Neurocomputing 260 (2017) 69–78

Contents lists available at ScienceDirect

Neurocomputing

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

Positive influence maximization in signed social networks based on simulated annealing

Dong Li[a,*], Cuihua Wang[b,*], Shengping Zhang[b], Guanglu Zhou[b], Dianhui Chu[b], Chong Wu[a]

[a] School of management, Harbin Institute of Technology, Harbin, China
[b] School of computer science and technology, Harbin Institute of Technology, Weihai, China

ARTICLE INFO

Article history:
Received 24 August 2016
Revised 12 December 2016
Accepted 1 March 2017
Available online 9 March 2017

Communicated by: Peng Cui

Keywords:
Influence maximization
Signed social networks
Polarity
Positive
Simulated annealing

ABSTRACT

Current studies of influence maximization focus almost exclusively on unsigned social networks ignoring the polarities of the relationships between users. Influence maximization in signed social networks containing both positive relationships (*e.g.*, friend or like) and negative relationships (*e.g.*, enemy or dislike) is still a challenging problem which remains much open. A few studies made use of greedy algorithms to solve the problem of positive influence or negative influence maximization in signed social networks. Although greedy algorithm is able to achieve a good approximation, it is computational expensive and not efficient enough. Aiming at this drawback, we propose an alternative method based on Simulated Annealing (SA) for the positive influence maximization problem in this paper. Additionally, we also propose two heuristics to speed up the convergence process of the proposed method. Comprehensive experiments results on three signed social network datasets, Epinions, Slashdot and Wikipedia, demonstrate that our method can yield similar or better performance than the greedy algorithms in terms of positive influence spread but run faster.

1. Introduction

In recent years, online social networks represented by Twitter, Weibo and Facebook are developing rapidly. The increasing availability of online data promotes social networks analysis and mining research where influence maximization is a key problem which is formally defined as following: given a social network modeled as a graph G and a nonnegative number k, find a set of k seed nodes, such that by activating them initially, the influence propagation scale of the seed node set is maximum under a certain diffusion model.

Influence maximization has gained tremendous attention motivated by widespread application scenarios (*e.g.*, such as viral marketing, expert finding, disease outbreak detection *etc.*). Kempe et al. [1] were the first to formulate influence maximization as a discrete stochastic optimization problem and attempted to solve the problem using a greedy algorithm to based on independent cascade (IC) model and linear threshold (LH) model. Then for solving the scalability issue of the solution of Kempe et al., improved greedy algorithms and heuristic methods are proposed in [2–7]. Research communities also have explored the extensive studies of influence maximization problems from various aspects, such as competitive influence maximization [8–12], time-critical influence maximization [13,14], influence maximization in heterogeneous networks [15], topic-aware influence maximization [16] and so on.

All the above studies are conducted in unsigned social networks which ignore the polarities of relationships between users. Actually, however, the relationships own the property of the polarity, which can divided into positive relationships (*e.g.*, friend or like) and negative relationships (*e.g.*, enemy or dislike). The social networks including both positive relationships and negative relationships simultaneously are called as signed social networks. In the study of influence maximization, ignoring the relationship polarities between users and treating signed social networks as unsigned ones roughly lead to over-estimation of positive influence which will cause bad effect in practical applications. Influence maximization in signed social networks is a critical problem that remains pretty much open.

The goal of influence maximization in signed social networks is usually to find the seed node set with maximum positive influence, which is also called as positive influence maximization (PIM). There is a few research work related with influence maximization in signed social networks. Li et al. [17] extended the most widely used independent cascade model to signed social networks, and attempt to solve the PIM problem based on the extended model. Wang et al. [18] and Shen et al. [19] also

* Corresponding authors.
E-mail addresses: hitlidong@hit.edu.cn, superdongli@gmail.com (D. Li), vipwch@126.com (C. Wang).

http://dx.doi.org/10.1016/j.neucom.2017.03.003
0925-2312/

explored similar problems under different extensions of another popular linear threshold model. Influence maximization problem is composed of two parts: diffusion model and seed node selection strategy. All of above studies adopted the greedy algorithm as the seed node selection strategy to solve the PIM problem under different models. Although the greedy algorithm can approximates to the optimum within a high factor $(1-1/e)$, it is expensive computing and not effective enough. To address the inefficiency issue, we propose a novel seed node set selection strategy based on simulated annealing (SA) to solve the PIM problem in this paper, which runs faster than the greedy algorithm with similar or better positive influence performance. Specifically, we make the following contributions in this paper:

- We propose a novel seed node set selection strategy based on simulated annealing for finding the seed node set with maximum positive influence in this paper. To the best of our knowledge, this is the first work taking advantage of the simulated annealing mind for the influence maximization in signed social networks.
- To further improve the efficiency of the seed node set selection algorithm based on SA, we also propose the single node positive influence heuristic and the distance limitation heuristic which is favorable to accelerate the algorithms convergence process and achieve better performance.
- We conduct comprehensive experiments on Epinions, Slashdot and Wikipedia datasets, and experimental results validate that our method is more efficient than the state-of-the-art greedy algorithm, while the performance is comparable or better than the greedy algorithms in terms of positive influence propagation.

The rest of this paper is organized as follows: Section 2 summarizes related works; Section 3 presents problem statement and the Polarity-related Independent Cascade diffusion model; Section 4 introduces the new seed selection algorithm based on simulated annealing; Section 5 presents experimental results that validate the effectiveness of our methodology; Finally, we conclude our work and present future research directions in Section 6.

2. Related work

Influence maximization problem. Influence maximization as an algorithmic technique for viral marketing was first proposed by Domingos and Richardson [20]. Kempe et al. [1] were the first to formulate it as a discrete optimization problem. They proved that the problem is NP-hard, and proposed a greedy approximation algorithm under independent cascade (IC) model and linear threshold (LH) model. To address the inefficiency issue of the greedy method, Leskovec et al. [2] proposed a seed nodes seeking method named as "Cost-Effective Lazy Forward" (CELF) an optimization in selecting new seeds, which fully makes use of the sub modularity property of the objection function. Chen et al. proposed two heuristic methods, DegreeDiscountIC [4] and PMIA [5], which make use of local arborescence structures of each user to approximate the influence propagation. Jung et al. [3] proposed the IRIE algorithm integrating the advantages of influence estimation and influence ranking methods, which runs much faster than the PMIA method. Cheng et al. [6] presented a static greedy algorithm which can strictly guarantee the submodularity of influence function in the process of the seed selection. They also [7] proposed an iterative ranking framework (IMRank) to solve influence maximization problem.

Budak et al. [8] attempted to address the influence limitation problem where a bad campaign begins to spread from a certain user in the network and make use of limiting campaigns to counteract the impact of misinformation. Similarly, under the competitive linear threshold (CLT) model. He et al. [9] also studied competitive influence diffusion in social networks. To simulate multiple cascades on a network, Pathak et al. [10] proposed a generalized version of the linear threshold model while allows nodes to switch between different cascades. Borodin et al. [11] suggested several natural extensions to the linear threshold model that was used in the single-technology case. Chen et al. [12] extended the independent cascade model for incorporating the emergence and propagation of negative opinions. Liu et al. [13] and Chen et al. [14] considered the time factor of influence diffusion, and studied the time constrained influence maximization problem. Wang et al. [15] proposed the problem of influence maximization on the heterogeneous networks. Chen et al. [16] studied topic-aware influence maximization problem.

