Information Sciences 558 (2021) 34–49

Contents lists available at ScienceDirect

Information Sciences

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

Positive opinion maximization in signed social networks

Qiang He [a], Lihong Sun [a], Xingwei Wang [b,d,*], Zhenkun Wang [c], Min Huang [d,e], Bo Yi [b], Yuantian Wang [a], Lianbo Ma [f]

[a] College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, Liaoning 110169, China
[b] College of Computer Science and Engineering, Northeastern University, Shenyang 110169, China
[c] School of System Design and Intelligent Manufacturing, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China
[d] College of Information Science and Engineering, Northeastern University, Shenyang 110819, China
[e] State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China
[f] College of Software, Northeastern University, Shenyang 110169, China

ARTICLE INFO

Article history:
Received 2 January 2020
Received in revised form 20 December 2020
Accepted 31 December 2020
Available online 26 January 2021

Keywords:
Social network
Influence maximization
Opinion dynamics
Product promotion

ABSTRACT

Opinion maximization is a kind of optimization method, which leverages a subset of influential nodes in social networks to spread user opinions towards the target product and eventually obtains the largest opinion propagation. The current propagation models on the opinion maximization mainly focus on the activated nodes and the static opinion formation process. However, they neglect the combination between the activated nodes and the dynamic opinion formation process. Moreover, previous studies are more attentive to the positive relationships among users. In the real scenario, negative relationships among users may damage the product reputation. Therefore, in this paper, we study positive opinion maximization by using an Activated Opinion Maximization Framework (AOMF) in signed social networks. The proposed AOMF is composed of three phases: i) the selection of candidate seed nodes, ii) the activated opinion formation process and iii) the determination of seed nodes. We first use an effective heuristic rule to select candidate seed nodes. To model the activation and dynamic opinion formation process of network nodes, we devise the activated opinion formation model based on the multi-stage linear threshold model and the Degroot model. Then, we calculate the opinion propagation of each candidate seed node by using the activated opinion formation model. Based on the candidate seed nodes and the activated opinion formation process, seed nodes are further determined. Finally, experimental results on six social network datasets demonstrate that the proposed method has superior potential opinions and positive ratio than the chosen benchmarks.

1. Introduction

With the rapid development of social media (e.g., Facebook, Wechat and Weibo), social network analysis has emerged and flourished. So far, social network analysis has attracted great attention from academia and industry. As a typical problem of social network analysis, Influence Maximization (IM) [4,5] has been studied due to its widespread applications in product promotion, health interventions and rumour control [1–3]. The target of IM is to select some influential nodes, spread the target product by using the propagation model and eventually maximize the number of influenced nodes.

* Corresponding author at: State Key Laboratory of Synthetical Automation for Process Industries, College of Computer Science and Engineering, Northeastern University, Shenyang 110169, China.

https://doi.org/10.1016/j.ins.2020.12.091
0020-0255/

As a practical extension of the IM, the Opinion Maximization (OM) problem has been investigated [6,7] and widely utilized in product promotion or health interventions. However, each individual of the OM has a positive, neutral or negative opinion towards each product, while which of the IM only has a positive opinion. Thus, the goal of OM is to maximize the opinion propagation of influenced nodes rather than the number of influenced nodes. Correspondingly, the properties of objective function (i.e., the opinion propagation of influenced nodes) for the OM problem may be changed due to the introduction of negative opinions. Furthermore, the IM usually assumes that all users are potential customers and more influenced users will achieve a larger market share for the product. In the real scenario, some users may dislike the product and express negative opinions towards the product, which can damage the reputation of the product. Therefore, it is worthwhile for the company to maximize the spread of positive opinions and minimize the spread of negative opinions. For example, when a new product comes into the market, the company may employ some customers with positive opinions towards the product to spread positive comments to others and try to establish its good reputation. In a nutshell, the target of OM problem is to select some influential nodes, spread the desired opinions to others and maximize the desired opinion propagation within a specific time [8,9].

There are a substantial set of OM-related studies in the literature which mainly focus on the opinion propagation models and seeding algorithms. For opinion propagation models, there are two categories [10–12]: i) the opinion is known and fixed and ii) the opinion is changing dynamically over time. For seeding algorithms, there are greedy algorithms [13–16], heuristic algorithms [17] and their variants [18,40,41]. However, current studies on the OM problem are facing the following key challenges: i) the propagation models mainly concentrate on the activated nodes and the static opinion formation process, but neglect the combination between the activated nodes and the dynamic opinion formation process. ii) the OM problem is studied in unsigned social networks or signed networks with static user opinions [38,39]. However, in the real scenario, the OM problem in signed social networks with dynamic opinion formation process has not been investigated yet. iii) the propagation models usually assume an activated individual has a positive opinion towards the product. From the above analysis, we find both negative and neutral opinions can produce a significant influence on product promotion. For example, under the influence of word-of-mouth, one user likes to buy the Samsung mobile phone based on his preference. However, after the Samsung mobile phone explosion, the negative public opinions emerge. As a result, the Samsung phone gets a bad rap, which may affect the user's mind when deciding to buy a Samsung phone or others.

To settle the three issues mentioned above, in this paper, we propose an Activated Opinion Maximization Framework (AOMF) in signed social networks. The main contributions of this paper are summarized as follows.

1) We formulate the OM problem in signed social networks as an optimization model by considering positive opinions and negative opinions towards the product. We prove that the OM problem is NP-hard and the objective function does not satisfy the monotonicity and submodularity. To our best knowledge, this is the first effort to study the OM problem by considering dynamic opinion formation process and user preference (positive, neutral and negative opinions) at the same time in signed social networks.
2) We propose an AOMF in signed social networks, which is composed of three phases: i) the selection of candidate seed nodes, ii) the activated opinion formation process and iii) the determination of seed nodes. In particular, in the activated opinion formation process, we combine the activation process with dynamic opinion formation process, which can model the opinion dynamics of each individual effectively.
3) We leverage an effective heuristic rule to select candidate seed nodes. Moreover, based on the multi-stage linear threshold model and the Degroot model, we devise the activated opinion model to calculate the opinion spread of each candidate seed node. Seed nodes are further determined based on the selection of candidate seed nodes and activated opinion formation process.
4) We evaluate the effectiveness of our method on six social network datasets. The results demonstrate that our method achieves superior potential opinions and positive ratio.

The rest of this paper is organized as follows. Section 2 reviews the related work. In Section 3, the proposed OM problem is described, and the activated opinion model is devised. AOMF is proposed in Section 4. Section 5 shows the experimental results. Finally, this paper is concluded in Section 6.

2. Related work

2.1. Influence maximization

Kempe et al. [13] first proposed the IM problem and showed the problem was NP-hard. In addition, the objective function was monotonic and submodular. Based on the above properties, the authors proposed the provable approximation greedy algorithm in Independent Cascade (IC) model and Linear Threshold (LT) model to select the initial seed nodes. After that, many approximation algorithms were proposed to determine the initial seed nodes. For example, Zhu et al. [14] studied the Balanced Influence and Profit (BIP) maximization problem by considering the price in viral marketing and proposed Price Related (PR) framework that contained PR-I and PR-L models for classic independent cascade and linear threshold models. Two unbudgeted greedy pricing strategies (binary pricing (BYC) and PAnoramic Pricing (PAP)) were designed, in which BYC was offered free samples and PAP was offered different discounts. The authors proved that the BIP problem was NP-hard under PR-I and PR-L models. Moreover, the objective function was not monotone but submodular under certain conditions. Wang et al. [15] investigated how to economically select seeds within a given budget to maximize the diffusion process. The authors characterized each user with two distinct factors: the Susceptibility of being Influenced (SI) and Influential Power (IP). A convex price-demand curve-based model was utilized to make the individual adopt a new behavior successfully. Furthermore, a price-performance-ratio inspired heuristic scheme was proposed to select the seed nodes. Tang et al. [16] defined a general profit metric by leveraging the benefits of influence spread and the cost of seed selection. The authors showed that the profit metric was no longer monotone, which was significantly different from the conventional IM problem. New seed selection algorithm for profit maximization with strong approximation guarantees was further proposed. Aral et al. [17] specified a class of empirically motivated influence models and studied their implications for the IM. The optimal seeds under empirical influence models were relatively less well-connected and less central nodes, and they had more cohesive, embedded ties with their contacts. Hence, empirically motivated influence models had the potential to identify more realistic sets of key influencers in a social network. Kuhnle et al. [18] studied the Multiplex IM problem, in which each layer had its propagation model. This problem was a novel version of the conventional IM problem by incorporating the type of propagation on each layer of the multiplex. The authors showed that the propagation on the multiplex satisfied the submodularity when the propagation in each layer was submodular. To solve the Multiplex IM problem, the authors proposed a Knapsack Seeding of Networks (KSN) algorithm, which could be run on each layer of the multiplex in parallel.

