Containment of Rumor Spread in Complex Social Networks[⋆]

Lan Yang[a,d] Zhiwu Li[a,b] Alessandro Giua[c]

[a]SEME, Xidian University, Xi'an, China

[b]ISE, Macau University of Science and Technology, Taipa, Macau

[c]DIEE, University of Cagliari, Cagliari, Italy

[d]Aix Marseille University, Universite de Toulon, CNRS, LIS, Marseille, France

Abstract

Rumors can propagate at great speed through social networks and produce significant damages. In order to control rumor propagation, spreading correct information to counterbalance the effect of the rumor seems more appropriate than simply blocking rumors by censorship or network disruption. In this paper, a competitive diffusion model, namely Linear Threshold model with One Direction state Transition (LT1DT), is proposed for modeling competitive information propagation of two different types in a same network. The problem of minimizing rumor spread in social networks is explored and a novel heuristic based on diffusion dynamics is proposed to solve this problem under the LT1DT. Experimental analysis on four different networks shows that the novel heuristic outperforms pagerank centrality. By seeding correct information in the proximity of rumor seeds, the novel heuristic performs as well as the greedy approach in scale-free and small-world networks but runs three orders of magnitude faster than the greedy approach.

Key words: Social networks, Threshold models, Information propagation, Rumor containment.

1   Introduction

Propagation phenomena can be observed in many complex systems as diverse as diffusion of innovation, belief, and information in social networks, epidemics spread in human and animal physical contact networks, cascading failures in infrastructure networks, virus spread in computer networks, etc. Besides, the prevalence of online social media has been witnessed in the past decades where people are more and more likely to exchange ideas, share information, and even adopt innovations or new products. Thus a rich body of the research has been devoted to analyzing the propagation of information, opinion, social behavior, and innovation. This kind of research may help people to understand how the information or opinion propagates in social networks and how to convince people to adopt an innovation or follow a new idea. To model such propagation, Kempe et al. [20] firstly proposed two main discrete diffusion models, namely the *Linear Threshold Model* (LTM) and the *Independent Cascade Model* (ICM). Both models consist in directed graphs denoted by $G(V, E)$, where each node represents a user or an individual and has two alternative states at a given time step, namely active (if it has adopted the innovation) and inactive (if it has not adopted the innovation). Initially, all the nodes are inactive. At time step $t = 0$, a subset of nodes $S \subseteq V$ is activated in order to start the diffusion process and other nodes stay inactive.

1.1   Related work

The most widely studied problem considering propagation in social networks is the influence maximization problem. This is a problem of identifying a small subset of early adopters to maximize the total influence propagation for a given diffusion model and has a clear practical motivation in many applications including budget allocation in viral marketing [1, 11, 20, 27, 34, 41, 46], controlling the rumor spread [17, 40, 48], opinion formation [14, 33, 38, 49], information spread forecasting [26], etc. The influence maximization problem was firstly formalized by Kempe et al. [20] as an optimization problem and is proved to be NP-hard under both the ICM and LTM. The authors then presented a greedy algorithm to solve the problem that can achieve a good sub-optimal solution. Following this work, improved greedy algorithms [8, 9, 13, 24], community based approaches [7, 23, 36], new centrality measures [35, 37] and efficient heuristics [15, 25, 31, 45] have been proposed to solve the problem aiming to balance the time complexity of algorithms and the influence propagation and trying to make them scalable to large datasets. In recent years, researchers have studied the influence maximization problem in more complex networks by taking the heterogeneity of individual relationships [18] or multiplexity [42] into consideration.

Influence maximization problems are to maximize the spread of good things such as positive information and innovation. Undesirable things, such as failures, rumors, epidemics, etc., may also spread in a similar fashion and could produce significant damages to the society. For instance, the rumor "Two explosions in White House and Obama is injured" occurred in April 23, 2013 caused 10 billion USD losses before the rumor was clarified [32]. Therefore, the problem of containing or controlling the undesirable things spread is nonnegligible. Recently, Yang et al. [44] addressed two different classes of influence minimization problems. The first one, called loss minimization with disruption, aims to minimize the spread due to an unavoidable disturbance. The second one, called diffusion minimization with guaranteed target, aims to minimize the spread under the constraint that the spread should reach a given target set.

In this paper, we focus on containing the rumor spread in social networks. Rumor control strategies can be divided into two categories: network disruption strategies and counterbalance strategies. Network disruption strategies can be carried out by removing some critical nodes and edges of the underlying network and were used in [21, 22] to suppress the misinformation spread. However, such a strategy is not viable in practice since (1) interfering with the network structure may be impossible for a control agent and (2) removing critical nodes is a form of censorship that may violate ethical standards. Counterbalance strategies, on the contrary, aim to reduce the diffusion of rumors by spreading correct information. This is the approach that we will consider in this paper.

The basic diffusion models –such as the one we considered in [44]– assume that there is only one item propagating in the network. For this reason they are not suitable to model the competitive propagation

phenomenon that characterizes the use of a counterbalance strategy to control the rumor spread. Several competitive diffusion models have been proposed [3, 4, 5, 16, 29, 47] being different in the mechanisms describing how an individual chooses an item to adopt when two competitive items reach him/her at the same step. Under the competitive diffusion models, a meaningful problem, i.e., competitive influence maximization, is that of choosing a suitable set of nodes adopting an item initially, so as to maximize the influence spread of such an item. It is usually assumed that the seeding strategy of the competitors has already been fixed and is known. Nash equilibrium analysis has been proposed in the case in which all competitors may adapt their strategy in a game theoretic perspective [10].

In this paper, we also consider the competitive propagation between two types of information, e.g., rumor and truth (from now on for the sake of simplicity we will use "truth" instead of "correct information") on a topic or an activity. Instead of studying the competitive influence maximization problem, we address the problem of minimizing rumor spread by choosing a suitable set of nodes to spread the truth once the rumor is detected.

Similar work has been done in [5] and [16]. Specifically, in [5], minimizing rumor spread was studied considering an independent cascade model where it is assumed that if truth and rumor reach an individual at the same step then truth prevails. In [16], He et al. formalized the influence blocking maximization problem under an extension of the linear threshold model where rumor prevails if an individual can be activated by both rumor and truth at the same time step.

However, the above two models are limited in two aspects.

- The probability of a user accepting the truth when rumor and truth reach him/her simultaneously is either 0 or 1. However, in realistic cases this probability of individuals accepting the truth varies with many complex social factors. As an example, sometimes he/she may doubt the source of the truth.
- In addition, all the mentioned competitive diffusion models assume that once an individual is activated by one type of information, he/she keeps like that forever and never changes his/her mind. This assumption works well for product adoption that is associated with purchase behaviors which typically are not easily reversible, but not for information spread or opinion formation since people's attitude towards a politic activity or a news event changes according to new information gathered from the network.

1.2   Contributions

In this work, there are three main contributions.

First, a Linear Threshold model with One Direction state Transition (LT1DT) is presented to overcome the shortcomings of previous models.

- We introduce two thresholds for each individual: influence threshold and decision threshold which take effect in different stages when activating an inactive node or convincing a rumor activated node to switch its belief to the truth.
- The model allows a reconsideration for nodes that have already accepted the rumor but not for nodes that have already accepted the truth.

Such a model is consistent with behavioral patterns that many of us have personally observed: people who have unintentionally contributed to the rumor spread may later apologize for the mistakes. On the other hand, an individual which knowingly accepts the truth will most likely disregard the rumor [40].

The LT1DT turns out to be non-progressive due to the reconsideration process. However, by analyzing the steady state, we can provide an upper bound of the number of nodes that will eventually adopt the rumor and a lower bound of the number of nodes that will eventually adopt the truth.

Second, under the framework of the LT1DT model, we address the issue of minimizing rumor spread adopting a counterbalance strategy. We show that this problem is NP-hard and propose a novel heuristic

called *ContrId* which is based on the diffusion dynamics to solve the problem. We compare ContrId with two different heuristics which are adaptations of well known ones. One is solely based on the network structure and uses a *PageRank* algorithm to select nodes with higher centrality score and the other one uses a greedy framework called *MinGreedy*. ContrId outperforms PageRank and has the same level of computation complexity, which grows linearly with the number of nodes n in the network. ContrId performs not as well as MinGreedy in terms of rumor containment but runs three orders of magnitude faster than MinGreedy.

Third, we study the proximity effect, i.e., the effect of seeding truth in the proximity of rumor seeds, in combination with the algorithms ContrId, PageRank, and MinGreedy. The corresponding heuristics are called ProxContrId, ProxPageRank, and ProxMinGreedy. Such restriction significantly reduces the computation complexity of the heuristics especially for MinGreedy and brings better effectiveness of rumor containment. The experimental results also show that ProxContrId performs as well as ProxMinGreedy in scale-free and small-world networks and runs three orders of magnitude faster than ProxMinGreedy. Furthermore, ContrId and ProxContrId can guarantee that adding nodes to the truth seed set will not increase the rumor spread. This cannot be guaranteed if the truth seed set is selected by other approaches such as PageRank and ProxPageRank.

Some preliminary results on this topic were presented in [43]. This paper is substantially different: an analysis of the steady state is provided; a greedy approach and its constrained version are presented as well to solve the problem of minimizing rumor spread in order to compare ContrId and the greedy approach; complete analysis of computation complexity and monotonicity guarantee of all presented approaches are clarified; the issue of scalability of the approaches are discussed; a more comprehensive simulation on both synthetic datasets generated by some popular models (e.g., Barabási-Albert model [2] and Watts-Strogatz model [39]) and real-world datasets is given.

1.3 Organization

The organization of this paper is the following. Section 2 proposes a new competitive diffusion model, i.e., Linear Threshold model with One Direction state Transition. Section 3 presents the minimizing rumor spread problem (MRS) and its hardness. Section 4 lists two categories of heuristics to solve the MRS problem. A series of experimental study are presented in Section 5. Section 6 concludes the paper and discusses the future work.

2 Linear Threshold model with One Direction state Transition

In this section, we generally introduce the Linear Threshold model with One Direction state Transition (LT1DT for short) to describe the competitive influence propagation of two items (e.g., two similar products with different qualities, two opposite beliefs on a topic, etc.), say R and T, in the same network.