Although influence maximization has attracted tremendous attention, a few of current studies focus on influence maximization in signed social networks. Li et al. [17] proposed polarity-related influence maximization (PRIM) problem in signed social networks. They also developed Polarity-related Independent Cascade (IC-P) diffusion model and made use of a greedy algorithm to solve the PRIM problem under the IC-P model. Wang et al. [18] and Shen et al. [19] extended another classic linear threshold model from different perspectives, then also adopted a greedy method to select the seed node set with maximum positive influence. Srivastava et al. [21] proposed a method to estimate the expect positive influence, and then made use of the greedy mind to select seed nodes. However, the greedy algorithm used in current studies is not efficient enough, we will focus on solving this problem in this paper based on simulated annealing mind. Our idea is inspired by the literature [22] where Jiang et al. attempted to make use of simulated annealing to solve the traditional IM problem.

Signed social networks. Song and Meyer [23] optimized the GAUC loss for positive relationship recommendation in signed social networks. They [23] also attempted to forecast the signs of the links and rank links in signed networks. Tang et al. [24] claimed that negative relationships are helpful for various analytical tasks, and modeled positive and negative links simultaneously to solve the node classification task. Fan et al. [25] prosed an extension of the classic epidemic model (Susceptible Infected Recovered, SIR) to simulate the process of information propagation over signed social networks. Shi et al. [26] attempted to model the evolution of opinions in a signed graph, and found that the relative strengths and structures of polarity links impact opinion convergence deeply. Tang et al. [27] exploited positive and negative relationships to recommend items in signed social networks. Hassan et al. [28] leveraged the natural language processing technology to extract signed social networks form text. However, above studies did not conduct influence maximization research on signed social networks.

3. Problem formulation and IC-P model

In this section, we first present the positive influence maximization (PIM) problem [17] in signed social networks we aims to solve, and then introduce the adopt polarity related independent cascade model (IC-P) in this paper.

3.1. Positive influence maximization problem

We define $f_+(\cdot)$ be the positive influence function. Given a node set S, $f_+(S)$ returns the expected number of nodes activated to be positive state by S which is considered as the positive influence of S. Similarly, let $f_-(\cdot)$ be the the negative influence function. $f_-(S)$ returns the expected number of nodes activated to be negative state by S which is considered as the negative influence. In this paper, we adopt Polarity-related Independent Cascade (named IC-P) diffusion [17] to estimate the values of $f_+(S)$ and $f_-(S)$, which is an extension of Independent Cascade in signed social networks.

Given a signed social network, a diffusion model and a $k > 0$, the goal of the positive influence maximization problem is seeking a set S containing k seed nodes such that the expected number of nodes activated to be positive state by S is maximized. The PIM problem can be formalized as,

$$S^{+} = \arg \max_{S \subseteq V, |S|=k} f_{+}(S), \tag{1}$$

In PIM problem, all initial seed nodes in the set S are set to be positive state. This design is decided by the particular application scenarios of the PIM problem. The more explanations can be found in the literature [17].

3.2. Modeling information diffusion in signed social networks

In this paper, we adopt polarity related independent cascade model (IC-P) [17] to estimate the polarity influence of candidate node sets. Here, we first introduce how to model a signed social network to be a directed and signed graph, and then present the details of IC-P model.

A signed social network can be modeled as a directed and signed graph $G = (V, E, A, P)$, where V and E represent the set of nodes and the set of directed and signed edges, respectively. Nodes correspond to users and relationships correspond to relationships. A is a non-negative weighted adjacency matrix whose element $A_{u,\, v}$ is the weight of the edge $(u,\, v) \in E$. P is another matrix whose element $P_{u,\, v}$ represents the sign of edge $(u,\, v)$ in the graph. $P_{u,\, v}$ has three values (1, -1, 0) which denote positive relationship, negative relationship and no relationship from u to v, respectively.

The IC-P model is an extension of the classic IC model which integrates the polarities of user relationships in signed social networks. In the IC-P model, each node owns three state choices: positive, negative, or inactive. We define $S(u)$ as the state of node u. Values 1, -1, 0 of $S(u)$ represent u's positive state, negative state and inactive state, respectively. For a node u, three states signify the corresponding user supports, opposes or does not care the the propagating information, respectively.

As Li et al. introduced in [17], "*in the IC-P model, the propagation process begins with an initial node set S. Each node in set S can be positive or negative state, and all other nodes not in S are inactive in the graph at the beginning. Here, the diffusion rules of the IC-P model are as following: for a node u, if it is activated by its neighbors in time* $t - 1$, *the state of node u will become positive or negative in next time step* (t). *we assume* $N^{t}_{active}(v)$ *as the neighbor set of node v who become positive or negative in time t. The new activated user u only have a single chance to activate each node currently inactive neighbor v in time step t. In time t, each node* $u \in N^{t}_{active}(v)$ *will attempt to activate v successfully with the probability*$A_{v,\, u}$ *in an arbitrary order. In the IC-P model, a node vcan only be activated once in a time step. Once the node v is activated by a node in* $N^{t}_{active}(v)$, *other nodes in* $N^{t}_{active}(v)$ *cannot activate node v any more. The diffusion process continues until no newly node will be activated.*"

When a node v is activated, how to determine the state of the node is key problem. The IC-P model incorporates the social principles that "the friend of my enemy is my enemy" and "the enemy of my enemy is my friend". The state of the newly activated node is determine by the node u which activated v and the polarity of relationship from node u to v, *i.e.* $S(v) = S(u) \times P_{u,v}$. The more details of the IC-P model can be found in the literature [17].

4. Simulated annealing based algorithm for positive influence maximization

In this section, we first propose the seed node set selection algorithm based on simulated annealing (SA) for solving the PIM problem, and then present two heuristic methods to speed up the convergence process of our method.

