v_i \rangle$ in the shortest path. This path is defined as

$$P(s_0 = 1) = pp_{l_{min},i} * \prod_{j=1}^{l_{min}-1} pp_{j,j+1}, \tag{25}$$

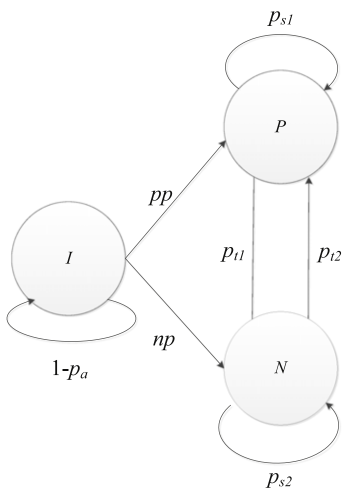

Fig. 2. Probabilities used in information spreading.

where $pp_{l_{min},i}$ denotes positive propagation probability between users $v_{l_{min}}$ and v_i, and $pp_{j,j+1}$ denotes positive propagation probability between users v_j and v_{j+1}.

The probability $P(s_0 = -1)$ that the initial opinion of user v_i is negative is defined as

$$P(s_0 = -1) = np_{l_{min},i} * \prod_{j=1}^{l_{min}-1} pp_{j,j+1}, \tag{26}$$

where $np_{l_{min},i}$ denotes the negative propagation probability between users $v_{l_{min}}$ and v_i, and $pp_{j,j+1}$ denotes the positive propagation probability between users v_j and v_{j+1}.

After user v_i obtains an initial opinion, user v_i updates this opinion when his in-neighbors update their views. We calculate probability $P(s_{t+1} = 1)$, in which user v_i holds positive opinion at time step $t + 1$ after he forms an initial opinion. Two cases lead to this situation. First, user v_i holds positive opinion at time step t and still keeps positive at time step $t + 1$. Second, user v_i holds negative opinion at time step t and changes to be positive at time step $t + 1$. This process is explained by

$$P(s_{t+1} = 1) = \begin{cases} P(s_t = 1) * p_{s1}, & s_t = 1 \\ P(s_t = -1) * p_{t2}, & s_t = -1, \end{cases} \tag{27}$$

where $P(s_t = 1)$ denotes the probability that user v_i holds a positive opinion at time step t, $P(s_t = -1)$ denotes the probability that user v_i holds a negative opinion at time step t, p_{s1} denotes the stubborn probability for positive opinions, and p_{t2} denotes the transforming probability from negative to positive.

Similarly, we calculate probability $P(s_{t+1} = -1)$ that the opinion of user v_i is negative at time step $t + 1$. Two cases lead to this situation. First, user v_i holds negative opinion and retains the negative view at time step $t + 1$. Second, user v_i holds positive opinion at time step t, but changes to be negative at time step $t + 1$. Specifically, this probability is defined as

$$P(s_{t+1} = -1) = \begin{cases} P(s_t = 1) * p_{t1}, & s_t = 1 \\ P(s_t = -1) * p_{s2}, & s_t = -1, \end{cases} \tag{28}$$

where $P(s_t = -1)$ and $P(s_t = 1)$ denote the probability that user v_i holds negative and positive opinions at time step t, respectively, p_{s2} denotes the stubborn probability that user v_i keeps the negative opinion, and p_{t1} denotes the transforming probability that user v_i changes the opinion from positive to negative opinion.

The algorithm is presented as Algorithm 1. Based on influence graph $G = (V, E)$, the number of influential users is given as k. The number of influential users who have maximum positive influence is then calculated. Based on the class to which user v belongs, we could now predict the class to which user v will belong at the next time step.

4.2.2. Properties of the IMIC-OC model

The IMIC-OC model for influence maximization is NP-hard. Here, we mainly describe some properties of IMIC-OC model, such as monotonicity and submodularity.