2.1 Network Structure

A *LT1DT* network N is a 4-tuple $(G(V,E), W, \theta, \theta^R)$ where $G(V,E)$ is a directed graph with $V = \{1, 2, ..., n\}$ denoting the set of nodes and $E \subseteq V \times V$ denoting a set of directed arcs, i.e., we write $(u,v) \in E$ when there is a directed arc from node u to v. Functions $\theta, \theta^R : V \to (0,1]$ are mappings that assign an *influence threshold* $\theta_u \in (0,1]$ and a *decision threshold* value $\theta_u^R \in (0,1]$ to each node $u \in V$. Function $W : V \times V \to (0,1]$ is a mapping that assigns a weight $W(u,v) \in (0,1]$ to each directed arc $(u,v) \in E$ such that $W(u,v) = 0$ if $(u,v) \notin E$ and for all $v \in V$ it holds $\sum_{u \in V} W(u,v) = 1$.

The influence threshold θ_u intuitively represents the tendency of node u to be influenced to participate into an activity or adopt a type of information or innovation when their neighbors do [20] (the smaller θ_u is, the easier node u can be influenced) and the decision threshold θ_u^R represents the tendency of node u to adopt item R after being influenced (the smaller θ_u^R is, the easier node u will adopt item R). We define the *in-neighbor* set of node $u \in V$ as $N_u^{in} = \{v | (v, u) \in E\}$ and *out-neighbor* set as $N_u^{out} = \{v | (u, v) \in E\}$. Arc (u,v) represents that node u has direct influence on node v.

In this paper we assume that the parameters of the network are given. In reality, it may not be easy to learn them from real social network data. This identification problem goes beyond the scope of this paper but we refer to [12] for the studies on learning network parameters.

2.2 Diffusion Dynamics

In the LT1DT model, each node has three alternative states at each step: inactive, R-active (adopting R), and T-active (adopting T) and one transient state: influenced. Let S_R denote the *seed set* of R which represents a set of nodes that adopt R at time step $t = 0$. Such an adoption starting from the seed set propagates to the network in discrete steps. We denote by ϕ_t^T the set of nodes that adopt T at step t. The set of nodes T-*active* at step t, i.e., those which adopt T in the interval $[0, t]$, is denoted by $\Phi_t^T = \bigcup_{k=0}^{t} \phi_k^T$.

The notations S_R, ϕ_t^R will be similarly adopted for item R. Let $\phi_t^{R \to T}$ denote the set of nodes changing from R-active to T-active at step t. Then, the set of nodes R-*active* at step t, i.e., those which adopt R and do not change the decision in the interval $[0, t]$, is denoted by $\Phi_t^R = \bigcup_{k=0}^{t} \phi_k^R \setminus \bigcup_{k=0}^{t} \phi_k^{R \to T}$. The set of nodes that are active (no matter if R-active or T-active) at step t, is denoted by $\Phi_t = \Phi_k^R \cup \Phi_t^T$. By definition, we have $\Phi_0 = S_R \cup S_T$.
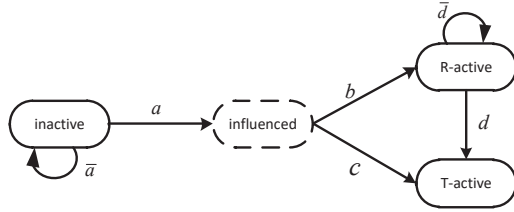

Fig. 1. LT1DT model: Node-level automaton. Events $a, \bar{a}, b, c, d, \bar{d}$ represent different possible state transition. When a node is inactive, event a (the node is influenced) occurs if Eq. (1) holds, otherwise event $\bar{a}$ (remains inactive) occurs. When a node is influenced, event b (the node adopts R) occurs if Eq. (2) holds, otherwise event c (the node adopts T) occurs. When a node is R-active, event $\bar{d}$ (the node remains R-active) occurs if Eq. (2) holds, otherwise event d (the node changes its mind to T-active) occurs.

Different with the LTM, in the LT1DT at each step, each node uses its own node-level automaton to decide to which state it transits. The state transition depends on its current state and the influence it received and compared to its thresholds. The node-level automaton is shown in Fig. 1 where each state is indicated by a circle with label and the influenced state framed by the dotted line is transient.

Precisely, at step $t = 1, 2, \cdots$, two processes occur concurrently, i.e., *activation process* (for each inactive node $u \in V \setminus \Phi_{t-1}$) and *reconsideration process* (for R-active nodes except the seeding nodes $u \in V \setminus (\Phi_{t-1}^T \cup S_R)$) [1].

2.2.1 Activation process

The process of activating an inactive node takes place in two stages: *influence stage* which depends on the influence threshold and *decision stage* which depends on the decision threshold. In the influence stage, an inactive node u is influenced to participate into the activity if the total weight from its active in-neighbors is greater than or equal to θ_u, i.e., $\forall u \in V \setminus \Phi_{t-1}$,

$$u \text{ is influenced} \iff \sum_{v \in N_u^{in} \cap \Phi_{t-1}} W(v, u) \geq \theta_u. \tag{1}$$

In the decision stage a node's decision depends on its decision threshold θ_u^R. Node u will adopt R if the ratio of the sum of the influence weights from its R-active in-neighbors to the total weight from its active

---

[1] The reconsideration of the R's seeds is not allowed since normally initial spreaders are given free samples or benefits and it is hard to change their mind.

in-neighbors is greater than or equal to θ_u^R, i.e.,

$$u \in \phi_t^R \iff \frac{\sum_{v \in N_u^{in} \cap \Phi_{t-1}^R} W(v,u)}{\sum_{v \in N_u^{in} \cap \Phi_{t-1}} W(v,u)} \geq \theta_u^R. \tag{2}$$

Otherwise, it will adopt T, i.e., $u \in \phi_t^T$.

2.2.2 *Reconsideration process*

At each step t, an R-active node $u \in V \setminus (\Phi_{t-1}^T \cup S_R)$ reconsiders its active states due to the influence of its in-neighbors. It remains R-active if the ratio of the sum of the influence weights from its R-active in-neighbors to the total weight from its active in-neighbors is greater than or equal to θ_u^R, i.e., Eq. (2) holds. Otherwise, it becomes T-active, i.e., $u \in \phi_t^{R \to T}$.

The evolution propagates until no more individual can be influenced or reconsider his decision. In this case the network enters a steady state. The sets of *R-final adopters* and *T-final adopters* are denoted by $\Phi_R^*(S_R, S_T)$ and $\Phi_T^*(S_R, S_T) = \bigcup_{t=0}^{\infty} \phi_t^T$, respectively. The set of *final adopters* including R-final adopter and T-final adopter is denoted by $\Phi^*(S_R, S_T)$ and its cardinality $|\Phi^*(S_R, S_T)|$ is called the *adoption function*. Analogously, the cardinality of R-final adopters, i.e., $|\Phi_R^*(S_R, S_T)|$, is called the *R-adoption function*, and $|\Phi_T^*(S_R, S_T)|$ is called the *T-adoption function*.

Note that the LT1DT can be directly reduced to the simple LTM model when there is only one type of item propagating through the network.

In the following two parts, we discuss the properties of the LT1DT model in terms of the transient phase and the steady state.

2.3 *Analysis of the System's Dynamics*

We say that a set function $f(S) : 2^V \to \mathbb{R}$ is *non-decreasing* (resp., *non-increasing*) if for any subsets $S \subseteq S' \subseteq V$, it holds $f(S) \leq f(S')$ (resp., $f(S) \geq f(S')$). Function $f(S)$ is said to be *monotone* if it is non-increasing or non-decreasing. An analogous notion of *monotonicity with respect to argument S_1* can be given for a function $f(S_1, S_2, \ldots, S_n)$ with multiple arguments.

The propagation of the two items in the LT1DT model is a complex phenomenon where competition and complementarity may occur.

Let us first consider the timed evolution during the transient phase. The total propagation is *progressive* in the network since any active node never switches back to inactive. The propagation of T is *progressive* as well since T-active is a sink state. However, the propagation of R is *non-progressive* since a node that is R-active at a certain step may switch to T-active in a future step.

Concerning the final steady state, while both adoption function $|\Phi^*(S_R, S_T)|$ and T-adoption function $|\Phi_T^*(S_R, S_T)|$ are obviously *non-decreasing* w.r.t. S_T given S_R, the R-adoption function $|\Phi_R^*(S_R, S_T)|$ is *not non-increasing* w.r.t. S_T given S_R, that is, adding nodes to the seed set of T may not always decrease the spread of its competitor R. This property means that in this setting the choice of the seed set of item T requires particular consideration if the objective is that of minimizing the number of adopters of item R. This phenomenon cannot be avoided in some scenarios because spreading a competitive product boosts the awareness of such type of product or innovation, while the type of product adopted depends on the human decision making. However, we will show that if S_T is selected by some particular strategies (e.g., ContrId and ProxContrId presented in Section 4), $|\Phi_R^*(S_R, S_T)|$ is guaranteed to be non-increasing w.r.t. S_T.

We show the non-monotonicity in the general case by giving an example. Consider the network in Fig. 2. Assume that $S_R = \{1\}$, $W(u,v) = 1/|N_v^{in}|$, $\theta_2, \theta_3 \in (0.5, 1], \theta_5 \in (0, 0.5)$, and $\theta_2^R, \theta_3^R \in (0, 0.5), \theta_5^R \in (0.5, 1]$. Without the competition with T, i.e., only R spreads in the network, $\Phi_R^*(S_R, \emptyset) = \{1, 2, 5, 6\}$. By
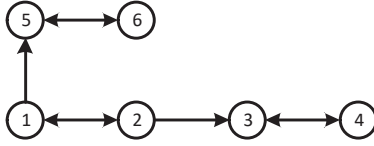
Fig. 2. A counterexample for monotonicity.

spreading T from node 6 (i.e., $S_T = \{6\}$), node 5 will reconsider its decision to adopt T, which results in $\Phi_R^* = \{1, 2\}$ and $\Phi_T^* = \{5, 6\}$. Such an addition of S_T decreases the spread of R. However, if set $S_T = \{4\}$, we have $\Phi_R^* = \{1, 2, 5, 6, 3\}$ and $\Phi_T^* = \{4\}$. Node 3 gets an additional influence from node 4 but adopts R based on its decision threshold. This implies that adding nodes to S_T may increase the spread of R.