Based on the above researches, we can observe that the IM methods have a hidden assumption that each individual has a positive opinion towards the product. In fact, the opinion of each individual towards the product should be positive, neutral or negative. Moreover, these mentioned methods are mainly leveraged in unsigned social networks. In the real scenario, the IM problem is more valuable and challenging in signed networks due to the existence of negative edge weights.

2.2. Opinion maximization

Influenced by user opinions, Chen et al. [9] proposed the extended independent cascade model, called IC-N by incorporating the spread of negative opinions. IC-N introduced a parameter based on the traditional IC model. However, the parameter was the same to each individual, which was not reasonable in the real scenario. Zhang et al. [10] proposed an Opinion-based Cascading (OC) model based on the LT model. Specifically, the authors studied the OM problem, not only considering the individual opinion but also capturing the change of opinions at the same time. However, the opinion of each individual only changed once after they were activated. Abebe et al. [19] studied the maximization of the total sum of opinions both without budget and with budget. For the former, the authors proved that the maximization problem could be solved in polynomial time. For the latter, the objective function was not sub-modular, and the authors proposed a heuristic method with the largest marginal gain for the maximization problem. Chen et al. [20] explored the negative-aware influence maximization (NIM) problem by maximizing positive users as well as minimizing negative ones. In addition, the objective function did not satisfy submodularity and monotonicity, which was different from the IM problem. Liu et al. [21] studied the AcTive Opinion Maximization (ATOM) problem to find some initial seed nodes and maximized the overall opinion propagation in a multi-round campaign. The authors incorporated the historical opinion and rating information into the OM problem. However, the authors assumed that the node opinion was unchanged after being determined. Shen et al. [38] studied the OM in signed social networks and proposed a new diffusion model called LT-S (LT model to signed networks), which was an extension of LT model incorporating both positive and negative opinions. The authors proved that the influence spread function under LT-S model was submodular and proposed an improved restricted greedy algorithm to solve the problem. However, the authors neglected the dynamics of user opinions.

From the above analysis on the OM, we can observe that the existing methods concentrate on unsigned social networks or signed networks with static user opinions, but neglect the dynamic opinion formation process in signed networks. Therefore, in this paper, we propose an AOMF in signed social networks to achieve the positive opinion maximization for product promotion.

3. Model formation

3.1. Problem model

We model the social network as a directed connected graph, denoted by $G = (V, E, W)$, where $V = \{v_1, v_2, \cdots, v_N\}$ and $E = \{e_1, e_2, \cdots, e_M\}$ are the node set and the edge set respectively. Each node represents an individual, and each edge represents the connection relationship between two adjacent individuals. The weight matrix $W = \{w_{ij}, 1 \leqslant i \leqslant N, 1 \leqslant j \leqslant N\}$ ($0 \leqslant w_{ij} \leqslant 1$ denotes the weight between v_i and v_j).

In general, the IM is formulated as a problem on how to select k ($1 \leqslant k \leqslant N$) seed nodes from G and eventually maximize the number of influenced nodes. However, the target of OM is to maximize the opinion propagation of influenced nodes rather than the number of influenced nodes. To describe the OM problem more easily, we give the following five definitions.

Definition 1. (Seed nodes) [39]. Seed nodes are the initial k nodes in set V used to activate or affect other nodes.

Definition 2. (Active nodes) [39]. Active nodes have a chance to activate their neighbor nodes from inactive status to active status.

Definition 3. (Node opinion). For each node $i \in V$, the opinion of node i is defined as o_i, which is classified as three categories: 1) positive opinion, 2) neutral opinion, and 3) negative opinion. The opinion varies between $[-1, 1]$, where the range of positive opinion is $(0, 1]$, the range of neutral opinion is 0 and the range of negative opinion is $[-1, 0)$.

Definition 4. (Potential opinion propagation). For the seed node set S, the potential opinion propagation is defined as the difference between the positive opinions and the negative opinions, i.e., $\Gamma(S) = \sum_{v_i \in C^+} o_i + \sum_{v_j \in C^-} o_j$, where C^+ is the active node set with positive opinions and C^- is the active node set with negative opinions.

Definition 5. (Opinion maximization). Given the seed node set S $(S \subset V)$, we select the seed node set S with the specific opinions, propagate the opinion towards the product to their neighbors by the opinion formation model, and eventually maximize the potential opinion propagation.

Therefore, the OM problem can be expressed mathematically, as follows.

-maximize:

$$\Gamma(S) = \sum_{v_i \in C^+} o_i + \sum_{v_j \in C^-} o_j \tag{1}$$

-subject to:

$$S = \bigcup_{i=1}^{N} \{\gamma_i v_i\} \tag{2}$$

$$\gamma_i = \begin{cases} 1, & \text{if} \qquad v_i \text{ is seed node,} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

$$\sum_{1 \leqslant i \leqslant N} \gamma_i \leqslant k \tag{4}$$

$$o_i \in [-1, 1], \quad \forall i \in I^* \tag{5}$$

Here, $\Gamma(S)$ is called the potential opinions, which is the objective function to maximize positive nodes and minimize negative nodes after the specific time. The opinion value is a continuous value between -1 and +1. In this paper, we mainly focus on the selection of seed nodes. More concretely, we study the product promotion problem by selecting some influential individuals to obtain the maximum potential opinions towards the product.

3.2. Activated opinion model

3.2.1. Activated nodes

Our proposed Multi-stage Linear Threshold (MLT) is a multi-stage diffusion model. Similar to the conventional LT model [22,23], in which each user u_i in a weighted social network G is associated with a threshold $\theta_i \in [0, 1]$. Before the first stage of MLT, all users are inactive towards the product. At stage t of MLT, an inactive user v_i becomes active when $\sum_{v_j \in C^{(t-1)}} w_{ji} > \theta_i$, where $C^{(t-1)}$ is the set of active users before stage t starts and node j is in-degree neighbor of node i. Here, all active users will stay active and cannot go back to inactive status. Next, we will consider the dynamic opinion formation process of activated nodes.

3.2.2. Dynamic opinion

After activating active nodes, we need to model the dynamic opinion process of each node. In this paper, we leverage a variant of the well-known Degroot model [24,25] to produce the opinion dynamics of each node. In this model, each node has an inherent internal opinion s_i which remains unchanged during the process and an expressed overall opinion o_i which is dynamically changing through internal opinion of the node and expressed opinions of its neighbors. To be specific, node i updates its expressed opinion at time step t as follows:

$$o_i^{(t+1)} = \frac{s_i + \sum_{j \in N_i^{in}} I_{ji} w_{ji} o_j^t}{1 + \sum_{j \in N_i^{in}} I_{ji} w_{ji}} \tag{6}$$

$$I_{ji} = \begin{cases} +1, & \text{if node } j \text{ is positive to node } i, \\ -1, & \text{if node } j \text{ is negative to node } i. \end{cases} \tag{7}$$

Here, s_i denotes the inherent belief of an individual based on its background, preference, and education. o_i^t is the expressed opinion at time step t which changes dynamically over the time step. N_i^{in} is the set of nodes j and then (i,j) is a directed edge in which node j points to node i in the graph. I_{ji} is the sign in which node j points to node i. I_{ji} represents the positive or negative relationships between node j and node i (i.e., the trust relationship between node j and node i, such as like/hate, trust/suspicion, and friend/enemy). Hence, I_{ji} can be regarded as the sign of w_{ji}.