Theorem 1. *Given influence graph $G = (V, E)$, the maximum positive influence $\phi_m(S)$ is monotone, and is expressed as*

$$\phi_m(S \cup \{u\}) - \phi_m(S) \geq 0. \tag{29}$$

Algorithm 1 IMIC-OC

Input: $G = (V, E)$, an expected number k of influential users S
Output: Influential users S

Let $S = \emptyset, S^* = \emptyset$
for iter=1:k **do**
  for each $u \in V \setminus S$ **do**
    set $S^* = S \bigcup \{u\}$
    for each $v \in N_{out}(u)$ **do**
      if $v \in I$ **then**
        calculate $path_{min}(S, v)$
        generate random value r_1
        if $r_1 < P(s_0 = 1)$ **then**
          $P = P \bigcup \{v\}$
          $\phi(S^*) = \phi(S^*) \bigcup \{v\}$
        else
          generate random value r_2
          if $r_2 < P(s_0 = -1)$ **then**
            $N = N \bigcup \{v\}$
          end if
        end if
      end if
      if $v \in P$ **then**
        generate random value r_3
        if $r_3 < p_{t1}$ **then**
          $P = P \setminus \{v\}, N = N \bigcup \{v\}$
          $\phi(S^*) = \phi(S^*) \setminus \{v\}$
        end if
      end if
      if $v \in N$ **then**
        generate random value r_4
        if $r_3 < p_{t2}$ **then**
          $P = P \bigcup \{v\}, N = N \setminus \{v\}$
          $\phi(S^*) = \phi(S^*) \bigcup \{v\}$
        end if
      end if
    end for
  end for
  $S = S \bigcup \{argmax_{u \in V \setminus S} |\phi(S^*)|\}, S^* = S$
end for

Proof. We obtain a set of influential users S and user $u \in V \setminus S$. We calculate $\phi_m(S)$ at step 1 and $\phi_m(S \cup \{u\})$ at step 2. Here, v_{s1} denotes the state of user v at step 1 and v_{s2} denotes the state of user v at step 2. With such information, we obtain

$$\phi_m(S \cup \{u\}) - \phi_m(S) = \underbrace{|\{v|v_{s1} \in I, v_{s2} \in P\}| + |\{v|v_{s1} \in N, v_{s2} \in P\}|}_{1} - \underbrace{|\{v|v_{s1} \in P, v_{s2} \in N\}|}_{2}, \tag{30}$$

where $v \in I$ denotes that user v is inactive, $v \in P$ denotes that user v holds a positive opinion, and $v \in N$ denotes user v holds a negative opinion.

The result of Eq. (30) is based on which part is larger in the right side of this equation. During information spreading, the most important function is to activate neighbors. Stubborn weight adds difficulty to changing minds from negative to positive than to activating inactive neighbors. Thus, the number of users who change opinions is smaller than the number of users who are activated. In other words, part 1 is larger than part 2, that is, $\phi_m(S \cup \{u\}) - \phi_m(S) \geq 0$. Hence, if we enlarge the number of influential users S, positive influence increases in the network. In conclusion, $\phi_m(S)$ is monotone. □

Theorem 2. *Given any influence graph $G = (V, E)$, the maximum positive influence $\phi_m(S)$ is submodular, and is expressed as*

$$\phi_m(S \cup \{u\}) - \phi_m(S) \geq \phi_m(T \cup \{u\}) - \phi_m(T) \quad \text{if } S \subseteq T \subseteq V \text{ and } u \in V \setminus T. \tag{31}$$

Table 1
Statistical data of the three real-world networks.

Dataset	Facebook	HEP-PH	Epinions
Number of nodes	14,980	34,546	75,879
Number of edges	14,359	421,578	508,837
Average degree	4.58	13.11	8.43

Proof. We obtain two sets of influential users, S and T, $S \subseteq T \subseteq V$, and user $u \in V \setminus T$. Maximum positive influence $\phi_m(S \cup \{u\})$ and $\phi_m(T \cup \{u\})$ are expressed as

$$\phi_m(S \cup \{u\}) = \phi_m(S) + \phi_m(\{u\}) - \phi_m(S) \cap \phi_m(\{u\}), \tag{32}$$

$$\phi_m(T \cup \{u\}) = \phi_m(T) + \phi_m(\{u\}) - \phi_m(T) \cap \phi_m(\{u\}). \tag{33}$$