4.1. Seed nodes seeking algorithm based on simulated annealing

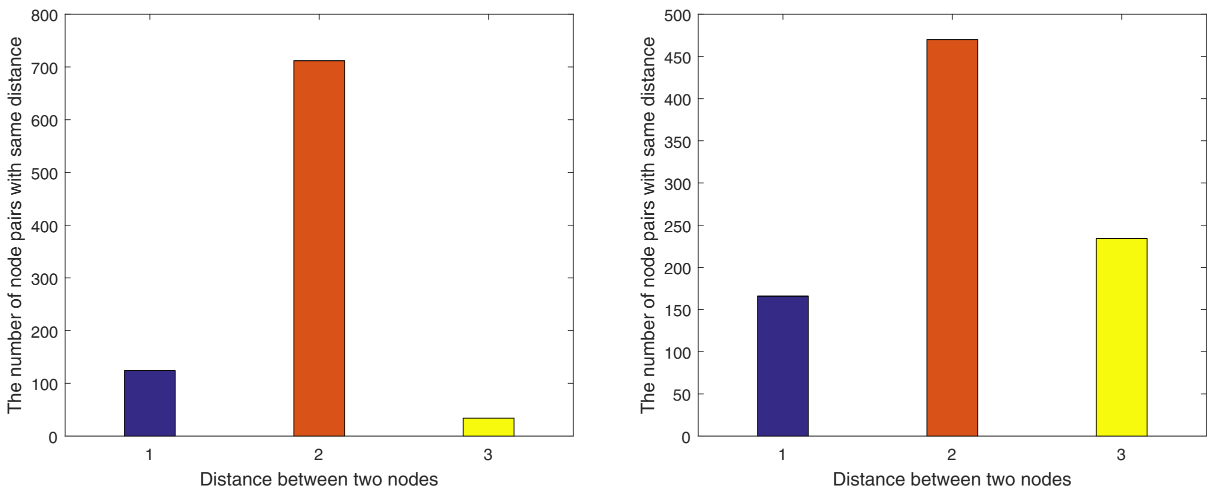
Simulated annealing was first proposed in 1953 by Metropolis et al. [29]. The name and inspiration originate from annealing in metallurgy. The basic iteration of simulated annealing is as following: in each step, the SA method takes into account both some neighboring state S' and the current state S, and probabilistically choose one from state S' or state S as the state of the next step. These probabilities ultimately drive the system to shift to states with lower energy. Typically the above process is repeated until a given computation budget has been exhausted or until the system reaches a state which can satisfy the requirement of the application.

Simulated annealing is a method that finds a good (not necessarily perfect) solution for an optimization problem. It is often applied in the situation where the search space is discrete (*e.g.*, all users that visit a given set of cities). When we want to maximize or minimize something, our problem can likely be tackled with simulated annealing. The positive influence maximization is exactly the problem which can be solved by simulated annealing. Nextly, we propose the seed node set seeking strategy based on simulated annealing.

Algorithm 1 states the details of the the seed node set seeking strategy based on simulated annealing for the PIM problem, which contains two levels of iterations: T_{f_+} and ΔT controls the outer lever iteration, and q controls the inner lever. At the beginning, we generate an initial seed node set $S = \{v_1, v_2, \dots v_k\}$ (line 4). In each iteration, we replace one node in solution set S with a node in $V - S$ (line 7) and then get a new solution S'. If $\Delta(f_+) = f_+(S') - f_+(S)$ is positive, this means that the new seed node set owns larger positive influence than the old one. Therefore, the new solution S' is better and is adopted (line 11). Or else, the new generated solution will be adopted if $exp(\Delta(f_+)/T_t) > random[0, 1]$ (lines 13–15). T_t is a outer loop parameter which will be updated when the number of inner interactions reaches the number q (line 17). We can see that our algorithm allows to replace the current solution with a worse one, which is different from methods based

Algorithm 1 Seed node set selection algorithm based on simulated annealing.

1: **Input:** a signed network graph G, and size of node set k, initial temperature T_0, termination temperature T_{f_+}, the number of inner loop q, the amount to cut down in the outer loop ΔT;
2: **Output:** the seed set of top-k positive influential nodes S;
3: $t \leftarrow 0$, $T_t \leftarrow T_0$, $count \leftarrow 0$;
4: select an initial seed set $S \subset V$, $|S| = k$;
5: **while** $T_t < T_{f_+}$ **do**
6:   calculate $f_+(S)$ using IC-P model;
7:   generate a new neighbor solution set S';
8:   $count \leftarrow count + 1$;
9:   calculate the positive influence change of the new solution$\Delta f_+ \leftarrow f_+(S') - f_+(S)$;
10:   **if** $\Delta f_+ > 0$ **then**
11:     $A \leftarrow S'$;
12:   **else**
13:     generate a random number $\lambda = random(0, 1)$;
14:     **if** $exp(\Delta f_+/T_t) > \lambda$ **then**
15:       $A \leftarrow S'$;
16:   **if** $count > q$ **then**
17:     $T_t \leftarrow T_t - \Delta T$, $t \leftarrow t + 1$, $count \leftarrow 0$;
18: return S.



(a) WC model (b) UN model

Fig. 1. Distance distribution among top 30 nodes selected from Epinions dataset by the greedy algorithm.

on the greedy strategy. Our method can escape the local optimum and achieve a larger probability of getting optimal solutions than the greedy algorithm.

How to ensure the algorithm based on SA for the PIM problem converges is a critical problem. Next we present the convergence proof of our method.

Theorem 1. *The proposed seed node set selection algorithm based on simulated annealing for positive influence maximization problem will converge towards the optimum with the increasing of the iteration number t.*

Proof. The Markov Chain corresponds to the PIM problem: we define a state i as a result node set S_i which contains k nodes in the signed graph G, *i.e.* $f_+(i) = f_+(S_i)$. The union set of S_i's neighbor solutions are denoted as $N(i)$. If the current state is i, the probability that another state $j \in N(i)$ is created is $g_{i,j} = 1/(k*(N-k))$ and the probability that state j is accepted is $a_{i,j} = min\{1, exp[-(f_+(j) - f_+(i))/T(t)]\}$. Thus, the transition probability from state i to state j is as following:

$$\forall i,\, j,\, p_{i,j}(t) = \begin{cases} g_{i,j}a_{i,j}(t), & if\ j \in N_i, j \neq i \\ 0, & if\ j \notin N_i, j \neq i \\ 1 - \sum_{k \in N_i} p_{i,k}(t), & j = i \end{cases} \tag{2}$$

The finite state Markov Chain corresponding to Algorithm 1 satisfies the following conditions:

(1) $\forall i, j \in \Omega$ (states union), $g_{i,j}(t)$ is not related to t, and $g_{i,j} = g_{j,i}$, while $\exists n \geq 1$, $s_0, s_1, \ldots, s_n \in \Omega$, $s_0 = i$, $s_n = j$, to make $g_{s_k, s_{k+1}}(t) > 0$, $k = 0, 1, \ldots, n-1$;
(2) $\forall i$, j, $k \in \Omega$, if $f_+(i) \leq f_+(j) \leq f_+(k)$, then $a_{i,k}(t) = a_{i,j}(t)a_{j,k}(t)$;
(3) $\forall i, j \in \Omega$, if $f_+(i) \geq f_+(j)$, then $a_{i,j}(t) = 1$; if $f_+(i) < f_+(j)$, then $0 < a_{i,j}(t) < 1$.

Thus, based on the work of Mitra et al. [30], for the stationary distribution of the Markov Chain $v = \{v_1, v_2, \ldots, v_{|V|}\}$.
□

$$\lim_{t \to 0} v_i(t) = \begin{cases} \frac{1}{|\Omega_{opt}|}, & i \in \Omega_{opt} \\ 0, & i \notin \Omega_{opt} \end{cases} \tag{3}$$

Ω_{opt} is the sum of optimal solutions. This result means that our propose method can tend to converge to optimum.