2.4 Analysis of the Steady State

In this part, we first present a characterization of the steady state based on the notion of cohesiveness which was formalized in [1] for the LTM. We can see from the state transition graph from Fig. 1 that, each influenced node will become active (either R-active or T-active) eventually. Therefore, the influence stage is exactly equivalent to the activation process of the LTM. Then, if we define the total seed set $S = S_R \cup S_T$, the set of final adopters can be characterized based on the notion of cohesiveness.

Definition 2.1 *[1] A subset $X \subseteq V$ is called a* cohesive set *if for all $u \in X$ it holds:*

$$\sum_{v \in X \cap N_u^{in}} W(v, u) > 1 - \theta_u \tag{3}$$

$\diamond$

This states that a set $X \subseteq V$ is cohesive if for all $u \in X$, the sum of the influence coming from set X exceeds $1 - \theta_u$, which is equivalent to say that the sum of the influence that does not come from set X is below its influence threshold, i.e.,

$$\sum_{v \in V \setminus X \cap N_u^{in}} W(v, u) < \theta_u. \tag{4}$$

Given a cohesive set X, if no node in X belongs to the total seed set $S = S_R \cup S_T$, then all nodes in X will never become active (or influenced) at the following steps. This leads to the following characterization of the set of final adopters.

Lemma 2.1 *[1] Given a LT1DT $N = (G(V, E), W, \theta, \theta^R)$ and seed sets $S_R, S_T \subseteq V$, let $S = S_R \cup S_T$ and $M \subseteq V \setminus S$ be the* maximal cohesive set *contained in $V \setminus S$. The set of final adopters is:*

$$\Phi^* = V \setminus M \tag{5}$$

$\diamond$

Example 2.1 *Consider again the simple LT1DT shown in Fig. 2, where set $X = \{1, 2, 3, 4\}$ is a cohesive set. Assume that $S_R = \{1\}$ and $S_T = \{6\}$. The maximal cohesive set M contained in $V \setminus S$ is $M = \{3, 4\}$ and then $\Phi^* = \{1, 2, 5, 6\}$.* $\triangle$

The set of final adopters Φ^* can be computed by simulating the evolution. Here, Lemma 2.1 provides a direct way to compute Φ^* without necessity to determine the evolution of the network. We refer to [44] for the computation of Φ^* by solving an integer linear programming.

Based on the characterization of the steady state, we can extract an *active subgraph* $\hat{G}(\hat{V}, \hat{E})$ from the original graph $G(V, E)$ by keeping all nodes in Φ^* and their relationships such that $\hat{V} = \Phi^* = V \setminus M$

and $\hat{E} \subseteq \hat{V} \times \hat{V} \subseteq E$. In the active subgraph $\hat{G}$, all nodes are active but we cannot distinguish which is R-active or T-active except for the seed nodes without studying the evolution of the network. However, we can provide a lower bound for the T-adoption function $|\Phi_T^*|$ and an upper bound for the R-adoption function $|\Phi_R^*|$.

Definition 2.2 *A node $v \in \hat{V} \setminus S_T$ is said to be reachable from S_R in $\hat{G}$ if there exists at least a directed path from any node in S_R to v without passing through a node in S_T.* ◇

Next two results provide for nodes in $\hat{G}$ a sufficient condition for becoming T-active and a necessary condition for becoming R-active.

Theorem 2.1 *In the active subgraph $\hat{G}$, a node v is reachable from S_R in $\hat{G}$ if it is R-active in the steady state.*

PROOF. For any R-active node v, Eq. (2) holds in a certain step t, which indicates that there exists at least one of its in-neighbor node u that is R-active before t. Repeating this argument, we can find a directed path from S_R to v without passing through any node in S_T. Then by Definition 2.2, node v is reachable from S_R in $\hat{G}$.

Theorem 2.2 *In the active subgraph $\hat{G}$, a node $v \in \hat{V}$ is T-active if v is not reachable from S_R.*

PROOF. By Theorem 2.1, we can infer that any node that is not reachable from S_R in $\hat{G}$ cannot be R-active in the steady state, which means that it must be T-active since every node in $\hat{G}$ is active.
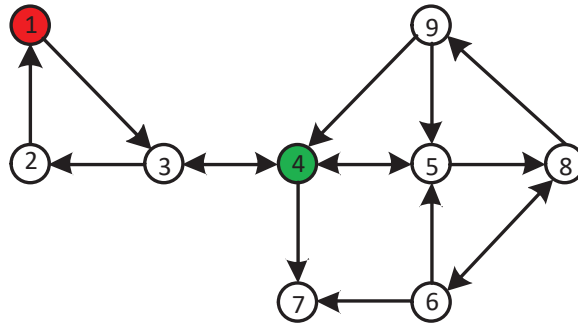
(a) A simple LT1DT
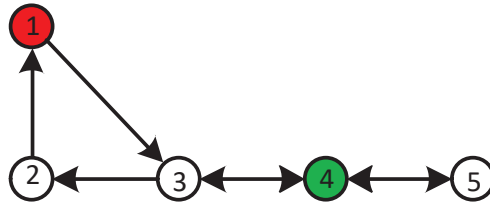
(b) Active subgraph

Fig. 3. Extraction for the active subgraph.

Example 2.2 *Consider the simple LT1DT network shown in Fig. 3a with 9 nodes whose influence thresholds are [0.6, 0.6, 0.5, 0.3, 0.2, 0.7, 0.8, 0.7, 0.6] and decision thresholds are [0.5, 0.5, 0.6, 0.3, 0.5, 0.5, 0.5, 0.5, 0.5] and the influence weight $W(u, v) = 1/|N_v^{in}|$. Assume that $S_R = \{1\}$ and $S_T = \{4\}$. The maximal cohesive set within $V \setminus (S_R \cup S_T)$ is $M = \{6, 7, 8, 9\}$. Then we have $\Phi^* = \{1, 2, 3, 4, 5\}$. The active subgraph $\hat{G}$ can be extracted as in Fig. 3b. In $\hat{G}$, node 5 is not reachable from S_R, therefore it is T-active. By simulating the evolution, we have $\Phi_R^* = \{1\}$ and $\Phi_T^* = \{2, 3, 4, 5\}$.* △

Theorems 2.1 and 2.2 provide an upper bound and a lower bound for $|\Phi_R^*|$ and $|\Phi_T^*|$, respectively.

Corollary 2.1 *Given a LT1DT $N = (G(V,E), W, \theta, \theta^R)$ with seed sets $S_R, S_T \subseteq V$ such that $S_R \cap S_T = \emptyset$, it holds $|\Phi_T^*| \geq |\bar{\mathcal{R}}_{S_R}(\hat{G})|$, where $\bar{\mathcal{R}}_{S_R}(\hat{G})$ denotes the set of nodes that are not reachable from S_R in the active subgraph $\hat{G}$.*

Corollary 2.2 *Given a LT1DT $N = (G(V,E), W, \theta, \theta^R)$ with seed sets $S_R, S_T \subseteq V$ such that $S_R \cap S_T = \emptyset$, it holds $|\Phi_R^*| \leq |\mathcal{R}_{S_R}(\hat{G})|$, where $\mathcal{R}_{S_R}(\hat{G})$ denotes the set of nodes that are reachable from S_R in the active subgraph $\hat{G}$.*

Example 2.3 *Consider again the LT1DT shown in Fig. 3a. By analyzing the reachability of each node in $\hat{G}$ shown in Fig. 3b, we have $\bar{\mathcal{R}}_{S_R}(\hat{G}) = \{4,5\}$ and $\mathcal{R}_{S_R}(\hat{G}) = \{1,2,3\}$, which indicates that the lower bound of $|\Phi_T^*|$ is 2 and the upper bound of $|\Phi_R^*|$ is 3.* $\triangle$

3 Problem definition

The LT1DT model is a competitive diffusion model that allows a state transition from R-active to T-active but not in the opposite direction. We believe that this model can be used to describe many different practical cases, e.g., product adoption, rumor spread, etc.

For example, in product adoption, items R and T can be interpreted as two similar products with low quality and high quality, respectively. A user adopting the low quality product may change its mind and pass to the high quality product when he compares them. However, it is hard to influence a user who uses a high quality product to adopt a low quality product. After studying a series of product choices, Macdonald and Sharp [28] found that consumers tend to select the high-quality brand among several totally unknown brands.

Similarly in the case of rumor spread, R represents the rumor and T represents the truth. An individual that adopts the rumor may reconsider his belief, while an individual that adopts the truth will become immune to the rumor [40]. We should mention here that in this case, the term rumor seeds, is not equivalent to rumor sources, which however means that the set of nodes has adopted the rumor when the rumor is detected.

In the following part, we focus on the application of rumor spread, a problem of rumor containment. Rumors may cause panic or lead people to wrong decisions, therefore it is vital to study how to control the rumor spread. Rather than cutting some connections in the network or censoring the opinions expressed by the individuals, we identify individuals to seed the truth so as to reduce people's vulnerabilities to rumor. Under the framework of the LT1DT, in this section we study the problem of minimizing rumor spread.

3.1 Minimizing rumor spread

The problem of rumor containment is to select a suitable set of nodes for spreading truth to minimize the number of nodes eventually accepting the rumor. We then formalize the problem as follows.

Problem 3.1 (MRS: Minimizing Rumor Spread) *Consider a LT1DT model represented by a network $N = (G(V,E), W, \theta, \theta^R)$ with rumor seed set S_R. Given an integer k, find in $V \setminus S_R$ a seed set for the truth $S_T = X$ of cardinality less than or equal to k such that the cardinality of R-final adopters is minimized, i.e.,*

$$min\ |\Phi_R^*(S_R, X)|$$
$$s.t.\ X \subseteq V \setminus S_R$$
$$|X| \leq k$$

9

The MRS problem has a combinatorial nature and finding the optimal solution requires an exhaustive search. In the following we show the hardness of the MRS problem.