According to Eqs. (32) and (33), we could obtain $\phi_m(S \cup \{u\}) - \phi_m(S) = \phi_m(\{u\}) - \phi_m(S) \cap \phi_m(\{u\})$ and $\phi_m(T \cup \{u\}) - \phi_m(T) = \phi_m(\{u\}) - \phi_m(T) \cap \phi_m(\{u\})$. Thus, the result depends on which is larger, $\phi_m(S) \cap \phi_m(\{u\})$ or $\phi_m(T) \cap \phi_m(\{u\})$. According to Theorem 1, $\phi_m(S)$ is monotone, such that $\phi_m(T) \geq \phi_m(S)$ if $S \subseteq T \subseteq V$. We obtain $\phi_m(S) \cap \phi_m(\{u\}) \leq \phi_m(T) \cap \phi_m(\{u\})$. Thus, $\phi_m(S \cup \{u\}) - \phi_m(S) \geq \phi_m(T \cup \{u\}) - \phi_m(T)$ if $S \subseteq T \subseteq V$ *and* $u \in V \setminus T$. □

5. Experiments

5.1. Datasets

We crawled the data in Facebook from March 18th to March 28th, 2014 on a specific topic. Our dataset includes 14,980 users, 1,955 posts, 9,721 comments, and 22,078 likes. We assume that if users like a post, then they have the same opinions as the authors of the original posts. We utilize the dataset from Facebook to validate the IC-OC model, and to find influential users S. We also use two other real networks to find influential users S. First, *HEP-PH*,[4] which is an undirected graph, is the co-authorship graph from the "high energy physics phenomenology" section in the e-print, arXiv.[5] Second, the dataset of Epinions,[6] which is a directed graph, is a who-trust-whom online social network of a general consumer review site, Epinions.[7] Table 1 illustrates the basic information of the three real networks.

5.2. Parameter estimation

The probabilities in the process of information spreading are estimated in this section. Fig. 3 exhibits the distribution of adoption probability p_a, which follows power-law distribution. Fig. 4 indicates the distribution of conforming probability p_c. Therefore, we obtain the distribution of positive propagation probability pp in Fig. 5, which also follows power-law distribution.

5.3. Validation of the IC-OC model

The comparison of the proposed model and baseline methods is made to prove that the proposed model performs better than the baseline methods.

5.3.1. Baseline methods

IC-N: Proposed by Chen et al. [3], this algorithm assumes that users would not change their opinions after they obtained initial opinions.

LT-IO: Proposed by Li et al. [20], this novel linear threshold model (LT) is used for modeling information spreading; it introduces broadcast and current attitude in considering user attitude modification.

5.3.2. Performance of the IC-OC model

We examine the performances of adoption and positive cascades. Adoption cascades denote the number of users who adopt the information, while positive cascades indicate the number of users who hold positive opinions in the final state.

Coefficient of determination (R^2), sum of square for errors (SSE), and root-mean-square error (RMSE) are adopted to determine the performance for adoption cascades (Table 2) and for positive cascades (Table 3). The IC-OC model has greater R^2 values as well as smaller SSE and RMSE values for both adoption and positive cascades. These results indicate that the

[4] http://snap.stanford.edu/data/cit-HepPh.html.
[5] http://www.arXiv.org.
[6] http://snap.stanford.edu/data/soc-Epinions1.html.
[7] http://www.epinions.com.

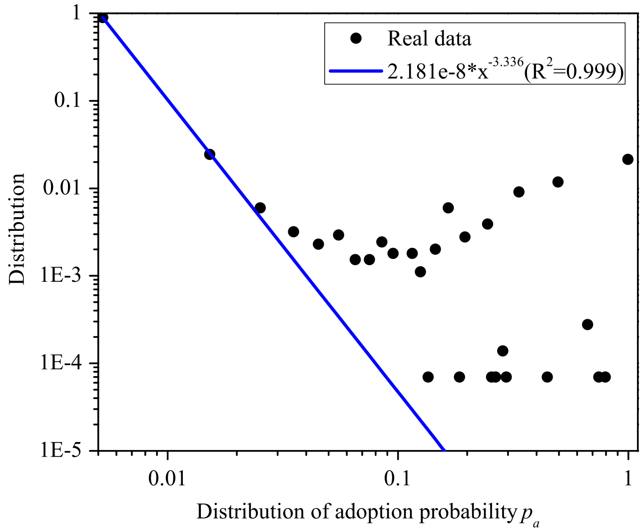

Fig. 3. Distribution of adoption probability p_a.

