Received April 8, 2019, accepted May 13, 2019, date of publication May 24, 2019, date of current version June 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2918810

Influence Maximization in Signed Social Networks With Opinion Formation

WENXIN LIANG[1], (Member, IEEE), CHENGGUANG SHEN[2], XIAO LI[3], RYO NISHIDE[4], (Member, IEEE), IAN PIUMARTA[5], (Member, IEEE), AND HIDEYUKI TAKADA[6]

[1]School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2]Meituan -Dianping, Beijing 100102, China
[3]School of Software, Dalian University of Technology, Dalian 116620, China
[4]Center for Data Science Education and Research, Shiga University, Hikone 522-8522, Japan
[5]Department of Mechanical and Electrical Systems Engineering, Kyoto University of Advanced Science, Kyoto 615-8577, Japan
[6]Department of Computer Science, Ritsumeikan University, Kyoto 525-8577, Japan

Corresponding author: Wenxin Liang (wxliang@cqupt.edu.cn)

This work was supported in part by the National Science Foundation of China under Grant 61632019.

ABSTRACT Influence maximization (IM) has been widely studied in recent years. Given fixed number of seed users and certain diffusion models, the IM problem aims to select proper seed users in a social networks such that they can achieve the maximal spread of influence. Most previous work assumes that there are only positive relationships between users, and thus users spread influence positively. However, negative relationships also universally exist in various social networks and are complementary to positive relationships in information diffusion. In this paper, the influence maximization problem is addressed in signed social networks that contain both positive and negative relationships. We propose a novel diffusion model called LT-S and two influence spread functions. The proposed LT-S model extends the classical linear threshold model with opinion formation that incorporates both positive and negative opinions and simulates information diffusion in real-world social networks. The influence spread functions under the LT-S model are neither monotone nor submodular which bring challenges to maximization. The RLP algorithm is proposed to tackle the issue, which is improved from R-Greedy algorithm by incorporating two proposed accelerating techniques, the live-edge based and propagation-path based techniques. The results of the extensive experiments on public real signed social network datasets demonstrate that our algorithm outperforms the baseline algorithms in terms of both efficiency and effectiveness.

INDEX TERMS Social network services, modeling, greedy algorithms, influence maximization, opinion formation.

I. INTRODUCTION

The popularity of social networking sites has rapidly increased over the past few years. Social networks such as Facebook, Twitter and Epinions provide many kinds of services and benefits to their users like helping them to connect with new people, share opinions with like-minded people and stay in touch with old friends and colleagues. The interactions among individuals make these social networks become important form for spreading information, ideas, opinions and influence, which provides great opportunities and challenges for large-scale viral marketing campaigns.

The associate editor coordinating the review of this manuscript and approving it for publication was Burak Kantarci.

Influence Maximization (IM) is one of the most interesting and important problems in viral marketing and has attracted much research interest in recent years. The IM problem expects to achieve commercial return as large as possible with the least cost of marketing under the word-of-mouth diffusion. In detail, IM is the problem of choosing a small set of influential users in a social network, initially targeted as seed users, so as to maximize their spread of influence after the diffusion of opinions in the network. Consider a company wants to promote a new product with limit budgets. The company can select a small set seed users by offering the discount or sample to make these users accept the product. These seed users then recommend the product to their friends and their friends continue to promote the product. The company wants to find the best seed users who make the final acceptance of the new product maximized with the given budget.

Dominigos and Richardson [1] were the first to study influence maximization in probabilistic settings. Kempe *et al.* [2] formulated the influence maximization as a directed optimization problem and proposed two basic influence diffusion models, independent cascade (IC) and linear threshold (LT). Both of the two models consider the social network as a directed graph, in which users are represented as vertices and directed edges reflect the social relationships between users. A user can be either *active* (an adopter of the product) or *inactive*. The initial seed users are active, and once a user is activated, he will stay active. In the IC model, active users will try only once to activate their inactive neighbors according to given probabilities. In the LT model, each user has a threshold and will become active when the sum of incoming influence from his active neighbors exceeds that threshold. They propose a greedy algorithm to solve the IM problem based on the monotonicity and submodularity of the influence spread function under both IC and LT models.

Much effort [3]–[6] has been devoted to solving the IM problem and additional influence diffusion models [7]–[10] have been developed. However, previous work only considers positive relationships, ignoring the fact that negative relationships also exist in real social networks and have effects on influence diffusion. For example, Slashdot[1] (a news discussion website) allows users to annotate others to be either ''friends'' or ''foes''. Social networks having a sign, positive or negative, associated with edges are called *signed social networks* [11], [12]. In signed networks, the sign of each edge characterizes whether the corresponding individuals are ''friends'' (positive links) or ''enemies'' (negative links). In previous studies, two nodes sharing a positive link are supposed to hold the same opinions [2]–[4], [10], as people mostly trust their friends and embrace the same opinions. In contrast, sharing a negative link means two nodes hold the opposite opinions [11]–[14].

Most previously proposed models for the IM problems assume that active users can only generate and uniformly express positive opinions to their neighbors. However, in reality, users are possible to hold negative opinions, and express their opinion nonuniformly. Chen et al. [8] proposed a new model called IC-N by extending the IC model with negative opinions. In IC-N, a parameter q (the same for all users) is incorporated into the spread of negative opinions, which is too primitive and can not always represent real-world behaviors. Besides, there are more problems need to be addressed in information propagation. For example, a girl bought a handbag and she likes it, then she recommends it to her best friends. She also did not want her foes to buy handbags of the same style, so she said bad words about the handbag to her foes. Some of her foes may trust her expressed opinion, then this relationship causes negative influence to the handbag, but some other foes may distrust her expressed opinion, although this is a negative relationship, it still causes positive influence to the handbag. People sometimes do not express their true opinions to enemies, and reversely, people also do not always trust the information expressed by their enemies. These phenomenons are very common in social networks.

Much work has also been developed in the sociology and economics literatures on modeling opinion dynamics in social networks. Some models [15]–[17] incorporate both innate opinions and expressed opinions. Innate opinions are fixed for a user representing user's personal preference and history. Expressed opinions, on the other hand, are the opinions that users finally spread after being influenced by both their own opinions and their neighbors' opinions. From the aspects of computer science, however, no related work has been done to combine opinion formation and its propagation into a signed social network.

In this article, we propose a novel influence diffusion model by extending the classic **L**inear **T**hreshold model to **S**igned networks, which we call the LT-S diffusion model. The status of each user in LT-S model can be either in *active* or *inactive*. A user will become active if the influence from his neighbors, both ''friends'' and ''enemies'', achieves his threshold. If a user is active, he will spread his opinion to his neighbors by expressing the *expressed opinion*. In addition, users have *innate opinion* which is fixed and originally generated according to users' own preference and history. The expressed opinion of an active user is the result of the linear combination of his own innate opinion and all the expressed opinions spread from his active neighbors. In LT-S model, a conformity α_u of user u is associated with each user, which reflects the importance of the user's innate opinion compared to the expressed opinions coming from active neighbors. We adopt the concept of *conformity* [16], [18] as a weighted parameter in the above convex combination. A high conformity indicates the user has an expressed opinion that is highly dominated by the expressed opinions of his neighbors, while a low conformity indicates that the user is more insistent to his own innate opinion. In the process of spreading opinions, the expresser has *consistency* to his opinion, where *consistency* will be 1 if he is expressing his true opinion and will be -1 if he is expressing opposite opinion. The accepter has *trust* to the received opinion, where *trust* will be 1 for trusting the received opinion, or -1 for distrusting the received opinion. The difference of LT-S compared to the models proposed in [14] and [19] is obvious. Both Li *et al.* [14] and this article study influence diffusion in signed networks. Li *et al.* extend the classic voter model to signed networks, whereas LT-S is based on the linear threshold model. In addition Zhang *et al.* [19] propose an opinion-based cascading model, which takes individual opinions in account, but ignore the negative relationships between users.

Based on the LT-S model, we formulate the Influence Maximization (IM) problem in signed social networks and prove the problem NP-hard. The traditional IM problem simply

[1] http://slashdot.org/

focuses on maximizing the total number of active users, in this article we propose two different influence spread functions under the LT-S model which have two different focuses. The first is to maximize only positive opinions of all activated users, and the other is to maximize the overall opinions of all activated users. We prove that both the two influence spread functions are neither monotone nor submodular, which causes the classical greedy algorithm [2] not applicable for the maximization. Therefore the R-Greedy [9] algorithm is adopted in this article to solve the IM problem. In this article, we propose the **R**-Greedy with **L**ive-edge and **P**ropagation-path (RLP), which is improved from classic R-Greedy algorithm with the two accelerating techniques. Extensive experiments conducted on real signed social network datasets demonstrate that the RLP algorithm can achieve competitive high influence spread with much less execution time comparing with the baseline algorithms.

The main contributions of this article are summarized as follows.

- We formulate the influence maximization problem in signed social networks and propose a new model called LT-S by extending the classical linear threshold (LT) model. In the LT-S model, we propose two influence spread functions, positive influence spread function and all influence spread function, which are more effective for simulating the information diffusion in real-world social networks.
- We prove the influence maximization problem in signed social networks under the proposed LT-S model is NP-hard and the two proposed influence spread functions under LT-S model are non-monotone and non-submodular.
- We utilize the R-Greedy algorithm to solve the influence maximization problem in signed social networks. However, the R-Greedy algorithm using Monte-Carlo simulation is very time-consuming. To improve the efficiency of the original algorithm, we propose a new algorithm called RLP (**R**-Greedy with **L**ive-edge and **P**ropagation-path), which combines two accelerating techniques, *live edges* and *propagation paths*, with the original R-Greedy algorithm.
- We evaluate the performance of RLP algorithm on public real world signed social networks, and conduct extensive experiments on real signed social network datasets. The experimental results demonstrate that our algorithm outperforms the baseline algorithms in terms of efficiency and effectiveness.

The remainder of this article is organized as follows. In Section II we present the LT-S model and the formal problem definition. Section III introduces the R-Greedy algorithm and our optimized form, RLP. In Section IV we describe our experiments and analyze the experimental results. Related work is reviewed in Section V. Finally, Section VI offers concluding remarks.

