Applied Intelligence (2019) 49:912–928
https://doi.org/10.1007/s10489-018-1303-2

Influence maximization on signed networks under independent cascade model

Wei Liu[1,2,3] · Xin Chen[1] · Byeungwoo Jeon[3] · Ling Chen[1] · Bolun Chen[2]

Published online: 10 October 2018

Abstract
Influence maximization problem is to find a subset of nodes that can make the spread of influence maximization in a social network. In this work, we present an efficient influence maximization method in signed networks. Firstly, we address an independent cascade diffusion model in the signed network (named SNIC) for describing two opposite types of influence spreading in a signed network. We define the independent propagation paths to simulate the influence spreading in SNIC model. Particularly, we also present an algorithm for constructing the set of spreading paths and computing their probabilities. Based on the independent propagation paths, we define an influence spreading function for a seed as well as a seed set, and prove that the spreading function is monotone and submodular. A greedy algorithm is presented to maximize the positive influence spreading in the signed network. We verify our algorithm on the real-world large-scale networks. Experiment results show that our method significantly outperforms the state-of-the-art methods, particularly can achieve more positive influence spreading.

Keywords Influence maximization · Independent cascade model · Signed networks

1 Introduction

With the rapid expansion of electronic devices and the growth of social media, people are more closely connected by social networks. The social network is made up of a set of social actors (such as individuals or organizations) and their relations. Usually the social network can be formalized as a graph where each node corresponds to a social actor, and each edge corresponds to the relationship between a pair of social actors. Such relations can be cooperators, friends, enemies, etc. Recently, with the popularization of many large social networks such as Facebook, Linkedin, Twitter, WeChat and Google+, social networks have become a hotspot of research.

In the research on the social network, the problem of how the information spreads through the interconnections of the people in the network has drawn more and more attention. From the marketing strategy such as "word of mouth" and "viral marketing" [1–5], it was found that the marketer can select a representative subset of individuals in the crowd so that they can produce a greater cascade effect. This is the influence maximization(IM) problem. When the IM problem is introduced into the social network research field, it becomes a research hotspot in recent years. Domingos et al. first [4] introduced the problem of influence maximization and presented a mathematical model for the problem. After that, various propagation models and methods to solve the problem have been proposed. In recent years, the IM problem has received extensive attention from scholars. The research on the influence of these

✉ Wei Liu
yzliuwei@126.com

Xin Chen
2548445680@qq.com

Byeungwoo Jeon
bjeon@skku.edu

Ling Chen
lchen@yzu.edu.cn

Bolun Chen
chenbolun1986@163.com

[1] College of Information Engineering of Yangzhou University, Yangzhou 225127, China

[2] The Laboratory for the Internet of Things and Mobile Internet Technology of Jiangsu Province, Huaiyin Institute of Technology, Huaiyin 223002, China

[3] School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon, South Korea

social networks is very important in advertising, marketing, messaging and so on.

A social network can be formulated as a directed graph $G = (V, E)$, where V is the set of nodes which represent the individuals in the network, and E is the set of links which reflect a collection of interactions and relationships between individuals in the social network. Let $|V| = n$, $|E| = m$, and assume that each directed e link in G is assigned a propagation probability $p(e) \in [0, 1]$. Let M be a probabilistic model that controls the nodes in G to influence its neighbors. Suppose k is a positive integer ($k<n$), the IM problem is to search for k nodes in G that can directly or indirectly influence the largest number of nodes in G. Such k nodes are called the seeds of influence spreading. For a seed set $S \subset V$ and $|S| = k$, let $I(S)$ be the number of nodes that are influenced by the seed set S under some probabilistic model. The result of maximizing social network influence can be represented as $S*$ of the following optimization:

$$S^* = \underset{S \subset V, |S|=k}{\arg\max}\ I(S) \tag{1}$$

In a probabilistic model of influence spreading, the seed nodes are first activated. At each time, after a node being activated, it will attempt to activate its inactivated neighbors. When the neighbor node is activated, it will attempt to activate its own inactive neighbor nodes. This process will repeat until no new node in the network can be activated. There are two widely used diffusion models for influence spreading, namely linear threshold (LT)model and independent cascade (IC)model.

The IC model is a kind of probability model simulating the viral marketing and rumor spreading. When a node v is activated in this model, it will attempt to activate its inactive neighbor node w under the probability p_{vw}. An active node v has only one chance to activate each of its inactive neighbor node w. Such attempts are mutually independent for different neighbors, namely, the activation of v to w will not be affected by the influences from other neighbors of w. Many studies on IM are based on the IC model. But since the IC model is a probabilistic model, its activation process is uncertain, and the result of influence spreading may vary widely for different simulations even in the same network with the same seed nodes.

The LT model is a value accumulation model. In this model, each node in the network is assigned a specific threshold for activation. For a given network $G(V, E)$, let $N(v)$ be the set of active neighbor nodes of v, $\theta_v \in [0, 1]$ be the threshold assigned to node v in V, b_{uv} be the influence weight between nodes u and v satisfying

$$\sum_{u \in N(v)} b_{uv} \leq 1$$

The influence spreading under this model is as follows: (a) For a node v, if

$$\sum_{u \in N(v)} b_{uv} \geq \theta_v$$

the node v will become active. (b) After node v becoming active, it will attempt to activate its inactive neighbor nodes. (c) repeat the process (a) and (b) until no more nodes can be activated

As mentioned above, most of the current models focus on the process of influence spreading in unsigned networks. However, the social relationships between actors might not be always friend (i.e., positive or trust) relationships. Many social networks can be described as a signed one, in which each edge is signed as positive or negative wherein, the positive and negative links represent positive and negative relationships, respectively. In the real world, there are many complex networks having antagonistic relations, particularly in the field of information, biological as well as the social networks.

Recently, link prediction in signed social networks [6] has drawn much attention from the researchers. In signed social networks, links between nodes can either be positive or negative. The positive link in the network is usually denoted by the sign of "+", while the negative link is denoted as "-". In undirected social networks, the signs on the edges indicate the relations between the individuals represented by the nodes, while in a directed social network, the sign on the edges indicates the rank of the individuals. In undirected social networks, the positive link indicates the trust, friendship, support or love among users, while the negative link indicates distrust, hostility, dislike and opposition among users. In directed social networks, such sign on a directed edge reflects the social status or prestige of the person the nodes represented. Node u has a lower status than v if there is a positive link from u to v or a negative link from v to u For example, users of the consumer review site Epinions [7] can decide to trust or distrust other users according to their comments. The social website Slashdot [8] mainly focuses on news concerning technology, and users can make friends (like) or foes (dislike) by clicking on the comments of articles.

The signed properties of the edges have great theoretical significance in analyzing, understanding and predicting the complex network topology structure, function and dynamics behavior. Besides, the signed network has widespread application for the personalized recommendation, public opinion monitoring, user personality analysis and clustering.

Recently, IM in signed networks has drawn much attention of the researchers [9, 10]. In a signed network, positive links propagate influence in a positive way, while negative links transmit influence in an opposite way, namely,

one is more likely to accept a friend's positive influence or its foe's negative influence. Social influence may become more complicated when multiple competing processes are launched on the network. For instance, several merchants may sell the same products using competing viral marketing strategy [11–13]. In a signed network, not only positive influences but also negative influences can spread. For instance, if you know from one of your friends that the food in a restaurant was too difficult to eat, you probably won't go to this restaurant again. In addition, you may discourage your other friends and relatives to go to the restaurant. Sometimes, negative influences often have stronger and wider impact. Therefore, it is very important to integrate negative opinions into the spreading model and compete for its influence with the positive one. Hence, it is not appropriate to directly apply the traditional IM algorithms in an unsigned network to a signed network. And it is necessary to advance effective approaches for influence maximization in signed networks which can make full use of negative links and negative opinions.

In this study, we first put forward an Independent Cascade diffusion model in Signed Network (named SNIC)for describing two opposite types of influence spreads in a signed network which can obtain better results. Specifically, we make the following contributions in this paper:

1. We propose a diffusion model named SNIC to simulate information diffusion of the two opposite influences in real signed network properly.
2. We propose a propagation function for estimating the spreading of a given seed set. We also prove that such propagation function is nonmonotonic and non-submodular, and hence the greedy algorithm can guarantee the approximation ratio.
3. We devise an efficient greedy algorithm to solve the IM problem under SNIC
4. We present the concept of the independent propagation path set and address an algorithm for constructing the set of independent propagation paths and computing their probability
5. We define the gain of a propagation path to maximize the number of positive-activated nodes in the path. We give a method to construct activating set for each node. The activating set consists of the nodes that can activate it with the maximum gain.
6. To reduce the computation time, we use a pruning technique to eliminate the paths which have very low propagation probabilities and are almost impossible to complete the influence spreading.
7. Experimental results on real-world networks validate that the proposed algorithm outperforms other methods in term of positive influence spreads in signed networks.

The remainder of this paper is organized as follows. In Section 2, we briefly review the related works in influence maximization. Section 3 defines the spreading model SNIC and independent propagation paths and presents an algorithm for constructing the set of independent propagation paths and computing their probability. Section 4 defines influence spreading set and spreading function, and proves that the spreading function is monotone and submodular. Section 5 presents a greedy-based algorithm for IM problem in the signed networks. Experimental results have been shown and analyzed in Section 6. In Section 7, we offer a conclusion and future work.

2 Related work

Recently, some new influence propagation models and algorithms for influence maximization have been presented based on the theories and techniques in the areas of mathematics, computer science, and data mining.

Kempe et al. [14] proposed the triggering model (TR) to generalize the aforementioned IC and LT models In 2005, Kempe et al. [15] advanced the decreasing cascade (DCM) model which considered the attenuation effect of the influence between nodes. Based on the IC model, Kimura et al. [16] presented a cascade model considering the influence of the shortest path (SPM).