4.2. Heuristics for speeding up convergence

Algorithm 1 randomly selects one unselected node to replace a node in current solution set to generate a new solution set. In fact, the random selection process is not unfavorable to the convergence of SA algorithm, which drives us to explore effective heuristics to accelerate the convergence. In this paper, we propose two heuristics based on analysis of results got by the greedy algorithm on Epinions, Slashdot and Wikipedia datasets. Three datasets are introduced in detail in Section 5.

Distance limitation heuristic. Figs. 1–3 present the distance distribution among top 30 nodes selected from Epinions, Slashdot and Wikipedia datasets by the greedy algorithm, respectively. In all of three figures, x-axis represents the length of the shortest path between two nodes, y-axis represents the number of node pairs with the same distance. From Figs. 1–3, we observe that the probability that two nodes from the set obtained by the greedy algorithm have a short distance is very large. Hence, in the process of selecting a new node to generate a new solution set S', we do not consider the nodes whose shortest path to nodes in the current solution set S are larger than a threshold d (the value of d is set to be 2 in our experiments). We define $len(u,\ v)$ as the length of the shortest path from u to v in a social network. Let $M(S, d) = \{u | u \in V - S, \forall v \in S, len(u, v) \leq d\}$. When we generate a new neighbor solution, we will select a node from $M(S, d)$.

Single node positive influence heuristic. In Epinions, Slashdot and Wikipedia datasets, for each node i, we first make use of IC-P model to estimated its single positive influence $f_+(\{i\})$. Then we sort these nodes with positive influence in descending order and select top 30 nodes in the sorted list. We also select top 30 nodes using the greedy algorithm based on IC-P model from two datasets. Then we compare two kinds of node sets got by different methods. Figs. 4–6 present the overlap ration distribution at different sizes of the node set on Epinions, Slashdot and Wikipedia datasets respectively. In all of three figures, x-axis represents the size of the seed node set, x-axis represents the overlap ration between two node sets. We find that, the overlap ration is usually very high at different node set sizes. So, in the process of selecting a new node to generate a neighbor solution set S', we consider the positive influence of single node such that the algorithm will be able to obtain a better solution in less time.

Epinions, Slashdot and Wikipedia are the three most popular and public signed social networks. To the best of our knowledge,

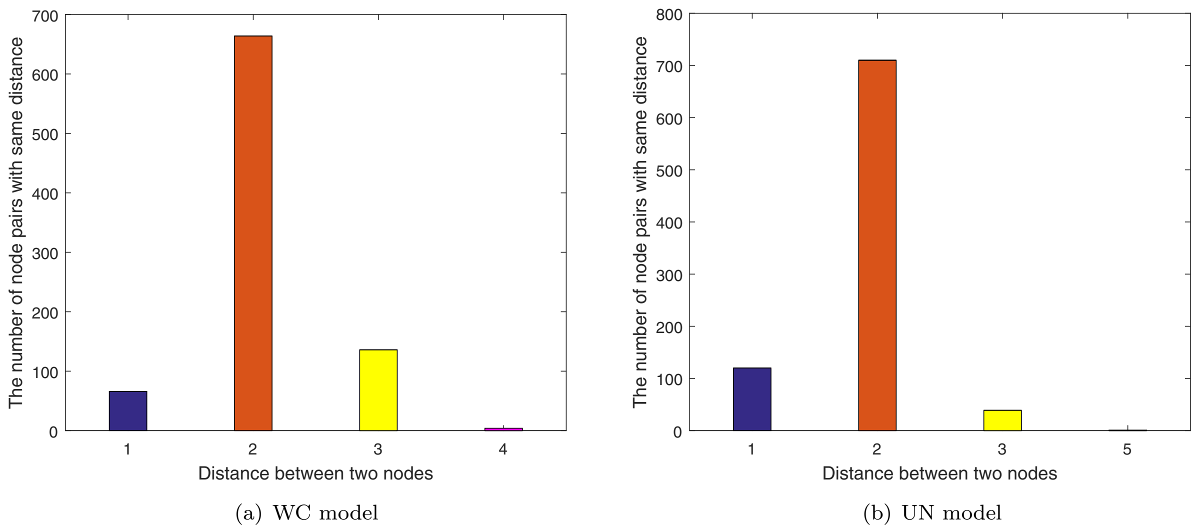


Fig. 2. Distance distribution among top 30 nodes selected from Slashdot dataset by the greedy algorithm.

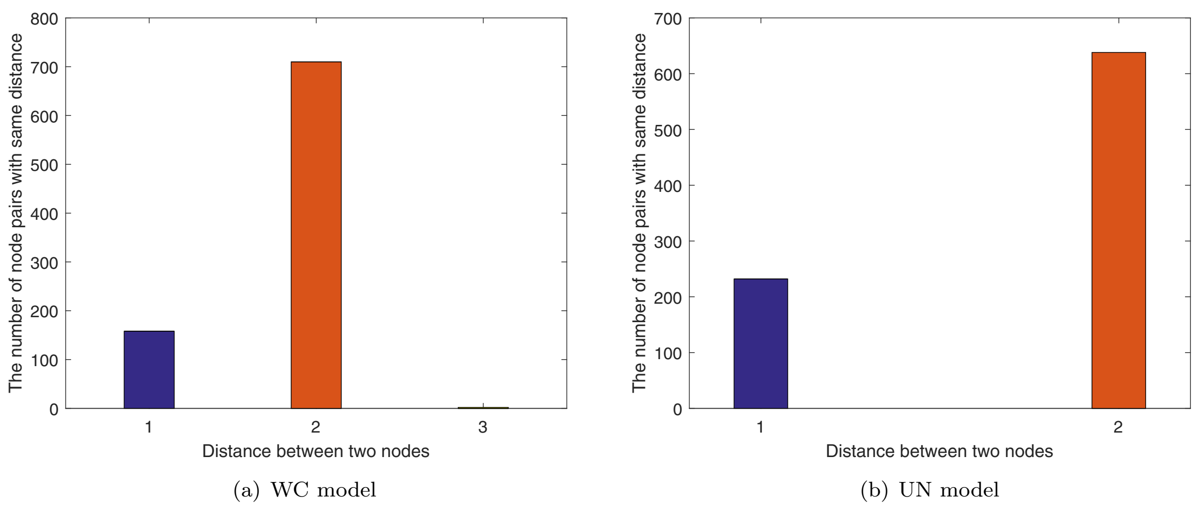


Fig. 3. Distance distribution among top 30 nodes selected from Wikipedia dataset by the greedy algorithm.