The preliminary version of this article was presented at "the 16th International Conference of Web Information Systems Engineering (WISE 2015)" [20]. We extend the following contents in this article. In Section I, we add supplementary backgrounds and explanation of the information maximization problem in signed social networks. In Subsection II-A, we deeply describe why LT-S model can handle complex interaction of negative relationship in real social networks. In Subsection II-B, considering real application of influence maximization, we propose another influence spread function, called "all influence spread function", which examines the summation of all opinions of active users. And the proof of the property of non-monotone and non-submodular is modified in Subsection II-C to be more logical and clear. In Section III, the proposed algorithm is explained more extensively. We extend abundant experiments in Section IV. We introduce an additional dataset, Wikipedia, and compare the statistical properties of all three datasets in Table 2. In Subsection IV-B, we conduct experiments on all three datasets with two influence spread functions and two weight assignment metrics to make fully comparison of six algorithms. We also make additional discussions about the difference between two influence spread functions and how the proposed LT-S model performs under non-signed social networks in Subsection IV-C.1 and Subsection IV-C.2, respectively. Section V is extended into four parts to detail the related work, which contains additional introduction to diffusion models, signed social networks and opinion formation.

II. PROPOSED FRAMEWORK

A. PRELIMINARIES

In this article, a signed social networks is described as a weighted, directed graph, $G = (V, E, W, R)$, where V is the set of nodes representing users and E is the set of directed edges representing relationships between users. The notation W is the influence weight defined by the function $W : E \leftarrow [0, 1]$, the weight $w_{u,v}$ associated with an edge $(u, v) \in E$ specifies the influence weight of node u when expressing opinions to influence v. For any $(u, v) \notin E$, $w_{u,v} = 0$, and for any node, the sum of incoming edge weights is no greater than 1, $\sum_{u\in V} w_{u,v} \leq 1$. The notation R is the matrix specifying the signed relationship influences. $r_{u,v}$ is derived from consistency and trust of edge (u, v), which is defined as: $r_{u,v} = cons_{u,v} \cdot trust_{u,v}$. Notation $cons_{u,v} \in \{-1, 1\}$ is the consistency when u express his opinion to v, and $trust_{u,v} \in \{-1, 1\}$ denotes if v trust the opinion expressed by u. If $r_{u,v} = +1$, then the relationship from node u to v causes positive influence; if $r_{u,v} = -1$, the relationship from node u to v causes negative influence. If there is no relationship from node u to v, then $r_{u,v} = 0$. Note negative relationships can also cause positive influence. A illustration of a simple signed social network between 5 people is presented in Figure 1.

In the LT model, the status of users can be either active or inactive. Initially, all users are in inactive status and once a user been activated, it will remain active forever. Besides, a threshold θ_u is uniformly assigned to each user u at random in the range $[0, 1]$. The diffusion process

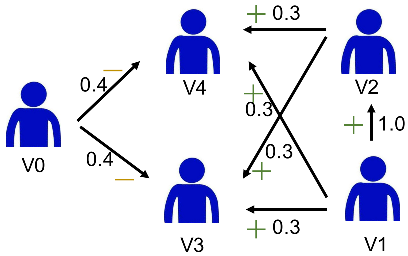

FIGURE 1. An example signed social network with influence weight. Edges are labeled with their influence weight, '+' (green) denotes positive influence, and '−' (orange) denotes negative influence.

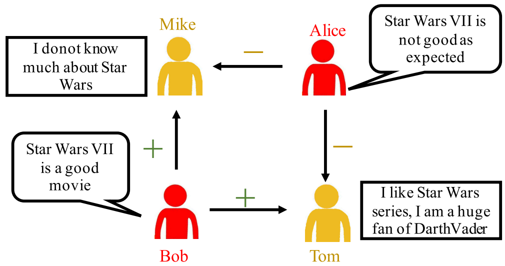

FIGURE 2. An example of opinion formation. The active users (Bob and Alice) are colored red, and the inactive user (Tom and Mike) are colored yellow.

TABLE 1. Notations in this article.

Symbols	Descriptions
G	signed social network
W	influence weight matrix of G
$w_{u,v}$	influence weight of link (u,v)
R	influence type matrix of G
$r_{u,v}$	influence type of link (u,v)
$cons_{u,v}$	consistency when u express opinion to v
$trust_{u,v}$	indicator of whether v trust the opinion from u
z_u	innate opinion of u
y_u	expressed opinion of u
θ_u	threshold for u being actived
α_u	conformity of u
S	seed set
K	number of seed users
A_k	set of activated user at step k
θ	threshold of influence probability
t	controlling the size of candidate seed set by $\lvert V\rvert/2^t$

proceeds in discrete time steps. At the initial step 0, users in the seed set $S \subseteq V$ are firstly activated and all other users are in inactive status. At some later step t, the user u will be activated if and only if the total weight of their active neighbors exceeds his threshold θ_u. Table 1 shows the notations in this article.

B. LT-S MODEL

Considering the opinion formation process in signed networks, we extend the LT model to LT-S in this article. In social networks, the opinions of an individual are often influenced by other people, like friends, neighbors, through social interactions between them. Thus, we introduce opinion formation into LT-S model and distinguish the *innate opinion* and *expressed opinion* of users. In LT-S model, each user, whether in active status or not, has an *innate opinion* (reflecting their preference or history) that is fixed and not amenable to external influences during the opinion diffusion. Users who are activated by their neighbors will generate an *expressed opinion*, which is expressed to their inactive neighbors differently according to different influence types ($r_{u,v}$) and different influence weights ($w_{u,v}$). For example, in Figure 2, Bob and Alice are in active status while Tom and Mike are inactive. Bob and Alice express their opinions about a movie to their neighbor Tom and Mike. Although Tom and Mike have not been activated, they have an innate opinion about movies based on their own preference and history.

In LT-S, we denote the *innate opinion* and *expressed opinion* of user u are as z_u and y_u respectively. The opinion is encoded as a real quantity, $z_u \in [-1, 1]$ and $y_u \in [-1, 1]$, where $z_u > 0$ ($z_u < 0$) and $y_u > 0$ ($y_u < 0$) indicate that user u holds a positive (negative) innate opinion and a positive (negative) expressed opinion about the product or information.

The LT-S model specifies the process of information diffusion in signed network as follows.The process is executed in discrete steps, $k = 0, 1, 2, \ldots$, where A_k denotes the set of users activated at step k. At initial step $t = 0$, a seed set $S \subseteq V$ is activated ($A_0 = S$) and each user in S has a positive expressed opinion valued constant 1. At any later step $t > 0$, all active users try to activate their inactive neighbors, and a user u will be activated if and only if the total weight of his active neighbors (both friends and enemies) exceeds his threshold θ_u, that is $\sum_{v\in\cup_{0\le i\le(t-1)}A_i} w_{v,u} \ge \theta_u$.

After a user be activated at a certain step t, his expressed opinion will be formed. The expressed opinion of a user depends on both his innate opinion and the expressed opinions of active neighbors. The expressed opinion of a user u is calculated by (1).

$$y_u = (1-\alpha_u)\cdot z_u + \alpha_u \cdot \sum_{v\in\cup_{0\le i\le(t-1)}A_i} y_v \cdot w_{v,u} \cdot r_{v,u} \quad (1)$$

In (1) we define y_u, the expressed opinion of u, as the sum of its innate opinion z_u and all the weighted incoming expressed opinion from its active neighbor. $\alpha_u \in [0, 1]$ is the conformity of u, which indicates how much u is influenced by others rather than its own opinion. A larger value of conformity α_u represents that the user u is more influenced by his neighbors rather than his own experience. More specifically,large α_u will cause user u form expressed opinion that more influenced by the expressed opinions of his neighbor, and a low α_u will have the opposite effect. The user's social actions will help decide the values of their *innate opinion* and *conformity*. In real application, we can use the fraction of a user's tweet and retweet frequency as a proxy for conformity value and obtain the innate opinion through opinion mining [21] of recent tweets on Twitter or posts on Facebook.

When an active user forms an expressed opinion, his expressed opinion will be propagated to inactive neighbors (possibly causing their activation). The influence propagation process terminates at step k when $A_k = \emptyset$.

Let $\sigma(S)$ denote the final influence spread of seed set S. In this article, we give two definitions of $\sigma(S)$. In a practical situation, a company wants to promote a new product in the online signed social network. With a limit budget, the company expects to select a small number of initial seed users by offering the discounts or samples to make these users as the first adoption users. These seed users recommend the product to their neighbors and their activated neighbors continue the propagation. The company may expect the campaign has the maximal influence in two ways. One is that, in final active users, the overall opinions (include both positive and negative) is maximized, the other is that the positive opinion is maximized. We give the definitions of these two kinds of $\sigma(S)$ as bellow.

- $\sigma(S)_a$ denotes the sum of all expressed opinions of all users activated by S, $\sigma(S)_a = \sum_{v \in \cup_{i\geq 0} A_i} y_v$;
- $\sigma(S)_p$ denotes the sum of positive expressed opinions of all users activated by S, $\sigma(S)_p = \sum_{v \in \cup_{i\geq 0} A_i} y_v$ (if $y_v > 0$).

$\sigma(S)$ is called the influence spread of seed set S, and $\sigma(\cdot)$ is called the influence spread function. Specially, $\sigma(\cdot)_a$ denotes the all influence spread function and $\sigma(\cdot)_p$ denotes the positive influence spread function for abbreviation.

C. PROBLEM DEFINITION AND PROPERTIES

The influence maximization problem in signed networks is formally defined based on the proposed LT-S model in this section, considering the process of opinion formation and both positive and negative influence types of relationships.

Definition 1 (Influence Maximization Problem): Given a signed social network $G(V, E, W, R)$ and a parameter K ($K < |V|$), the influence maximization problem in a signed network is to find a seed set of users $S \subseteq V$ ($|S| = K$), such that by activating these users with positive opinions, $\sigma(S)$ is maximized under the LT-S model, i.e., $S = \text{argmax}_{S\subseteq V,\ |S|=K}\sigma(S)$.

Theorem 1: The influence maximization problem in signed networks under LT-S model is NP-hard.

Proof: We can prove Theorem II-C by considering a specified instance of the problem. If each node v has innate opinion $z_v = 0$ and conformity $\alpha_v = 1$, and the expressed opinions of all users in active status are uniformly 1 and the edges are all signed as positive influence. Then the problem in this article to maximize the positive expressed opinions is equivalent to the classical influence maximization problem, which have been proved to be NP-hard by literature [2]. ■

Theorem 2: The influence spread function $\sigma(\cdot)$ under the LT-S model is non-monotone and non-submodular.