Compared with the conventional propagation models [10–12], our model combines the activated nodes with dynamic opinion formation process, which can reflect the social relationship in product promotion more effectively. First of all, different from the IC or LT model [13] in which the opinion of each individual towards the product is always positive, our model contains positive, neutral and negative opinions. In the real scenario, negative and neutral opinions may produce large influence on the decision of the consumer. As illustrated in Fig. 1, we consider a simple situation and assume the weight between any two connected nodes as 1. There are node a and its five in-degree neighbors in a signed social network. We can find that node a adopts positive opinion towards one product and its five in-degree neighbors adopt negative opinions towards the product. Through considering sign of the links and exchange of views with its in-degree neighbors, node a transforms its position into the reverse direction (from positive opinion to negative opinion). Specifically, the opinion of node a is calculated as follows: $o_i = (0.9 + 0.2 + 0.5 - 0.8 - 0.7 - 0.6)/(1 + 3 - 2) = -0.25$ by Eq. 6. Secondly, unlike the traditional opinion model [21] in which the opinion of each individual is fixed once being determined, our model considers the dynamic opinion formation process of each individual over the time step.

4. The generation of seed nodes

To cope with the OM problem, we propose a novel AOMF in signed social networks, which is composed of three phases: i) the selection of candidate seed nodes, ii) the activated opinion formation process and iii) the determination of seed nodes. We first use an effective heuristic rule to select candidate seed nodes. Then, we determine the seed nodes of each stage by using the activated opinion formation process. Finally, we obtain k seed nodes based on the candidate seed nodes and activated opinion formation process. The whole process of AOMF is summarized in **Algorithm 4**. To describe the proposed AOMF intuitively, we give the whole process of opinion maximization. Next, we first introduce the detailed information on selection of candidate seed nodes. After that, we devise the activated opinion formation process to model the dynamic change of each user's opinion. Based on candidate seed nodes and activated opinion formation process, seed nodes are further determined. Eventually, we analyze the related two theorems.

Before obtaining candidate seed nodes, we first analyse the classical greedy scheme summarized in **Algorithm 1**. Typically, greedy algorithm selects each seed node by traversing all remaining nodes and the greedy algorithm will be completed when k seed nodes are determined or there are no more activated nodes. Specifically, when a node is added to the seed node set S, the increased potential opinions are defined as follows.

$$\varepsilon(S) = \Gamma(S \cup \{v\}) - \Gamma(S) \tag{8}$$

where $\varepsilon(\cdot)$ denotes the added potential opinions of influenced nodes.

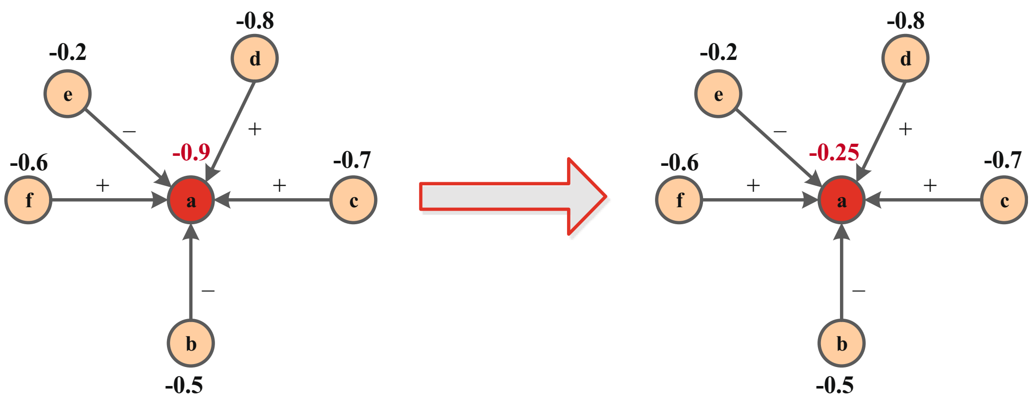


Fig. 1. The example of opinion change based on the activated opinion model.

4.1. Selection of candidate seed nodes

According to the analysis of **Algorithm 1**, we can observe that the greedy algorithm usually obtains superior influence propagation compared with other methods at the sacrifice of efficiency [26], that is, it consumes a large amount of running time. To decrease the running time and ensure superior opinion propagation, in this paper, we propose AOMF to settle the OM problem. Firstly, we select candidate seed nodes by **Algorithm 2**. The pseudo-code of selection of candidate seed nodes is composed of the following three steps: i) the potential influence of each node is calculated (lines 1–3); ii) we rank N network nodes in descending order (line 4); iii) we obtain the top $l * k$ candidate seed nodes (line 5). Through the above heuristic algorithm, we can reduce the number of candidate seed nodes effectively, which has a faster running speed than the conventional greedy algorithms.

Algorithm 1: Classical greedy algorithm

Input: Network node set $G = (V, E, W)$ and the number of seed nodes k;
Output: Seed node set S;

BEGIN
Initialize seed node set $S \leftarrow \varnothing$
for $i = 0 : k$, **do**
  for each $v \in V \setminus S$, **do**
    Calculate the optimal potential opinion $v^* \leftarrow argmax_{v \in V \setminus S}(\Gamma(S \cup v) - \Gamma(S))$;
  end for
  $S \leftarrow S \cup \{v^*\}$
end for
Generate seed node set S;
END

Here, the proposed potential influence is composed of the following three parts: i) the opinion of node v, ii) the in-degree neighbours of node v and iii) the sign and edge weight in which node v points its in-degree neighbour u. Thus, the potential influence of node v, denoted by p_v, is calculated as follows:

$$p_v = o_v + \sum_{u \in N_v^{in}} (o_u + o_v I_{vu} w_{vu}) \tag{9}$$

Algorithm 2: Selection of candidate seed nodes

Input: Network node set $G = (V, E, W)$, the number of seed nodes k and parameter l;
Output: Candidate seed node set;

BEGIN
1: **for** $v = 0 : N$, **do**
2:   $p_v = o_v + \sum_{u \in N_v^{in}} o_u + o_v I_{vu} w_{vu}$;
3: **end for**
4: Rank N network nodes in descending order by comparing their potential influence by Eq. (9);
5: Select the top $l * k$ as candidate seed nodes;
END

---

where o_v and o_u are the initial opinions of node v and node u respectively.

4.2. Activated opinion formation process

After selecting candidate seed nodes, we will determine the seed nodes from these candidate seed nodes. Therefore, an effective estimation scheme should be designed to evaluate the opinion and spread the potential opinion of each candidate seed node. Different from the previous opinion propagation methods in which they mainly concentrate on the activated nodes and the static opinion formation process in unsigned social networks, in this paper, we devise the activated opinion formation process by combining the activation process with dynamic opinion formation process in signed networks at the same time. The pseudo-code of the proposed activated opinion formation process is summarized in **Algorithm 3**, which is composed of three steps: i) we initialize the activated node set (line 1); ii) we judge whether network nodes are activated by the activated opinion model (lines 2–13); iii) we update the opinions of all activated nodes (lines 14–19).

4.3. Determination of seed nodes

Next, we determine the seed nodes by the above proposed selection of candidate seed nodes in **Algorithm 2** and the activated opinion formation process in **Algorithm 3**. The pseudo-code of AOMF is summarized in **Algorithm 4**, which is composed of four steps: i) we initialize the seed node set (line 1); ii) we select $l * k$ candidate seed nodes (line 2); iii) we update the activated opinion propagation of all candidate seed nodes and determine the seed nodes of each stage (lines 3–10); iv) we determine k seed nodes of R stages (lines 11–14).