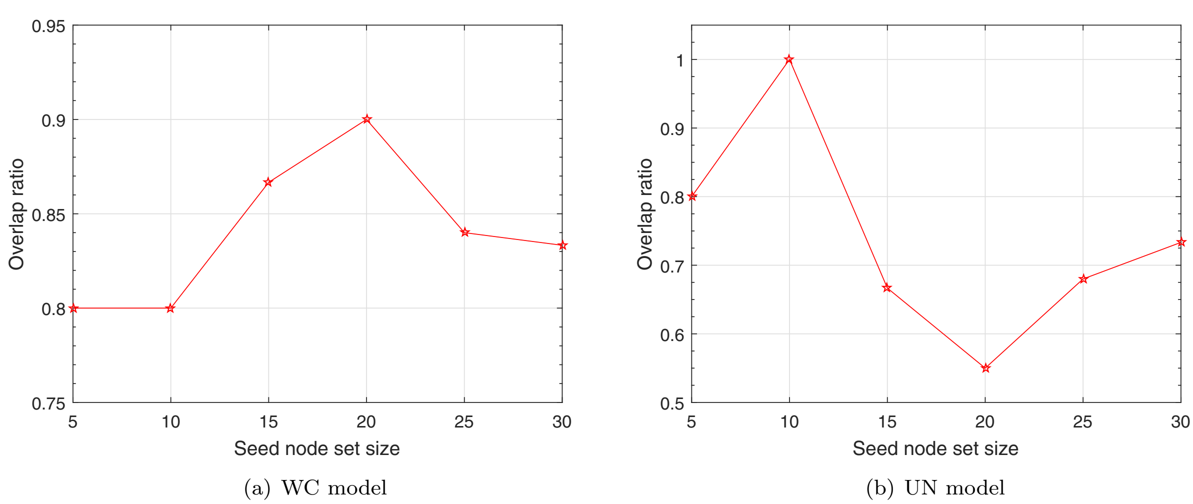
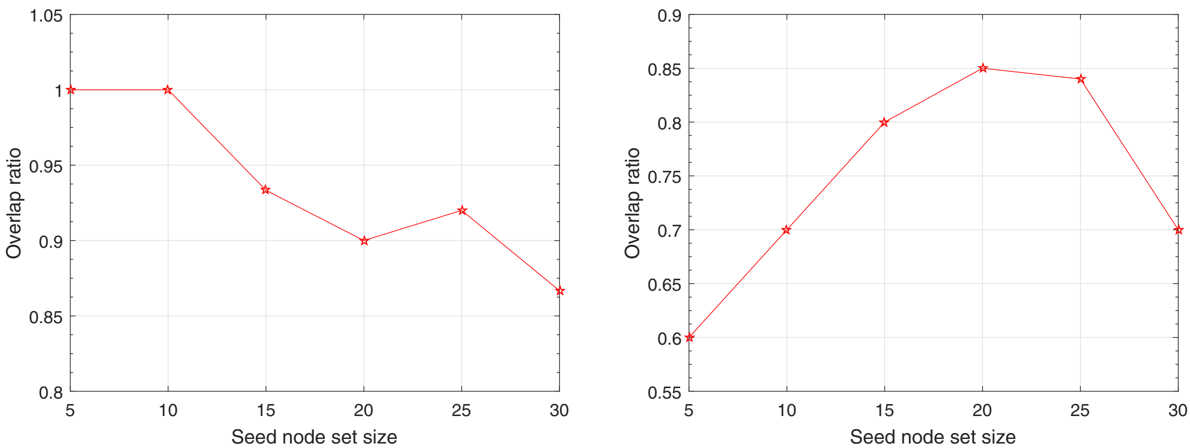
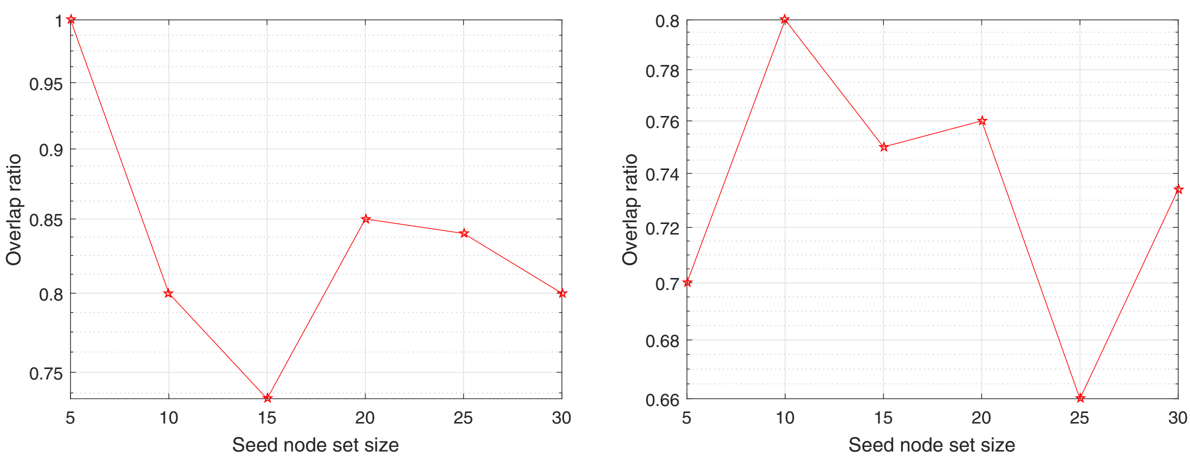


Fig. 4. Overlap ration between two node sets selected by the single node positive influence and the greedy algorithm on Epinions dataset.



(a) WC model (b) UN model

Fig. 5. Overlap ration between two node sets selected by the single node positive influence and the greedy algorithm on Slashdot dataset.



(a) WC model (b) UN model

Fig. 6. Overlap ration between two node sets selected by the single node positive influence and the greedy algorithm on Wikipedia dataset.

nearly all of the studies about signed social networks adopted one or more ones from above three networks to conduct experiments. In this paper, we find the common properties on all of three signed social networks and propose two effective heuristics based on the properties. The found properties come from all available signed networks, we consider that they should hold a fairly high credibility and the corresponding heuristics will be helpful to the SA method.

Based on two heuristics, we propose Algorithm 2 to construct a new neighbor set S'. Given an initial node set S, Algorithm 2 first calculates the individual positive influence of each node in G. Then we randomly choose a node belonging to the set S and replace it with another node in $V - A$ (lines 8–11). How to select the new candidate node from $V - A$ mainly follows two rules corresponding to two heuristics: (1) the bigger the individual positive influence of a node, the bigger the probability that the node is selected; (2) only if the distance from a node to the current node set A is less than 2, this node has chance to be selected.

Algorithm 2 The algorithm of creating a new neighbor solution.