A set function $f : 2^V \to R$, from sets to reals, is monotone if $f(S) \leq f(T)$ for all $S \subseteq T$. The function f is submodular if $f(S \cup \{w\}) - f(S) \geq f(T \cup \{w\}) - f(T)$ for all $S \subseteq T$, $w \in V$ and $w \notin T$. In contrast to the classical IM problem, the influence spread function under LT-S is non-monotone and non-submodular. To prove Theorem 2, we only need to present counter examples for proving non-monotonicity and non-submodularity respectively. Therefore, we first specify Figure 1 with an instance. The innate opinion of all users is fixed at 0.2, the influence threshold is set to 0.3, and conformity is set to 0.9. We then prove Theorem 2 by proving the situation of $\sigma(\cdot)_a$ and $\sigma(\cdot)_p$ respectively.

- The all influence spread function $\sigma(\cdot)_a$ is non-monotone and non-submodular.

  Proof: Non-monotonicity. Suppose that $S_1 = \{V_1\}$, $S_2 = \{V_1, V_0\}$, $S_3 = \{V_1, V_0, V_3\}$, then $\sigma(S_1) = 2.99$, $\sigma(S_2) = 2.78$, $\sigma(S_3) = 3.85$. As $\sigma(S_3) > \sigma(S_1) > \sigma(S_2)$, so $\sigma(S)$ is non-monotone.

  Non-submodularity. Suppose that $S_1 = \{V_1\}$, $S_2 = \{V_1, V_3\}$, then $\sigma(S_1 \cup \{V_0\}) - \sigma(S_1) = -0.21$, and $\sigma(S_2 \cup \{V_0\}) - \sigma(S_2) = 0.39$. Because $S_1 \subset S_2$ and $\sigma(S_1 \cup \{V_0\}) - \sigma(S_1) < \sigma(S_2 \cup \{V_0\}) - \sigma(S_2)$, $\sigma(S)$ is non-submodular. ■
- The positive influence spread function $\sigma(\cdot)_p$ is non-monotone and non-submodular.

  Proof: Non-monotonicity. Suppose that $S_1 = \{V_1\}$, $S_2 = \{V_1, V_0\}$, $S_3 = \{V_1, V_0, V_3\}$, then $\sigma(S_1) = 2.99$, $\sigma(S_2) = 2.92$, $\sigma(S_3) = 3.92$. As $\sigma(S_3) > \sigma(S_1) > \sigma(S_2)$, so $\sigma(S)$ is non-monotone.

  Non-submodularity. Suppose that $S_1 = \{V_1\}$, $S_2 = \{V_1, V_3\}$, then $\sigma(S_1 \cup \{V_0\}) - \sigma(S_1) = -0.07$, and $\sigma(S_2 \cup \{V_0\}) - \sigma(S_2) = 0.46$. Because $S_1 \subset S_2$ and $\sigma(S_1 \cup \{V_0\}) - \sigma(S_1) < \sigma(S_2 \cup \{V_0\}) - \sigma(S_2)$, $\sigma(S)$ is non-submodular. ■

III. PROPOSED ALGORITHM

A. R-GREEDY ALGORITHM

As mentioned in Section I, the traditional greedy algorithm proposed by Kempe et al. [2] is inapplicable for the influence maximization problem in this article, due to the non-monotonicity and non-submodularity of the influence spread function $\sigma(\cdot)$ under LT-S. Some investigation has been made for the influence functions that are submodular but non-monotone [22], [23], but there is relatively little work proposed to maximizing a function which is both non-monotone and non-submodular. In [9], Feng *et al.* proposed a restricted greedy (R-Greedy) algorithm to solve the problem.

The core idea of the R-Greedy algorithm is to select the first K nodes with maximal marginal influence, and then choose the set of seed nodes that have the largest influence spread. Algorithm 1 illustrates the main framework of R-Greedy algorithm, in which S_k denotes the set of selected seeds until round k, s_k is a single selected seed node at round k, and Inf^u_k denotes the influence after adding u to the selected seed set S_{k-1}. The queue Q_k is used in Algorithm 1 to store the nodes estimated in the round k and the elements in Q_k are in the form of (u, Inf^u_k). The dynamic pruning optimization is used in R-Greedy to skip the nodes whose influence is smaller than *maxMargin* (Line 6). For a node checked in round $(k-1)$, the upper bound of its marginal influence is $(Inf^u_{k-1} + \sigma(\{s_{k-1}\}) - \sigma(S_{k-1}))$. If the upper

Algorithm 1 R-Greedy Algorithm

Require: G, K

Ensure: $S, \sigma(S)$

1: **for** $v = 1$ to $|V|$ **do**
2: calculate $\sigma(v)$ and insert $(v, \sigma(v))$ into Q_0
3: **end for**
4: **for** $k = 1$ to K **do**
5: $maxMargin \leftarrow -\infty$
6: **for** $u \in V \backslash S_{k-1}, \sigma(u) \geq maxMargin$ **do**
7: **if** $u \in Q_{k-1}$ and $(Inf_{k-1}^{u} + \sigma(\{s_{k-1}\}) - \sigma(S_{k-1}) < maxMargin)$ **then**
8: Continue
9: **else**
10: Calculate Inf_k^u and insert (u, Inf_k^u) into Q_k
11: **if** $Inf_k^u - \sigma(S_{k-1}) > maxMargin$ **then**
12: $maxMargin \leftarrow Inf_k^u - \sigma(S_{k-1})$
13: $s_k \leftarrow u$
14: **end if**
15: **end if**
16: **end for**
17: $S_k \leftarrow S_{k-1} \cup \{s_k\}$
18: $\sigma(S_k) \leftarrow \sigma(S_{k-1}) + maxMargin$
19: **end for**
20: return $S, \sigma(S)$

bound is smaller than *maxMargin*, the node is also ignored (Line 7). In order to obtain the influence spread, Monte-Carlo simulation is utilized in R-Greedy, leading extreme time cost for convergence.

B. TECHNIQUES FOR IMPROVING R-GREEDY

In order to accelerate the R-Greedy algorithm, we propose two techniques in this article: the propagation-path based technique that removes users with small influential ability and the live-edge based technique that can reduces the times of Monte-Carlo simulations.

1) PROPAGATION-PATH BASED TECHNIQUE

Instead of considering all users in the signed network, with propagation-path based technique, we only consider those with high influence that could be potential seed users. The influence from user u can be seen as the information spread through paths starting from u, so we can calculate the influential ability of a user based on his propagation path and then choose the users with high influence ability as potential seed users.

Definition 2 (Propagation Path): A propagation path from node u to node v ($v \neq u$) is defined as $P = \langle v_1 = u, v_2, \ldots, v_m = v \rangle$, where $m > 1$. The influence probability of path P is given by $Pr(P) = \prod_{i=1}^{m-1} w_{v_i, v_{i+1}}$.

$Pr(P)$ gives the probability of node u to influence node v through the propagation path P. Therefore, the sum of influence probabilities of all propagation paths from node u to all reachable nodes can represent the influence ability of the node u. The problem of enumerating all simple paths is proved to be #P-hard in [24]. However, we find that the influence probability of the propagation path will diminish rapidly with the increasing of path length of P, since the influence weight w among edges is constrained in $[0, 1]$. Therefore, We can define a threshold θ and prune all the paths whose $Pr(P)$ is smaller than θ. The procedure to calculating of the influential ability of a user is presented in detail in Algorithm 2.

Algorithm 2 Calculate the Influence Ability of Users

Require: u, u', p, θ

Ensure: inf_u, influence ability of u

1: $visit_{u'} \leftarrow$ true
2: $inf_u \leftarrow inf_u + p$
3: **for** each $v \in Adj(u')$ **do**
4: $p' \leftarrow w_{u',v} \times p$
5: **if** $visit_v$ is false and $p' \geq \theta$ **then**
6: $calInf(u, v, p', \theta)$
7: $visit_v \leftarrow$false
8: **end if**
9: **end for**
10: return inf_u

Algorithm 2 recursively calculates the influence probability of user u until the propagation path that starts from u terminates or the influence probability of the propagation path is smaller than the threshold θ. The parameter θ is used to control the length of the propagation path. A smaller θ brings the longer propagation paths, and makes the calculation of influential ability more accurate, but also brings the problem of larger time cost. So in Algorithm 2, θ is used to make a trade off between accuracy and efficiency. $visit_{u'}$ marks whether the node u' is visited or not.

In Algorithm 3, we present the whole process of obtaining seed users through R-Greedy algorithm accelerated by propagation path-based technique. At Line 1 we initialize a max heap H to store the influence probability for all users. At Lines 2–6, the influence probability of all users in the graph G is calculated and inserted into H. At Line 3, where we refer to Algorithm 2, we initialize the influence probability to 1.0 as node u influences itself with probability 1.0. At Line 7, we use parameter t to control the size of candidate set by choosing the top $1/2^t \cdot |V|$ users from H as the candidate set. Then the R-Greedy algorithm is applied with the candidate set as input users to obtain the final seed set S (Line 8).

2) LIVE-EDGE BASED TECHNIQUE

Here, we propose the live-edge based technique to generate live-edge graphs which can be utilized to calculate the influence spread. In [2], Kempe *et al.* have proved that the LT model is equivalent to reachability in "live-edge" graphs. For each node $v \in V$, we select one of its incoming edges by the following rule. The incoming edge (u, v) of node v will be

Algorithm 3 R-Greedy Algorithm Adopting Propagation-Path Based Technique

```
Require: G, K, t, θ
Ensure: S, σ(S)
Initialize a max-heap H
for each u in V do
    inf_u ← calInf(u, u, 1.0, θ)
    insert inf_u into H
    visit_u ← false
end for
CandidateSet ← choose the top 1/2^t · |V| users from H
S, σ(S) ← use R-Greedy with CandidateSet as input users
return S, σ(S)
```

selected at the probability $w_{u,v}$ while no edge will be selected at the probability of $1 - \sum_u w_{u,v}$. We call the selected edges as *live edges* and all other edges as *blocked edges*. For all nodes $u \in V$ do the selection, we can obtain the 'live-edge' graph, which is equivalent to one Monte-Carlo simulation. Cheng *et al.* [4] also have proved that if the computation of influence spread is limited to a smaller number of live-edge graphs, the computational expense can be reduced without loss of accuracy. In this article, we dependently obtain N_g "live-edge" graphs and calculate the average $\sigma(S)$ of all "live-edge" graphs.