Kempe et al. [14] formatted the IM problem as a discrete optimization of selecting k nodes in the network as a seed set which can make the spread of influence maximization in the network. They proved that the problem of optimizing (1) is NP-hard under both the IC and LT propagation models. Moreover, the computation of the expectation of influence spreading function $E[I(S)]$ in this optimization problem is #P-hard in both IC and LT model. At the same time, they also proved that in the IC model and LT model, the function $E[I(S)]$ reflecting the influence spread is monotone and submodular. Due to these two properties, Kempe et al. proposed a greedy approximation algorithm KKT which maximized influence spread under these two propagation models. The algorithm first sets the initial seed set S as an empty set Φ. Then it iteratively chooses the node u that can maximize the increase in $E[I(S)]$ and adds u to S. The iterations are repeated until the size of S reaches k. In each iteration, the node u is selected by the following optimization:

$$u = \arg\max_{v \in V/S}(E[I(S \cup \{v\})] - E[I(S)]) \tag{2}$$

Their experimental results show that the greedy algorithm outperforms the heuristic algorithm. They also prove that the greedy algorithm can obtain (1-1/e)-approximate solution for the problem of optimizing (1).

Although the greedy method is conceptually easy and effective, it is difficult to implement since the computation of $E[I(S)]$ is #P-hard. In each iteration of the greedy algorithm, the increase of influence spread should be estimated for every inactive node in the network, and the one with the maximum increment of influence spread is selected to join the seed set. This is very time consuming, so it is not practical for large networks.

Leskovec et al. [17] presented an improvement of KKT which was called CELF (Cost-Effective Lazy Forward). CELF algorithm uses a new method for selecting initial seed for optimization. Using the submodularity of the influence spread function, in each iteration of CELF, many nodes do not need to reassess their increments of influence spread. This greatly reduces the computation time for calculating the nodes' influence spread. Experimental results show that the CELF algorithm, which selects the proper initial seeds, is 700 times faster than the greedy algorithm, and can achieve very close results with the greedy algorithm. Goyal et al. [18] further improved the CELF algorithm and proposed CELF++ method. Compared with CELF algorithm, the CELF++ algorithm can increase the efficiency by $35\% \sim 55\%$.

Estevez et al. [19] proposed an improved greedy algorithm SCG (Set Cover Greedy). Considering the overlapping of neighbor nodes, in the greedy selection of the most influence, SCG algorithm avoids selecting the nodes with overlapping neighbors. The algorithm initially sets all nodes as "uncover". Then, in each iteration, the node with the most uncover degree is selected as a seed, and all its neighbors are marked as "covered". Such seed selection will be iterated until k seed nodes are selected. Experimental results show that the computation time of SCG is much lower than that of the greedy algorithm.

Bharathi et al. [20] studied the competition between nodes and established an easier model of influence spreading. They advanced the first-mover strategy to maximize the influence when there exists a competition in the network. They also give an algorithm FPTAS (Full Polynomial Time Approximation Scheme) for influence maximization in a network of the tree structure.

Wu Peng et al. [21] studied the Influence Blocking Maximization (IBM) problem on two competitive propagation models describing competitive propagation processes in two classic situations in OSNs.

Chen Wei et al. [22] improved the efficiency of the maximization of influence in two aspects. Firstly, they presented an improved greedy algorithm NewGreedyIC. In each iteration, the algorithm removes the edge of unsuccessful spreading from network G, and gets a new network G'. The next seed will be selected in the new network G'. Experimental results show that the improved greedy algorithm can greatly reduce the computation time. In addition, they also proposed an improved heuristic algorithm Degree-Discount to select the initial seed node. Experiments show that the Degree-Discount algorithm can obtain more widely spreading than the heuristic algorithm based on degree and distance.

Chen Wei et al. [23] proposed a fast method for calculating the influence spread on the directed acyclic graph (DAG). The computation time of this method is linear with the size of the graph, and a scalable IM algorithm for the LT model was proposed based on this method. The algorithm can be applied to a large-scale network with millions of nodes and is much faster than the greedy algorithm.

Li et al. [24] presented an algorithm to solve the IM problem using community detecting in the network. The algorithm firstly partitions the network into k communities using a modified $k-$means algorithm. The nodes in the same community with higher similarity and the similarity between the nodes in different communities is much lower. Then the central node in each community is selected as a seed. The experimental results show that the central nodes selected by community detecting algorithm have higher influence spreading.

Zhu et al. [25] applied semi-definite programming to solve the problem of maximization of influence. Considering the timesensitive features of the information transmission in the realworld network, they proposed a new propagation model to transform the influence maximization problem into an optimization. They designed optimization algorithms by using semidefinite programming. When the size of node set S is not limited, the approximate performance ratio of the designed approximate algorithm is roughly 0.857. When the size of the initial node set S is limited to a certain range, the approximate performance ratio of the designed algorithm can reach 1-1/e.

In [26], the authors proposed a new algorithm to solve the minimum positive influence dominating set (MPIDS)problem and then applied it to the influence maximization in the online social network. In the proposed algorithm, each node is assigned a learning automaton and the algorithm tries to find the near-optimal positive influence dominating set in the input network with the help of learning automata. The results of the experiments performed have confirmed that the proposed algorithm has better performance in terms of the total influence spread.

Recently, many new influence maximization algorithms have been proposed, such as PPR(Personalized PageRank) algorithm [27] and STORIE (A holistic approach) [28], the hop-based algorithms [29] discrete particle swarm optimization based algorithm [30], community structure based algorithm [31], constrained simulated annealing based

algorithm [32], CascadeDiscount algorithm (Scalable influence maximization algorithm) [33] and so on. In addition, there are some new topics advanced on the influence maximization, such as the timing constraints in IM problem advanced by Nan Du et al. [34], the Misinformation Containment(MC) problem proposed by Huiyuan Zhang et al. [35] the Seed Activation Scheduling Problem (SASP) proposed by Mohammadreza Samadi [36] and so on Although several improvements [17–24] of the original greedy algorithm have been proposed, the greedy method is still inefficient. We believe that we need to find alternatives, such as new heuristic algorithms or an approximation of the influence spreading function, to solve the efficiency problem in IM.

In recent years, some works on IM in signed networks are reported. Ajitesh Srivastava et al. [37] studied the two-competing influence spreads in a signed network under the IC model. They presented an approximate approach to estimate the probability for a node be activated at a given time, and a heuristic method for influence maximization in the signed network. Chen Wei et al. [38] presented an extended IC model that combined and propagated negative opinions in a signed network. They first designed an algorithm to estimate the influence of treestructured networks. Then they extended the algorithm for influence maximization for general networks. Chen Shubo et al. [39] used an extended vote model to simulate the influence spread on signed networks and presented an integrated PageRank method for IM in signed networks. Siwar Jendoubi et al. [40] proposed an evidential approach for maximizing positive opinion influence. Dong Li et al. [10] addressed a novel strategy based on simulated annealing for finding the seed node set with maximum positive influence in a signed network. The study also proposed two heuristics which can speed up the convergence process and guarantee better performance Gerald Petz et al. [41] investigated the differences between social network services regarding opinion mining and evaluated the performance in terms of correctness of sentence splitting, stemming, POS tagging and parsing. Experimental results show that both extensive text preprocessing prior to sentiment analysis tasks and the improved algorithms that take noisy text into account seem to be reasonable in order to cope with texts published by users on social media platforms.

In the existing methods for IM in signed networks, the main difficulty is the estimation of expected influence spread for a given seed set. The commonly used way is using many simulations. In the IC model, even for a single simulation requires a large amount of computation time. Therefore, it is necessary to design a method for estimation of the influence spread through analytical computation to avoid numerous simulations.

3 Influence spreading model and independent propagation paths

In this section, we first introduce an independent cascade model named SNIC for describing two opposite types of influence spreads in a signed network. Then, we define the independent propagation path set which consists of m paths with the highest probability from a certain node to all the other nodes. With the help of independent propagation paths, the seed node selection process can be accelerated.

3.1 SNIC: the independent cascade diffusion model in signed network

A signed network can be formulated as a graph $G = (V, E)$, where V is the set of nodes, $E{=}E^p \cup E^n$is the set of two types of edges. Here, E^p and E^n are respectively the sets of positive and negative edges. Let $A \in R^{N\times N}$ be the adjacency matrix of G, where $A_{ij} = 1$, $A_{ij} = -1$ and $A_{ij} = 0$ indicate that the link from v_i to v_j is positive, negative and missing respectively.

Nodes connected by a positive link are likely to share the same point of view to a great extent, while nodes connected by a negative link probably have opposite views. Let P_{uv} be the propagation probability on the link (u, v), namely the probability of node v being influenced by u. In the signed network, there are two opposite influences representing opposite point of views for the same thing, such as agreeing or disagreeing, trust or distrust. Accordingly, there will be two different states for an activated node: 'positive-activated' and 'negative-activated' in the network, indicating the type of influence the node accepted.

We define the IC model for spreading of the two opposite influences in such signed network, which is named as SNIC, as follows:

1. Initially, k seed nodes are selected and activated in either positive-activated state or negative-activated state, while all the non-seed nodes are in the inactive state.
2. At each time step, every activated node attempts to activate its adjacent nodes. Suppose u has been activated in the state f, and v is a neighbor of u. If the sign on edge (u, v) is "+", u will activate v into the state f with the probability of P_{uv}. If the sign on edge (u, v) is "-", u will activate v into the state $\bar{f}$ with the probability of P_{uv}. Here, $\bar{f}$ denotes the opposite state of f. Once u fails to activate v, it has no more chance to activate v again.
3. The activation of u to v will not be affected by the influences from other neighbors of v. Namely, the attempts by the neighbors for activation v are

mutually independent. If an inactive node v is activated by multiple attempts by its active neighbors, it will randomly select one of such activations and change its state accordingly.

4. The newly activated node v attempts to activate its inactive neighbor nodes according to the above-mentioned rules. Such activation process will be repeated until no more nodes can be activated in the network.