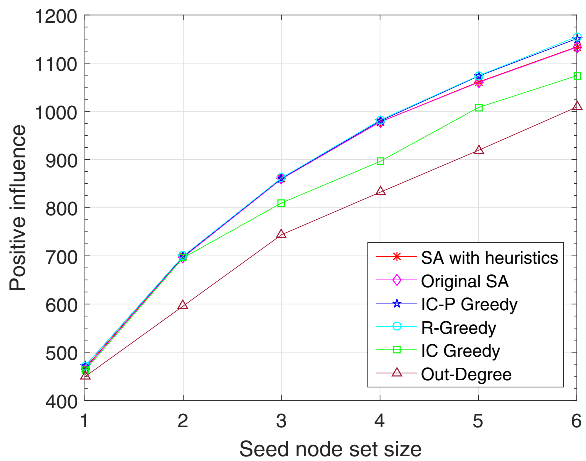
1: **Input:** a signed network graph G, and the initial seed node set S;
2: **Output:** a new neighbor solution set S';
3: calculate $f_+(v)$ for each node $v \in V$;
4: $total \leftarrow \sum_{v \in V}$, $p(0) \leftarrow 0$, $mark \leftarrow true$;
5: **for** i from 1 to $|V|$;
6: $p(i) \leftarrow p(i-1) + f_+(v_i)/total$;
7: **while** $mark$ **do**;
8: generate a random number $r \in [1, |S|]$, and a random probability $p \in [0, 1]$;
9: **for** each node i in $V - S$;
10: if $p(i-1) < p < p(i)$, $v_i \in M(S, 2)$;
11: then choose v_i to replace rth element of S to generate S', $mark \leftarrow false$;
12: return S';

5. Experiments

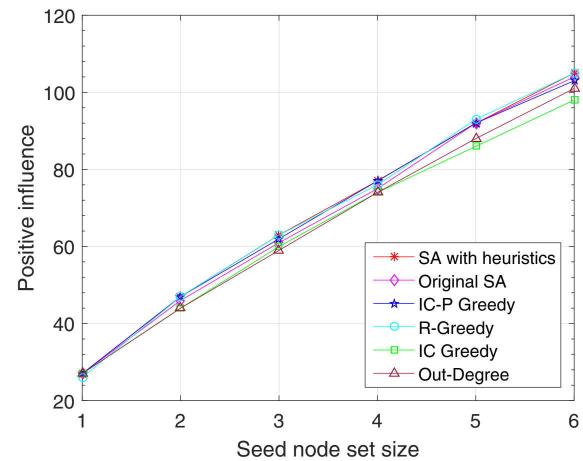
5.1. Experimental setup

Datasets. In the experiments, we adopt three large online signed social networks (Epinions, Slashdot and Wikipedia) as the datasets, which are downloaded from Standard Large Network Dataset Collection (http://snap.stanford.edu/). In all of three signed networks, positive polarity or negative polarity is tagged on each relationship between users directly.

- *Epinions.* It is a popular product review site. Users in Epinions can label trust (positive) or distrust (negative) to other



(a) WC model



(b) UN model

Fig. 7. Results on Epinions dataset.

users. This dataset contains 11,567 users and 93,204 relationships with polarities.

- *Slashdot.* It is a popular technical news site. Users in Slashdot can label friend (positive) or foe (negative) to other users. This dataset contains 10,966 users and 44,356 relationships with polarities.
- *Wikipedia.* It is a free online encyclopedia, written and modified by volunteers all over the world. This website develops a voting mechanism to generate administrators. Users can give positive votes or negative votes to the admin candidates. This dataset contains 7118 users and 107,080 relationships with polarities.

Generating influence probabilities. Because the interactive information used to calculate the influence diffusion probability $A_{u,v}$ of each link (u, v) in the graph G is not available, so we make use of three popular models [4,17] to generate diffusion probabilities for edges in the graph in our experiments.

- *Weighted Cascade (WC) model.* This model mainly make use of the in-degree of node to generate diffusion probabilities. For an edge (u, v), the diffusion probability of the edge $A_{u,v}$ is equal to $1/d(v)$, and $d(v)$ is the in-degree of v.
- *Uniformly (UN) model.* All edges in the graph are uniformly assigned a same probability value, which is set to be 0.01 in this paper.

Comparison methods. We compare our method (SA) with IC-P Greedy, IC Greedy and Out-Degree methods. We list the methods compared in our experiments as following:

- *Original SA.* This is our method presented by Algorithm 1 in Section.
- *SA with heuristics.* Different from Original SA, this method makes use of two heuristics to accelerate the convergence process following Algorithm 2.
- *IC-P Greedy.* This method makes use of the greedy algorithm with the lazy-forward optimization to solve the PIM problem under the IC-P model [17].
- *R-Greedy.* This method is mainly used to solve the problem whose influence function is non-monotone and non-submodular. Shen et al. [19] adopted this method to solve the influence maximization problem in signed social networks.
- *IC Greedy.* This method does not consider the relationship polarities and treat signed social networks as unsigned ones. In these unsigned social networks, we make use of the original greedy algorithm with the CELF optimization [1] to get the seed node set of size k owning biggest non-polar influence under the standard IC model.
- *Out-Degree.* This is a popular baseline method that selects top k nodes based on out degrees, which is also used in [1,5,17].

Because IC Greedy and out-degree methods cannot provide the positive influence of selected seed node sets, we adopt IC-P model to obtain them. For each seed node set, we first simulate the spread process for 20,000 times using the IC-P diffusion model, then take the average all these simulations as the positive influence. On the two datasets, the upper limitation of selected number k is set to be 30. We compare the positive influence of seed node with different sizes (*i.e.* $k = 5, 10, 15, 20, 25, 30$). All the experiments are implemented on a PC with 2.93 GHz Quad-Core Intel Core(TM)i7 870 and 8G memory.

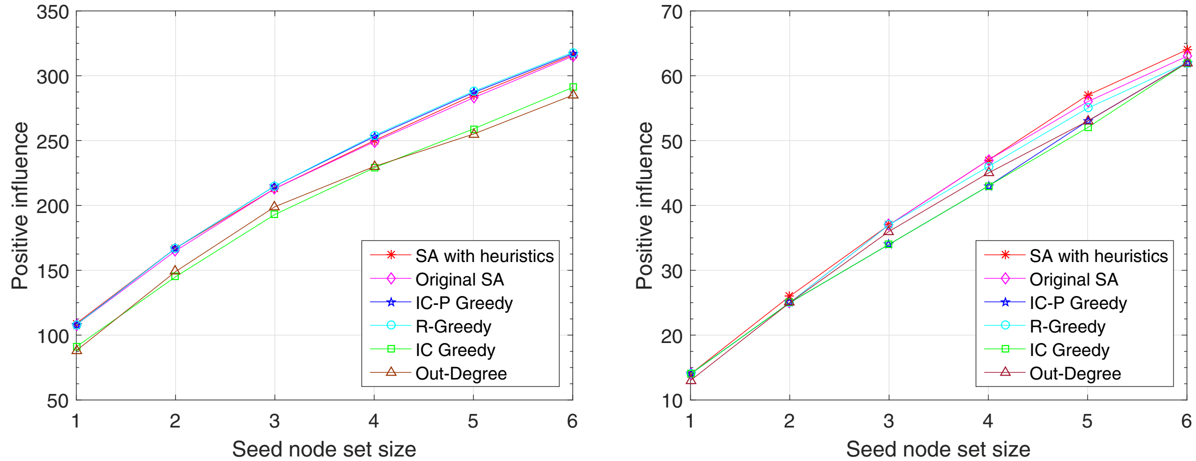
5.2. Experiment result

5.2.1. Accuracy comparison

Fig. 7 presents the performance of six different methods (Original SA, SA with heuristics, IC-P Greedy, R-Greedy, IC Greedy and Out-Degree) in terms of positive influence diffusion, with two kinds of diffusion probability (WC model, UN model) on the Epinions dataset. From Fig. 7, we can observe that our Original SA and SA with heuristics methods consistently get similar performance with IC-P Greedy and R-Greedy, and outperform significantly IC Greedy and Out-Degree methods. The bad performance of IC Greedy fully indicates that the necessary of influence maximization research taking into account polarities of relationships between users in social networks.