In LT-S model, when a user u becomes active, he will form an expressed opinion and diffuses his expressed opinion to neighbors. According to (1), the expressed opinion of a user is not only related with his innate opinion, but also with the expressed opinions of neighbors who are in active status. Thus the order of active users needed to be record when we calculate the opinions of users in the live-edge. In Algorithm 4, we adopt the breadth-first search (BFS) strategy to traverse the live-edge graph G and create an ordered record of active users. At lines 2–4, the expressed opinion and active status of users in S are initialized. When visiting the ith layer of G_r we use a queue Q_i to record the reachable active users in order. Seed users are firstly pushed into queue Q_0 (Line 3). Then node u is successively popped from queue in Line 7 and for each node v reachable from u in G_r, the expressed opinion of v is calculated in Line 10–16. As the expressed opinion of v is depended on his active neighbors, so we find those nodes whose are the activated neighbors of v in the Line 11–14 of Algorithm 4. In Line 17, node v is labeled in active status and pushed into queue Q_{i+1}.

C. ALGORITHMS ANALYSIS

Let n and m denote the number of nodes and edges in G respectively. The R-Greedy algorithm execute Monte-Carlo simulation to estimate and approximate the influence spread. Its time complexity is therefore $O(KnN_s(n + m))$, where K is the size of seeds set and N_s is the times of simulations, generally set to $10,000$.

Algorithm 4 R-Greedy Algorithm Adopting Live-Edge Based Technique

```
Require: G, S, N_g (# live-edge graphs)
Ensure: σ(S)
1:  for r = 1 to N_g do
2:    for u in S do
3:      y_u ← 1,   active_u ← true,   enqueue u into Q_0
4:      σ(S) ← σ(S) + y_u
5:    end for
6:    while Q_i ≠ ∅ do
7:      while Q_i is not empty do
8:        u ← dequeue from Q_i
9:        for v in users reachable from u in G_r do
10:         y ← 0
11:         for each t ∈ Adj(v) do
12:           if active_t then
13:             y ← y + y_t × w_{t,v} × s_{t,v}
14:           end if
15:         end for
16:         y_v ← α_v × z_v + (1 − α_v) × y
17:         active_v ← true,   enqueue v into Q_{i+1}
18:         σ(S) ← σ(S) + y_v
19:       end for
20:     end while
21:     i ← i + 1
22:   end while
23: end for
24: σ(S) ← σ(S)/N_g
25: return σ(S)
```

The R-Greedy algorithm adopting the propagation-path technique contains two parts. The first part calculates the influential ability of users and chooses the most influential users as potential seed users; then in the second part, R-Greedy is applied to select certain number of the most influential users. The total time complexity is $O(n\bar{n} + KnN_s(n' + m))$, where $\bar{n}$ is the average number of nodes in the local region within threshold θ, and n' is the number of 'most influential' users.

The R-Greedy algorithm adopting the live-edge technique also contains two parts. Firstly, generating N_g live-edge graphs has the time complexity of $O(Rm)$; secondly, it costs $O(KnN_g m')$ for R-Greedy to select seed nodes in N_g live-edge graphs, where m' is the average number of live-edges. Therefore, the total time complexity is $O(N_g m + KnN_g m')$, where N_g (as suggested in [4]) is 100.

IV. EXPERIMENTS

In this section we conduct experiments on several real-world, public signed social networks to evaluate the performance of different algorithms. All algorithms are implemented in C++ and measured on a server with Intel i7-3770 (3.9 GHz) and 32GB main memory. Subsection IV-A describes the experiment setup, includes the datasets, the compared

TABLE 2. Statistics of datasets.

Datasets	#Nodes	#Links	+Links	-Links	Avg. Degree
Wikipedia	138,592	740,397	87.9%	12.1 %	5.64
Epinions	131,828	841,372	85.3%	14.7%	12.765
Slashdot	77,350	516,575	76.7%	23.3%	13.028

algorithms and the influence models used in our experiments. Subsection IV-B illustrates the performances of our algorithms from two aspects: (a) the achieved maximal influence spread comparing to baseline algorithms; (b) the running time comparing to baseline algorithms. Subsection IV-C compares and analyses (a) different influence spread functions; (b) influence maximization problem in signed and non-signed social networks; (c) the tuning of parameters in our algorithms.

A. EXPERIMENT SETUP

1) DATASETS

In our experiments, three large online social network data sets Wikipedia[2], Epinion[3], Slashdot[3] are adopted, which have been previously used as benchmarks in research for signed social networks. In these three data sets, users in data set are represented as nodes and interactions are represented as links which are labeled as positive relationships or negative relationships.

- **Wikipedia** is a who-votes-for-whom network in which users can vote for or against others to be administrators in Wikipedia. In this data set, the users that have edited pages votes for or against each other, which makes the interaction be positive or negative respectively.
- **Epinions** is a general consumer product review site in which users can either trust or distrust other's reviews. User u Trust (distrust) a user's reviews means there is a link with positive (negative) relationship from u to the user.
- **Slashdot** is a technology-related news web site known for its specific user community. In this web site, users can tag each other as 'friend' or 'foe' which makes the directed links be positive or negative respectively.

Table 2 shows the comparison between the data sets where "+Links" represents for "Links with positive relationships" and "−Links" represents for "Links with negative relationships". The statistical difference demonstrates that these three data sets can discrepantly represent the signed social networks.

2) ALGORITHMS FOR COMPARISON

We compare our proposed algorithm (RLP) with two state-of-art original greedy algorithms and modified versions of R-Greedy with proposed techniques. The following is a list of algorithms we evaluate in our experiments.

- **CELFGreedy [CELF]**. The greedy algorithm with CELF optimization [6], denoted as **CELF**. Following the literature we set $R = 10,000$ which means that, for each seed set S, Monte-Carlo simulations are conducted 10,000 times to obtain an accurate result.
- **R-Greedy [RG]**. The restricted greedy algorithm proposed in [9], designed for an influence spread function which is non-monotone and non-submodular. R is set to be the same value as in **CELF**.
- **R-Greedy with Propagation-Path [RP]**. R-Greedy using the propagation-path technique proposed in this article. We set t to 5 and θ to 0.003.
- **R-Greedy with Live-Edge [RL]**. R-Greedy with the live-edge technique for influence spread estimation proposed in this article. R is set to 100, as suggested in [4].
- **R-Greedy with Live-Edge and Propagation-Path [RLP]**. R-Greedy incorporating both live-edge and propagation-path techniques, which uses the candidate seed set provided by RP as the input seed set of RL. R is the same as in RL; t and θ are the same as in RP.
- **Max Weight Degree [Degree]**. Select the K nodes with the largest degrees which are the total influence weights on the outgoing links (either negative signs or positive signs).

To obtain the accurate influence spread of each algorithm, for each seed set, we run the simulation on the networks 10,000 times and take the average of the influence spread.

3) WEIGHTS ASSIGNMENTS

In LT-S model, each edge is assigned with influence weight, in experiment, we adopt the weighted model and the trivalency model to generate the influence weight.

- **Weighted Model** [2], [4], [5], [25] sets the weight of every incoming edge of v to be $1/d_v$, where d_v is the indegree of v.
- **Trivalency Model** [3], [5] sets the weight of edges randomly from $\{0.1, 0.01, 0.001\}$ then normalizes the weights of all incoming edges to each node so that they sum to 1.

4) INNATE OPINION, CONSISTENCY AND TRUST SETTINGS

To simulate real social interaction, we set the innate opinion z_u according to Gaussian distribution. We follow the general social principles that the relationship between friends are straight-out. Similar to [2]–[4], [10], we assume positive relationships carry the opinions in a positive manner between users, as people are more likely trust their friends. Thus, we set $cons_{u,v}$, $trust_{u,v}$ as 1 when user u and v have positive relationship, so that the influence type $r_{u,v} = 1$. Conversely, negative relationships influence the opinion ambiguously. Considering the fact that negative relationship can also cause positive influence to opinion spread, instead of uniformly setting negative relationships as negative influence, we randomly set $cons_{u,v}$ and $trust_{u,v}$ as 1 or -1 to represent real social network activities, so that the influence type $r_{u,v}$ could change accordingly.

[2] http://konect.uni-koblenz.de/networks/wikisigned-k2
[3] http://snap.stanford.edu/data/index.html

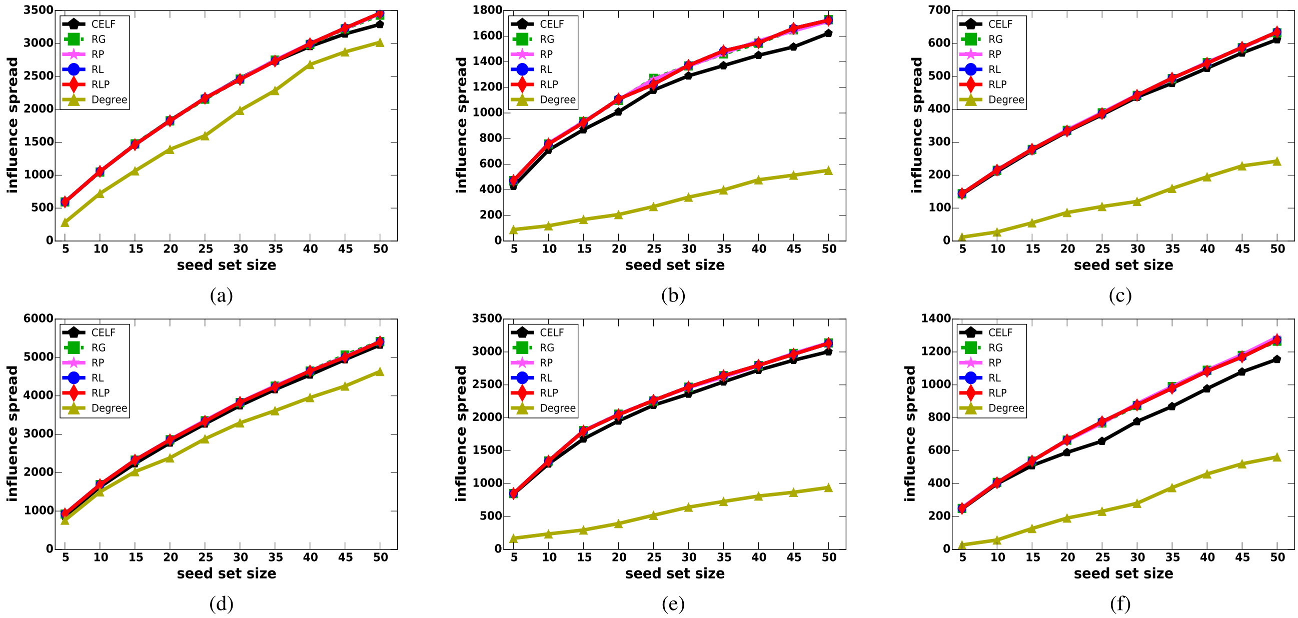

FIGURE 3. Influence spread achieved by various algorithms on different datasets under all influence function $\sigma(\cdot)_a$. (a) Wikipedia with Weighted Model. (b) Epinions with Weighted Model. (c) Slashdot with Weighted Model. (d) Wikipedia with Trivalency Model. (e) Epinions with Trivalency Model. (f) Slashdot with Trivalency Model.