Algorithm 3: Activated opinion formation process

Input: Network node set $G = (V, E, W)$, the number of time step T and the number of stage R;
Output: The updated opinions of activated nodes;

BEGIN
Initialize activated node set $C^{(0)} \leftarrow \varnothing$;
for $r = 1 : R$, **do**
  for $i = 1 : |V|$, **do**
   if $\sum_{u_j \in C^{(r-1)}} w_{ji} > \theta$, **then**
    Node i is activated;
   else
    Continue;
   end if
   $C^{(r)} \leftarrow C^{(r)} \cup \{i\}$;
   $V \leftarrow V \setminus \{i\}$;
  end for
end for
$c \leftarrow |C|$;
for $i = 1 : c$, **do**
  for $t = 1 : T$, **do**
   Calculate the opinion of node i at time step t by Eq. (6);
  end for
end for
Obtain the updated opinions of activated nodes;
END

The application example of AOMF is illustrated in Fig. 2. As depicted, the signed network consists of 8 nodes and 13 directed edges. Network node represents user and edge represents the relationship between users. In addition, the stage number is set to 2. Our target is to select some influential users (i.e., seed nodes) to spread user's positive opinions towards the product with specific propagation model, and eventually achieve the maximum potential opinions towards the product. Firstly, we initialize the opinions of network nodes. Then, we start to select candidate seed nodes (i.e., nodes 1, 4, 5, 7) and leverage AOMF by two stages to determine the seed nodes (i.e., nodes 4 and 1 respectively). Finally, we leverage the activated opinion model to select the activated node set (nodes 2, 3, 5, 6, 7 and 8), update the initial opinions and obtain new opinions within specific time. The generated opinion set is $V = \{1, 0.1, -1, 1, 0.7, -0.7, 0.6, 0.3\}$ and the potential opinion equals to 2. We subsequently introduce two Theorems in this paper, as follows.

Theorem 1. *The opinion maximization problem in definition 5 is NP-hard.*

Proof. If a problem in a specific instance is NP-hard, the problem is also NP-hard [10]. The traditional IM problem is NP-hard [36,37]. Therefore, to prove that the OM problem is NP-hard, we should show that a specific instance of the OM is equivalent to the conventional IM problem.

We prove the theorem by reducing the conventional IM problem to a specific instance of OM problem. Given the conventional IM problem, we construct an instance of the OM problem. In OC model, we assume that the opinion of each node is +1 and keeps unchanged over time. Moreover, the sign in signed network is +1. We can observe that maximizing the potential opinions of influenced nodes for the OM problem is equivalent to maximizing the number of influenced nodes for the conventional IM problem. Therefore, the conventional IM problem is equivalent to a specific instance of the OM. The OM problem is NP-hard.

Theorem 2. *The objective function in this paper no longer satisfies the monotonic and submodule properties.*

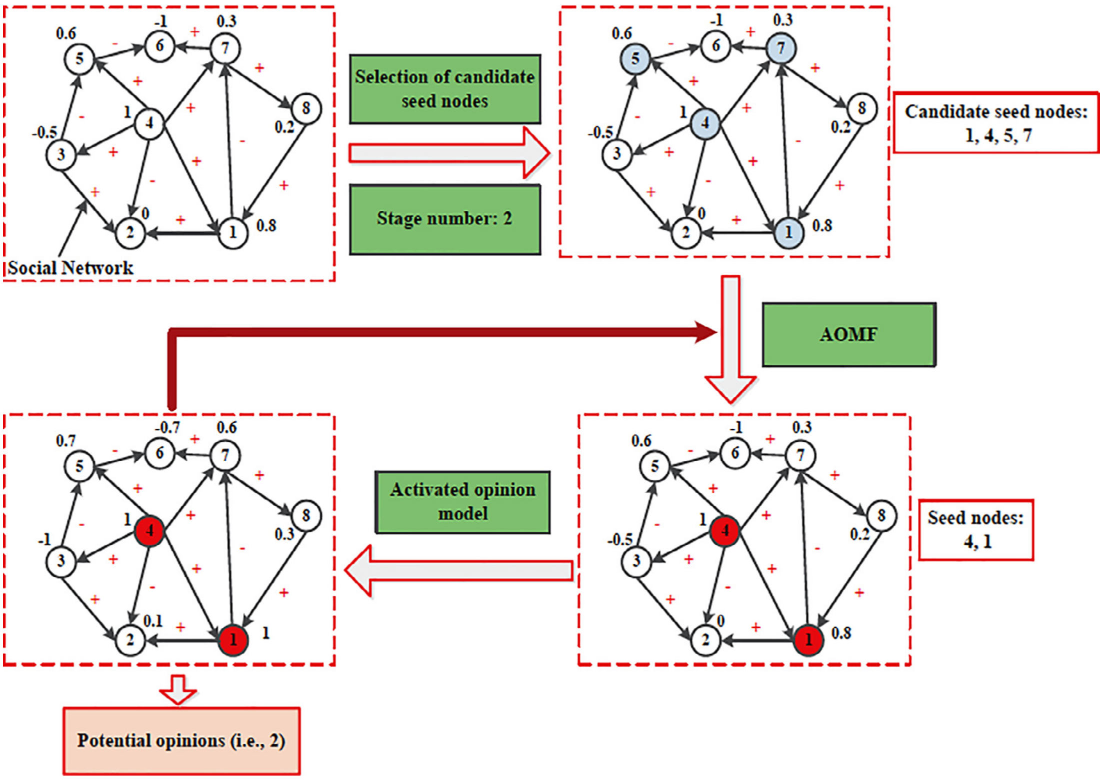


Fig. 2. The application example of AOMF.

Proof. Here, we use reduction to absurdity to prove the Theorem 2. We assume that the objective function satisfies the monotonic and submodule properties. The objective function is defined as the maximization of potential opinion propagation, i.e., $\Gamma(S) = \sum_{v_i \in C^+} o_i + \sum_{v_j \in C^-} o_j$. When we add one node with negative opinions (its opinions is -1 and the edge weight is $+1$), the objective function will be reduced and does not satisfy monotonic properties. In addition, we assume that there are sets A and B such that $A \subset B \subset V$ and a node v such that $v \notin A$ and $v \notin B$. When node v is added to the set A, the added positive opinions equal to negative opinions. When node v is added to the set B, the added potential opinion of node v is $+1$. If the objective function satisfies the submodule properties, we have,

$$\Gamma(A \cup \{v\}) - \Gamma(A) \geqslant \Gamma(B \cup \{v\}) - \Gamma(B) \tag{10}$$

Obviously, we can observe that $\Gamma(A \cup \{v\}) - \Gamma(A) = 0, \Gamma(B \cup \{v\}) - \Gamma(B) = 1$, which is a contradiction.

To understand the non-submodularity of objective function more clearly, we give a specific example. As illustrated in Fig. 3, this is a signed social network with six nodes and four directed edges. The opinion of node 1, node 2, node 3, node 4, node 5 and node 6 is 0.9, 0.9, 0.3, 0.3, 0.3, and 0.3 respectively. Activation probability of node 1 to node 3 equals that

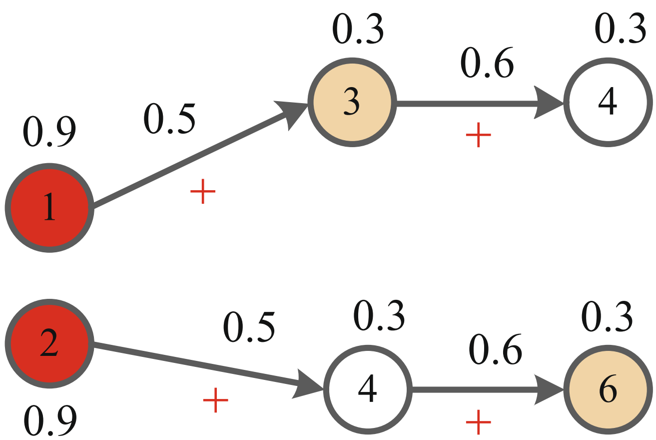


Fig. 3. The example of non-submodularity.

of node 2 to node 4 (both side are 0.5). Activation probability of node 3 to node 5 equals that of node 4 to node 6 (both side are 0.6). Seed node set S includes node 1 and node 2, i.e., S={1, 2}. When we add node 6 into set S, we can find that node 6 cannot activate any nodes. Hence, $\Gamma(S \cup \{6\}) - \Gamma(S)$ =0.3. Then, we add node 3 into new set S ($S = \{1, 2, 6\}$), we can find that node 3 has a chance to activate node 5 and $\Gamma(S \cup \{3\}) - \Gamma(S) > 0.3$. In summary, the example cannot satisfy the inequality in Eq. (10). Thus, the objective function is not submodule.