Suppose k is a positive integer ($k<n$), the IM problem in the signed network G is to search for k seed nodes with their signs that can positively influence the largest number of nodes, and negatively influence the least number of nodes in G. For a seed set $S \subset V$ and $|S| = k$, and an assignment function $F : S \to \{+, -\}$ to set the active state of the nodes in S let $N_p(F, S)$ and $N_n(F, S)$ respectively be the number of nodes positively and negatively influenced by the seed set S under the SNIC model mentioned above. The problem of maximizing influence in the signed network is to find a seed set S satisfying $|S| = k$, and assignment function F which can maximize $I(F, S) = N_p(F, S) - N_n(F, S)$ Namely, our goal is to find a seed set S^* and assignment function F to optimize

$$S^* = \underset{S \subset V, |S|=k, F}{\arg\max} \; I(F, S) \tag{3}$$

It can be shown that the problem of optimizing (3) is NP-hard under the SNIC propagation model mentioned above. Although the greedy method can be used, it is difficult to implement since the computation of $N_p(F, S) - N_n(F, S)$ by the Monte-Carlo simulations is #P-hard. Since it is very time consuming, it is necessary to find an approach without using expensive simulations to estimate the spread.

3.2 Independent propagation paths and spreading probability

Since estimating the influence spreading of a seed set S is #P-hard, most methods employ Monte-Carlo simulations. However, such simulation is very time consuming, we need a new model to describe the influence spreading. The difficulty in computing the influence spreading is that the nodes can indirectly activate another inactive node through different paths which may partially overlap. Therefore, we propose an influence spreading path set which consists of the paths with important attributes, and mutually independent.

For two paths from node u to v, if they do not share any edge, we call the two paths mutually independent. Let l_1 and l_2 be two independent paths from node u to v, they represent two different influence spreads which have no effect on each other.

Let u be a seed node, and it successfully activates a node v through a propagation path l_{uv} from node u to v. Suppose $l_{uv} = (e_1, e_2, \ldots, e_m)$ is a path formed by the connections of edges $e_1, e_2, \ldots, e_m$. Then the probability of seed u successfully activating node v through path l_{uv} is:

$$P(l_{uv}) = \prod_{i=1}^{m} P(e_i) \tag{4}$$

where $P(e_i)$ is the activation probability on the edge e_i

Since $P(l_{uv})$ indicates the probability for a seed node u successfully activating v through propagation path l_{uv}, the paths with very low probabilities are almost impossible to complete the influence spreading. Therefore we ignore these paths to reduce the computation. We set a threshold $\theta \in (0, 1)$. For a path l_{uv} from node u to v, if $P(l_{uv} < \theta)$, we set $P(l_{uv}) = 0$. Namely, we only consider the paths with activation probability greater than threshold θ.

In fact, there may exist multiple independent paths from node u to v. We only consider several independent paths with the highest probability. Those paths can be obtained from a single source shortest path detection on a weighted graph G' where each edge e is assigned a weight - $\ln[p(e)]$. The algorithm for finding m independent paths with the highest probability is as follows.

4 The influence spreading set and spreading function

Since each seed in S is assigned a positive or negative activation state, each node u in S is denoted as (u, f) which means the activation state of seed u is f, here $f \in \{+, -\}$. Suppose the seed (u, f) can activate v through l_{uv} and we denote this propagation path as l_{uv}^f. Then all the nodes on l_{uv}^f are also activated in different activation states according to the propagation rules in the SNIC model. Let $N_p(l_{uv}^f)$ and $N_n(l_{uv}^f)$ be respectively the number of positive-activated and negative-activated nodes in the propagation path l_{uv}^f. Since our goal is to maximize the number of positive-activated nodes and minimize the number of negative-activated nodes, we define the gain of the path l_{uv}^f as follows:

$$Gain(l_{uv}^f) = [N_p(l_{uv}^f) - N_n(l_{uv}^f)] \cdot p(l_{uv}) \tag{5}$$

If there are more positive-activated nodes than the negative-activated nodes in the path l_{uv}^f, then $Gain(l_{uv}^f) > 0$, we call the propagation path l_{uv}^f an effective one. Noticing that the nodes in the propagation paths l_{uv}^f and $l_{uv}^{\overline{f}}$ are always in opposite activate states, we know that $N_n(l_{uv}^f) = N_p(l_{uv}^{\overline{f}})$ and $N_p(l_{uv}^f) = N_n(l_{uv}^{\overline{f}})$. Therefore, we have $Gain(l_{uv}^f) = -Gain(l_{uv}^{\overline{f}})$. When, $Gain(l_{uv}^f) < 0$, l_{uv}^f is ineffective and $l_{uv}^{\overline{f}}$ will be effective.

In order to get the initial seed set S which can obtain the maximal positive influence, we construct activating set $R(v)$ for each node v. $R(v)$ consists of the nodes that can activate v with the maximum *Gain*:

$$R(v)=\{(u,f)|Gain(l_w^f)>\rho,\ l_{uv}^f \text{ is effective}, l_{uv}\in L(u,v)\} \tag{6}$$

Here, ρ is a threshold for the gain.

Based on the activating set $R(v)$ for all the nodes, we define an influence spreading set $I(u,f)$ for each potential seed (u,f):

$$I(u,f)=\{v|(u,f)\in R(v)\} \tag{7}$$

From (7), we can see that $I(u,f)$ is the set of nodes that can be effectively positive influenced by seed (u,f). Based on (7), we define the influence spread function of a seed set S as

$$I(S)=\underset{(u,f)\in S}{U} I(u,f) \tag{8}$$

Using $I(S)$ in (8) to replace the $I(FS)$ in (3) as the objective function, our goal is to find a seed set S^* to optimize

$$S^*=\underset{S\subset V,|S|=k}{\arg\max}|I(S)| \tag{9}$$

Theorem 1 *Given a signed graph $G=(V,E)$, the influence spread function $|I(S)|$ defined in (8) is monotone and submodular.*

Proof 1) Since $|I(S\cup\{x\})|=|\underset{(u,f)\in S\cup\{x\}}{\bigcup} I(u,f)|\geq |\underset{(u,f)\in S}{\bigcup} I(u,f)|=|I(S)|$. |I(S)| is monotone.

2) Let S and Q be two seed sets and $S\subset Q$. From the monotone property of $|I(S)|$, it is obvious that

$$I(S)\subseteq I(Q)$$

From the definition of the influence spread function $I(S)$, we can see that

$$\begin{aligned}&|I(S\cup\{x\})|-|I(S)|=|I(x)\backslash I(S)|\\&|I(Q\cup\{x\})|-|I(Q)|=|I(x)\backslash I(Q)|\end{aligned} \tag{10}$$

Since $I(S)\subseteq I(Q)$
we know $I(x)\backslash I(S)\supseteq I(x)\backslash I(Q)$,
and $|I(x)\backslash I(S)|\geq|I(x)\backslash I(Q)|$
By (10), we get
$|I(S\cup\{x\})|-|I(S)|\geq|I(Q\cup\{x\})|-|I(Q)|$
Therefore, $|I(S)|$ is submodular. □

5 The greedy algorithm

Using these two properties, we propose a greedy approximation algorithm Greedy-SIM (Greedy algorithm for Influence Maximization problem in Signed network) to maximize influence spread under the SNIC propagation model. The algorithm first constructs a set of independent paths $L(u,v)$ for each node pair (u,v) in G by calling the algorithm *Independent-paths*(u,v). Then the activating set $R(v)$ for each node v is constructed according to (6).Based on the activating sets, the influence spreading sets $I(u,+)$ and $I(u,-)$ are constructed according to (7) for each node u in V. We set the initial seed set S as an empty set θ, and denote the set of non-seed nodes as V_f. Initially $V_f=V\times\{+,-\}$ Then it iteratively chooses the node (u,f) from V_f which can maximize the increment in $|I(S)|$, and adds (u,f) into S. In each iteration, the node (u,f) is selected by the following optimization:

$$(u^*,f^*)=\underset{(u,f)\in V_f}{\arg\max}(|I(S\cup\{u,f\})|-|I(S)|) \tag{11}$$

To maximize (11) in each iteration, the node (u,f) in V_f with the largest $I(u,f)$ is selected and added to the seed set S. Once node (u,f) is selected as a seed, it will be deleted from V_f, and $(u,\overline{f})$ is also deleted from V_f since it is impossible to be selected as a seed. For every other node (v,f) in V_f, its spreading set $I(v,f)$ is modified by removing the overlapping nodes in $I(u,f)$ from it, and its increment of spreading $|I(v,f)|$ is also be updated. The iterations are repeated until the size of S reaches k.

The framework of the algorithm Greedy-SIM is as follows:

Suppose there are n vertices in the network, the time complexity of algorithm *Independent-paths* (u,v) is $O(m^2)$. Therefore in algorithm2, the first step costs $O(n^2)$, the second step costs $O(n)$ and the fourth step costs $O(k*n^2)$ On the whole, if k is regarded as a constant, the complexity of the whole algorithm is $O(n^2)$.

Although algorithm Greedy-SIM is designed for signed networks, it can be treated as a general framework for both signed and unsigned networks. The problem for influence maximization in unsigned networks is a special case in applying algorithm Greedy-SIM. In this case, since there is one type of influence spreading in the network, each node may be activated in one state. Accordingly, each seed is simply a node without a positive or negative state assigned to it. Also, there is no positive or negative sign attached to each independent spreading path l_{uv}. Instead of using (5), Gain of the unsigned independent path l_{uv} can be computed using

$$Gain(l_{uv})=N(l_{uv}).p(l_{uv}) \tag{12}$$

Here, $N(l_{uv})$ is the number of influenced nodes in the path l_{uv}.

6 Experimental results

To evaluate the quality of the results by our algorithm Greedy-SIM, we test it on some real-world datasets and compare the quality of the results with the other similar methods. All the tests are conducted on Pentium IV processor with 1.7G memory, under the Windows XP operating system. The algorithms are coded in VC++ 6. 0.

6.1 Tests on unsigned networks

To validate the effectiveness of algorithm GREEDY-SIM, we first apply it on three real unsigned networks: USAir 97, Email and PGP. USAir 97 [42] is a network of the US air transportation system, it consists of 322 airports and the airline between two airports can be represented by an edge between two nodes in the network. Email [43] is the network of e-mail interchanges between members of the University Rovirai Virgili. PGP [44] is an encrypted communication network. Table 1 shows the basic topological features of the test datasets including the number of nodes (n), the number of edges (m), the average degree ($<k>$), the maximum degree ($k_{\max}$), and the clustering coefficient (C).