B. EXPERIMENTAL RESULTS

We evaluate the algorithms with two metrics including influence spread and running time. The influence spread reveals the effectiveness of algorithms, and the running time reflects the efficiency of algorithms. For the seed set obtained by different algorithms, the same criterion is required to compare the influence spread. Monte Carlo simulation is the common standard in the influence maximization problem to assess the influence spread. In this article, we run Monte-Carlo simulation $10,000$ times for each seed set. The seed set size K is in the range of 1 to 50 and we compare running time using the case where $K = 50$.

1) INFLUENCE SPREAD

In this article, we propose two different influence spread function: positive influence spread function, $\sigma(\cdot)_p$, which maximizes only the positive opinions of all active users; all influence spread function, $\sigma(\cdot)_a$, which maximizes the overall opinions of all active users. Figure 3 and Figure 4 show the influence spread under all influence spread function and positive influence spread function, respectively. The horizontal axis of experimental results is the size of seed set, the vertical one is the influence spread. In both Figure 3 and Figure 4, we can observe among all datasets, the performance of **CELF** algorithm is the worst in all five greedy algorithms and very unstable. The reason is that **CELF** is designed for the traditional influence maximization which has the monotone and submodular influence spread function. In addition, **RG** and our improved algorithms, **RP**, **RL** and **RLP**, achieve much more stable and accurate results. The algorithm **RG** has the best results among all datasets and influence models because of the non-monotone and non-submodular influence spread function. Our algorithms, **RP**, **RL** and **RLP**, have very close influence spread to **RG**.

Compared with **RG**, **RP**, **RL** and **RLP**, the results of **Degree** indicate that it is not effective by simply choosing high-degree nodes. Among all datasets, Wikipedia achieves the largest influence spread in both weighted model and trivalency model. In trivalency models, all algorithms achieve more influence spread than weighted models. The reason is that the influence weight in trivalency models is larger than that in weighted models.

Besides, we can learn that the final influence spreads of three signed networks are significantly different. The reason is that the influence spread of a network is actually related to multiple factors about the properties of the network, such as the size of the network, the density of the network or the number of communities in the network. Different network properties may cause tremendous difference among final influence spreads.

2) RUNNING TIME

Table 3 shows the running time of different algorithms applying two kinds of weight assignments and two kinds of influence spread function on the Wekipedia, Epinions and Slashdot datasets. Running time reflects the scalability of algorithms. From the results, we can learn that both **CELF** and **RG** are very time-consuming and not suitable for the large-scale datasets. Algorithm **RP** is faster than **CELF** and **RG**, but the improvement is not very significant. Algorithms **RL** and **RLP** are several orders of magnitude more efficient than **RG**. Compare weighted model with trivalency model, we find that algorithms in trivalency model cost more

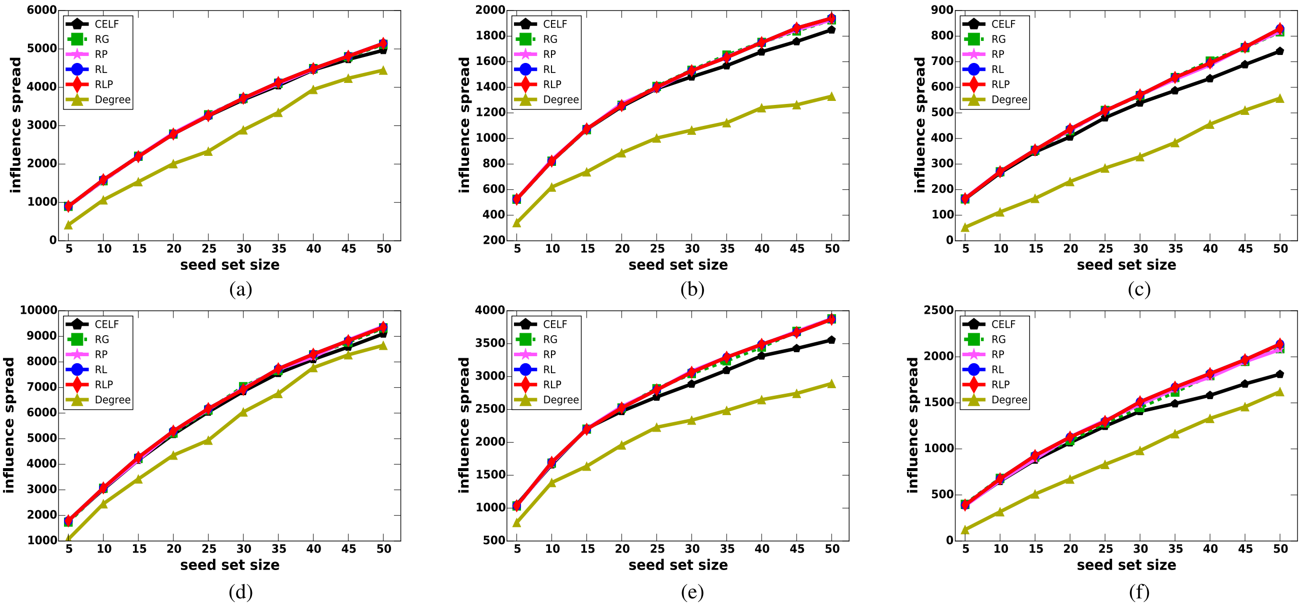

FIGURE 4. Influence spread achieved by various algorithms on different datasets under positive influence function $\sigma(\cdot)_p$. (a) Wikipedia with Weighted Model. (b) Epinions with Weighted Model. (c) Slashdot with Weighted Model. (d) Wikipedia with Trivalency Model. (e) Epinions with Trivalency Model. (f) Slashdot with Trivalency Model.

TABLE 3. Running time of different algorithms with different influence function and weight assignments (in hours and minutes). (a) Weighted model under all influence function $\sigma(\cdot)_a$. (b) Trivalency model under all influence function $\sigma(\cdot)_a$. (c) Weighted model under positive influence function $\sigma(\cdot)_p$. (d) Trivalency model under positive influence function $\sigma(\cdot)_p$.

(a)

Datasets	CELF	RG	RP	RL	RLP
Wikipedia	74.8h	71.3h	18.4h	20.9m	**19.1m**
Epinions	36.8h	35.3h	9.1h	13.4m	**11.1m**
Slashdot	14.9h	13.1h	4.4h	6.7m	**5.3m**

(b)

Datasets	CELF	RG	RP	RL	RLP
Wikipedia	209.3h	190.3h	59.4h	79.7m	**70.1m**
Epinions	98.3h	96.3h	29.1h	37.8m	**32.2m**
Slashdot	50.1h	43.1h	14.4h	20.8m	**18.1m**

(c)

Datasets	CELF	RG	RP	RL	RLP
Wikipedia	85.8h	80.3h	20.1h	32.7m	**29.9m**
Epinions	57.9h	55.3h	11.1h	11.7m	**10.2m**
Slashdot	28.1h	24.1h	7.4h	9.95m	**8.6m**

(d)

Datasets	CELF	RG	RP	RL	RLP
Wikipedia	249.8h	238.2h	66.4h	225.9m	**191.4m**
Epinions	127.9h	123.8h	23.1h	71.5m	**67.5m**
Slashdot	89.1h	88.1h	15.8h	69.8m	**65.1m**

time than in weighted model because the weights assigned by trivalency model are uneven. Besides, algorithms under positive influence function $\sigma(\cdot)_p$ are slightly more time-consuming than those under all influence function $\sigma(\cdot)_a$.

In conclusion, our improved R-Greedy algorithm with two speed-up techniques **RLP** can achieve the best influence spread with lowest time cost in signed social networks.

C. DISCUSSIONS

1) DIFFERENT INFLUENCE SPREAD FUNCTIONS

In this article, we consider two different influence spread functions, positive influence function $\sigma(\cdot)_p$ and all influence function $\sigma(\cdot)_a$. We compare the difference between $\sigma(\cdot)_p$ and $\sigma(\cdot)_a$ when setting seed size $k = 50$. Figure 5 and Figure 6 show the differences in six areas: the positive opinions of all active users, negative opinions of all active users, all opinions of all active users, the number of positive active users, the number of negative active users, and the number of all active users, where the blue bars are the results obtained by applying positive influence spread function $\sigma(\cdot)_p$, and the red bars are the results obtained by applying all influence spread function $\sigma(\cdot)_a$. From Figure 5 and 6 we obtain the results that the influence spreads of the positive opinions and negative opinions of influence spread function $\sigma(\cdot)_p$ are both larger than those of $\sigma(\cdot)_a$ in all datasets. The reason is that $\sigma(\cdot)_p$ is to maximize the positive opinions of all active users, so the final positive opinions and final negative opinions are larger than those of $\sigma(\cdot)_a$. As the influence spread function $\sigma(\cdot)_a$ is to maximize the overall opinions of active users, so in almost all datasets (except Epinions with trivalency model), the all opinions of influence spread function $\sigma(\cdot)_a$ are larger than those of $\sigma(\cdot)_p$. Figure 5 and 6 also show the results of the number of (positive, negative, all) active users. In almost all datasets (except Wikipedia with weighted model), $\sigma(\cdot)_p$ achieves more active users than $\sigma(\cdot)_a$. The reason is that the efforts for maximizing influence spread of positive active users tend to increase the number of positive active users.