Theorem 3.1 *The MRS problem is NP-hard under the Linear Threshold model with One Direction state Transition.*

PROOF. To prove the NP-hardness of the MRS problem, it is sufficient to show that the classical Maximum Coverage Problem (MCP) which is known to be NP-hard can be reduced to the MRS problem under the LT1DT. To do this, we employ a similar approach in [47] to construct an instance of MRS following the definition of MCP.

Problem 3.2 (Maximum Coverage Problem (MCP)) *Consider a set of q elements $O = \{o_1, o_2, \cdots, o_q\}$, and a collection of sets $\mathcal{X} = \{X_1, X_2, \cdots, X_p\}$ where X_i is a set of elements $o_j \in O$. The sets may have some elements in common. Given a positive constant l, find a subset $\mathcal{X}' \subset \mathcal{X}$ of cardinality less than or equal to l such that the number of elements $|\cup_{X_i \in \mathcal{X}'} X_i|$ covered in O is maximized, i.e.,*

$$max \; |\cup_{X_i \in \mathcal{X}'} X_i|$$
$$s.t. \; \mathcal{X}' \subset \mathcal{X}$$
$$|\mathcal{X}'| \leq l$$

The decision version of the MCP is to answer whether there exists a subset $\mathcal{X}' \subset \mathcal{X}$ of size l which can cover at least λ elements in O where $\lambda \leq q$ is a positive integer. The decision version of the MRS problem is to answer whether there exists a subset $S_T \subset V$ of size k such that the number of R-active nodes is at most ρ where ρ is a positive integer. In the following we show that the MCP problem can be reduced to the MRS problem. To prove the result, we show that for any instance of the MCP, an equivalent instance of the MRS can be constructed.
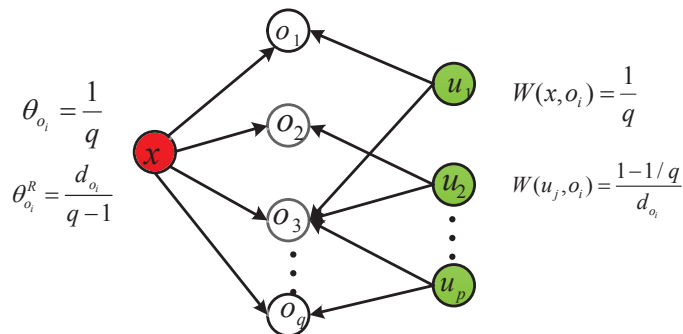

Fig. 4. Construction of an instance of the MRS.

Given an instance $\mathcal{I} = (O, \mathcal{X}, l, \lambda)$ of the MCP, we construct a corresponding instance $\mathcal{M} = (V, E, \theta, \theta^R, W, S_R, k, \rho)$ of the MRS as follows. A simple example shown in Fig. 4 is used to explain the construction. Beginning with 0 node, we add a node u_i for each subset $X_i \in \mathcal{X}$, a node o_j for each element in O, and a special node x. Then we connect x to each node o_j and add an edge (u_i, o_j) for each $o_j \in X_i$. Take node o_3 as an example. It has at least three incoming arcs from u_1, u_2, u_p which means that o_i exists in the corresponding subsets X_1, X_2, X_p at least. All nodes o_i are given the same influence threshold $\theta_{o_i} = 1/q$ and the decision threshold $\theta_{o_i}^R = \frac{d_{o_i}}{q-1}$ where the positive integer d_{o_i} is the number of subsets X_i to which element o_i belongs. We do not care about the thresholds of the source nodes u_i and x. The weights of arcs from node x to node o_i are set to $1/q$, i.e., $W(x, o_i) = 1/q$. The weights of arcs from u_j to o_i are homogeneous and set to $(1 - 1/q)/d_{o_i}$, i.e., $W(u_j, o_i) = (1 - 1/q)/d_{o_i}$. Then, for each o_i, the sum of the weights incoming from x and u_j is 1. Finally, set $S_R = \{x\}$, $k = l$, and $\rho = q - \lambda + 1$. We then show the equivalence between the instances $\mathcal{I}$ and $\mathcal{M}$.

10

Assume that $\mathcal{X}^*$ is a solution to the instance $\mathcal{I}$, i.e., $|\mathcal{X}^*| \leq l$ and it can cover at least λ elements in O. By our construction, we can select all the nodes u_j corresponding to subset $X_j \in \mathcal{X}^*$ as truth seed set S_T. Therefore, $|S_T| = k = l$. Since $\mathcal{X}^*$ can cover at least λ elements o_i in O, then from S_T there will be at least λ nodes in O believing the truth. This is because for any covered o_i, let $d_{o_i}^*$ be the number of subsets $X_j \in \mathcal{X}^*$ to which o_i belongs, then we have $1 \leq d_{o_i}^* \leq d_{o_i}$. Then, the total influence from both rumor and truth is $1/q + d_{o_i}^*(\frac{1-1/q}{d_{o_i}}) \geq \theta_{o_i} = 1/q$. The sum of weights from its R-active neighbors is $1/q$. Then, o_i will believe the truth since $\frac{1/q}{1/q+d_{o_i}^*(\frac{1-1/q}{d_{o_i}})} \leq \frac{1/q}{1/q+\frac{1-1/q}{d_{o_i}}} = \frac{d_{o_i}}{d_{o_i}+q-1} < \theta_{o_i}^R = \frac{d_{o_i}}{q-1}$. Therefore, the cardinality of T-final adopter set is at least $\lambda + k$ in the MRS problem and the cardinality of R-final adopter set is at most $\rho = q - \lambda + 1$.

Conversely, assume that S_T^* is a solution to the instance $\mathcal{M}$, i.e., $|S_T^*| = k$ such that the cardinality of R-final adopter set is at most ρ. This implies that the cardinality of T-final adopter set is at least $q - (\rho - 1) + k$. Then $\mathcal{X}^*$ can be a collection of subset X_i corresponding to those $u_i \in S_T^*$. Therefore, $\mathcal{X}^*$ has size k since $l = k$ according to the assumption and it can cover at least $q - \rho + 1$ elements in O.

4   Heuristic solutions for the MRS

Given that the MRS problem is NP-hard, we present some heuristic approaches to solve it. They can be classified into two categories, *unconstrained* and *constrained*, according to the search space for choosing truth seeds. We first present three different unconstrained heuristics and then give their constrained versions where we restrict the search space to be a subnetwork instead of the whole network.

4.1   Unconstrained approaches

The unconstrained approaches can select any node to be a truth seed in set $V \setminus S_R$ according to specific judgements.

4.1.1   *MinGreedy*

A *greedy* approach (also called *hill-climbing* approach) is a heuristic optimization paradigm that iteratively makes the locally optimal choice with the hope of finding a global optimal solution.

In the MinGreedy shown in Algorithm 1, at each iteration we add to the truth seed set S_T the node u which maximally decreases rumor spread (line 5). The node selection process stops when k nodes have been selected or when the addition of a node to S_T does not reduce the rumor spread (line 7). In each iteration, for each candidate node $w \in V \setminus S_R$ we need to compute the corresponding set of final adopters which may require to check every node if it is activated and this can be done in time $O(nd)$. Therefore, the time complexity of MinGreedy is $O(kdn^2)$.

4.1.2   *PageRank*

In network science, it is of interest to determine the relative importance of nodes which can be measured by centrality scores [6]. There are varieties of centrality scores, e.g., degree centrality, eigenvector centrality, pagerank centrality, etc. In this paper, we use pagerank centrality which is a well-known algorithm adopted by GOOGLE for ranking web pages. It is also widely used to solve influence maximization problems in social networks. We use the teleportation model [19] to compute pagerank values of all nodes in a network according to Eq. (6) where $\boldsymbol{p} \in (0,1)^n$ denotes the pagerank vector such that $\sum_{i=1}^{n} p_i = 1$, p_i denoting the pagerank value of node i. And $\hat{A}$ in Eq. (6) is the scaled adjacency matrix in which $\hat{A}(i,j) = A(i,j)/\mathbf{1}^T A(:,j)$ where A is the adjacency matrix of a given graph. In the model, the restart probability γ is set to 0.15.

$$\boldsymbol{p} = (1-\gamma)\hat{A}\boldsymbol{p} + \frac{\gamma}{n}\mathbf{1}, \tag{6}$$

In order to contain the rumor spread, PageRank chooses k nodes in $V \setminus S_R$ with highest pagerank values

---

Algorithm 1 MinGreedy

1: **Input:** A LT1DT network $N = (G(V,E), W, \theta, \theta^R)$, constant k, rumor seed set S_R, R-adoption function $\Phi_R^*(\cdot, \cdot)$.
2: **Output:** Truth seed set S_T.
3: Let $S_T \leftarrow \emptyset$.
4: **while** $|S_T| < k$ **do**
5:     $u = \underset{i \in V \setminus S_R}{argmin} |\Phi_R^*(S_R, S_T \cup \{i\})|$.
6:     $\Delta = |\Phi_R^*(S_R, S_T)| - |\Phi_R^*(S_R, S_T \cup \{u\})|$.
7:     **if** $\Delta > 0$ **then**
8:         $S_T \leftarrow S_T \cup \{u\}$.
9:     **else**
10:        break.
11:     **end if**
12: **end while**
13: Return S_T.

---

to serve as the truth seeds.

Since the scaled adjacency matrix $\hat{A}$ is a sparse matrix with average d non-zero elements in each row (so called average degree of the network), the multiplication of the matrix and a vector can be done in time $O(nd)$. Therefore, the computation of pagerank values requires time $O(n(d+1)I)$ where I is the number of iterations for p converging to a reasonable approximation. Selecting the first k nodes with highest pagerank values in $V \setminus S_R$ requires time $O(kn)$. Overall, the time complexity of PageRank is $O(In(d+1) + kn)$.

4.1.3 ContrId

PageRank, regardless of the diffusion dynamics of the considered diffusion model, solely depends on the network structure. Here, we present a novel heuristic ContrId (short for Contributors Identification) which, based on the dynamics of rumor spread, tries to identify which nodes contribute most (called *contributors*) to the rumor spread and interrupt the important paths through which the rumor will spread from the rumor seeds to the remaining network.