Algorithm 1 Independent-paths(u, v)

Input: $G(V, E)$: the network;
u, v: the source and destination nodes of the paths;
m: the number of independent paths with the highest probability;
Output: $L(u, v)$: the set of the independent paths between u and v with the highest probability;
Begin
1. /*Construct a weighted graph G' from G*/
   For every edge$e \in E$ **do**
   Assign a weight - ln[$p(e)$] on e;
   Endfor;
2. Let $G_1 = G'$; $L(u, v) = \Phi$;
3. **For** i =1 **to** m **do**
   Find the shortest path l^i_{uv} between u and v on G_i;
   Compute $p(l^i_{uv})$ according to (4);
   If $p(l^i_{uv} > \theta)$ **then**
   $L(u, v) = L(u, v) \cup \{l^i_{uv}\}$;
   $G_{i+1} = (V, E \setminus \cup^i_{j=1} l^i_{uv})$;
   endif
   End for i;
End

We compare the influence spreads by GREEDY-SIM with that of four algorithms. Three of them are centrality-based algorithms: degree centrality (DC) [45], betweenness centrality (BC) [45] and closeness centrality (CC) [46].

Table 1 The basic topological features of the three realworld networks

Network	n	m	$<k>$	$k_{\max}$	C
USAir97	332	2126	12.81	139	0.3545
Email	1133	5451	9.62	71	0.2202
PGP	10680	24328	4.55	206	0.2659

In these algorithms, k nodes with the largest centralities are chosen as seed nodes. The other algorithm is named PIDS [26] which is a heuristic algorithm based on learning automata. It selected the minimum positive influence dominating set (MPIDS)as the seed set. Figures 1, 2 and 3 show the comparison of the numbers of activated nodes by the seeds selected by GREEDY-SIM and other algorithms in different propagation time t.

Figure 1 shows the numbers of activated nodes by GREEDY-SIM and the other algorithms in dataset PGP. From Fig. 1a we can see that, when time $t<3$, the number of activated nodes by DC is almost the same as the number of GREEDY-SIM method. But when$t>3$, the number of activated nodes by GREEDY-SIM is significantly larger than that of DC. Similarly, we can see from Fig. 1b that there is a big gap in the number of spreading between the CC method and GREEDY-SIM at all the time steps. In Fig. 1c we can observe that when time $t<5$, the curve of BC method almost coincides with that of GREEDY-SIM, which means that the two methods activate almost the same number of nodes. Nonetheless, when $t>5$, we can have a penetrating insight that the spreading of GREEDY-SIM is significantly higher than that of BC, indicating that our GREEDY-SIM method can influence much more nodes than BC. In Fig. 1d, it can be observed that when $t<3$, PIDS performs slightly better than GREEDY-SIM. However, after $t>=3$, GREEDY-SIM outperforms PIDS, especially after $t>8$, the advantage of GREEDY-SIM is more obvious. From the comprehensive comparisons in Fig. 1, it can be concluded that although the GREEDY-SIM method shows no difference with other methods in the early stages in PGP, the final number of nodes that are activated in GREEDY-SIM is significantly larger than that of the other four methods.

Figure 2 shows the numbers of activated nodes by GREEDY-SIM and other algorithms in dataset USAir97. From Fig. 2a we can see that the performance of the GREEDY-SIM method is almost the same as DC in activated nodes when time $t<3$. But when time $t>3$, GREEDY-SIM has much larger spread than the DC method. Figure 2b and c also show that the spread of GREEDY-SIM method is obviously larger than CC and BC at all the spreading steps. And from Fig. 2d, it can be investigated that when $t<4$, PIDS is slightly better than GREEDY-SIM.

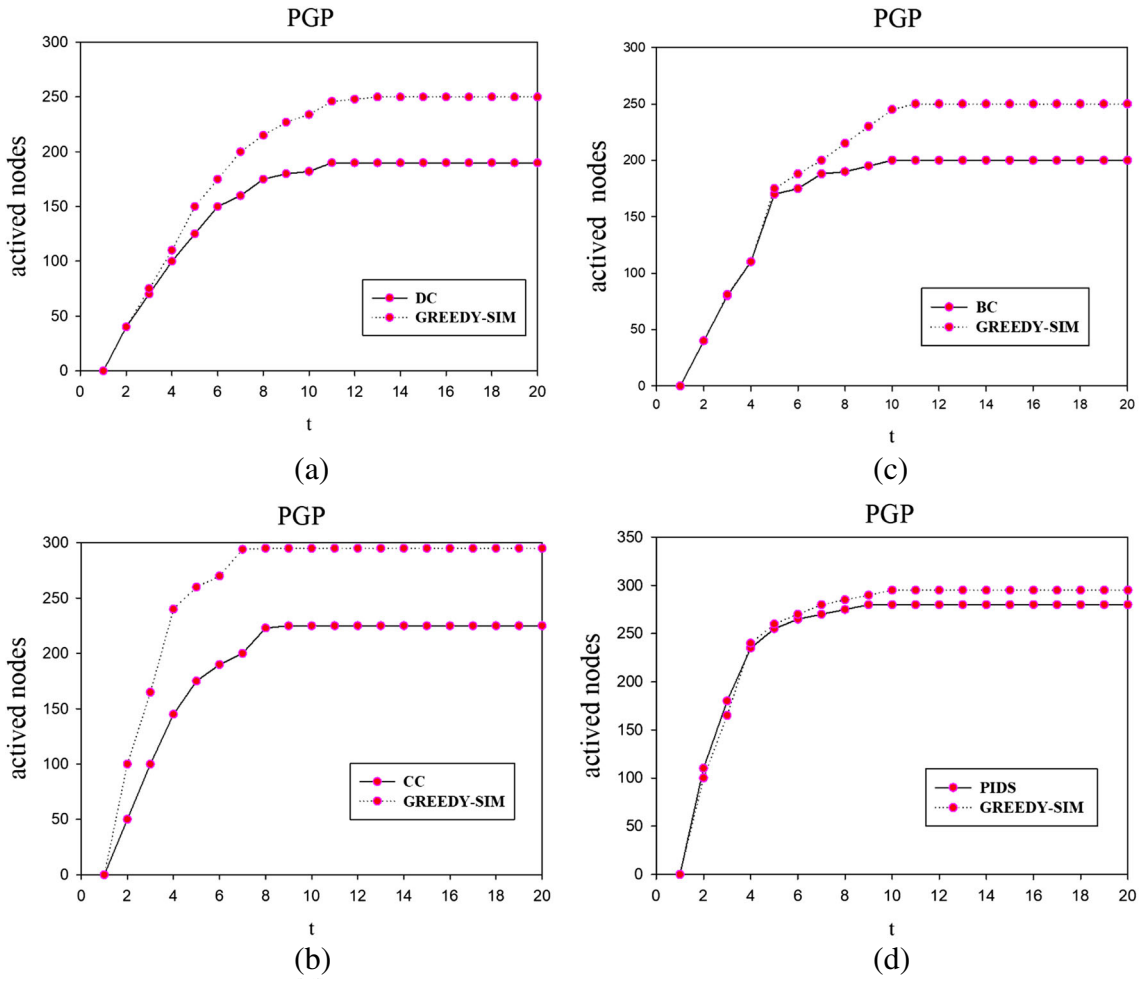

Fig. 1 Comparison of the numbers of activated nodes by GREEDY-SIM and other algorithms in PGP

When t ranges from 4 to 6, the two algorithms achieve the same performance. After $t>6$, GREEDY-SIM is superior to PIDS. From the comparisons in Fig. 2, we can see that our approach GREEDY-SIM can obtain much larger spreading than other methods in USAir97.

Figure 3 shows the numbers of activated nodes by GREEDY-SIM and other algorithms in dataset Email. From Fig. 3b, it can be noticed that our algorithm has more absolute advantages than CC. Meanwhile, it can be shown from Fig. 3a, c and d, when $t <= 6$, the other algorithms are slightly better than GREEDY-SIM or have similar performance with our proposed algorithm. However, while $t>6$, GREEDY-SIM outperforms the other algorithms. We can see from Fig. 3 that GREEDY-SIM can obtain much larger influence spreading on dataset Email than other methods.

Through the comprehensive analysis of Figs. 1–3, we can see that it is undoubtedly GREEDY-SIM outperforms other methods and achieves wider influence spreading in unsigned networks.

6.2 Tests on signed networks

6.2.1 Data set

We use Epinions, Slashdot and Wikipedia to validate the experiment. Both of the three networks are large signed online social networks. Epinions [7] is a product review site. Based on their reviews and ratings of products, users can choose to trust or distrust each other. The network has 131828 users and 841372 relationships. Slashdot [8] is a news site where users can be friends or enemies. We treat the relationship of friends as a positive link, while the enemy's relationship as a negative link. The network has 77350 users and 516575 relationships. The relationship among each user is explicitly marked as positive or negative. Data sets of these two networks can be downloaded from website http://snap.standford.edu/data/index.html. Wikipedia [47] is a website for adminship elections which can be downloaded from website http://snap.stanford.edu/data/wiki-Elec.html

The network includes the administrator elections and vote history data of Wikipedia users and consists of about 700 users and about 100000 "for" vote and "against" vote relationships between them. Table 2 shows the basic topological features of those two networks where n denotes the number of nodes, m denotes the number of edges, $\langle k_{out}\rangle$ represents the average out-degree, $\langle k_{\max\ out}\rangle$ represents the maximal out-degree, $\langle k_{pout}\rangle$ denotes the average positive out-degree, $\langle k_{\max\ pout}\rangle$ denotes the maximal positive out-degree, $\langle k_{nout}\rangle$ denotes the average negative out-degree, $k_{\max nout}$ denotes the maximal negative out-degree, and C is the clustering coefficient.