These results demonstrate that in the practical applications, if the company wants to maximize the overall opinions of active users and minimize the negative opinions, then the

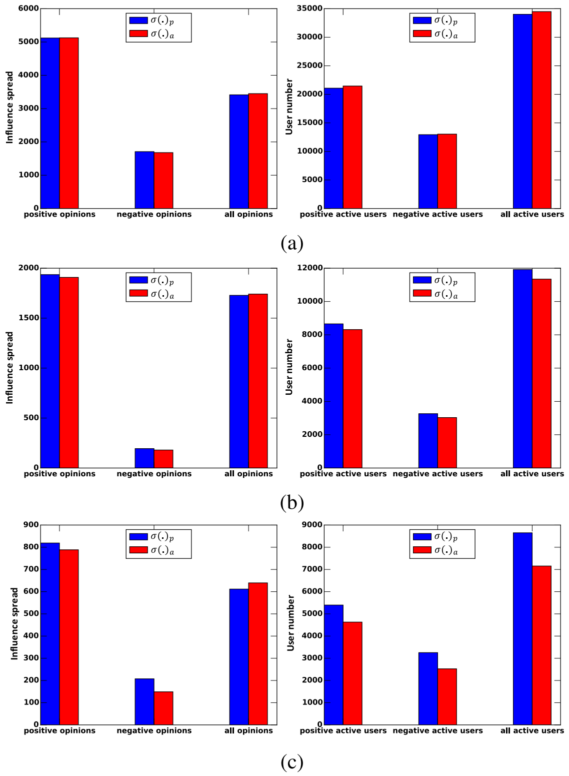

FIGURE 5. Influence spread and active users achieved by various algorithms under the weighted model. (a) Wikipedia with Weighted Model. (b) Epinions with Weighted Model. (c) Slashdot with Weighted Model.

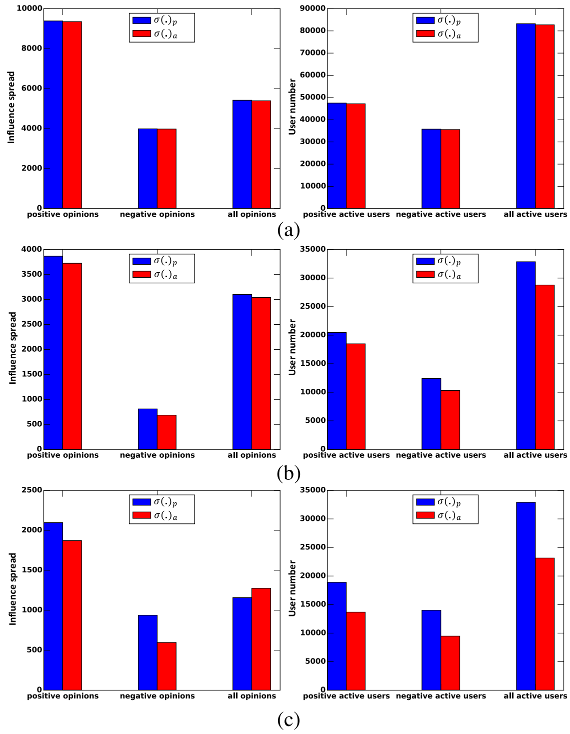

FIGURE 6. Influence spread achieved by various algorithms under the trivalency model. (a) Wikipedia with trivalency model. (b) Epinions with trivalency model. (c) Slashdot with trivalency model.

better choice is $\sigma(\cdot)_a$; if the company wants to maximize the positive opinions of active users, then the $\sigma(\cdot)_p$ will satisfy the demand.

2) LT-S MODEL IN SOCIAL NETWORKS WITHOUT SIGNED RELATIONSHIPS

In this article, we consider the influence maximization problem in signed social networks and propose the LT-S model incorporating opinion formation with signed relationships. We conduct experiments on datasets to compare the differences of LT model and LT-S model in influence maximization without signed relationships. The datasets Wikipedia, Epinions and Slashdot contain the signed relationships, we ignore the signed relationships to conduct the experiments of influence maximization without signed relationships. The influence maximization without signed relationships is the traditional influence maximization problem, so CELF algorithm [6] is carried out in the LT model. In the traditional influence maximization problem, users in active status means that they have generated their own opinion and are starting to influence their neighbors. The influence spread is the active users activated by diffusion models. In this traditional influence maximization problem, we also ignore the impact of *consistency* and *trust* of users in LT-S model. Figure 7 shows the influence spread achieved by LT model and LT-S model in social networks without signed relationships. As shown in Figure 7, the number of active users of LT-S model (the pink line) is larger than those of LT model (the black line), which indicates that LT-S model can activate more users than LT model. Besides, we can observe that active users in LT-S model are more than the users with positive opinions in LT-S model (the blue line). The reason is that users in active status cannot be positively influenced and always hold positive opinions.

3) PARAMETER SETTINGS

1) Effect of propagation probability threshold θ on **RLP**.
Figure 8 shows the effect of different values of θ on the **RLP** algorithm. In Algorithm 2 the parameter θ controls the length of propagation paths, and Figure 8(a) shows the running time taken to obtain influential users using **RLP**. We notice that varying θ has little effect on influence spread in Figure 8(b), but reducing θ significantly decreases the running time in Figure 8(a). We apply $\theta = 0.003$ throughout the experiments to balance the influence spread and running time.
2) Effect of candidate user set size $|V|/2^t$ on **RLP**.
Figure 9 shows the effect of different values of t on the **RLP** algorithm. In Algorithm 3 the parameter t controls the size of the candidate users set by $|V|/2^t$ where $|V|$ is the number of all users. As t increases, the candidate set becomes smaller and we can observe in Figure 9(a) that the running time decreases. However, Figure 9(b) shows higher t will damage the influence spread, so we

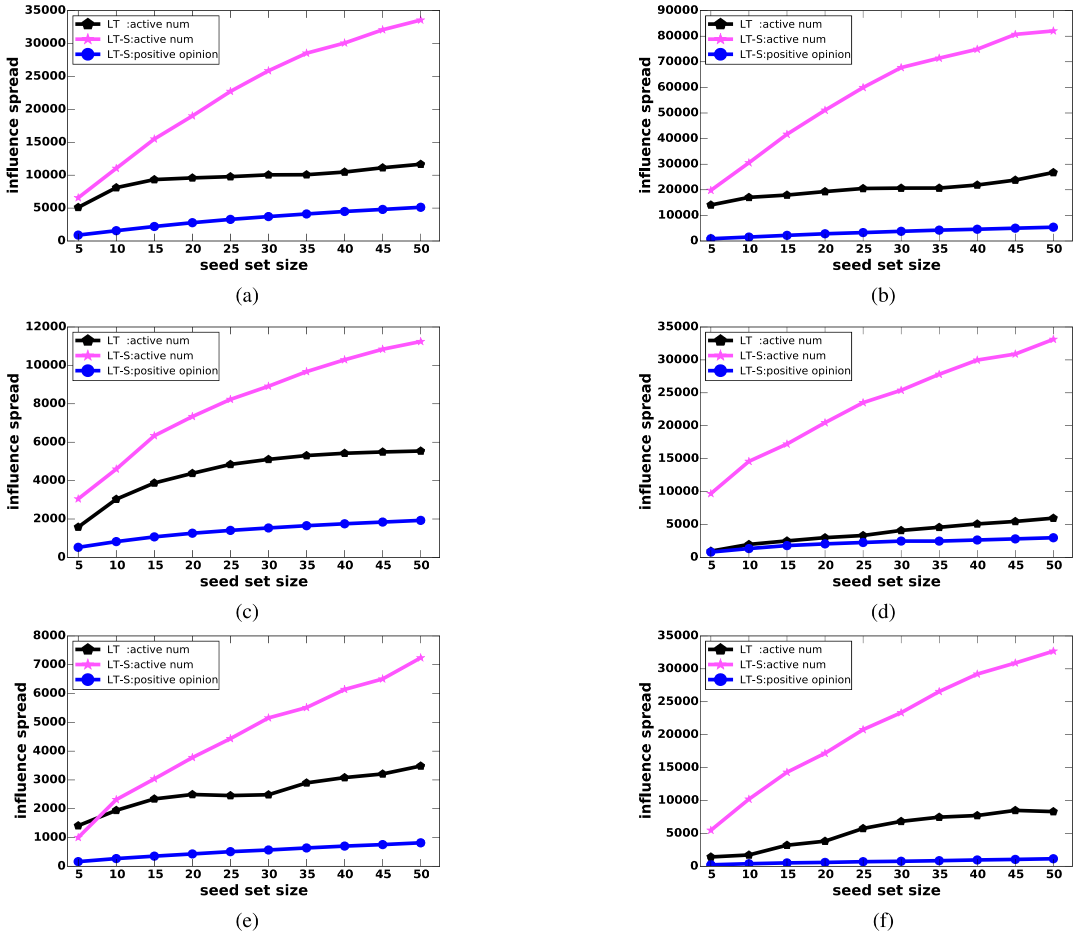

FIGURE 7. The influence spread achieved by LT model and LT-S model in soical networks without signed relationships. (a) Wikipedia with Weighted Model. (b) Epinions with Weighted Model. (c) Slashdot with Weighted Model. (d) Wikipedia with Trivalency Model. (e) Epinions with Trivalency Model. (f) Slashdot with Trivalency Model.

choose $t = 5$ as a compromise between influence spread and running time.

V. RELATED WORK

A. INFLUENCE MAXIMIZATION

Social influence has been a widely accepted phenomenon in social networks for decades. Many applications have been built based on the implicit social influence between people, such as marketing, advertisement and recommendations. Recently, there are many efforts have been put to understand influence qualitatively and quantitatively. [26]–[28] try to distinguish influence and homophily in social networks, [29], [30] quantifying influence and selection, [31]–[33] measure social influence quantitatively. Reference [34], [35] learn the influence probabilities from social data.

Influence maximization is one of the most interesting problems in the study of social influence and has received much research interest in recent years.

Dominigos and Richardson [1] are the first to study influence maximization in probabilistic settings. Kempe *et al.* [2] formulate influence maximization problem as a discrete optimization problem. They prove that the problem is NP-hard and the influence spread function is monotone and submodular under both the IC and LT models. Given these properties, a greedy algorithm which guarantees $(1 - 1/e)$ approximation ratio is proposed to solve the influence maximization problem. The fundamental idea of greedy algorithm is to repeatedly select the user with largest marginal influence spread as seed uses and add it into the seed set until the budget number is reached. Through experiments they show that the greedy algorithm significantly outperforms the classic degree and centrality-based heuristics in influence spread. The main time-cost part is to compute exact influence spread under both IC and LT models, which is proved to be #P-hard in [3]. Thus, Monte-Carlo (MC) simulations are used in each iteration to effectively estimate the influence spread. Monte-Carlo simulations can slightly accelerate the influence spread, but the greedy algorithm is still very time-consuming in practice and not scalable for large social networks. A number of studies are devoted to addressing this efficiency issue.