Algorithm 4: Activated Opinion Maximization Framework (AOMF)

Input: Network node set $G = (V, E, W)$, $l * k$ candidate seed nodes,and the number of stage R;
Output: Seed node set S;

BEGIN
1: Initialize the seed node set $S \leftarrow \varnothing$;
2: Calculate the $l * k$ candidate seed nodes by Eq. (9), denoted as L;
3: **for** $r = 1 : R$, **do**
4:   **for** $t = 1 : k/R$, **do**
5:     **for** each $v \in L \setminus S$, **do**
6:       Calculate the optimal potential opinion $v^* \leftarrow argmax_{v \in L \setminus S}(\Gamma(S \cup v) - \Gamma(S))$;
7:     **end for**
8:     $S^{(r)} \leftarrow S^{(r)} \cup \{v^*\}$
9:     $L \leftarrow L \setminus \{v^*\}$;
10:   **end for**
11:   $S \leftarrow S \cup S^{(r)}$
12:   Update the opinions of network nodes at stage r;
13: **end for**
14: Obtain the seed node set S;
END

5. Experimental results

In this section, we first show the adopted social network datasets and parameter settings. Then, we introduce the baseline algorithms. At last, we analyse the experimental results.

5.1. Data setup

Table 1 demonstrates the basic information of six social networks (i.e., Bitcoin Alpha [27] and Bitcoin OTC [28], Wikipedia elections [29], Slashdot [30], Epinions trust [31] and WikiSigned [32]). Those network datasets are signed, which can be downloaded from Standard Large Network Dataset Collection (http://snap.stanford.edu/) and Koblenz Network Collection (http://konect.uni-koblenz.de/networks/). Moreover, similar to reference [38], the weight of every incoming edge of node i is defined as $1/d_i$, where d_i is the number of in-degree neighbors of node i. It should be noted that the edge weight on Bitcoin Alpha [27] and Bitcoin OTC [28] ranges in a scale of -10 (total distrust) to $+10$ (total trust) in steps of 1. Since our edge weight varies between 0 and 1, in the simulation, we use the sign (+1 or -1) and the edge weight on Bitcoin Alpha and Bitcoin OTC is calculated by the above mentioned method in reference [38].

For the generation of initial user opinions, we leverage the random method in which the opinion of each user is randomly generated between [-1, 1]. Since the user opinions are randomly generated, we obtain the simulation results based on 20 experiments. In addition, the parameter is set as follows. l is the proportion parameter, which determines the number of candidate seed nodes. In general, $l > 1$. Since the number of network nodes is greater than 1000, the increase of l will increase the computational complexity of our method dramatically. In simulation process, l is set to 1.2. Furthermore, we adopt the

Table 1
Statistics of six social networks.

Network	Nodes	Links	+Links	−Links	Average degree
Bitcoin Alpha	3783	24,186	93%	7%	12.787
Bitcoin OTC	5881	35,592	89%	11%	12.104
Wikipedia elections	7118	103,675	78.4%	21.6%	29.130
Slashdot	77,350	516,575	76.7%	23.3%	13.028
Epinions trust	131,828	841,372	85.3%	14.7%	12.765
WikiSigned	138,592	740,397	87.9%	12.1%	10.685

multi-stage heuristic scheme to select the seed nodes. Therefore, the stage number $R > 1$. In the related researches on multi-stage algorithms, similar to the selection of seed nodes, R is adjusted within a certain range. We mainly consider the change of potential opinions over the number of seed nodes and the time steps. Thus, to obtain the stage number of seed nodes in **Algorithm 3**, R is set to a fixed value (i.e., 5). The selected number of seed nodes for each stage is distributed equally (i.e., $\frac{k}{R}$).

5.2. Baseline algorithms

We evaluate the performance of our AOMF and the three baseline algorithms (Random [33], Degree [34], and RISNIM [20]). The three baselines are classical and currently prevalent optimization approaches, which have been widely utilized in dealing with the OM problem. Therefore, we select them as the baselines compared with our proposed AOMF. To be specific, the basic information of the three baseline algorithms is described as follows:

Random [33]: k seed nodes are randomly selected from network nodes.

Degree [34]: Degree is the number of one-hop neighbour nodes. The nodes with high degree mean large influence power.

RISNIM [20]: It selects the optimal seed nodes with Reverse Influence Set [35] based algorithm for Negative-aware Influence Maximization (RISNIM).

Moreover, to better show the performance of our proposed method, we select a prevailing greedy approach (denoted as Greedy-L) [21] as the baseline for the experimental part. We also define the positive ratio to measure the proportion of the number of positive nodes in network nodes. The mathematical form is defined as follows:

$$Pr = \frac{|V^+|}{N} \tag{11}$$

where V^+ is the node set with positive opinions.

5.3. Result analysis

We evaluate the performance of our proposed method compared with the three baseline algorithms in terms of potential opinions and positive ratio. Next, we will show the potential opinions over the number of seed nodes and the time steps. In addition, we illustrate the positive ratio on six network datasets. We also compare AOMF with a prevailing greedy algorithm.

5.3.1. Potential opinions over the number of seed nodes

Fig. 4 demonstrates the potential opinions over the number of seed nodes. The time step is set to 20 and the number of seed nodes ranges from 0 to 200. We observe that the potential opinions of four algorithms constantly increase over the number of seed nodes. In particular, AOMF obtains the largest potential opinions on six network datasets. Specifically, Random acquires the smallest potential opinions and RISNIM obtains superior results than Random and Degree on all network datasets. Owing to the emergence of negative opinions, the objective function does not satisfy the monotonicity. Thus, Random may have a decreasing tendency over the increase of seed nodes. In addition, on WikiSigned, the results of AOMF and RISNIM are approaching. From Fig. 4, we can find that the number of seed nodes can decide the size of potential opinions, and our proposed AOMF can effectively select influential seed nodes and acquire more superior potential opinions compared with the other three baseline algorithms.

5.3.2. Potential opinions over the time steps

Fig. 5 demonstrates the potential opinions over the time steps. The number of seed nodes is 100 and time step ranges from 0 to 50. We observe that potential opinions of four algorithms constantly increase over the number of seed nodes. In particular, AOMF obtains the largest potential opinions on almost each time step on the six network datasets. Specifically, the results of RISNIM are superior to those of the other three baseline algorithms on six network datasets. Similarly, Random obtains worse results than the other methods due to its randomness and irregularity of seed node selection. From Fig. 5, we can find that AOMF can select more superior seed nodes and produce larger potential opinions.

To evaluate the efficiency of AOMF, we select the average results of six networks over the number of seed nodes and time steps in Tables 2 and 3 respectively. For example, in Table 2, we calculate the average of potential opinions for each algorithm over the number of seed nodes between [0, 200] based on 20 experiments. From the results, it can be observed that our proposed AOMF outperforms the baseline algorithms (especially Random and Degree). Moreover, we adopt a Student's t test to assess the performance of our approach and the baselines. We first make the following assumptions.

H0: the performance of our approach and the baselines are similar.

H1: the performance of our approach and the baseline algorithms is significantly different.

The threshold is set to 0.05. To indicate whether the difference between our adopted threshold and the p-values is significant, we add an asterisk (*) to the right of the significant values in Tables 2 and 3. We can find that the p-values of Degree and RISNIM are larger than the threshold except for Random. We can conclude that our approach is significantly better than Random, and our approach is not significantly better than Degree and RISNIM.