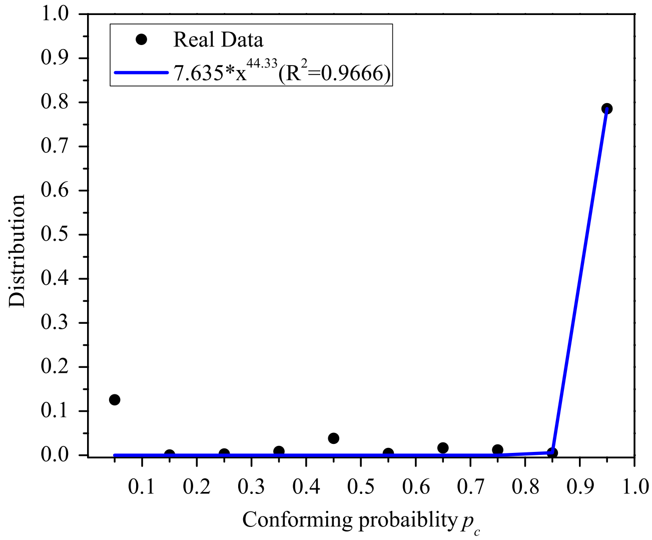

Fig. 4. Distribution of conforming probability p_c.

IC-OC model has better performance, compared with the baseline methods. Relative to adoption cascades, the IC-OC model has 10.5% and 11.6% improvements compared with the IC-N and LT-IO models, respectively. Relative to positive cascades, the IC-OC model has 3.1% and 20.8% improvements compared with the IC-N and LT-IO models, respectively.

Performance metrics are defined as

$$R^2 = \frac{\sum_{i=0}^{n} (\hat{x}_i - \bar{x})^2}{\sum_{i=0}^{n} (x_i - \bar{x})^2}, \tag{34}$$

$$SSE = \sum_{i=0}^{n} (x_i - \hat{x}_i)^2, \tag{35}$$

$$RMSE = \sqrt{\frac{\left(\sum_{i=0}^{n} x_i - \hat{x}_i\right)^2}{n}}, \tag{36}$$

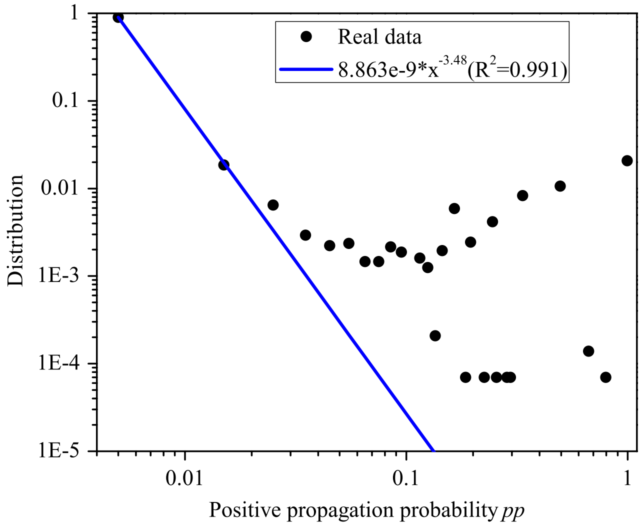

Fig. 5. Distribution of positive propagation probability *pp*.

Table 2
Performance for adoption cascades.

Model	R^2	SSE	RMSE
LT-IO	0.5923	163,502	11.8467
IC-N	0.6035	159,045	11.6841
IC-OC	**0.7086**	**116,864**	**10.0156**

Table 3
Performance for positive cascades.

Model	R^2	SSE	RMSE
LT-IO	0.5231	142,056	11.0425
IC-N	0.7000	89,380	8.7591
IC-OC	**0.7311**	**80,110**	**8.2924**

where X is a series of values to be predicted, x_i is the ith value in series X, $\bar{x}$ is the mean of series X, and $\hat{x}_i$ is the predicted value of x_i.

5.4. Influence maximization

First, we investigate how parameter c affects positive influence. Second, influential users are found in three real networks, which are Facebook, HEP-PH, and Epinions. The results of maximum positive influence, which are obtained using different methods, are shown in this section. Experimental results manifest that the IMIC-OC model calculated the largest maximum positive influence.

5.4.1. Parameter analysis

Parameter c has an impact on stubborn weight w_{ii}, which represents how users keep their initial opinions. A greater parameter c value indicates that users have stronger willingness to keep their initial opinions. Fig. 6 illustrates maximum positive influence with respect to parameter c by varying the number of influential users from 1 to 50. This figure shows that positive influence increases as parameter c rises. Thus, increasing parameter c has greater positive influence and smaller marginal increments.