In [6], Leskovec *et al.* propose an optimization algorithm, called "Cost-Effective Lazy Forward"(CELF) algorithm. The CELF algorithm fully exploits the submodularity

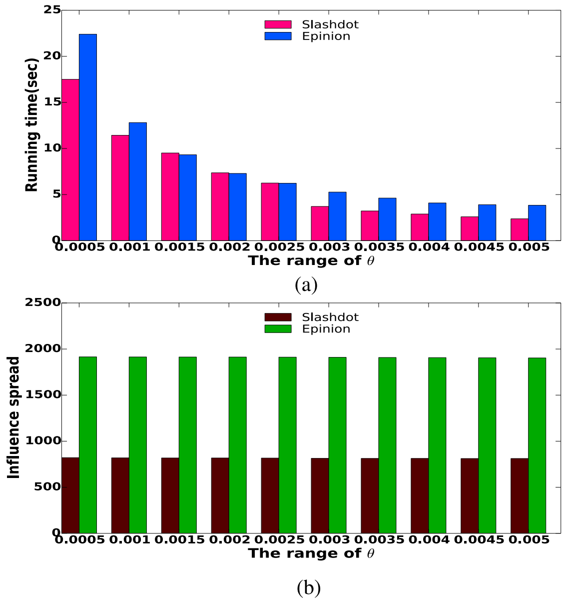

FIGURE 8. **The effect of θ on running time and influence spread under the weighted model. (a) Running time to obtain influential users. (b) Influence spread.**

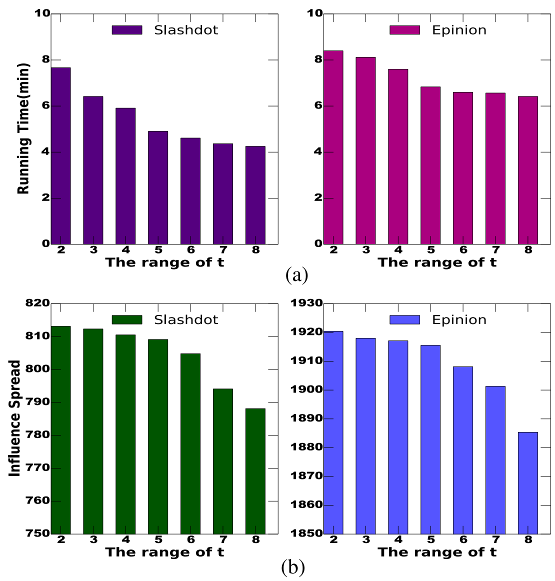

FIGURE 9. **The effect of t on running time and influence spread under the weighted model. (a) Running time to obtain influential users. (b) Influence spread.**

of influence spread functions to greatly reduce the number of evaluations during influence spread. Their results show that CELF algorithm achieves as much as 700 times speedup in selecting seed users. Although CELF algorithm is faster than original greedy algorithm, it still not scalable to large networks with hundreds of thousands of nodes and edges. Chen *et al.* [25] generate a new smaller graph by removing the unreachable edges and nodes for every Monte-Carlo simulation. In the equivalent smaller social graph, they use a linear scan of the new graph by BFS or DFS to speed up the greedy algorithm. This algorithm is named as NewGreedy. Based on NewGreedy algorithm, they also propose a MixGreedy algorithm, which combines NewGreedy algorithm and CELF optimization. In MixGreedy algorithm, NewGreedy is used in selecting the first seed user and CELF optimization is used in the selection of rest seed uses. A static greedy algorithm is creatively proposed by Cheng *et al.* [4], which reuses the generated subgraphs to guarantee the submodularity property of the influence spread function.

In substitution for Monte-Carlo simulations, several kinds of heuristic algorithms are proposed. Different from the greedy-based algorithms, heuristic algorithms fully exploit the properties of models and the structures of networks to avoid Monte-Carlo simulations. DegreeDiscount algorithm [25] is firstly proposed by Chen et al. The main idea of DegreeDiscount algorithm is to discount the degrees of neighbors when considering to add the node of the largest degree into seed set. Chen *et al.* [3] propose to restrict the computations to the local influence regions of nodes and maximum influence paths is adopted to estimate influence spread. Jung *et al.* [5] propose the IRIE algorithm that integrates influence ranking with influence estimation to avoid the disadvantages of primitive influence ranking methods.

The heuristic algorithms are fast and greatly improve the efficiency problem of greedy algorithms. But without theoretical guarantees, and the solution quality of them is very unstable. Arora *et al.* [36] proposed a platform for benchmarking various IM techniques to help users choose the best one given specific scenarios.

Recently, researchers have observed the dynamic nature of social networks due to the social interaction and data transmission. Wang *et al.* [37] define Stream Influence Maximization (SIM) query to address the task of the real-time influence maximization in dynamic social networks, and propose Influential Checkpoint framework for SIM query processing. Tong *et al.* [38] state there are uncertainness in diffusion process in dynamic social networks because of high-speed data transmission and large population of participants. They propose a dynamic independent cascade model and a greedy adaptive seeding strategy to solve the problem.

B. DIFFUSION MODELS

Recently, several models have been proposed that extend IC and LT. Chen *et al.* [8] consider both positive and negative opinions in real social networks and extend the IC model into IC-N model. In IC-N model, the active status of users can be positive and negative and they introduce the parameter q to control the spread of negative opinion. Although IC-N model has a extension of IC model and considers negative opinions, the parameter q is the same for all users is too simplistic and not always realistic.

Reference [7], [39]–[41] focus on the case when multiple innovations are competing within a social network and propose the diffusion models in competitive settings. This scenario exists frequently in real world where multiple companies with comparable products to run for competition.

Li *et al.* [42] propose the polarity-related influence maximization (PRIM) problem which aims to find the seed node set with maximum positive influence or maximum negative influence in signed social networks, and they extend standard IC model to the proposed Polarity-related Independent Cascade (IC-P) model.

Chen *et al.* [43] study the time-constrained influence maximization problem. They notice that in real diffusion, the influence diffused one user to other one can be delayed and also diffusion has the constraint of time. For example, one company wants to promote new product in three days, so a user A can influence user B needs they will meet each other in three days. According to these, they propose the IC-M model. In [44], Liu *et al.* also consider the time-constrained and propose the LAIC model.

In [10], the IM problem has been extended into continuous-time diffusion networks, Feng *et al.* [9] consider the problem of influence maximization with novelty decay, [7] extends the problem to competitive settings. In [45] they distinguish the influence from adoption and propose the LT-C model.

Litou *et al.* [46] proposed Correlated Contagions Dynamic Linear Threshold (CCDLT) model to address the problem that the correlation of multiple contagions simultaneously cascade in the social network and analyze how these affect the users' decisions regarding the adoption of a contagion.

C. SIGNED SOCIAL NETWORKS

Signed network analysis was first proposed in Heider [12] and was formalized by Carwright and Harary [11]. With the rapid development of online social networks, the signed networks have already attracted much attention from computer scientists. Researchers in [47]–[49] attempt to predict the sign of the relationship between two given entities in a signed social network. The problem was first considered by Guha *et al.* [47] Kunegis *et al.* [48] focus on studying the problem with varied similarity functions, and Leskovec *et al.* [49] study this problem based on machine learning. Li *et al.* [14] extend the classic voter model to signed networks and analyze the dynamics of influence diffusion of two opposite opinions.

The other studies in signed social networks are in the direction of community detection. An agent-based method proposed by Yang *et al.* [50] performs a random walk on positive link. Chiang *et al.* [51] propose an effective low-rank modeling approach are. Li *et al.* [14] study influence diffusion in signed networks and extend the classic voter model by incorporating negative relationships.

D. OPINION FORMATION

Not only in real-world communication, but also in online social networks, most people hold opinions about hot topics, events and so on. The opinions can be formed either through interactions with other people or the result of reflection. Opinion formation tries to study the opinion formation of people and describe the process of opinion diffusion. Many opinion formation models have been presented in the sociology and statistics literature. The notable model in opinion formation is the one proposed by Degroot. In [15], Degroot propose the generation of consensus that individuals' opinions are updated by averaging the opinions of their neighborhood. Friedkin and Johnsen [17] are the first to extend the Degroot model, which take both disagreement and consensus into consideration. Clifford et al. [52] propose another famous model, namely the voter model. In the voter model, at each step a selected node randomly pick one of its neighbors at uniform probability and adopts the opinion of the picked neighbor as its own. Srivastava et al. [53] study the problem of competing cascades on signed networks which explore the progressive propagation of two competing cascades in a signed network under the IC model and LT model.

VI. CONCLUSION

In this article, we studied the influence maximization problem in signed social networks with opinion formation. Considering social interactions in real world, we proposed a new diffusion model called LT-S that incorporates both opinion formation and signed relationships. Based on the LT-S model, we formulated the influence maximization problem in signed social networks. We also proved that the influence spread functions under LT-S model are neither monotone nor submodular and proposed an improved R-Greedy algorithm, namely **R**-Greedy with **L**ive-edge and **P**ropagation-path (RLP), which combines R-Greedy with the two effective speed-up techniques. We conducted extensive experiments on datasets taken from large, real-world, signed social networks, and presented results that demonstrate the superior effectiveness and efficiency of the RLP algorithm.

REFERENCES

[1] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proc. KDD*, 2001, pp. 57–66.
[2] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proc. KDD*, 2003, pp. 137–146.
[3] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *Proc. KDD*, 2010, pp. 1029–1038.
[4] S. Cheng, H. Shen, J. Huang, G. Zhang, and X. Cheng, "Staticgreedy: Solving the scalability-accuracy dilemma in influence maximization," in *Proc. CIKM*, 2013, pp. 509–518.
[5] K. Jung, W. Heo, and W. Chen, "IRIE: Scalable and robust influence maximization in social networks," in *Proc. ICDM*, 2012, pp. 918–923.
[6] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proc. KDD*, 2007, pp. 420–429.
[7] S. Bharathi, D. Kempe, and M. Salek, "Competitive influence maximization in social networks," in *Proc. WINE*, 2007, pp. 306–311.
[8] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, and Y. Yuan, "Influence maximization in social networks when negative opinions may emerge and propagate," in *Proc. SDM*, 2011, pp. 379–390.
[9] S. Feng, X. Chen, G. Cong, Y. Zeng, Y. M. Chee, and Y. Xiang, "Influence maximization with novelty decay in social networks," in *Proc. AAAI*, 2014, pp. 37–43.
[10] G. M. Rodriguez and B. Schölkopf, "Influence maximization in continuous time diffusion networks," in *Proc. ICML*, 2012, pp. 313–320.
[11] D. Cartwright and F. Harary, "Structural balance: A generalization of Heider's theory," *Psychol. Rev.*, vol. 63, no. 5, pp. 277–293, 1956.
[12] F. Heider, "Attitudes and cognitive organization," *J. Psychol.*, vol. 21, no. 1, pp. 107–112, 1946.