We summarize the reasons, as follows. First of all, AOMF uses multi-stage heuristic scheme to determine seed nodes. In addition, AOMF considers the negative opinions and the signed networks. It is obvious that the OM problem in this paper is

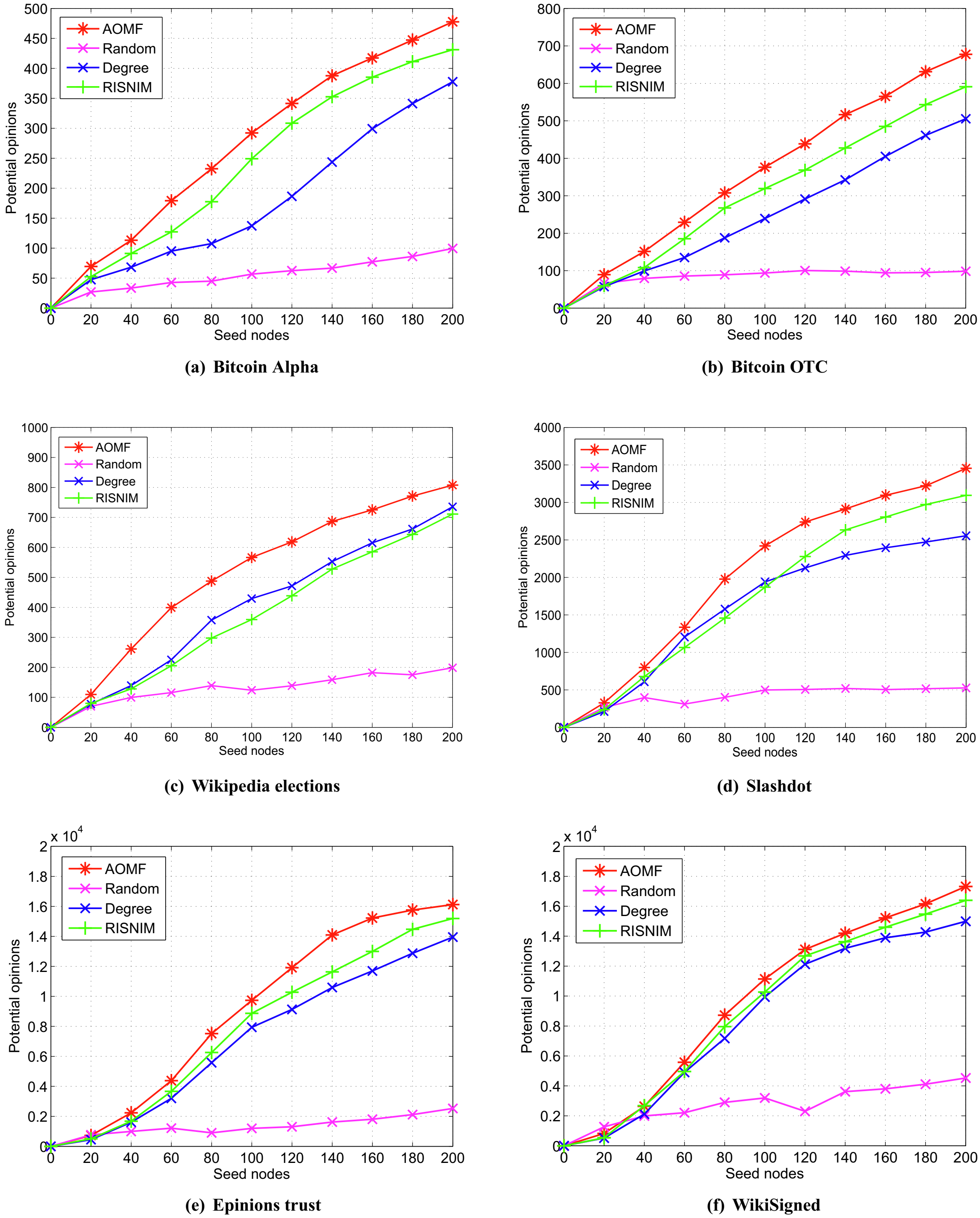
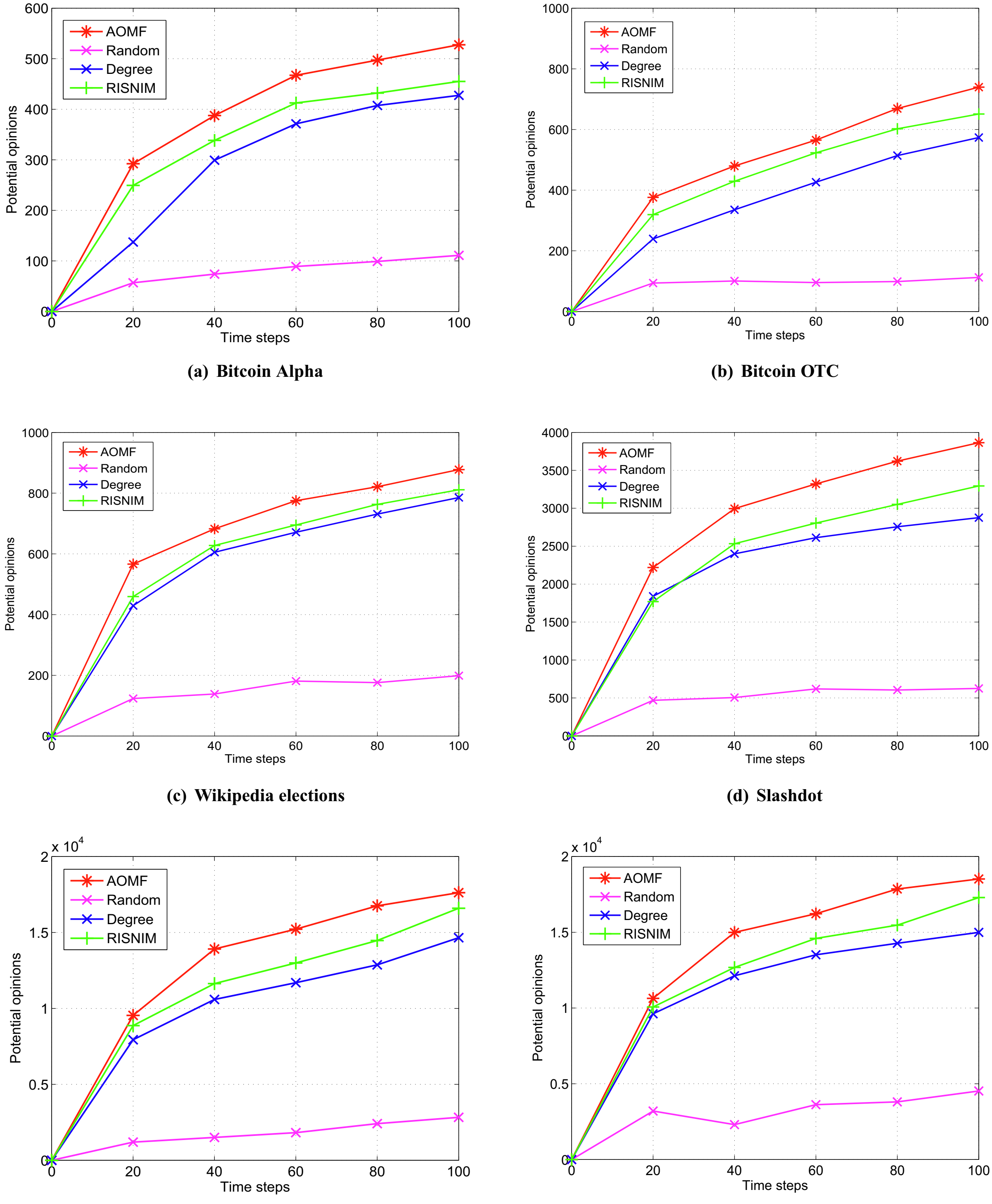


Fig. 4. Potential opinions over the number of seed nodes on six network datasets.

profoundly challenging. Conversely, for the baseline algorithms, Random determines the seed nodes by randomly selecting any network nodes. Degree leverages the number of one-hop neighbour nodes. RISNIM selects the optimal seed nodes with reverse influence set. The three baseline algorithms do not explore the negative opinions in signed networks, which decrease their efficiency of seed selection. The three baseline algorithms adopt single-stage rule to select seed nodes. Therefore, AOMF



(a) **Bitcoin Alpha**

(b) **Bitcoin OTC**

(c) **Wikipedia elections**

(d) **Slashdot**

(e) **Epinions trust**

(f) **WikiSigned**

Fig. 5. Potential opinions over the time steps on six network datasets.

produces more stable and superior potential opinions than the three baseline algorithms. Moreover, from Tables 2 and 3, we can get that AOMF is not significantly better than Degree and RISNIM. We can get that AOMF obtains about 78.0% larger average potential opinions than Random, 23.9% than Degree, 14.6% than RISNIM. In this case, this result shows the effectiveness of AOMF.