Algorithm 2 Greedy-SIM

Input: $G(V, E)$: the network;
ρ: threshold of the*Gain* value;
k: the size of the seed set;
Output: S: the seed set;
Begin
1. **For** each node pair (u, v) in G **do**
/*find the m independent path set $L(u, v)$ with the highest probability*/
Independent-paths(u, v);
For each path $l_{uv} \in L(u, v)$ **do**
Compute $Gain(l_{uv}^{+})$ according to (5);
If $Gain(l_{uv}^{+})>\rho$ **then** $R(v) = R(v)\cup\{(u, +)\}$
Else If $Gain(l_{uv}^{-})>\rho$ **then** $R(v) = R(v) \cup\{(u, -)\}$;
Endfor;
Endfor;
2. **For** every node u in V **do**
Construct $I(u, +)$ and $I(u,-)$ according to (7);
Endfor;
3. $S = \theta$; $V_f = V \times \{+, -\}$;
4. **While** $|S|<k$ **do**
$(u*, f^*) = \arg\max_{(u,f)\in V_f} |I(u, f)|$;
$S = S\cup\{(u*, f^*)\}$;
$V_f = V_f\backslash\{(u*, f^*), (u*, \overline{f}^*)\}$;
For every $(v, f) \in V_f$ **do**
$I(v, f) = I(v, f)\backslash I(u*, f^*)$;
If $I(v, f) = \theta$ **then** $V_f = V_f\backslash\{(v, f)\}$;
Endfor;
Endwhile
end

6.2.2 Experimental results and analysis

We test GREEDY-SIM method on the signed online social networks Epinions and Slashdot and compare its spreading results with that of other seven methods: Positive Out-Degree (POD) [48], Effective Degree [14], OSSUM$^{\pm}$[47], Positive Degree Discount [22], RLP [49], SA with heuristic [10] and IC-P Greedy [52] Positive Out-Degree is the heuristic algorithm which chooses k nodes with respectively the maximal positive out-degree. Effective degree is defined as the number of negative out-degree subtracted from the number positive out-degree. Effective Degree method selects the k nodes with the maximum effective degree to form the seed set. OSSUM$^{\pm}$ is an online seed set selecting method using a unified model on the signed network Positive Degree Discount is a heuristic algorithm designed for ICM, that implements a weighted discount form based on parameter p RLP is a greedy algorithm to study the influence maximization problem in signed social networks with opinion formation SA with heuristic uses simulated annealing for finding the seed node set with maximum positive influence maximization problem in the signed network. IC-P Greedy algorithm extends the classic IC model to signed social networks and solves the polarity related influence maximization (PRIM) problem. It selects the seed node set with maximum positive influence or maximum negative influence in signed social networks. Table 3 lists what information diffusion models the evaluated algorithms are based on, IC models or LT models. From Table 3 we can observe that besides the RLP algorithm, the other algorithms for comparison are all based on the IC model. We conduct experiments to study the influence of spread achieved by the eight methods with different seed set size. In these datasets, there are much less negative edges than positive edges. Due to the conclusion obtained from [38], we flipped the signs on all edges in the three original networks. The experimental results are shown in Figs. 4, 5 and 6.

Figure 4 shows the numbers of positive influenced nodes of different seed sizes by GREEDY-SIM and compares with the other seven algorithms on Epinions. From Fig. 4 we can see that GREEDY-SIM outperforms the other seven methods in Epinions network. Although when the size of the seed node set $k<6$, the number of positive influenced nodes by our method has no significant difference with that of other seven methods, when $k>6$, the GREEDY-SIM can achieve much more spread than all the other methods. This indicates that the seed set selected by our GREEDY-SIM method can spread much wider influence than other methods on Epinions.

Figure 5 is a comparison of the numbers of positive influenced nodes of different seed sizes by GREEDY-SIM with the other seven algorithms on Slashdot. We find that the seed set selected by our GREEDY-SIM method can spread much wider influence than other methods on Slashdot when the size of the seed node set $k>10$ This indicates that GREEDY-SIM has the ability to find more hidden influential users than the other algorithms Similar to results on the Epinions dataset, GREEDY-SIM also outperforms the other seven methods on the Slashdot dataset.

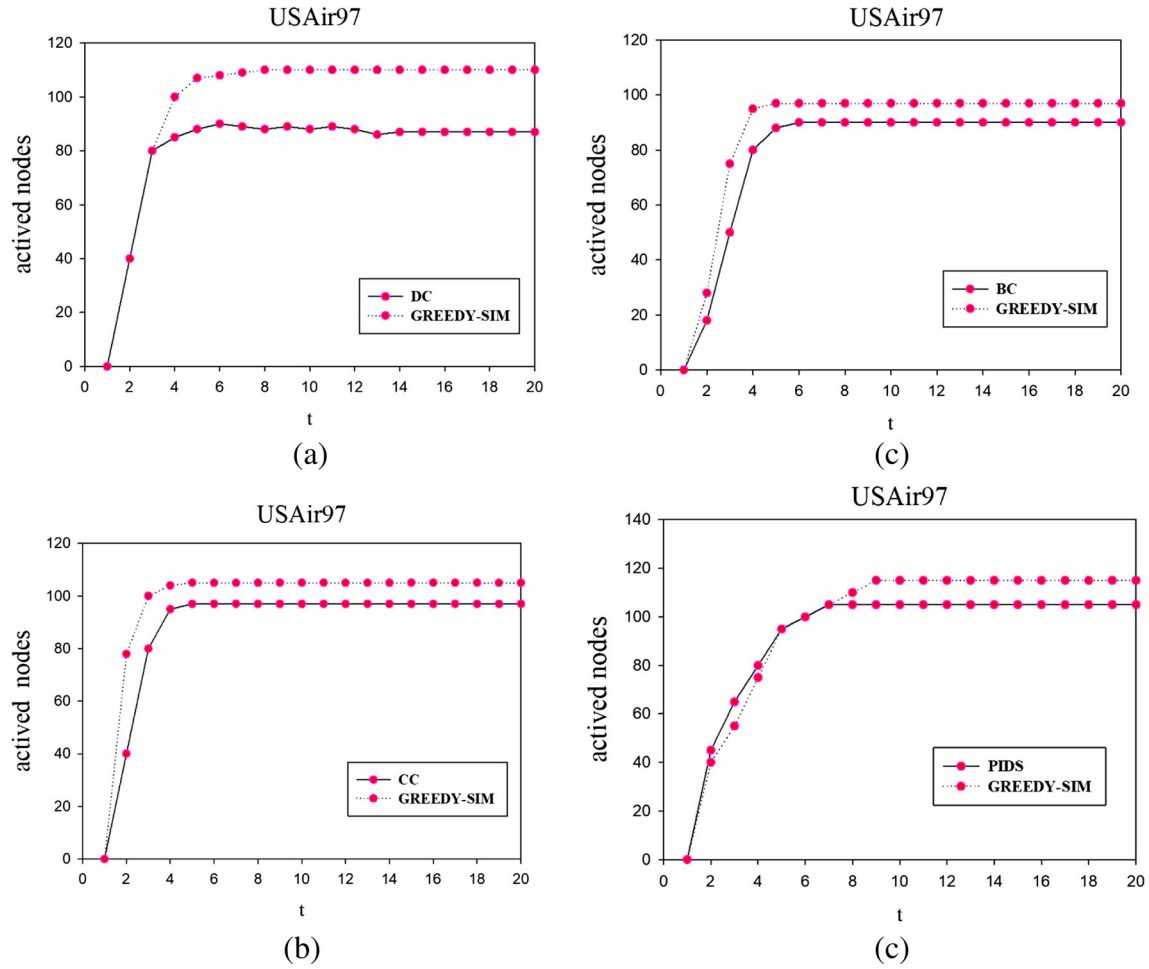

Fig. 2 Comparison of the numbers of activated nodes by GREEDY-SIM and other algorithms in USAir97

Figure 6 is a comparison of the numbers of positive influenced nodes of different seed sizes by GREEDY-SIM with the other seven algorithms on Wikipedia. From Fig. 6 we can observe that our GREEDY-SIM method consistently get similar performance with the other algorithms when $k<15$. However, when $k>15$, the GREEDY-SIM performs better than all the other methods. This indicates that the seed set selected by our GREEDY-SIM method can spread much wider influence than other methods on Wikipedia.

Based on the comprehensive analysis of Figs. 4–6 it can be seen that regardless of the number of seeds selected, our GREEDY-SIM method always has a larger number of positive activated nodes than other methods. It is obvious that GREEDY-SIM method has better seed identification efficiency and strong stability. The experimental results show the correctness as well as the effectiveness of applying GREEDY-SIM to the influence maximization in signed networks.

Table 4 shows the comparison of the influence spread of the results by different algorithms. For the convenience of calculation, we normalize the influence spread results by different algorithms over Wikipedia. We use t-test to show that the amount of influence spread by GREEDY-SIM is statistically much different from that of the other algorithms. Table 5 shows the results of t-values of the influence spread by GREEDY-SIM with other algorithms.

Because each group of data has 15 samples, the degree of freedom in the t-test is (15-1)*2=28. We set the significance level $\alpha = 0.05$, then the confidence level p is 97.5%. From t-distribution table, we can obtain $t_{0.975}(28)= 2.04841$. From Table 5, we can see that all the t values are greater than $t_{0.975}(28)$. This indicates that there are much different significances between the influence spread by GREEDY-SIM and the other algorithms. Therefore, the quality of results by our algorithm GREEDY-SIM is significantly higher than that by other algorithms.