The following proposed way to measure a node's contribution to the rumor spread is motivated by observations under the LTM model. The spread of rumor evolves in discrete steps till no more nodes can be activated. A node activated at time step t immediately spreads its influence to its out-neighbors which may be activated in later steps. Therefore, the contribution of node v which is activated at time t_v can be defined as the number of its out-neighbors that are activated at any time $t > t_v$.

We first simulate the rumor spreads from S_R with $S_T = \emptyset$ and record the set of nodes activated at every step from step 1 to step T_s at which the network enters a steady state, i.e., ϕ_t^R for all $t = 1, 2, \cdots, T_s$. Let L denote the length of the longest simple path in the network, we then have $T_s \leq L$ since the diffusion process requires at most L time steps. Now, for each node $v \in \phi_t^R$, its contribution denoted as $Ctr(v)$ can be formalized as

$$Ctr(v) = |(\cup_{i=t+1}^{T_s} \phi_i^R) \cap N_v^{out}|. \tag{7}$$

For the completeness of the definition on node's contribution, we set the contribution of all inactive nodes as 0.

ContrId selects the first k contributors ranked by their contributions to rumor spread $Ctr(v)$ and this can be done in time $O(kn)$. Computation time $O(nd)$ is required to simulate the rumor spread from S_R. Therefore, the time complexity of ContrId is $O(n(d+k))$.

4.2 Constrained approaches

In the previous part, we present three heuristic approaches that select k truth seeds in $V \setminus S_R$. In this part, we also consider their constrained versions (called ProxMinGreedy, ProxPageRank, and ProxContrId) by restricting the search space to the proximity of the rumor seed set. We call this space S_R-*proximity space* and denote it as $\mathcal{N}^{out}(S_R) = (\cup_{v \in S_R} N_v^{out}) \setminus S_R$. The idea is that of blocking the influential paths from the rumor seeds to the remaining network. In other words, this measure may reduce the number of nodes which are reachable from S_R in corresponding active subgraphs. For instance, consider Example 2.3, if $S_T = \{3\}$ which is the out-neighbor set of $S_R = \{1\}$ then $\mathcal{R}_{S_R}(\hat{G}) = \{1\}$ instead of $\{1,2,3\}$ (when $S_T = \{4\}$). We will see from experimental results presented in Section 5 that the constrained versions perform better than the unconstrained approaches in terms of both computation speed and rumor containment.

4.2.1 ProxMinGreedy

ProxMinGreedy iteratively selects k nodes in $\mathcal{N}^{out}(S_R)$ with the minimal gain of rumor spread to serve as the truth seeds. We use Eq. (8) to replace line 5 in Algorithm 1, i.e.,

$$u = \underset{i \in \mathcal{N}^{out}(S_R)}{argmin} \; |\Phi_R^*(S_R, S_T \cup \{i\})|. \tag{8}$$

The resulting ProxMinGreedy has lower time complexity than MinGreedy. The time complexity in average should be $O(k|S_R|d^2 n)$ where $|S_R|$ denotes the size of rumor seed set and d denotes the average degree of the network, therefore $|S_R|d$ is the size of the S_R-proximity space.

4.2.2 ProxPageRank

ProxPageRank selects k nodes in $\mathcal{N}^{out}(S_R)$ with highest pagerank values to serve as the truth seeds, which results in time complexity $O(Rn(d+1) + k|S_R|d)$.

4.2.3 ProxContrId

ProxContrId selects k nodes in $\mathcal{N}^{out}(S_R)$ with highest contributions to serve as the truth seeds, which results in time complexity $O(nd + k|S_R|d)$.

Table 1
Comparison on the time complexity

Type	Approach	Complexity
Unconstrained	MinGreedy	$O(kdn^2)$
	PageRank	$O(In(d+1) + kn)$
	ContrId	$O(n(d+k))$
Constrained	ProxMinGreedy	$O(k|S_R|d^2 n)$
	ProxPageRank	$O(Rn(d+1) + k|S_R|d)$
	ProxContrId	$O(nd + k|S_R|d)$

The constrained approaches have lower time complexity than their unconstrained version thanks to the smaller search space. The time complexity of all the above proposed approaches is summarized in Table 1. Besides MinGreedy, the others have the same level of time complexity which grows linearly with n. Introducing the proximity effect can reduce the time complexity of greedy approach from $O(n^2)$ to $O(n)$. The idea of restricting the search space to S_R-proximity space can be also applied to restricting the search space to any possible subset of nodes, e.g., the set of nodes that can be easily convinced to adopt the truth.

4.3 Analysis of monotonicity guarantee

In Section 2.3, we have shown that, given a rumor seed set S_R, the R-adoption function may increase as nodes are added to the truth seed set S_T. In this subsection, we show that the R-adoption function is guaranteed to be non-increasing by some particular truth seed set selection strategies.

Theorem 4.1 *For any given LT1DT $N = (G(V, E), W, \theta, \theta^R)$ and seed set $S_R \subseteq V$, the R-adoption function $|\Phi_R^*(S_R, S_T)|$ is non-increasing w.r.t S_T if S_T is selected by MinGreedy or ProxMinGreedy.*

PROOF. This result follows from the constraint shown in line 7 of Algorithm 1 which ensures that a candidate seed can only be added to S_T if adding this seed decreases the rumor spread.

Theorem 4.2 *For any given LT1DT $N = (G(V, E), W, \theta, \theta^R)$ and seed set $S_R \subseteq V$, the R-adoption function $|\Phi_R^*(S_R, S_T)|$ is non-increasing w.r.t S_T if S_T is selected by ContrId or ProxContrId.*

PROOF. What we need to prove is that for any two seed sets S_T^1 and S_T^2 selected by ContrId (or ProxContrId) such that $S_T^1 \subseteq S_T^2 \subseteq V \setminus S_R$, it holds that $|\Phi_R^*(S_R, S_T^2)| \leq |\Phi_R^*(S_R, S_T^1)|$. Assume that $\hat{G}^0(\hat{V}^0, \hat{E}^0)$ denotes the active subgraph extracted from $G(V, E)$ given S_R and $S_T = \emptyset$. Since ContrId (or ProxContrId) chooses truth seeds in $\hat{G}^0$, the addition of a seed set S_T selected by ContrId will not change the active subgraph, i.e., $\Phi^*(S_R, \emptyset) = \Phi^*(S_R, S_T)$. Therefore, it is equivalent to prove that for any $S_T^1 \subseteq S_T^2 \subseteq V \setminus S_R$, it holds that $|\Phi_T^*(S_R, S_T^1)| \leq |\Phi_T^*(S_R, S_T^2)|$, which is obviously true since $|\Phi_T^*(S_R, S_T)|$ is non-decreasing w.r.t. S_T. This completes the proof.

However, such monotonicity guarantee cannot be achieved by PageRank and ProxPageRank. This will be confirmed by the experimental results in Section 5.

5 Evaluations

5.1 Experimental set up

We performed a series of experiments on two synthetic datasets and two real-world datasets to evaluate the effectiveness of the heuristics solving the MRS problem under the proposed LT1DT. The synthetic datasets are the following.

- *Scale-free network*: A scale-free network is a network with power-law degree distribution. A few examples of networks claimed to be scale-free include collaboration networks, protein-protein interaction networks, etc. We adopt the Barabási-Albert model [2] to generate a scale-free network with 500 nodes.
- *Small-world network*: A small-world network is a network where most nodes are not linked with others, but the neighbors of a certain node are likely to be linked with each other. Besides, most nodes in the network can reach any other node through a few steps or hops. Many real-world networks exhibit small-world properties, e.g., electronic power grids, neural networks, social influence networks, etc. A small-world network with 500 nodes is generated by adopting the Watts-Strogatz model [39] and setting the mean degree of a regular lattice to 4 and the rewiring probability $\beta = 0.2$.

The real-world datasets are the following.

- NetScience: NetScience is the co-authorship of researchers who work on network theory and experiments, compiled by Newman [30].
- US-power: US-power is the topology of the Western States power grid of USA, compiled by Watts and Strogatz [39].

We summarize in Table 2 the statistics of the benchmarks. The networks are all undirected, and thus we considered the arcs in both directions. For achieving a good estimate of the effectiveness of the tested approaches, we run each experiment on 100 different randomly generated LT1DT networks with the same network structure $G(V, E)$ and influence weights W but different influence and decision threshold values θ_u and θ_u^R that are selected in $(0, 1]$ uniformly at random. We set the influence weight $W(u, v) = 1/|N_v^{in}|$ for all $u \in N_v^{in}$ and $v \in V$. Then we take the average cardinality of Φ_R^* and Φ_T^* over the 100 runs.

Table 2
Summary of the datasets

Dataset	Number of nodes	Number of edges	Average degree	Clustering coefficient
Scale-free	500	498	2.0	0
Small-world	500	1,000	4	0.29
NetScience	1,589	2,742	3.5	0.88
US-power	4,941	6,594	2.7	0.11

In addition, the set of original spreaders for the rumor should be given. On all networks, we set $|S_R| = 3$ and let $|S_T|$ range from 1 to 10. To determine the rumor seeds, we adopt the following two different selection strategies.

- **Random**: rumor spreads from three nodes selected randomly.
- **Degree**: rumor spreads from three nodes with highest degrees.

All experiments are performed on a 2.5 GHz Intel(R) Core(TM) i5-6300U CPU and 8GB memory and all code is written in C++.