Table 2
Average of potential opinions on six networks over the number of seed nodes. (asterisk (*) represents that the difference between our adopted threshold and *p*-value is significant)

Network	Random	Degree	RISNIM	AOMF
Bitcoin Alpha	54.3 (*)	173.1	235.1	268.9
Bitcoin OTC	82.0 (*)	247.7	305.1	362.1
Wikipedia elections	127.4 (*)	387.7	361.4	493.9
Slashdot	403.8 (*)	1580.2	1735.4	2025.5
Epinions trust	1312.9 (*)	7002.0	7777.2	8882.1
WikiSigned	2722.0 (*)	8467.5	9016.3	9536.6

Table 3
Average of potential opinions on six networks over the time steps. (asterisk (*) represents that the difference between our adopted threshold and *p*-value is significant)

Network	Random	Degree	RISNIM	AOMF
Bitcoin Alpha	71.6 (*)	273.8	314.6	362.0
Bitcoin OTC	83.4 (*)	348.0	420.9	471.7
Wikipedia elections	136.4 (*)	537.1	559.4	620.5
Slashdot	470.5 (*)	2080.5	2242.1	2670.1
Epinions trust	1627.3 (*)	9625.5	10760	12175
WikiSigned	2908.8 (*)	10754	11685	13038

5.3.3. Positive ratio

Moreover, to better reflect the effectiveness of AOMF, we adopt the positive ratio to measure the performance of four algorithms. Here, the time step is set to 50 and the number of seed nodes is set to 200. To describe the six networks more simply in Fig. 6, Bitcoin Alpha, Bitcoin OTC, Wikipedia elections, Slashdot, Epinions trust and WikiSigned are denoted as BA, BO, WE, SL, ET and WS respectively. The initialized positive ratios are the proportion of the number of positive nodes in network nodes in initial time. From the results in Fig. 6, we can observe that AOMF and the three baseline algorithms (i.e., Random, Degree and RISNIM) obtain larger positive ratio than the initialized ratio, and AOMF has larger positive ratio than the three baseline algorithms especially. For example, on BA, AOMF obtains about 32.8% larger average positive ratio than that of the initialized positive ratio, 27.1% than Random, 18.4% than Degree, 7.2% than RISNIM. On SL, AOMF obtains about 17.2% larger average positive ratio than that of the initialized positive ratio, 15.5% than Random, 12.5% than Degree, 6.3% than RISNIM.

5.3.4. Comparison with greedy algorithm

To further evaluate the efficiency of AOMF, we compare it with a prevailing greedy algorithm (represented as Greedy-L) [21]. Fig. 7(a) and (b) demonstrate the potential opinions of AOMF and Greedy-L on Bitcoin Alpha (BA) and Wikipedia elections (WE). In Fig. 7, the *p*-value equals to 0.85 and 0.95 on BA and WE respectively. In Fig. 8, the *p*-value equals to 0.045. We can conclude that the potential opinions of Greedy-L are not significantly better than those of our approach (CAOM), and the

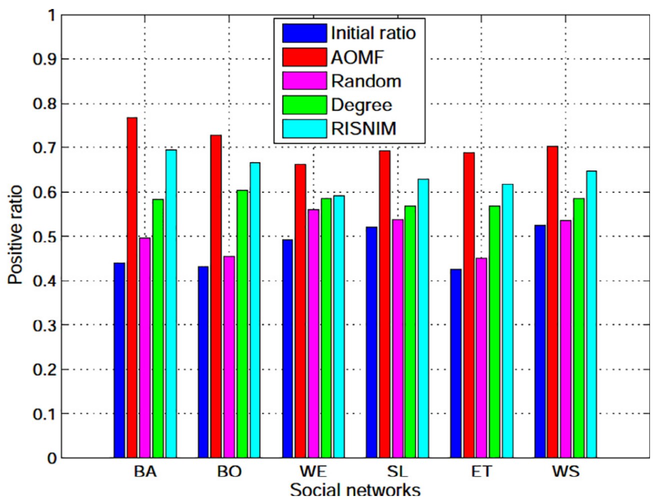


Fig. 6. Positive ratio on six network datasets.

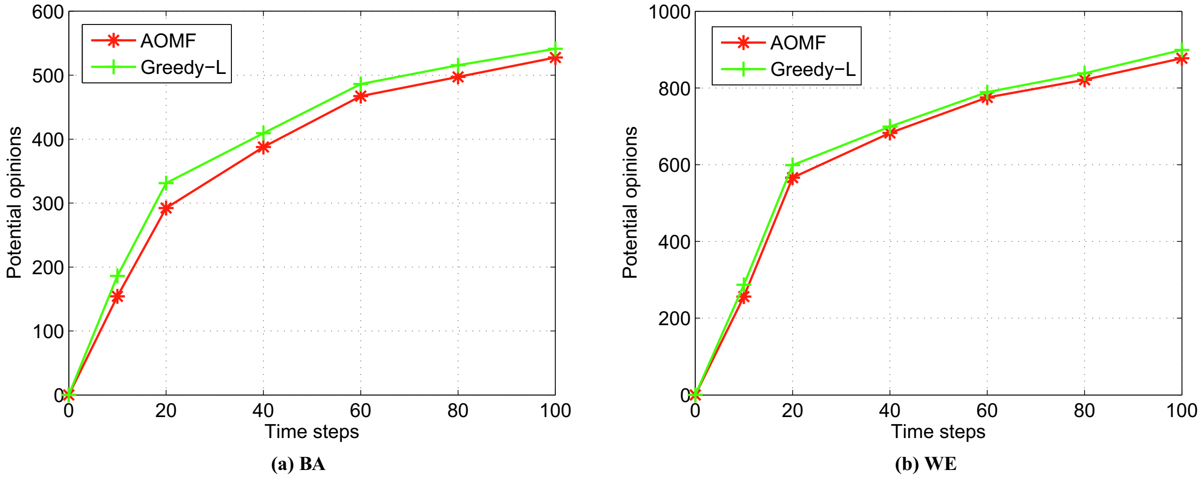


Fig. 7. Potential opinions between AOMF and Greedy-L on BA and WE.

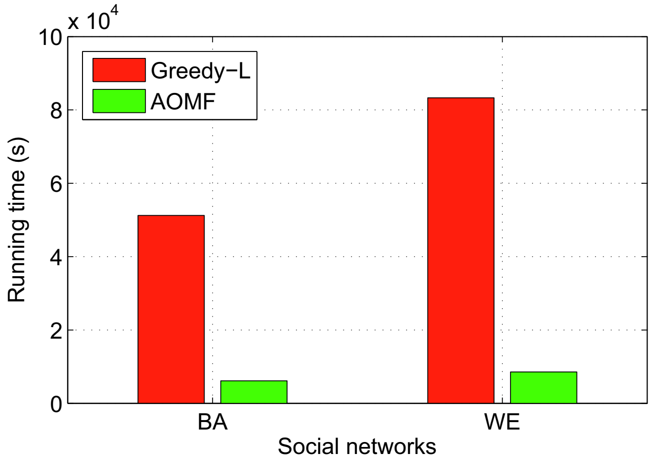


Fig. 8. Comparisons on running time between AOMF and Greedy-L.

running time of CAOM is significantly smaller than that of Greedy-L. The results in Fig. 7 show that Greedy-L performs slightly better than CAOM. However, in Fig. 8, the running time of Greedy-L is considerably larger than that of CAOM. This is because Greedy-L selects each seed node by traversing all remaining nodes, which improves the potential opinions but greatly increases the computational complexity. In contrast, AOMF adopts the multi-stage heuristic scheme to dynamically determine the seed nodes, which decreases the computational complexity as well as obtains competitive potential opinions.

6. Conclusions

In this paper, we study a novel OM problem in signed social networks for product promotion. We show the NP-hardness of OM problem, the non-monotonicity and non-submodularity of objective function. To cope with the challenges of this problem, we propose an approach called AOMF. We first use an effective heuristic rule to select candidate seed nodes. Then, we calculate the opinion propagation of each candidate seed node by using the activated opinion formation model. Different from the previous opinion propagation methods that mainly concentrate on the activated nodes and the static opinion formation process in unsigned social networks, the proposed activated opinion formation model combines the activation process with the dynamic opinion formation process in signed social networks to estimate the potential opinion of each individual. Based on the selection of candidate seed nodes and the activated opinion formation process, seed nodes are further determined. Experimental results on six social network datasets demonstrate that our method outperforms the baseline algorithms (i.e., Random, Degree, and RISNIM) on potential opinions and positive ratio. Our proposed method also obtains competitive potential opinions as well as lower computational complexity compared with the prevailing greedy algorithm. For future works, we will study the OM problem in signed social networks under a competitive scenario.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant No. 61872073; the Major International(Regional) Joint Research Project of NSFC under Grant No. 71620107003; LiaoNing Revitalization Talents Program under Grant No. XLYC1902010.