To further verify the performance of GREEDY-SIM, we test the numbers of activated nodes in different propagation time t by the seeds selected by GREEDY-SIM on datasets Epinions and Slashdot. We also compare the numbers of activated nodes in different time t by GREEDY-SIM with other seven algorithms: Rand, PageRank [50], d^++d [51]

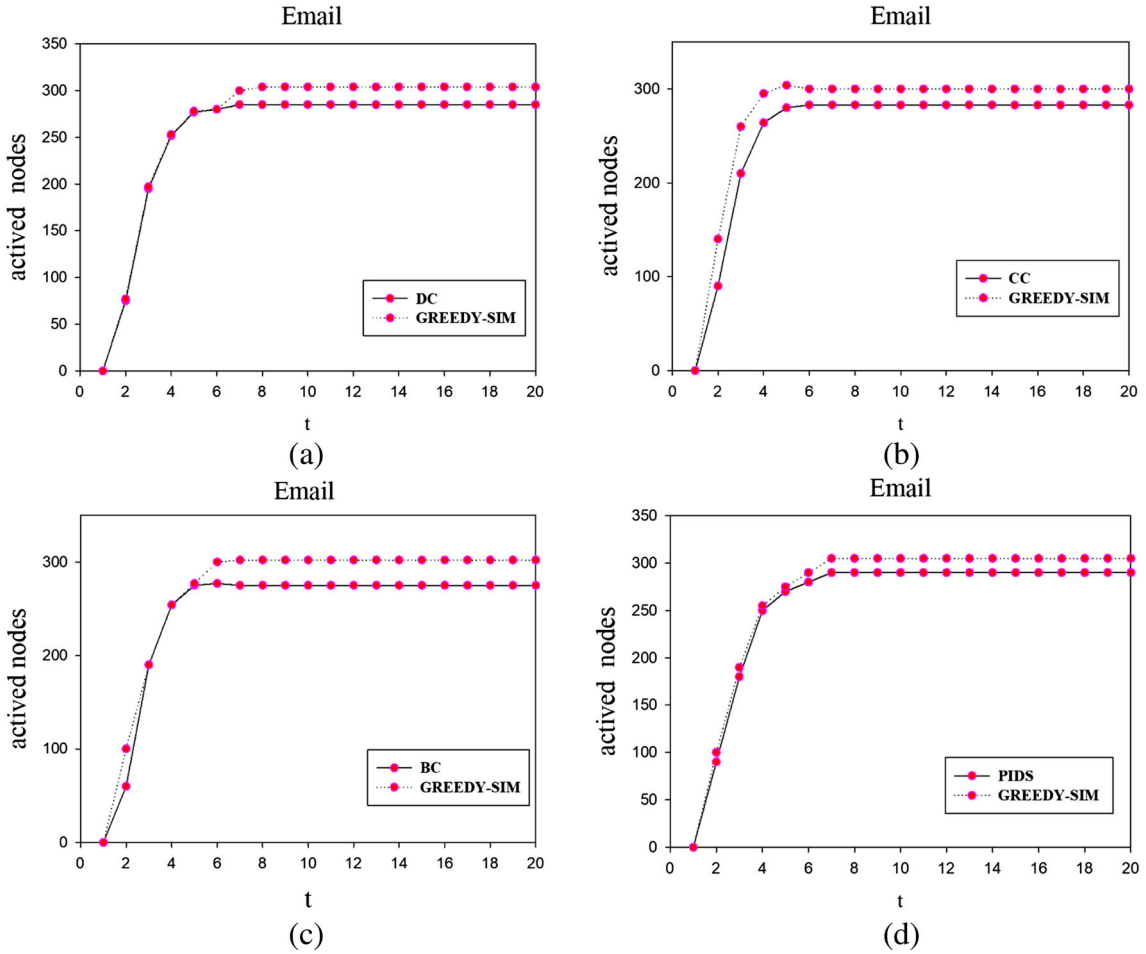

Fig. 3 Comparison of the numbers of activated nodes by GREEDY-SIM and other algorithms in Email

SSVIM [51], RLP [49] SA with heuristic [10] and IC-P Greedy [52]. Rand is a stochastic method which selects the seeds randomly. PageRank is an iterative algorithm which can be used to rank the quality of the web pages that are connected by hyperlinks in social networks. Method d^++d selects seeds with the highest in-going degrees. SSVIM selects seeds based on the theory of status in the signed network. RLP chooses those with higher influence ability as seed users. SA with heuristic proposes seed nodes seeking algorithm based on simulated annealing and two effective

Table 2 The basic topological features of three real networks

Network	Epinions	Slashdot	Wikipedia
n	131828	77350	7000
m	841372	516575	100000
$\langle k_{out} \rangle$	6.38	6.68	14.2
$k_{\max\ out}$	2070	2532	2301
$\langle k_{pout} \rangle$	5.44	5.12	12.8
$k_{\max\ pout}$	2070	2502	2301
$\langle k_{nout} \rangle$	0.94	1.56	1.25
$k_{\max\ nout}$	1562	495	1272
C	0.1279	0.0549	0.1746

Table 3 The diffusion models of different evaluated influence maximization algorithms

Algorithm	Diffusion Model
Positive Out-Degree (POD)	IC
Effective Degree	IC
$OSSUM^{\pm}$	IC
Positive Degree Discount	IC
RLP	LT
SA with heuristic	IC
Rand	IC
PageRank	IC
d^++d	IC
SSVIM	IC
GREEDY-SIM	IC
IC-P Greedy	IC

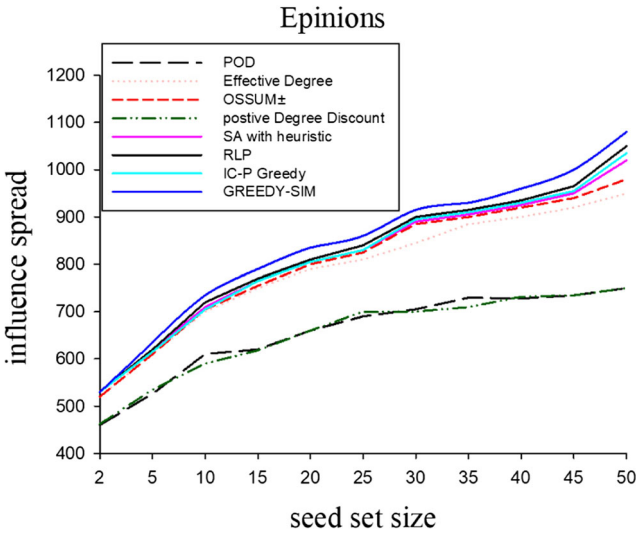

Fig. 4 Results on Epinions

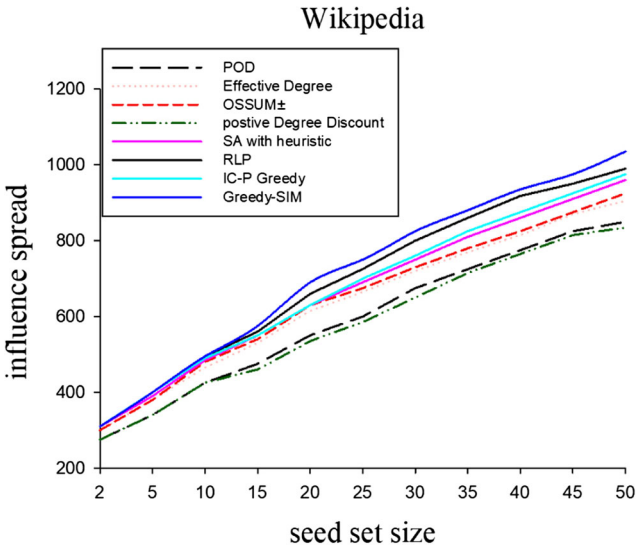

Fig. 6 Results on Wikipedia

heuristics to accelerate the convergence. IC-P Greedy aims to solve the positive influence maximization (PIM) problem and negative influence maximization (NIM) problem under IC-P model. Figure 7 shows the numbers of activated nodes in different propagation time t by the methods on dataset Slashdot with seed set sizes k =800, 1500 and 5000 It can be observed from Fig. 7 that the curves of the other traditional algorithms are almost below that of GREEDY-SIM. This means the seed sets selected by GREEDY-SIM can achieve wider influence spread at all propagation time. This is because some of the other algorithms such as PageRank and RLP only consider trust relationships in the social network while GREEDY-SIM considers both trust and distrust relationships. Without distrust (or foe, hostile) relationships, there may be part of information missed and thus lead to unreal situations on the social network. In fact, apart from trust relationships, there is also a distrust between users in the real world. From the figure, it can be concluded that the negative relationships are as important as the positive relationships, and it is necessary to consider both trust and distrust relationships in IM on the signed social network. Except for PageRank and RLP, for the other four algorithms we have insight into an interesting phenomenon. When k =800 or 1500 over Slashdot data, the curve of SSVIM is constantly below the curve of d^++d. However, when k =5000, we can notice that the SSVIM performs better than d^++d. This is because the SSVIM method assumes that every node was active at first, and the performance will get better and better as the number of the seeds increases. It may be superior to GREEDY-SIM when the seed size k is very close to $|V|$, but it is not going to happen in realworld applications, and SSVIM may cost huge computational time when k is very large. Similarly, with the increase of the seed size, especially at k =5000, the volatility of the IC-P Greedy curve is relatively large, and its performance is not as good as k =800 or k =1500 Moreover, when k =5000, the performance of SA with heuristic outperforms that k = 800 or 1500 and catches up with IC-P Greedy That's because with the increase of the seed nodes SA with heuristic method has the ability to find more hidden influential users that greedy algorithm cannot seek. In conclusion, GREEDY-SIM performs better than other algorithms under normal seed size.