5.2 Experimental results and discussions

We evaluate the performance of the approaches presented in Section 4 in terms of rumor containment, computation time and scalability. Note that since the MinGreedy is time-consuming in large networks, we do not test it on network US-power.
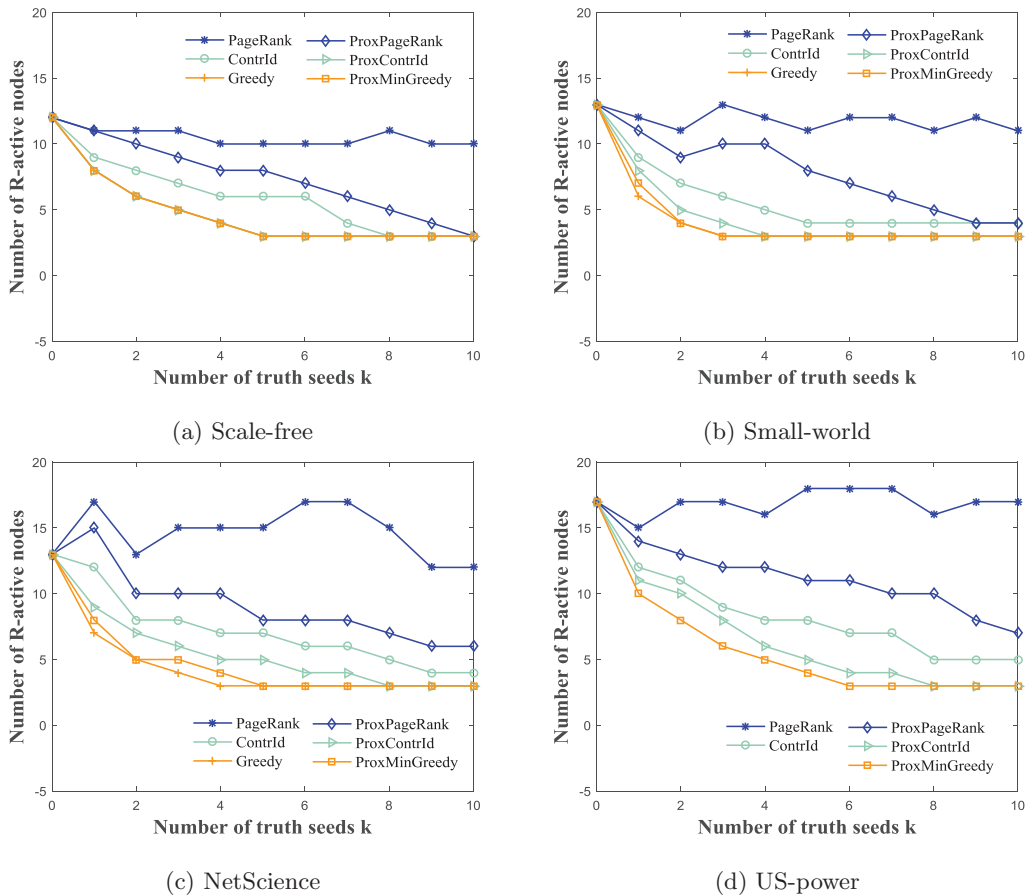

(a) Scale-free

(b) Small-world

(c) NetScience

(d) US-power

Fig. 5. Rumor spread in different networks (S_R is selected by Random).

15

(a) Scale-free   (b) Small-world
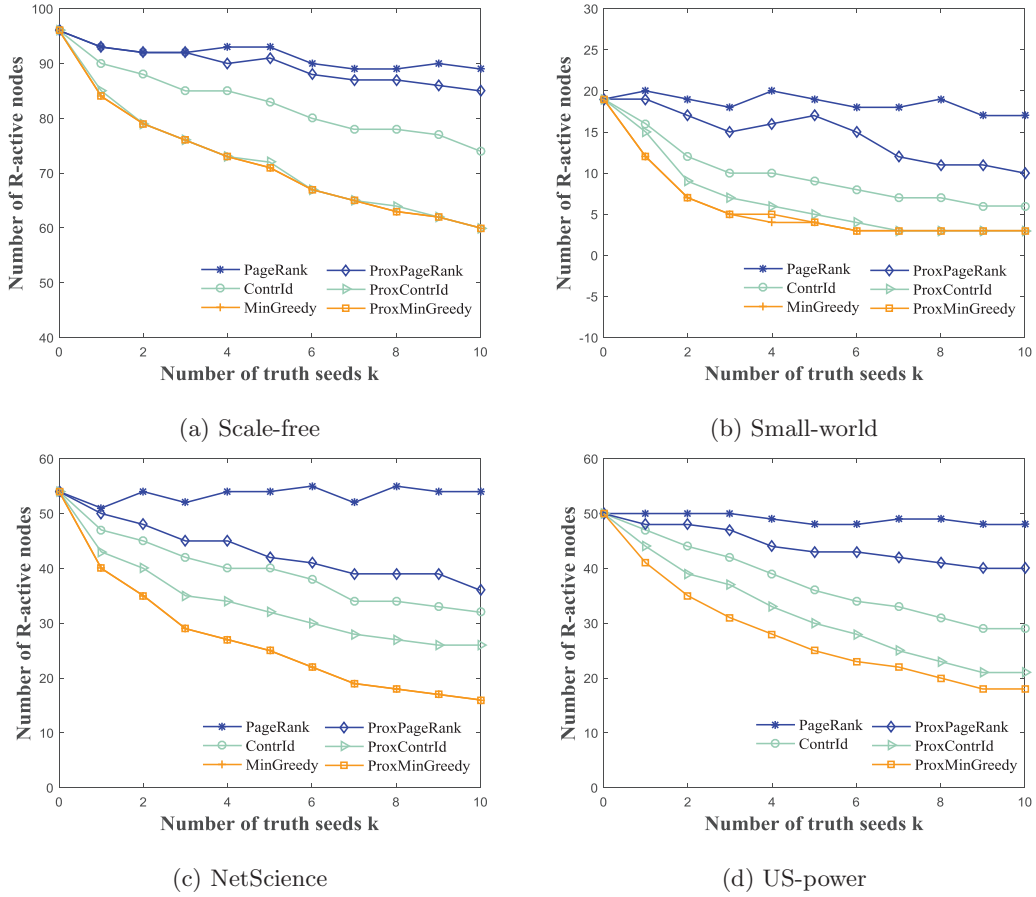
(c) NetScience   (d) US-power

Fig. 6. Rumor spread in different networks (S_R is selected by Degree).

5.2.1   Analysis of rumor containment

Table 3

Comparisons of all approaches on the performance of rumor containment

Outperformance Dataset Compared approaches	Scale-free	Small-world	NetScience	US-power
ProxPageRank v.s. PageRank	29% / 2%	37% / 21%	41% / 21.7%	36% / 10.8%
ProxContrId v.s. ContrId	22% / 12%	26% / 16.5%	28% / 11.9%	30% / 12.6%
MinGreedy v.s. ProxMinGreedy	0% / 0%	1.4% / 2%	5.7% / 0%	- / -
ContrId v.s. PageRank	47% / 10%	56% / 47%	55% / 28.9%	54% / 25.2%
ProxContrId v.s. ProxPageRank	40% / 19.6%	46% / 42.5%	45% / 19.1%	50% / 27%
ProxMinGreedy v.s. ProxContrId	0 / 0	5.7% / 5%	16.6% / 13.5%	14% / 8%

(The outperformance on the left of / is for the case when S_R is selected by Random, while the outperformance on the right is for the case when S_R is selected by Degree.)

The rumor spread, truth spread on the four networks under the proposed heuristic approaches are shown in Figs. 5, 6 and Figs. 7, 8. From the rumor spread in Figs. 5 and 6 we have the following observations.

Proximity effect works well: ProxPageRank, and ProxContrId outperform their unconstrained versions in terms of rumor containment, which well reflects the good property of the proximity effect.

• When S_R is selected by Random, in Scale-free (resp., Small-world, NetScience, US-power), ProxPageR-

16

(a) Scale-free   (b) Small-world

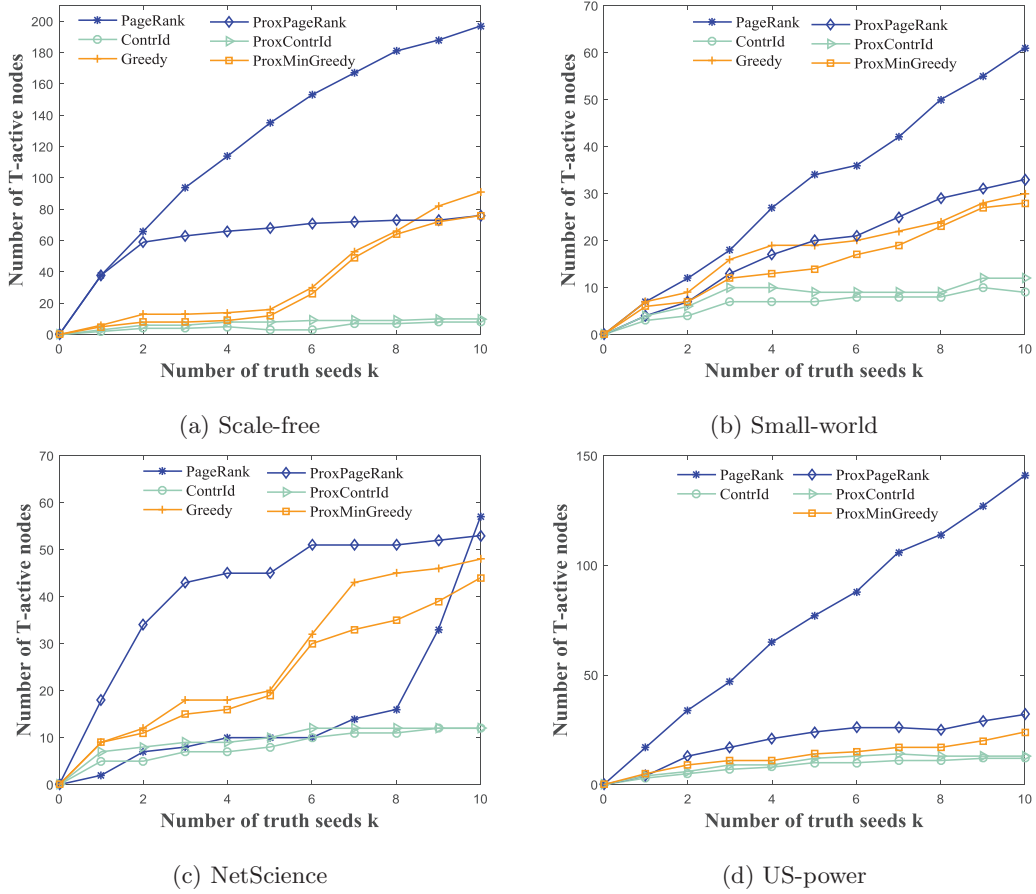(c) NetScience   (d) US-power

Fig. 7. Truth spread in different networks (S_R is selected by Random).

ank outperforms PageRank by 29% (resp., 37%, 41%, 36%) and ProxContrId outperforms ContrId by 22% (resp., 26%, 28%, 30%).