References

[1] E. Bakshy, I. Rosenn, C. Marlow, L. Adamic, The role of social networks in information diffusion, International Conference on World Wide Web, ACM, 2012, pp. 519–528.
[2] A. Guille, H. Hacid, C. Favre, D.A. Zighed, Information diffusion in online social networks: a survey, SIGMOD Rec. 42 (2) (2013) 17–28.
[3] B. Wilder, L. Onasch-Vera, J. Hudson, J. Luna, N. Wilson, R. Petering, E. Rice, End-to-end influence maximization in the field, International Conference on Autonomous Agents and MultiAgent Systems (2018) 1414–1422.
[4] Y. Tang, Y. Shi, X. Xiao, Influence maximization in near-linear time: a martingale approach, ACM SIGMOD International Conference on Management of Data, ACM, 2015, pp. 1539–1554.
[5] T. CAI, J. Li, A.S. Mian, R. li, T. Sellis, J.X. Yu, Target-aware holistic influence maximization in spatial social networks, IEEE Trans. Knowl. Data Eng. 2020. [In Press]..
[6] M. Jalili, Social power and opinion formation in complex networks, Phys. A 392 (2013) 959–966.
[7] Q. He, X. Wang, Z. Lei, M. Huang, Y. Cai, TIFIM: a two-stage iterative framework for influence maximization in social networks, Appl. Math. Comput. 354 (2019) 338–352.
[8] O. He, X. Wang, B. Yi, et al, Opinion Maximization Through Unknown Influence Power in Social Networks Under Weighted Voter Model, IEEE SYSTEMS JOURNAL 14 (2) (2020) 1874–1885.
[9] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, Y. Yuan, Influence maximization in social networks when negative opinions may emerge and propagate, SIAM International Conference on Data Mining (SDM) (2011) 379–390.
[10] H. Zhang, N.T. Dinh, M.T. Thai, Maximizing the spread of positive influence in online social networks, IEEE International Conference on Distributed Computing Systems (ICDCS) (2013) 317–326.
[11] A. Gionis, E. Terzi, P. Tsaparas, Opinion maximization in social networks, SIAM International Conference on Data Mining (SDM) (2013) 387–395.
[12] Q. He, X. Wang, M. Huang, J. Lv, L. Ma, Heuristics-based influence maximization for opinion formation in social net-works, Appl. Soft Comput. 66 (2018) 360–369.
[13] D. Kempe, J. Kleinberg, é. Tardos, Maximizing the spread of influence through a social network, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003, pp. 137–146.
[14] Y. Zhu, D. Li, R. Yan, W. Wu, Y. Bi, Maximizing the influence and profit in social networks, IEEE Trans. Comput. Social Syst. 4 (3) (2017) 54–64.
[15] Y. Wang, A.V. Vasilakos, Q. Jin, J. Ma, PPRank: economically selecting initial users for influence maximization in social networks, IEEE Syst. J. 11 (4) (2015) 2279–2290.
[16] J. Tang, X. Tang, J. Yuan, Profit maximization for viral marketing in online social networks: algorithms and analysis, IEEE Trans. Knowl. Data Eng. 30 (6) (2017) 1095–1108.
[17] S. Aral, P.S. Dhillon, Social influence maximization under empirical influence models, Nat. Human Behav. 2 (6) (2018) 375–382.
[18] A. Kuhnle, M.A. Alim, X. Li, H. Zhang, M.T. Thai, Multiplex influence maximization in online social networks with heterogeneous diffusion models, IEEE Trans. Comput. Social Syst. 5 (2) (2018) 418–429.
[19] R. Abebe, J. Kleinberg, D. Parkes, C.E. Tsourakakis, Opinion dynamics with varying susceptibility to persuasion, Social Inf. Networks (2018) 1089–1098..
[20] Y. Chen, H. Li, Q. Qu, Negative-aware influence maximization on social networks, AAAI Conference on Artificial Intelligence (2018) 1–2.
[21] X. Liu, X. Kong, P.S. Yu, Active opinion maximization in social networks, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2018) 1840–1849.
[22] F. Gursoy, D. Gunnec, Influence maximization in social networks under deterministic linear threshold model, Knowl. Based Syst. 161 (2018) 111–123.
[23] X. He, G. Song, W. Chen, Q. Jiang, Influence blocking maximization in social networks under the competitive linear threshold model, SIAM International Conference on Data Mining, SIAM, 2012, pp. 463-474.
[24] R.L. Berger, A necessary and sufficient condition for reaching a consensus using DeGroot's method, J. Am. Stat. Assoc. 76 (374) (1981) 415–418.
[25] A. Das, S. Gollapudi, K. Munagala, Modeling opinion dynamics in social networks, ACM International Conference on Web Search and Data Mining (2014) 403–412.
[26] Y. Wang, G. Cong, G. Song, K. Xie, Community-based greedy algorithm for mining top-k influential nodes in mobile social networks, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2010, pp. 1039–1048.
[27] S. Kumar, F. Spezzano, V.S. Subrahmanian, C. Faloutsos, Edge weight prediction in weighted signed networks, IEEE 16th International Conference on Data Mining (ICDM) (2016) 221–230.
[28] S. Kumar, B. Hooi, D. Makhija, M. Kumar, C. Faloutsos, V.S. Subrahmanian, Rev2: fraudulent user prediction in rating platforms, in: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018, pp. 333–341.
[29] J. Leskovec, D. Huttenlocher, J. Kleinberg, Governance in social media: a case study of the Wikipedia promotion process, International AAAI Conference on Weblogs and Social Media, 2010..
[30] J. Kunegis, A. Lommatzsch, C. Bauckhage, The Slashdot Zoo: mining a social network with negative edges, International Conference on World Wide Web (2009) 741–750.
[31] P. Massa, P. Avesani, Controversial users demand local trust metrics: an experimental study on epinions.com community, International Conference on American Association for Artificial Intelligence, 2005, pp. 121–126..
[32] S. Maniu, T. Abdessalem, B. Cautis, Casting a web of trust over Wikipedia: an interaction-based approach, International Conference on World Wide Web Posters (2011) 87–88.
[33] P. Natenzon, Random choice and learning, J. Polit. Econ. 127 (1) (2019) 419–457.
[34] T. Opsahl, F. Agneessens, J. Skvoretz, Node centrality in weighted networks: generalizing degree and shortest paths, Social Networks 32 (3) (2010) 245–251.
[35] Y. Tang, Y. Shi, X. Xiao, Influence maximization in near-linear time: a martingale approach, ACM SIGMOD International Conference on Management of Data, ACM, 2015, pp. 1539-1554.
[36] H.T. Nguyen, M.T. Thai, T.N. Dinh, A billion-scale approximation algorithm for maximizing benefit in viral marketing, IEEE/ACM Transactions on Networking (TON) 25 (4) (2017) 2419–2429.

[37] J. Li, T. Cai, K. Deng, X. Wang, T. Sellis, F. Xia, Community-diversified influence maximization in social networks, Inf. Syst. 101522 (2020), In Press.
[38] C. Shen, R. Nishide, I. Piumarta, H. Takada, W. Liang, Influence maximization in signed social networks, International Conference on Web Information Systems Engineering (2015) 399–414.
[39] W. Liang, C. Shen, X. Li, R. Nishide, I. Piumarta, H. Takada, Influence maximization in signed social networks with opinion formation, IEEE Access 7 (2019) 68837–68852.
[40] L. Ma, R. Wang, S. Chen, X. Wang, S. Cheng, Y. Shi, A novel many-objective evolutionary algorithm based on transfer learning with Kriging model, Inf. Sci. 9 (2020) 437–456.
[41] Q. He, X. Wang, F. Mao, J. Lv, Y. Cai, M. Huang, Q. Xu, CAOM: a community-based approach to tackle opinion maximization for social networks, Inf. Sci. 513 (2020) 252–269.