5.4.2. Baseline methods

Random: An algorithm selects k influential users randomly from entire users V.

MaxDegree: An algorithm selects top-k users who have maximum out-degrees as influential users.

Greedy: An algorithm obtains approximate solution of influence maximization.

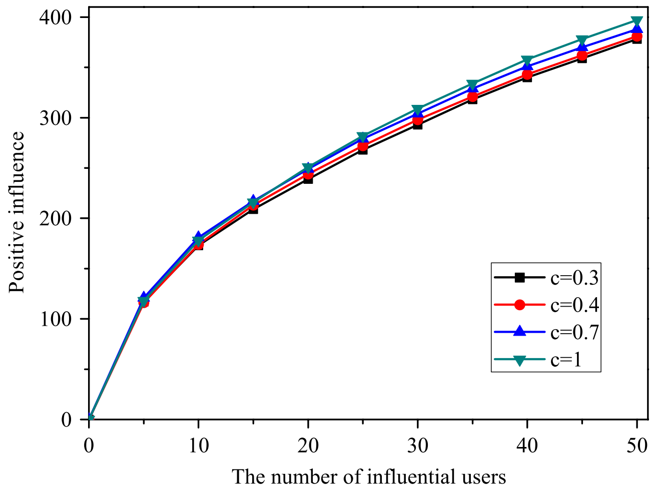

Fig. 6. How positive influence changes with different parameter c.

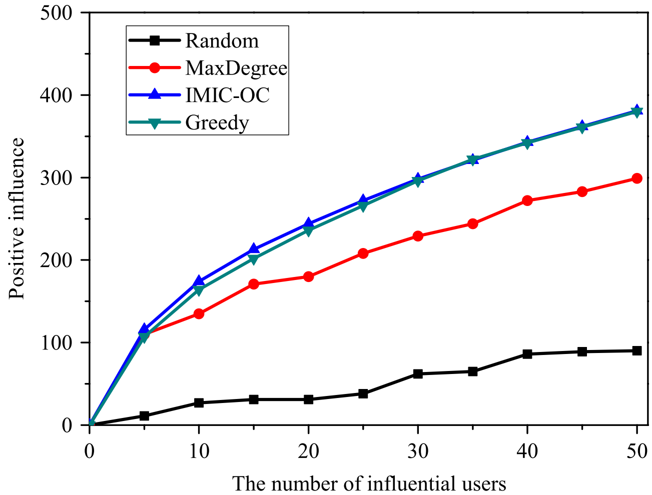

Fig. 7. The positive influence in Facebook.

5.4.3. Comparisons among algorithms

The performances of the IMIC-OC model and baseline methods are discussed in this sub-section. Maximum positive influence $\phi_m(S)$ is calculated by varying the number of influential users from 1 to 50.

Fig. 7 shows that on Facebook, the maximum positive influence of the IMIC-OC model is close to the Greedy model. In addition, the Random model has the smallest maximum positive influence among these algorithms, and the positive influence of the MaxDegree model is smaller than the IMIC-OC model.

Fig. 8 exhibits the maximum positive influence on NEP-PH. The result is similar with that from the Facebook dataset.

On Epinions, the maximum positive influence of the MaxDegree model is close to the IMIC-OC model in Fig. 9. This result depends on the structure of the network, suggesting that users who have higher degrees are more influential in this network.

In conclusion, the Random model has the worst performance among the three real networks; hence, this model is not feasible in calculating maximum positive influence. The result of the MaxDegree model depends on the structure of the network. If influential users have high degrees, the result of the MaxDegree model is close to the IMIC-OC model. The result of the IMIC-OC model is close to the Greedy model in all three networks, indicating that the IMIC-OC model is feasible in calculating maximum positive influence.

6. Conclusion

The proposed IMIC-OC model for influence maximization aims to find a set of users who maximize the adoption of information. Further, the IC-OC model for information spreading is proposed to explain how users build their opinions.