- When S_R is selected by Degree, in Scale-free (resp., Small-world, NetScience, US-power), ProxPageRank outperforms PageRank by 2% (resp., 21%, 21.7%, 10.8%) and ProxContrId outperforms ContrId by 12% (resp., 16.5%, 11.9%, 12.6%).

However, the MinGreedy and ProxMinGreedy almost perform the same. We can explain this phenomenon by comparing the truth seeds placement between MinGreedy and ProxMinGreedy. Let us take network Scale-free as an example on selecting 10 truth seeds, set all $\theta_v = 0.5$, $\theta_v^R = 0.3$, and let S_R be selected by Degree. By ProxMinGreedy, we have $S_T = \{85, 319, 5, 9, 16, 54, 87, 95, 100, 147\}$ where the number is node's label. Then by MinGreedy, we have $S_T = \{85, 319, 5, 9, 16, 54, 87, 95, 100, 147\}$ which is the same as selected by ProxMinGreedy. The result shows that MinGreedy prefers to locate the truth seeds at the S_R-proximity space as well.

Diffusion dynamics based approach performs better than centrality based approach: PageRank performs not as well as ContrId and MinGreedy and even sometimes does not help to contain the rumor spread. This fact can be seen from up-and-down lines for PageRank in Figs. 5 and 6, which reflects the non-monotonicity of the LT1DT model. This is because PageRank directly exploits the network structure and ignores the location of the rumor seeds. The nodes with high pagerank values sometimes are far away from the rumor seeds and then rumor and truth spread in different parts of the network.

- When S_R is selected by Random, ContrId performs better than PageRank, i.e., 47% (resp., 56%, 55%, 54%) in Scale-free (resp., Small-world, NetScience, US-power), since it tries to target the nodes that play important role in rumor spreading. ProxContrId outperforms ProxPageRank by 40% (resp.,
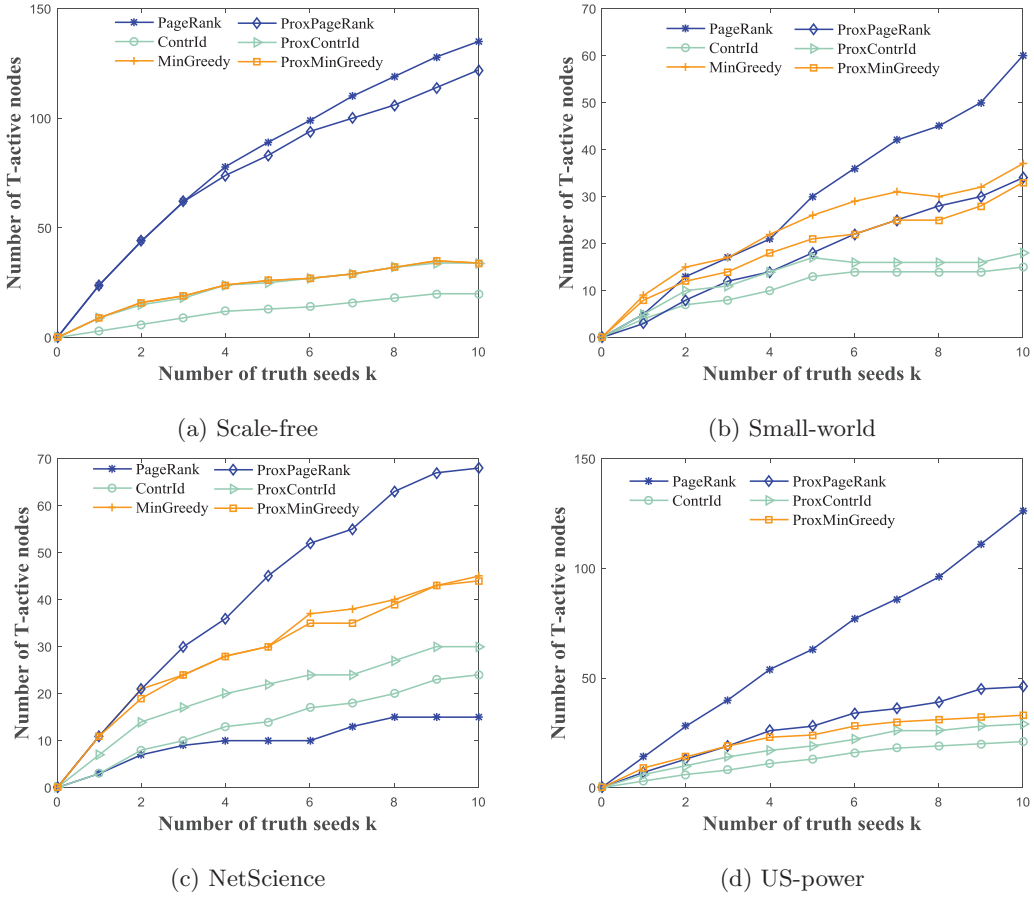
(a) Scale-free

(b) Small-world

(c) NetScience

(d) US-power

Fig. 8. Truth spread in different networks (S_R is selected by Degree).

46%, 45%, 50%) in Scale-free (resp., Small-world, NetScience, US-power). In Scale-free, ProxContrId performs as well as ProxMinGreedy. In Small-world (resp., NetScience, US-power) ProxMinGreedy outperforms ProxContrId 5.7% (resp., 16.6%, 14%).

- When S_R is selected by Degree, ContrId performs better than PageRank, i.e., 10% (resp., 47%, 28.9%, 25.2%) in Scale-free (resp., Small-world, NetScience, US-power). ProxContrId outperforms ProxPageRank by 19.6% (resp., 42.5%, 19.1%, 27%) in Scale-free (resp., Small-world, NetScience, US-power). In Scale-free, ProxContrId performs as well as ProxMinGreedy. In Small-world (resp., NetScience, US-power) ProxMinGreedy outperforms ProxContrId 5% (resp., 13.5%, 8%).

The comparisons of the approaches on the performance of rumor containment are summarized in Table 3.

From Figs. 5 and 6, it can also be observed that it is much harder to contain rumor when the rumor spreads from nodes with large degrees compared with rumor spreading from nodes randomly selected. As we can see in Fig. 5a, spreading truth from 5 nodes selected by ProxContrId and ProxMinGreedy can completely contain the rumor spread. However, in Fig. 6a, the rumor spread can only be reduced by about 37.5% spreading truth from 10 nodes selected by ProxContrId and ProxMinGreedy with the best performance.

From the truth spread in Figs. 7 and 8, spreading truth from the nodes selected by PageRank normally produces the largest set of nodes that adopt the truth, which reflects the fact that PageRank performs well in solving influence maximization problem. Note, however, that the goal of our study is not truth maximization but rumor minimization. We can also observe that information can spread further in networks with lower clustering coefficient, which is consistent with the results in [1].

5.2.2 Analysis of running time and scalability

Table 4

Average running time over 100 runs for selecting 10 truth seeds

Time(sec)　　　Dataset Approach	Scale-free	Small-world	NetScience	US-power
PageRank	0.012 / 0.03	0.006 / 0.015	0.13 / 0.25	3.0 / 2.2
ProxPageRank	0.014 / 0.028	0.006 / 0.015	0.14 / 0.21	3.0 / 2.23
ContrId	0.0002 / 0.003	0.0005 / 0.002	0.001 / 0.015	0.017 / 0.05
ProxContrId	0.0002 / 0.003	0.0005 / 0.002	0.001 / 0.015	0.017 / 0.05
MinGreedy	1.5 / 6.2	3.2 / 6.6	17 / 200	- / -
ProxMinGreedy	0.012 / 1.1	0.034 / 0.17	0.18 / 5.1	1.2 / 11

(The time on the left of / is for the case when S_R is selected by Random, while the time on the right is for the case when S_R is selected by Degree.)

The average running time over 100 runs for all the heuristics are illustrated in Table 4. Selecting truth seeds in S_R-proximity space obviously speeds up the seed location process, for instance, ProxMinGreedy runs 40 times faster than MinGreedy in NetScience when the rumor seed set is selected by Degree. We believe that the gap must increase with n. In terms of rumor containment, the ProxMinGreedy outperforms the ProxContrId a bit in some networks, but it is three orders of magnitude slower than the ProxContrId.

Finally, an experimental study on synthetic small-world networks with different number of nodes is carried out to test the scalability of the constrained approaches which outperform their unconstrained versions in terms of both the computation speed and the effectiveness of rumor containment. The networks are generated by Watts-Strogatz model [39] with the number of nodes from 0.5K to 4K where the mean degree and rewiring probability are set to 4 and 0.2, respectively. We select 3 rumor seeds selected by Degree in every network and choose 10 nodes as the truth seeds to contain the rumor spread. We average the running time over 100 runs with different threshold values. The result is shown in Fig. 9, which suggests that the ProxContrId are two and three orders of magnitude faster than the ProxPageRank and ProxMinGreedy.
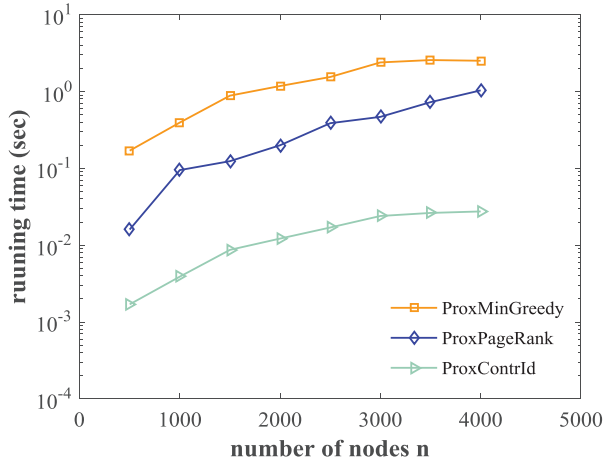

Fig. 9. Scalability of the heuristic algorithms.