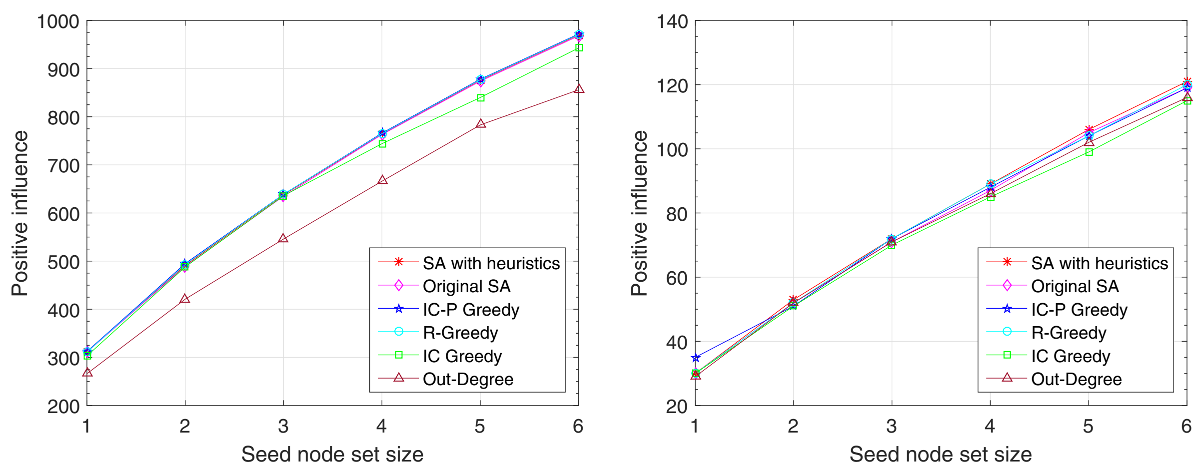
Fig. 8 presents the performance of six different methods in terms of positive influence diffusion, with two kinds of diffusion probability on the Slashdot dataset. We find that our Original SA and SA with heuristics methods yield similar performance with IC-P Greedy and R-Greedy under WC model, while achieve better performance than them under UN model. This means that our methods have the ability to find the hidden influential users that greedy algorithm cannot seek. Similar with results on the Epinions dataset, SA method and SA with heuristics also outperform significantly IC Greedy and Out-Degree methods on the Slashdot dataset.

Fig. 9 presents the performance of six different methods in terms of positive influence diffusion, with two kinds of diffusion probability on the Wikipedia dataset. We can see that Original SA, SA with heuristics, IC-P Greedy and R-Greedy achieve the similar



(a) WC model (b) UN model

Fig. 8. Results on Slashdot dataset.



(a) WC model (b) UN model

Fig. 9. Results on Wikipedia dataset.

Table 1
Running time of different methods on three datasets (*seconds*).

	Epinions dataset		Slashdot dataset		Wikipedia dataset	
	WC model	UN model	WC model	UN model	WC model	UN model
SA with heuristics	58874.639	11825.243	12314.620	14772.448	27729.096	10237.120
IC-P Greedy	198216.991	62485.413	52732.248	100315.747	48744.710	33280.951
R-Greedy	630858.691	168333.381	71505.395	1087110.724	140528.724	169095.315
Original SA	127845.656	34910.221	23187.387	50162.881	41039.561	24195.307

performance of positive influence diffusion. The rest two methods (IC Greedy and Out-Degree) perform worse than the above four methods.

5.2.2. Running time comparison

We compare the running time of SA with heuristics, IC-P Greedy, R-Greedy and Original SA when $k = 30$, with WC model and UN model, on Epinions, Slashdot and Wikipedia datasets. Table 1 shows the running time results. We find that our two SA methods runs faster than IC-P Greedy and R-Greedy. SA with heuristics needs less running time than Original SA. This proves that our proposed heuristics indeed speed up the coverage process. R-Greedy runs more slowly than IC-P greedy, and consumes the most running time among six methods. This is because that R-Greedy is designed to solve the problem whose influence function is non-monotone and non-submodular. It cannot adopt lazy-forward strategy [2] to speed up the selection process.

On the Epinions dataset, our SA with heuristics method runs faster 3.37 times and 5.28 times than IC-P Greedy, runs faster 10.71 times and 14.23 times than R-Greedy, with WC model and UN model respectively. On the Slashdot dataset, our SA method runs faster 4.28 times and 6.79 times than IC-P Greedy, runs faster 5.81 times and 73.59 times than R-Greedy Greedy, with WC model and UN model respectively. On the Wikipedia dataset, our SA method runs faster 1.76 times and 3.25 times than IC-P Greedy, runs faster 5.07 times and 16.52 times than R-Greedy, with WC model and UN model respectively.

6. Conclusions

In this paper, we propose a new seed node set seeking method based on simulated annealing to solve the positive influence maximization problem. Besides that, we also develop two heuristics, distance limitation and single node positive influence, to speed up the convergence process of our method. We conduct various experiments on three signed social networks, Epinions, Slashdot and Wikipedia. Experiment results present that our method and the greedy algorithm perform similarly in term of accuracy (positive influence scope), however, our method outperforms the greedy algorithm in term of efficiency (running time).

Several challenges and future directions remain. This paper study the PIM problem based on an extension of the classic Independent Cascade model (IC-P), which is able to estimate positive influence of node sets but is not efficient enough. There, one of our future directions is to explore new diffusion models in signed social network. Meanwhile, influence maximization in dynamic networks and influence maximization in location-based networks are also interesting problems.

Acknowledgments

This work is supported by the Hong Kong Scholar project of China (No. ALGA4131016116), the China Postdoctoral fund (No. 2016M600250), the Natural Science Foundation of China (No. 61602128), the Natural Science Foundation of Shandong Province (No. ZR2016FQ13) and the major science and technology project of Shandong Province (No. 2015ZDXX0201B02).

References

[1] 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, ACM, 2003, pp. 137–146.

[2] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, N. Glance, Cost-effective outbreak detection in networks, in: Proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2007, pp. 420–429.

[3] K. Jung, W. Heo, W. Chen, Irie: scalable and robust influence maximization in social networks, in: Proceedings of the Twelfth IEEE International Conference on Data Mining, IEEE, 2012, pp. 918–923.

[4] W. Chen, Y. Wang, S. Yang, Efficient influence maximization in social networks, in: Proceedings of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2009, pp. 199–208.

[5] W. Chen, C. Wang, Y. Wang, Scalable influence maximization for prevalent viral marketing in large-scale social networks, in: Proceedings of the Sixteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2010, pp. 1029–1038.