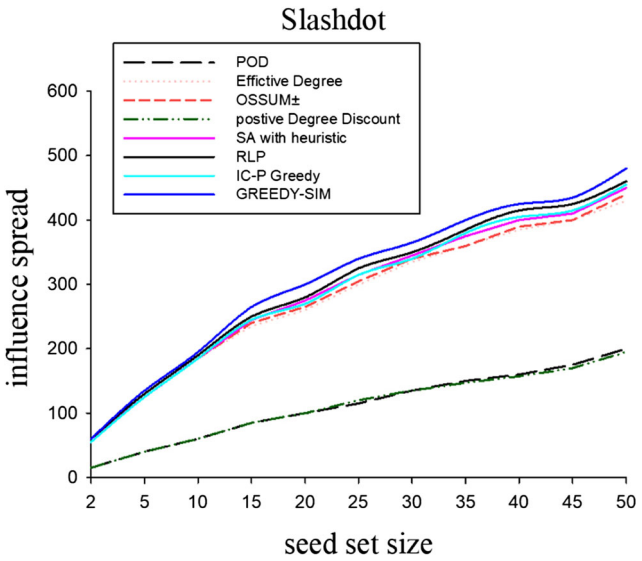

Fig. 5 Results on Slashdot

Figure 8 shows the numbers of activated nodes in different propagation time t by the methods on dataset Epinions with seed set sizes k =800 and 5000 From Fig. 8, we can observe that all the spreads by the eight algorithms are almost identical due to the topological structure of Epinions network. Since the component of the network we

Table 4 Comparison of the influence spread of the results by different algorithms over Wikipedia

No.	POD	Effective Degree	OSSUM±	Positive Degree Discount	SA with heuristic	RLP	IC-P Greedy	GREEDY- SIM
1	0.545	0.604	0.613	0.622	0.627	0.651	0.636	0.681
2	0.552	0.612	0.619	0.536	0.628	0.665	0.642	0.689
3	0.543	0.600	0.611	0.531	0.625	0.655	0.635	0.681
4	0.537	0.599	0.605	0.525	0.622	0.651	0.627	0.675
5	0.557	0.615	0.628	0.544	0.641	0.667	0.622	0.694
6	0.526	0.589	0.601	0.520	0.610	0.645	0.617	0.672
7	0.567	0.618	0.639	0.554	0.648	0.692	0.652	0.705
8	0.559	0.620	0.626	0.545	0.641	0.675	0.648	0.696
9	0.525	0.584	0.595	0.512	0.602	0.634	0.616	0.661
10	0.574	0.636	0.643	0.555	0.655	0.685	0.663	0.713
11	0.502	0.565	0.569	0.487	0.585	0.617	0.591	0.639
12	0.557	0.615	0.618	0.538	0.636	0.666	0.647	0.692
13	0.560	0.624	0.617	0.545	0.648	0.671	0.659	0.700
14	0.549	0.613	0.629	0.543	0.638	0.673	0.645	0.695
15	0.514	0.566	0.576	0.498	0.588	0.620	0.592	0.647

Table 5 t-values of the influence spread by GREEDY-SIM with other algorithms

Algorithm	GREEDY- SIM With POD	GREEDY- SIM With Effective Degree	GREEDY- SIM With OSSUM±	GREEDY- SIM With Positive DegreeDiscount	GREEDY- SIM With SA with heuristic	GREEDY- SIM With RLP	GREEDY-SIM With IC-P Greedy
t-values	24.3919	14.8519	12.9595	18.3956	10.1503	4.3948	8.8046

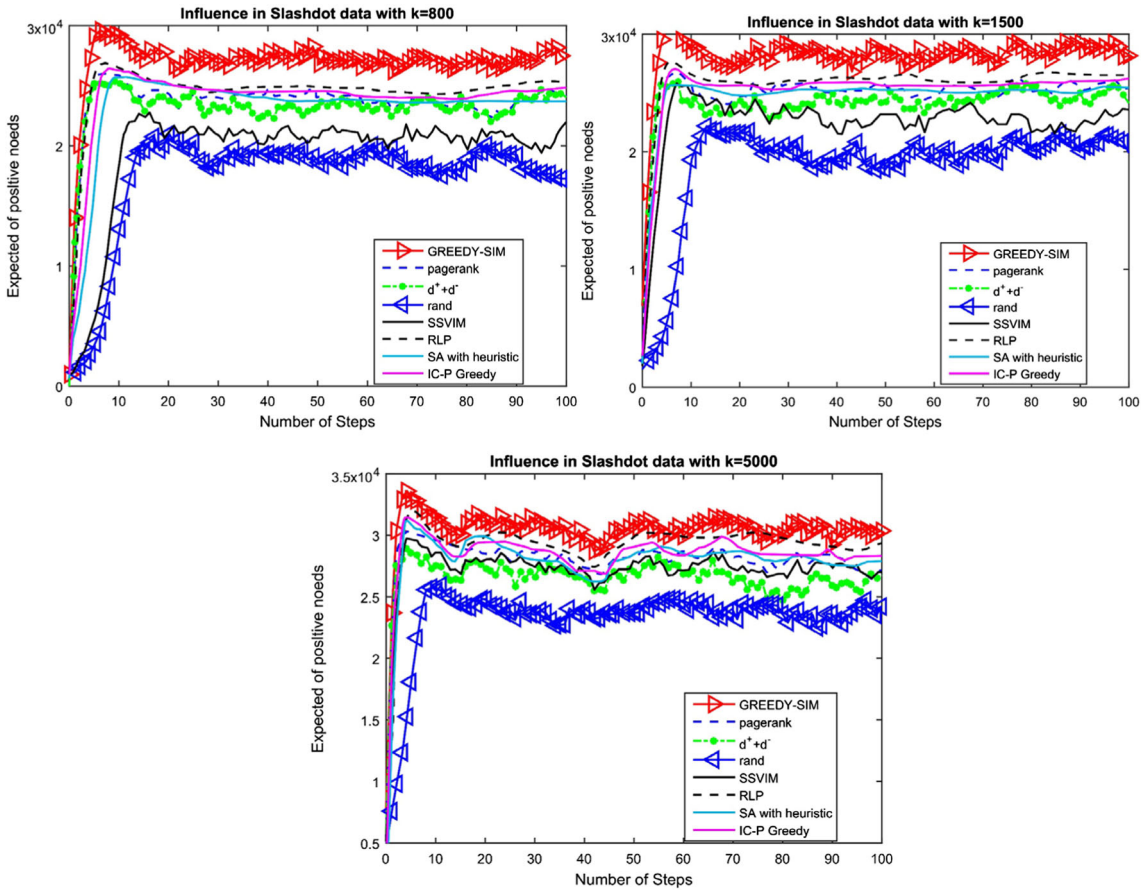

Fig. 7 Influence in Slashdot with seed set sizes $k = 800$, $k = 1500$ and $k = 5000$

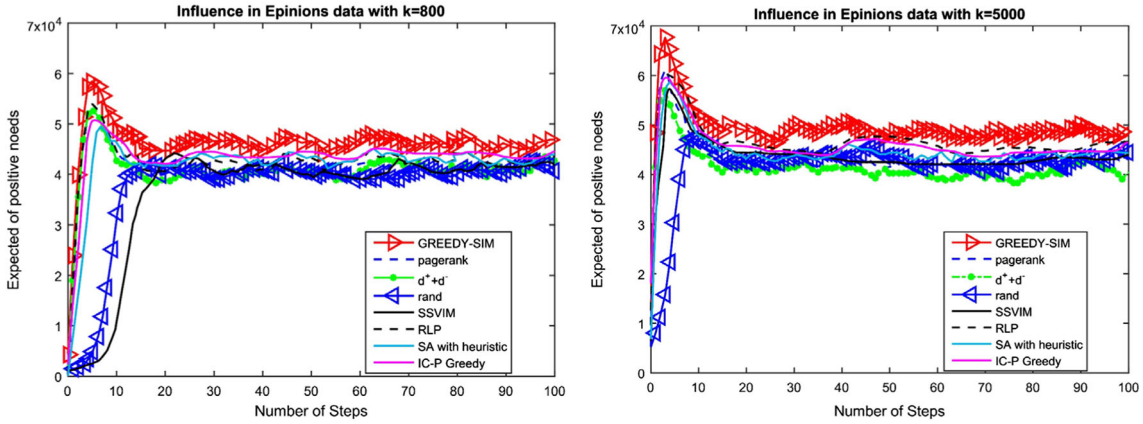

Fig. 8 Influence in Epinions with seed

tested is densely connected, the final number of expected positive nodes is always around 4700 for all the different seed sizes. Through the comprehensive analysis of the network structure, we find that the nodes in the connected component we tested have almost influence and spreading mostly happened in the component. This is the reason we cannot clearly distinguish the spread results by these methods. Nevertheless, as Fig. 8 shows, GREEDY-SIM still has a larger spread than other seven algorithms in most of the time steps.

7 Conclusion and future work

In this paper, we propose an algorithm GREEDY-SIM for influence maximization in signed networks. We defined a SNIC model for describing two opposite types of influence spreading in a signed network. Based on the SNIC model, we present an algorithm for constructing the set of independent propagation paths and computing their probability. Based on the independent propagation paths, we define the influence spreading function for a seed set, and prove that such a spreading function is monotone and submodular. Particularly, a greedy algorithm is presented to maximize the positive influence spreading in the signed network. Experimental results on real data sets show that our GREEDY-SIM method can achieve much larger positive influence spreading in signed networks than other methods.

Although our method has achieved high-quality results in some real data sets, the sizes of these data sets are still small. Moreover, those datasets are very close to the proposed SNIC model. However, the success of our method in practical application actually depends on the distance between the proposed SNIC model and the actual information diffusion model in real social network data. As a future work, we intend to apply our method on some larger size real data sets to prove the applicability of our approach.

Acknowledgments This research was supported in part by the National Natural Science Foundation of China under Grant Nos. 61702441, 61602202, Natural Science Foundation of Jiangsu Province under contracts BK20160428.