The IC-OC model introduces stubborn and transforming probabilities. These probabilities are used to predict whether users would change initial opinions when exposed to different opinions. At the beginning, information spreaders activate

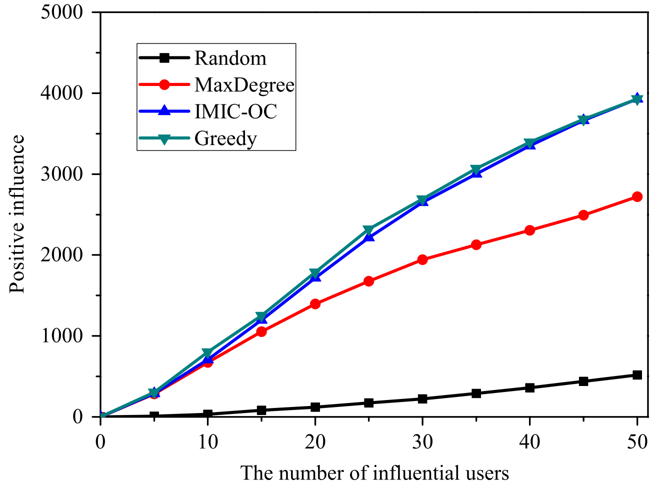

Fig. 8. The positive influence in HEP-PH.

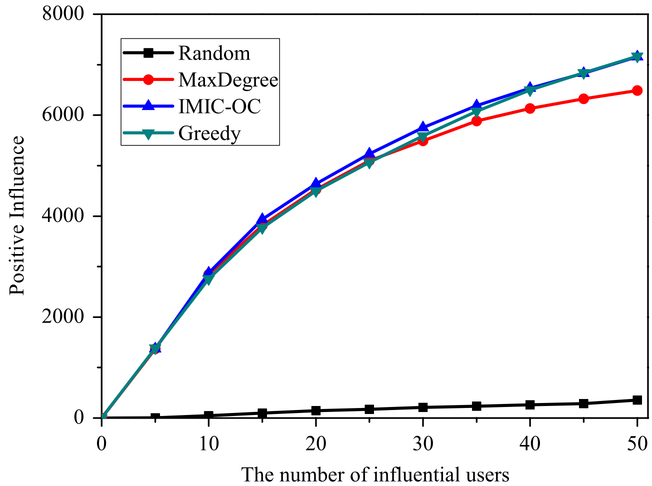

Fig. 9. The positive influence in Epinions.

their neighbors to adopt information based on adoption probability. Conforming probability denotes how neighbors keep the same opinions as the users who spread information. As time passes, different opinions emerge in the network. Users could then possibly change their minds by transforming probability or keep the same opinions by stubborn probability.

Next, we verify the IC-OC model on the Facebook dataset. Results show that the IC-OC model has better performance than the baseline methods. Relative to adoption cascades, the IC-OC model has 10.5% and 11.6% improvements compared with the IC-N and LT-IO models, respectively. Meanwhile, relative to positive cascades, the IC-OC model has 3.1% and 20.8% improvements compared with the same baseline methods, respectively.

The IMIC-OC model for influence maximization is proposed to find influential users. The minimum path to calculating initial opinions of users is obtained first. Then, we update user opinions at each step until no neighbor changes opinion. Ultimately, users who activate maximum positive influence are selected as influential users. Positive influence is calculated based on three real networks. The analyses of Facebook and NET-PH data yield similar results. Meanwhile, the Random model has the worst performance, while the IMIC-OC model has the largest maximum positive influence. The IMIC-OC model is close to the Greedy model, which proves that the IMIC-OC model is a feasible method in calculating positive influence. However, results from Epinions exhibit a difference. The MaxDegree model is close to the IMIC-OC model, indicating that more influential users have more degrees.

Acknowledgments

This work was supported in part by National Key Basic Research and Department (973) Program of China (No. 2013CB329606), National Key Technology R&D Program (No. 2014BAH23F03), Fundamental Research Funds for the Central

Universities of China (Grant No. 2014RC0501), Natural Science Foundation of China (No. 61402045), and Specialized Research Fund for Doctoral Program of Higher Education (No. 20130005110011).