[6] S. Cheng, H. Shen, J. Huang, G. Zhang, X. Cheng, Staticgreedy: solving the scalability-accuracy dilemma in influence maximization, in: Proceedings of the Twenty-second ACM International Conference on Information & Knowledge Management, ACM, 2013, pp. 509–518.

[7] S. Cheng, H. Shen, J. Huang, W. Chen, X. Cheng, Imrank: influence maximization via finding self-consistent ranking, in: Proceedings of the Thirty-second International ACM SIGIR Conference on Research & Development in Information Retrieval, ACM, 2014, pp. 475–484.

[8] C. Budak, D. Agrawal, A.E. Abbadi, Limiting the spread of misinformation in social networks, in: Proceedings of the Twentieth International Conference on World Wide Web, ACM, 2011, pp. 665–674.

[9] X. He, G. Song, W. Chen, Q. Jiang, Influence blocking maximization in social networks under the competitive linear threshold model, in: Proceedings of the 2012 SIAM International Conference on Data Mining (SDM), SIAM, 2012, pp. 463–474.

[10] N. Pathak, A. Banerjee, J. Srivastava, A generalized linear threshold model for multiple cascades, in: Proceedings of the 2010 IEEE International Conference on Data Mining, IEEE, 2010, pp. 965–970.

[11] A. Borodin, Y. Filmus, J. Oren, Threshold models for competitive influence in social networks, in: Proceedings of the 2010 International Workshop on Internet and Network Economics, Springer, 2010, pp. 539–550.

[12] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, Y. Yuan, Influence maximization in social networks when negative opinions may emerge and propagate, in: Proceedings of the 2011 SIAM International Conference on Data Mining (SDM), 11, SIAM, 2011, pp. 379–390.

[13] B. Liu, G. Cong, D. Xu, Y. Zeng, Time constrained influence maximization in social networks, in: Proceedings of the Twelfth IEEE International Conference on Data Mining, IEEE, 2012, pp. 439–448.

[14] W. Chen, W. Lu, N. Zhang, Time-critical influence maximization in social networks with time-delayed diffusion process, in: Proceedings of the Twenty-sixth AAAI Conference on Artificial Intelligence, AAAI, 2012, pp. 439–448.

[15] G. Wang, Q. Hu, P.S. Yu, Influence and similarity on heterogeneous networks, in: Proceedings of the Twenty-first ACM International Conference on Information and Knowledge Management, ACM, 2012, pp. 1462–1466.

[16] S. Chen, J. Fan, G. Li, J. Feng, K.-l. Tan, J. Tang, Online topic-aware influence maximization, Proc. VLDB Endow. 8 (6) (2015) 666–677.

[17] D. Li, Z.-M. Xu, N. Chakraborty, A. Gupta, K. Sycara, S. Li, Polarity related influence maximization in signed social networks, PloS One 9 (7) (2014) e102199.

[18] H. Wang, Q. Yang, L. Fang, W. Lei, Maximizing positive influence in signed social networks, in: Proceedings of the 2015 International Conference on Cloud Computing and Security, Springer, 2015, pp. 356–367.

[19] C. Shen, R. Nishide, I. Piumarta, H. Takada, W. Liang, Influence maximization in signed social networks, in: Proceedings of the 2015 International Conference on Web Information Systems Engineering, Springer, 2015, pp. 399–414.

[20] P. Domingos, M. Richardson, Mining the network value of customers, in: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2001, pp. 57–66.

[21] A. Srivastava, C. Chelmis, V.K. Prasanna, Social influence computation and maximization in signed networks with competing cascades, in: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, ACM, 2015, pp. 41–48.

[22] Q. Jiang, G. Song, G. Cong, Y. Wang, W. Si, K. Xie, Simulated annealing based influence maximization in social networks, in: Proceedings of the Twenty-fifth AAAI Conference on Artificial Intelligence, 11, 2011, pp. 127–132.

[23] D. Song, D.A. Meyer, Recommending positive links in signed social networks by optimizing a generalized AUC, in: Proceedings of the Twenty-ninth AAAI Conference on Artificial Intelligence, 2015, pp. 290–296.

[24] J. Tang, C. Aggarwal, H. Liu, Node classification in signed social networks, in: Proceedings of the 2016 SIAM International Conference on Data Mining, 2016, pp. 54–62.

[25] P. Fan, H. Wang, P. Li, W. Li, Z. Jiang, in: Analysis of opinion spreading in homogeneous networks with signed relationships, 2012, 2012, p. P08003.

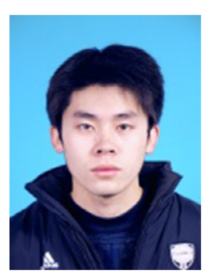
[26] G. Shi, A. Proutiere, M. Johansson, J.S. Baras, K.H. Johansson, The evolution of beliefs over signed social networks, Oper. Res. 64 (3) (2016) 585–604.

[27] J. Tang, C. Aggarwal, H. Liu, Recommendations in signed social networks, in: Proceedings of the Twenty-fifth International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2016, pp. 31–40.

[28] A. Hassan, A. Abu-Jbara, D. Radev, Extracting signed social networks from text, Workshop Proceedings of the 2012 TextGraphs-7 on Graph-based Methods for Natural Language Processing, Association for Computational Linguistics, 2012, pp. 6–14.

[29] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (6) (1953) 1087–1092.

[30] D. Mitra, F. Romeo, A. Sangiovanni-Vincentelli, Convergence and finite-time behavior of simulated annealing, in: Proceedings of the Twenty-fourth IEEE Conference on Decision and Control, IEEE, 1985, pp. 761–767.



Dong Li received his M.S. degree and Ph.D. degree from the Department of Computer Science and Technology at Harbin Institute of Technology in 2010 and 2015, respectively. He spent one year visiting Carnegie Mellon University during his Ph.D. period. He is currently a post doctoral at Hong Kong Baptist University. His research interests include information diffusion, link prediction and recommendation systems.

CuiHua Wang is currently a lecturer with the School of Computer Science and Technology, Harbin Institute of Technology at Weihai, China. Her main research interests include network security and social computing.

ShengPing Zhang received the Ph.D. degree in computer science from the Harbin Institute of Technology, in 2013. He is currently an Associate Professor with the School of Computer Science and Technology, Harbin Institute of Technology at Weihai, China. He has been a Post-Doctoral Research Associate at Brown University, and a Visiting Student Researcher at UC Berkeley. He has authored or co-authored over 30 research publications in refereed journals and conferences. His research interests focus on moving object detection, tracking, and social Computing.

GuangLu Zhou is currently an Associate Professor with the School of Computer Science and Technology, Harbin Institute of Technology at Weihai, China. His research interests include social media mining and computer version.

DianHui Chu received his M.S. degree and Ph.D. degree in Computer Science from Harbin Institute of Technology. He is now a Professor of Harbin Institute of Technology. His research interests are within the domain of service computing, service engineering, and social computing.

Chong Wu received his Ph.D. degree from Harbin Institute of Technology. He is now a Professor of Harbin Institute of Technology. His main research interests include statistics and management science and engineering.