[13] P. Fan, H. Wang, P. Li, W. Li, and Z. Jiang, "Analysis of opinion spreading in homogeneous networks with signed relationships," *J. Stat. Mech., Theory Exp.*, vol. 2012, no. 8, 2012, Art. no. P08003.
[14] Y. Li, W. Chen, Y. Wang, and Z.-L. Zhang, "Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships," in *Proc. WSDM*, 2013, pp. 657–666.
[15] M. H. DeGroot, "Reaching consensus," *J. Amer. Stat. Assoc.*, vol. 69, no. 345, pp. 118–121, 1974.
[16] P. M. DeMarzo, D. Vayanos, and J. Zwiebel, "Persuasion bias, social influence, and unidimensional opinions," *Quart. J. Econ.*, vol. 118, no. 3, pp. 909–968, 2003.
[17] N. E. Friedkin and E. C. Johnsen, "Social influence and opinions," *J. Math. Sociol.*, vol. 15, nos. 3–4, pp. 193–206, 1990.
[18] J. Tang, S. Wu, and J. Sun, "Confluence: Conformity influence in large social networks," in *Proc. KDD*, 2013, pp. 347–355.
[19] H. Zhang, T. N. Dinh, and M. T. Thai, "Maximizing the spread of positive influence in online social networks," in *Proc. ICDCS*, 2013, pp. 317–326.
[20] C. Shen, R. Nishide, I. Piumarta, H. Takada, and W. Liang, "Influence maximization in signed social networks," in *Proc. WISE*, 2015, pp. 399–414.
[21] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
[22] N. Buchbinder, M. Feldman, J. Naor, and R. Schwartz, "Submodular maximization with cardinality constraints," in *Proc. SODA*, 2014, pp. 1433–1452.
[23] U. Feige, V. S. Mirrokni, and J. Vondrák, "Maximizing non-monotone submodular functions," *SIAM J. Comput.*, vol. 40, no. 4, pp. 1133–1153, 2011.
[24] L. G. Valiant, "The complexity of enumeration and reliability problems," *SIAM J. Comput.*, vol. 8, no. 3, pp. 410–421, Aug. 1979.
[25] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proc. KDD*, pp. 199–208, 2009.
[26] S. Aral, L. Muchnik, and A. Sundararajan, "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks," *Proc. Nat. Acad. Sci.*, vol. 106, no. 51, pp. 21544–21549, 2009.
[27] T. La Fond and J. Neville, "Randomization tests for distinguishing social influence and homophily effects," in *Proc. 9th Int. Conf. World Wide Web*, 2010, pp. 601–610.
[28] A. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and correlation in social networks," in *Proc. KDD*, 2008, pp. 7–15.
[29] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri, "Feedback effects between similarity and social influence in online communities," in *Proc. KDD*, 2008, pp. 160–168.
[30] J. Scripps, P.-N. Tan, and A.-H. Esfahanian, "Measuring the effects of preprocessing decisions and network forces in dynamic network analysis," in *Proc. KDD*, 2009, pp. 747–756.
[31] J. Tang, J. Sun, C. Wang, and Z. Yang, "Social influence analysis in large-scale networks," in *Proc. KDD*, 2009, pp. 807–816.
[32] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "TwitterRank: Finding topic-sensitive influential twitterers," in *Proc. WSDM*, 2010, pp. 261–270.
[33] M. Cha, H. Haddadi, F. Benevenuto, and P. Krishna Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *Proc. Int. AAAI Conf. Weblogs Social Media*, 2010, pp. 1–8.
[34] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan, "Learning influence probabilities in social networks," in *Proc. WSDM*, 2010, pp. 241–250.
[35] K. Saito, R. Nakano, and M. Kimura, "Prediction of information diffusion probabilities for independent cascade model," in *Proc. KES*, vol. 3, pp. 67–75, Feb. 2008.
[36] A. Arora, S. Galhotra, and S. Ranu, "Debunking the myths of influence maximization: An in-depth benchmarking study," in *Proc. SIGMOD Conf.*, 2017, pp. 651–666.
[37] Y. Wang, Q. Fan, Y. Li, and K.-L. Tan, "Real-time influence maximization on dynamic social streams," *Proc. VLDB Endowment*, vol. 10, no. 7, pp. 805–816, 2017.
[38] G. Tong, W. Wu, S. Tang, and D. Du, "Adaptive influence maximization in dynamic social networks," *IEEE/ACM Trans. Netw.*, vol. 25, no. 1, pp. 112–125, Feb. 2017.
[39] X. He, G. Song, W. Chen, and Q. Jiang, "Influence blocking maximization in social networks under the competitive linear threshold model," in *Proc. 12th SIAM Int. Conf. Data Mining*, Anaheim, CA, USA, 2012, pp. 463–474.
[40] C. Budak, D. Agrawal, and A. El Abbadi, "Limiting the spread of misinformation in social networks," in *Proc. 20th Int. Conf. World Wide Web*, Hyderabad, India, 2011, pp. 665–674.
[41] A. Borodin, Y. Filmus, and J. Oren, "Threshold models for competitive influence in social networks," in *Proc. WINE*, 2010, pp. 539–550.
[42] D. Li, Z.-M. Xu, N. Chakraborty, A. Gupta, P. K. Sycara, and S. Li, "Polarity related influence maximization in signed social networks," *PLoS ONE*, vol. 9, no. 7, 2014, Art. no. e102199.
[43] W. Chen, W. Lu, and N. Zhang, "Time-critical influence maximization in social networks with time-delayed diffusion process," in *Proc. AAAI*, 2012, pp. 1–7.
[44] B. Liu, G. Cong, D. Xu, and Y. Zeng, "Time constrained influence maximization in social networks," in *Proc. ICDM*, 2012, pp. 439–448.
[45] S. Bhagat, A. Goyal, and L. V. S. Lakshmanan, "Maximizing product adoption in social networks," in *Proc. 5th Int. Conf. Web Search Web Data Mining*, Seattle, WA, USA, 2012, pp. 603–612.
[46] J. Litou, V. Kalogeraki, and D. Gunopulos, "Influence maximization in a many cascades world," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2017, pp. 911–921.
[47] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, "Propagation of trust and distrust," in *Proc. Int. Conf. World Wide Web*, 2004, pp. 403–412.
[48] J. Kunegis, A. Lommatzsch, and C. Bauckhage. "The slashdot zoo: Mining a social network with negative edges," in *Proc. Int. Conf. World Wide Web*, 2009, pp. 741–750.
[49] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," in *Proc. Int. Conf. World Wide Web*, 2010, pp. 641–650.
[50] B. Yang, W. Cheung, and J. Liu, "Community mining from signed social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 10, pp. 1333–1348, Oct. 2007.
[51] K.-Y. Chiang, C.-J. Hsieh, N. Natarajan, I. S. Dhillon, and A. Tewari, "Prediction and clustering in signed networks: A local to global perspective," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1177–1213, 2014.
[52] P. Clifford and A. Sudbury, "A model for spatial conflict," *Biometrika*, vol. 60, no. 3, pp. 581–588, 1973.
[53] A. Srivastava, C. Chelmis, and V. K. Prasanna, "Computing competing cascades on signed networks," *Social Netw. Anal. Mining*, vol. 6, p. 82, Dec. 2016.

WENXIN LIANG (M'08) received the B.E. and M.E. degrees from Xi'an Jiaotong University, China, in 1998 and 2001, respectively, and the Ph.D. degree in computer science from the Tokyo Institute of Technology, Japan, in 2006. He was a Postdoctoral Research Fellow with the CREST of Japan Science and Technology Agency and a Guest Research Associate with GSIC, Tokyo Institute of Technology, from 2006 to 2009. He was an Associate Professor with the School of Software, Dalian University of Technology, China, from 2009 to 2018. He is currently with the School of Software Engineering, Chongqing University of Posts and Telecommunications, China. His current research interests include data engineering, artificial intelligence, and social networks. He is a Senior Member of the China Computer Federation (CCF), and a member of the CAAI, ACM, ACM SIGMOD Japan Chapter, and the Database Society of Japan.

CHENGGUANG SHEN received the B.S. and M.S. degrees in software engineering from the Dalian University of Technology, China, in 2013 and 2016, respectively. He is currently with Meituan-Dianping, Beijing, China, as an Algorithm Engineer. His research interests include machine learning and data mining for advertising display and recommendation.

XIAO LI received the B.S. degree in software engineering from the Dalian University of Technology, China, in 2016, where he is currently pursuing the master's degree in computer science with the School of Software. His research interests include machine learning and data mining.

RYO NISHIDE (M'03) received the B.S., M.S., and Ph.D. degrees in informatics from Kansai University, Japan, in 2002, 2004, and 2008, respectively. He has been a Postdoctoral Research Associate with Kwansei Gakuin University, since 2008, also been a Research Associate/Assistant Professor with Ritsumeikan University, since 2010, and also been an Assistant Professor with the Graduate School of System Informatics, Kobe University, since 2017. He is currently an Associate Professor with the Center for Data Science Education and Research, Shiga University. His research interests include data engineering, mobile computing, networking, and smart agriculture. He is a member of the IPSJ and ACM.

IAN PIUMARTA (M'19) received the Ph.D. degree in computer science from Manchester University, U.K., in 1992. From 1993 to 2004, he worked in Paris, France, at IRCAM, INRIA, and the Computer Science Laboratory, University of Paris. Since 2004, he has been with the Viewpoints Research Institute, CA, USA. In 2010, he was a Visiting Researcher with Kyoto University, and in 2012, a JSPS Visiting Scholar with Ritsumeikan University, Japan, where he was an Associate Professor, from 2014 to 2019. He is currently a Professor with the Department of Mechanical and Electrical Systems Engineering, Kyoto University of Advanced Science. His research interests include collaborative and distributed systems, programming languages, meta-programming, and the IoT. He is a member of the ACM.

HIDEYUKI TAKADA received the B.E. and M.E. degrees in information science from Kyoto University, in 1991 and 1993, respectively, and the Ph.D. degree in social informatics from Kyoto University, in 2001. He was with Mitsubishi Electric Corporation, from 1993 to 2003. He was a Researcher with Kyoto University, from 2004 to 2006. He is currently a Professor with the College of Information Science and Engineering, Ritsumeikan University. His research interests include distributed systems, collaboration support systems, and learning support systems. He is a member of the IPSJ, IEICE, JSET, the IEEE Computer Society, and ACM.