References

[1] S. Bhagat, A. Goyal, L. Lakhmanan, Maximizing product adoption in social networks, in: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, 2012, pp. 603–612.
[2] D. Kempe, J. Kleinberg, E. 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, 2003, pp. 137–146.
[3] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, Y. Yuan, Efficient influence maximization in social networks, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge discovery and data mining, 2009, pp. 199–208.
[4] W. Chen, Y. Wang, S. Yang, Influence maximization in social networks when negative opinions may emerge and propagate, in: Proceedings of the 11th SIAM International Conference on Data Mining, 2011, pp. 379–390.
[5] Y. Tang, X. Xiao, Y. Shi, Influence maximization: Near-optimal time complexity meets practical efficiency, in: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, 2014, pp. 75–86.
[6] E.D. Demaine, M. Hajiaghayi, H. Mahini, D.L. Malec, S.R.A. Sawant, M. Zadimoghadam, How to influence people with partial incentives, in: Proceedings of the 23rd International Conference on World Wide Web, 2014, pp. 937–948.
[7] E. Cohen, D. Delling, T. Pajor, R.F. Werneck, Sketch-based influence maximization and computation: Scaling up with guarantees, in: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, 2014, pp. 629–638.
[8] H. Zhang, D.T. Nguyen, H. Zhang, M.T. Thai, Least cost influence maximization across multiple social networks, IEEE/ACM Trans. Netw. 99 (2015) 1–11.
[9] A. Goyal, W. Lu, L.V. Lakshmanan, Celf++: Optimizing the greedy algorithm for influence maximization in social networks, in: Proceedings of the 20th International Conference Companion on World Wide Web, 2011, pp. 47–48.
[10] M. Heidari, M. Asadpour, H. Faili, Smg: Fast scalable greedy algorithm for influence maximization in social networks, Physica A 420 (2015) 124–133.
[11] G. Li, S. Chen, J. Feng, K. Tan, W. Li, Efficient location-aware influence maximization, in: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, 2014, pp. 87–98.
[12] Z. Wang, H. Wang, Q. Liu, E. Chen, Influential nodes selection: a data reconstruction perspective, in: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, 2014, pp. 879–882.
[13] S. Cheng, H. Shen, J. Huang, W. Chen, X. Cheng, Imrank: Influence maximization via finding self-consistent ranking, in: Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2014, pp. 475–484.
[14] C. Zhou, P. Zhang, W. Zang, L. Guo, Maximizing the long-term integral influence in social networks under the voter model, in: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, 2014, pp. 423–424.
[15] Y. Wu, Y. Yang, F. Jiang, S. Jin, J. Xu, Coritivity-based influence maximization in social networks, Physica A 416 (2014) 467–480.
[16] H. Li, S.S. Bhowmick, A. Sun, J. Cui, Conformity-aware influence maximization in online social networks, VLDB J. 24 (2015) 117–141.
[17] T. Zhu, B. Wang, B. Wu, C. Zhu, Maximizing the spread of influence ranking in social networks, Inform. Sci. 278 (2014) 535–544.
[18] J. Guo, P. Zhang, C. Zhou, Y. Cao, L. Guo, Personalized influence maximization on social networks, in: Proceedings of the 22nd ACM International Conference on Conference on Information and Knowledge Management, 2013, pp. 199–208.
[19] J.-R. Lee, C.-W. Chung, A query approach for influence maximization on specific users in social networks, IEEE Trans. Knowl. Data Eng. 27 (2015) 340–353.
[20] S. Li, Y. Zhu, D. Li, D. Kim, H. Ma, H. Huang, Influence maximization in social networks with user attitude modification, in: IEEE International Conference on Communications, 2014, pp. 3913–3918.
[21] X. Si, Y. Liu, F. Xiong, Y. Zhang, F. Ding, H. Cheng, Effects of selective attention on continuous opinions and discrete decisions, Physica A 389 (18) (2010) 3711–3719.
[22] X.M. Si, Y. Liu, Empirical analysis of interpersonal interacting behavior in virtual community, Acta Phys. Sin. 60 (7) (2011) 78903.
[23] A. Das, S. Gollapudi, K. Munagala, Modeling opinion dynamics in social networks, in: ACM Conference on Web Search and Data Mining, 2014, pp. 403–412.
[24] B. Chazelle, Natural algorithms and influence systems, Commun. ACM 55 (12) (2012) 101–110.
[25] P. Clifford, A.W. Sudbury, A model for spatial conflict, Biometrika 60 (1973) 581–588.
[26] R.A. Holley, T.M. Liggett, Ergodic theorems for weakly interacting infinite systems and the voter model, Ann. Probab. 3 (4) (1975) 643–663.
[27] Z. Wang, Y. Liu, Y. Zhang, Z. Wang, Freezing period strongly impacts the emergence of a global consensus in the voter model, Sci. Rep. 3597 (2014) 1–10.