References

1. Brown JJ, Reingen PH (1987) Social ties and word-of-mouth referral behavior. J Consum Res 14(3):350–362
2. Goldenberg J, Libai B, Muller E (2001) Talk of the network: a complex systems look at the underlying process of word-of-mouth. Mark Lett 12(3):211–223
3. Goldenberg J, Libai B (2001) Using complex systems analysis to advance marketing theory development: modeling heterogeneity effects on new product growth through stochastic cellular automata. Academy Market Sci Rev 9(3):1–18
4. Domingos P, Richardson M (2001) Mining the network value of customers. In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 57–66
5. Richardson M, Domingos P (2002) Mining knowledge-sharing sites for viral marketing. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 61–70
6. Ahmed NM, Chen L, Wang YL, Li B, Li Y, Liu W (2018) DeepEye: link prediction in dynamic networks based on non-negative matrix factorization. Big Data Mining Anal 1(1):19–33
7. Kunegis J, Preusse J, Schwagereit F (2013) What is the added value of negative links in online social networks. In: Proceedings of the 22nd International Conference on World Wide Web, pp 727–736
8. Lampe CAC, Johnston E, Resnick R (2007) Follow the reader: filtering comments on slashdot. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 1253–1262
9. Wang Q, Jin Y, Lin Z, Cheng S, Yang T (2016) Influence maximization in social networks under an independent cascade-based model. Phys A: Stat Mech Appl 444:20–34
10. Li D, Wang C, Zhang S, Zhou G, Chu D, Wu C (2017) Positive influence maximization in signed social networks based on simulated annealing. Neurocomputing 260:69–78
11. Bharathi S, Kempe D, Salek M (2007) Competitive influence maximization in social networks. In: Proceedings of the Internet and Network Economics, pp 306–311
12. Borodin A, Filmus Y, Oren J (2010) Threshold models for competitive influence in social networks. In: Proceedings of International Workshop on Internet and Network Economics, pp 539–550

13. He X, Song G, Chen W, Jiang Q (2012) Influence blocking maximization in social networks under the competitive linear threshold model. In: Proceedings of the 12th SIAM International Conference on Data Mining, pp 463–474
14. Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pp 137–146
15. Kempe D, Kleinberg J, Tardos E (2005) Influential nodes in a diffusion model for social networks. In: Proceedings of the 32nd International Conference on Automata, Languages and Programming, vol 3580, pp 1127–1138
16. Kimura M, Saito K (2006) Tractable models for information diffusion in social networks. In: Proceedings of the European Conference on Principles of Data Mining and Knowledge Discovery, pp 259–271
17. Leskovec J, Krause A, Guestrin C et al (2007) Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 420–429
18. Goyal A, Lu W, Lakshmanan LVS (2011) CELF++: optimizing the greedy algorithm for influence maximization in social networks. In: Proceedings of the 20th International Conference Companion on World Wide Web, pp 47–48
19. Estevez PA, Vera P, Saito K (2007) Selecting the most influential nodes in social networks. In: Proceedings of the International Joint Conference on Neural Networks, pp 2397–2402
20. Bharathi S, Kempe D, Salek M (2007) Competitive influence maximization in social networks. In: Proceedings of the International Workshop on Web and Internet Economics, pp 306–311
21. Wu P, Pan L (2017) Scalable influence blocking maximization in social networks under competitive independent cascade models. Comput Netw 123:38–50
22. Chen W, Wang Y, Yang S (2009) Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 199–208
23. Chen W, Yuan Y, Zhang L (2010) Scalable influence maximization in social networks under the linear threshold model. In: Proceedings of the 10th IEEE International Conference on Data Mining, pp 88–97
24. Li J, Yu Y (2012) Scalable influence maximization in social networks using the community discovery algorithm. In: Proceedings of the 2012 Sixth International Conference on Genetic and Evolutionary Computing, pp 284–287
25. Zhu Y, Wu W, Bi Y et al (2015) Better approximation algorithms for influence maximization in online social networks. J Comb Optim 30(1):97–108
26. Mohammad M, Daliri K, Alireza R, Negin B, Mohammad RM (2018) Minimum positive influence dominating set and its application in influence maximization: a learning automata approach. Appl Intell 48:570–593
27. Zeynep ZA, Sule GO (2018) Identifying topical influencers on twitter based on user behavior and network topology. Knowl-Based Syst 141:211–221
28. Sumith N, Annappa B, Swapan B (2018) A holistic approach to influence maximization in social networks: STORIE. Appl Soft Comput 66:533–547
29. Tang J, Tang X, Yuan J (2018) An efficient and effective hopbased approach for influence maximization in social networks. Soc Netw Anal Mining 8:10
30. Gong M, Yan J, Shen B, Ma L, Cai Q (2016) Influence maximization in social networks based on discrete particle swarm optimization. Inform Sci 367-368:600–614
31. Arastoo B, Hassan H, Mohammad SZ, Mojtaba R (2016) INCIM: A community-based algorithm for influence maximization problem under the linear threshold model. Inf Process Manag 52:1188–1199
32. Zeng Y, Chen X, Cong G, Qin S, Tang J, Xiang Y (2016) Maximizing influence under influence loss constraint in social networks. Expert Syst Appl 55:255–267
33. Lu F, Zhang W, Shao L, Jiang X, Xu P, Jin H (2017) Scalable influence maximization under independent cascade model. J Netw Comput Appl 86:15–23
34. Du N, Liang Y, Maria FB, Manuel GR, Zha H, Song L (2017) Scalable influence maximization for multiple products in continuous-time diffusion networks. J Mach Learn Res 18:1–45
35. Zhang H, Zhang H, Li X, My T (2015) Limiting the spread of misinformation while effectively raising awareness in social networks. In: Proceedings of the International Conference on Computational Social networks (CSonet), vol 9197. LNCS, pp 35–47
36. Samadi M, Nagi R, Semenov A, Nikolaev A (2018) Seed activation scheduling for influence maximization in social networks. Omega 77:96–114
37. Ajitesh S, Chelmis C, Viktor K (2015) Prasanna. Social Influence Computation and Maximization in Signed Networks with Competing Cascades. Proc IEEE/ACM Int Conf Advan Social Netw Anal Mining 89(14):41–48
38. Chen W, Collins A, Cummings R et al (2011) Influence maximization in social networks when negative opinions may emerge and propagate. In: Proceedings of the Eleventh SIAM International Conference on Data Mining, pp 379–390
39. Chen S, He K (2015) Influence maximization on signed social networks with integrated PageRank. In: Proceedings of the IEEE International Conference on Smart City, pp 289–292
40. Siwar J, Arnaud M, Ludovic L et al (2016) Maximizing positive opinion influence using an evidential approach[J]. Social Inform Netw:168–174
41. Petz G et al (2015) Computational approaches for mining user's opinions on the Web 2.0. Inform Process Manag 51(4):510–519
42. Guimera R, Danon L, Diaz-Guilera A, Giralt F, Arenas A (2003) Self-similar community structure in a network of human interactions. Phys Rev E Stat Nonlin Soft Matter Phys 68(6 Pt 2):065103
43. Michalski R, Palus S, Kazienko P (2011) Matching organizational structure and social network extracted from email communication. Proc Int Conf Business Inform Syst 87:197–206
44. Bogu M, Pastor-Satorras R, Díaz-guilera A, Arenas A (2004) Models of social networks based on social distance attachment. Phys. Rev. E 70(5):056122
45. Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry 40(1):35–41
46. Sabidussi G (1966) The centrality index of a graph. Psychometrika 31(4):581–603
47. Leskovec J, Huttenlocher D, Kleinberg J (2010) Signed networks in social media. In: Proceedings of the 28th ACM Conference on Human Factors in Computing Systems, pp 1–9
48. Chen W, Wang C, Wang Y (2010) Scalable influence maximization for prevalent viral marketing in large-scale social networks. In: Proceedings of the 16th ACM SIGKDD Conference on Knowledge Discovery. Springer, pp 1029–1038
49. Shen C, Nishide R, Piumarta I, Takada H, Liang W (2015) Influence maximization in signed social networks. In: Proceedings of the 2015 International Conference on Web Information Systems Engineering. Springer, pp 399–414
50. Page L, Brin S (1998) The anatomy of a large-scale hypertextual web search engine. Comput Netw 30(1-7):107–117
51. Li Y, Chen W, Wang Y, Zhang Z (2013) Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In: Proceedings of the sixth ACM international conference on Web search and data mining, pp 657–666
52. Li D, Xu Z, Chakraborty N et al (2014) Polarity related influence maximization in signed social networks. Plos One 9(7):e102199

Wei Liu was born in Jiangyin, Jiangsu Province, P.R.China, in July 1, 1982. She received B. Sc degree and M. Sc degree in computer science from Yangzhou University, P.R. China in 2004 and 2007 respectively. In 2010, she received Ph.D degree in the department of computer science from Nanjing University of Aeronautics and Astronautics. She is currently an associate professor and Master's Supervisor in the Institute of Information Science and Technology, Yangzhou University, Yangzhou, P.R.China. Her research interest includes complex network, data mining, and bioinformatics. She has published more than 50 papers in journals and conferences. (Corresponding author, Email: yzliuwei@126.com).

Xin Chen was born in Huaian, China, in 1994. She is now perusing the Master Degree at the College of Information Engineering of Yangzhou University, Yanzghou, China. Her research topic is influence maximization.

Byeungwoo Jeon received the B.S. degree (Magna Cum Laude) in 1985, the M.S. degree in 1987 in electronics engineering from Seoul National University, Seoul, Korea, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, in 1992. From 1993 to 1997, he was in the Signal Processing Laboratory, Samsung Electronics, Korea, where he conducted research and development into video compression algorithms, the design of digital broadcasting satellite receivers, and other MPEG-related research for multimedia applications. Since September 1997, he has been with the faculty of the School of Electronic and Electrical Engineering, Sungkyunkwan University, Korea, where he is currently a full Professor. He served as Project Manager of Digital TV and Broadcasting in the Korean Ministry of Information and Communications from 3/2004 to 2/2006, where he supervised all digital TV-related R&D in Korea. He has authored many papers in the areas of video compression, pre/post processing, and pattern recognition. His research interests include multimedia signal processing, video compression, statistical pattern recognition, and remote sensing. Dr. Jeon is a member of Tau Beta Pi and Eta Kappa Nu. He is a member of SPIE, IEEK, KICS, and KSOBE. He was a recipient of the 2005 IEEK Haedong Paper Award in Signal ProcessingSociety, Korea. (Email: bjeon@skku.edu).

Ling Chen was born in 1951. He graduated in the Mathematics department of Yangzhou Teachers' College. Currently he is a professor of computer science, in the Information Technology College, Yangzhou University. He is a member of IEEE and ACM. His research interests include artificial intelligence, data mining, system optimization, complex network analysis. He has published more than 200 papers in journals and conferences. He has also authored/co-authored 6 books. He has 15 research projects supported by Chinese Natural Science Foundation and other organizations. He has received 5 Awards of Progress in Science and Technology from the Government of Anhui and Jiangsu Province. He was awarded the Government Special Allowance by the State Council.

Bolun Chen is the Professor in Huaiyin Institute of Technology. He received his Ph.D. degree from Nanjing University of Aeronautics and Astronautics (NUAA) in the year 2016. He was a visiting scholar at the University of Fribourg. His research interests include link prediction, recommender systems, data mining, and so on. At present, he is a reviewer of the Journal of Information Science,World Wide Web Journal and Reliability Engineering & System Safety.