6 Conclusion and future work

We have proposed a competitive and generalized version of the simple Linear Threshold model which captures practical situations in which rumor and truth compete in a same network. Based on the notion

of cohesiveness, we have provided a lower bound for the *T*-adoption function and an upper bound for the *R*-adoption function. We have addressed the problem of minimizing rumor spread (MRS) which is proved to be NP-hard under our generalized model. We have also presented three different heuristics and defined their constrained versions to highlight the proximity effect for solving the MRS problem. The novel heuristic approaches ContrId and ProxContrId can guarantee the monotonicity for the LT1DT. Finally, simulations were carried out under four different networks. Experimental results have shown that diffusion dynamics based approach (ContrId) outperforms centrality based approach (PageRank) with the same level of computation complexity which grows linearly with the number of nodes n, and therefore can scale well to large networks. By introducing the proximity effect, ContrId performs as well as MinGreedy but runs much faster than MinGreedy.

Among future directions for maximizing rumor containment we mention the problem of containing the rumor spread in finite time. This is motivated by the fact that the propagation of influence from one person to his friends may have a certain time delay due to the heterogeneity in human activities and the influence between individuals may also decay with time passing. Another interesting topic or trend is to analyze diffusion dynamics in signed social networks, i.e., networks where arcs are associated positive or negative signs. Social individuals are likely to make the same decision with their friends (connected with positive arcs) and make the opposite decision with their foes (connected with negative arcs).

References

[1] Acemoglu, D., Ozdaglar, A., & Yildiz, E. (2011). Diffusion of innovations in social networks. *In Proc. of the 50th IEEE Conf. on Decision and Control and European Control Conf.*, Orlando, USA (pp. 2329–2334).

[2] Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509–512.

[3] Bharathi, S., Kempe, D., & Salek, M. (2007). Competitive influence maximization in social networks. In *Proc. of the Int. Workshop on Web and Internet Economics*, San Diego, USA (pp. 306–311).

[4] Borodin, A., Filmus, Y., & Oren, J. (2010). Threshold models for competitive influence in social networks. In *Proc. of the Int. Workshop on Internet and Network Economics*, Stanford, USA (pp. 539–550).

[5] Budak, C., Agrawal, D., & El Abbadi, A. (2011). Limiting the spread of misinformation in social networks. In *Proc. of the 20th Int. Conf. on World Wide Web*, Hyderabad, India (pp. 665–674).

[6] Bullo, F. (2016). Lectures on network systems. *Online at http://motion. me. ucsb. edu/book-lns*, with contributions by J. Cortes, F. Dorfler and S. Martinez.

[7] Banerjee, S., Jenamani, M., & Pratihar, D. K. (2019). ComBIM: A community-based solution approach for the Budgeted Influence Maximization Problem. *Expert Systems with Applications*, 125, 1–13.

[8] Chen, W., Wang, Y., & Yang, S. (2009). Efficient influence maximization in social networks. In *Proc. of the 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, Paris, France (pp. 199-208).

[9] Chen, W., Yuan, Y., & Zhang, L. (2010). Scalable influence maximization in social networks under the linear threshold model. In *Proc. of IEEE 10th Int. Conf.on Data Mining*, Sydney, Australia (pp. 88–97).

[10] Etesami, S. R., & Basar, T. (2016). Complexity of equilibrium in competitive diffusion games on social networks. *Automatica*, 68, 100–110.

[11] Fazeli, A., Ajorlou, A., & Jadbabaie, A. (2014). Optimal budget allocation in social networks: Quality or seeding? In *Proc. of the 53rd IEEE Conf. on Decision and Control*, Los Angles, USA (pp. 4455–4460).

[12] Goyal, A., Bonchi, F., & Lakshmanan, L. V. (2010). Learning influence probabilities in social networks. In *Proc. of the third ACM Int. Conf. on Web Search and Data Mining*, New York, USA (pp. 241–250).

[13] Goyal, A., Lu, W., & Lakshmanan, L. V. (2011). Simpath: An efficient algorithm for influence

maximization under the linear threshold model. In *Proc. of IEEE 10th Int. Conf.on Data Mining*, Vancouver, Canada (pp. 211–220).

[14] Ghaderi, J., & Srikant, R. (2014). Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. *Automatica*, 50(12), 3209–3215.

[15] Gong, M., Yan, J., Shen, B., Ma, L., & Cai, Q. (2016). Influence maximization in social networks based on discrete particle swarm optimization. *Information Sciences*, 367, 600–614.

[16] He, X., Song, G., Chen, W., & Jiang, Q. (2012). Influence blocking maximization in social networks under the competitive linear threshold model. In *Proc. of the 2012 SIAM Int. Conf. on Data Mining*, Anahelm, USA (pp. 463–474).

[17] He, Z., Cai, Z., & Wang, X. (2015). Modeling propagation dynamics and developing optimized countermeasures for rumor spreading in online social networks. In *Proc. of 2015 IEEE 35th Int. Conf. on Distributed Computing Systems, Columbus, USA (pp. 205–214)*.

[18] Hosseini-Pozveh, M., Zamanifar, K., & Naghsh-Nilchi, A. R. (2019). Assessing information diffusion models for influence maximization in signed social networks. *Expert Systems with Applications*, 119, 476–490.

[19] Ishii, H., & Tempo, R. (2014). The PageRank problem, multiagent consensus, and web aggregation: A systems and control viewpoint. *IEEE Control Systems*, 34(3), 34–53.

[20] Kempe, D., Kleinberg, J., & Tardos, E. (2003). Maximizing the spread of influence through a social network. In *Proc. of the 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data mining*, Washington, USA (pp. 137–146).

[21] Kimura, M., Saito, K., & Motoda, H. (2009). Blocking links to minimize contamination spread in a social network. *ACM Transactions on Knowledge Discovery from Data*, 3(2), Art. 9.

[22] Khalil, E., Dilkina, B., & Song, L. (2013). CuttingEdge: influence minimization in networks. In *Proc. of Workshop on Frontiers of Network Analysis: Methods, Models, and Applications at NIPS*.

[23] Ko, Y. Y., Cho, K. J., & Kim, S. W. (2018). Efficient and effective influence maximization in social networks: a hybrid-approach. *Information Sciences*, 465, 144–161.

[24] Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., & Glance, N. (2007). Cost-effective outbreak detection in networks. In *Proc. of the 13th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, San Jose, USA (pp. 420-429).

[25] Liu, X., Li, M., Li, S., Peng, S., Liao, X., & Lu, X. (2014). IMGPU: GPU-accelerated influence maximization in large-scale social networks. *IEEE Transactions on Parallel and Distributed Systems*, 25(1), 136–145.

[26] Li, C. T., Lin, Y. J., & Yeh, M. Y. (2018). Forecasting participants of information diffusion on social networks with its applications. *Information Sciences*, 422, 432–446.

[27] Liu, Y., Liu, A., Liu, X., & Huang, X. (2019). A statistical approach to participant selection in location-based social networks for offline event marketing. *Information Sciences*, 480, 90–108.

[28] Macdonald, E. K., & Sharp, B. M. (2000). Brand awareness effects on consumer decision making for a common, repeat purchase product: A replication. *Journal of business research*, 48(1), 5-15.

[29] Mei, W., & Bullo, F. (2014). Modeling and analysis of competitive propagation with social conversion. In *Proc. of the 53rd IEEE Conf. on Decision and Control*, Los Angles, USA (pp. 6203–6208).

[30] Newman, M. E. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3), Id: 036104.

[31] Narayanam, R., & Narahari, Y. (2011). A shapley value-based approach to discover influential nodes in social networks. *IEEE Transactions on Automation Science and Engineering*, 8(1), 130–147.

[32] Peter, F. (2013). 'bogus' AP tweet about explosion at the white house wipes billions off US markets. *The Telegraph, Finance/Market, Washington*, Apr. 2013.

[33] Proskurnikov, A. V., Matveev, A. S., & Cao, M. (2016). Opinion dynamics in social networks with hostile camps: Consensus vs. polarization. *IEEE Transactions on Automatic Control*, 61(6), 1524–1536.

[34] Rosa, D., & Giua, A. (2013a). On the spread of innovation in social networks. In *Proc. of the 4th IFAC Workshop on Distributed Estimation and Control in Networked Systems*, Rhine-Moselle-Hall, Germany (pp. 322–327).

[35] Riquelme, F., Gonzalez-Cantergiani, P., Molinero, X., & Serna, M. (2018). Centrality measure in

social networks based on linear threshold model. *Knowledge-Based Systems*, 140, 92–102.

[36] Song, G., Zhou, X., Wang, Y., & Xie, K. (2015). Influence maximization on large-scale mobile social network: a divide-and-conquer method. *IEEE Transactions on Parallel and Distributed Systems*, 26(5), 1379–1392.

[37] Saito, K., Kimura, M., Ohara, K., & Motoda, H. (2016). Super mediator-A new centrality measure of node importance for information diffusion over social network. *Information Sciences*, 329, 985–1000.

[38] Urena, R., Kou, G., Dong, Y., Chiclana, F., & Herrera-Viedma, E. (2019). A review on trust propagation and opinion dynamics in social networks and group decision making frameworks. *Information Sciences*, 478, 461–475.

[39] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684), 440–442.

[40] Wen, S., Jiang, J., Xiang, Y., Yu, S., Zhou, W., & Jia, W. (2014). To shut them up or to clarify: Restraining the spread of rumors in online social networks. *IEEE Transactions on Parallel and Distributed Systems*, 25(12), 3306–3316.

[41] Yang, L., Giua, A., & Li, Z. (2017). Minimizing the influence propagation in social networks for linear threshold models. In *Proc. of the 20th World Congress of the International Federation of Automatic Control*, Toulouse, France (pp. 14465–14470).

[42] Yang, Y., Xu, Y., Wang, E., Lou, K., & Luan, D. (2018). Exploring influence maximization in online and offline double-layer propagation scheme. *Information Sciences*, 450, 182–199.

[43] Yang, L., Li, Z., & Giua, A. (2019). Rumor Containment by Spreading Correct Information in Social Networks. In *Proc. of the 2019 American Control Conference*, Philadelphia, USA.

[44] Yang, L., Li, Z., & Giua, A. (2019). Influence minimization in linear threshold networks. *Automatica*, 100, 10–16.

[45] Yang, W., Brenner, L., & Giua, A. (2019). Influence maximization in independent cascade networks based on activation probability computation. *IEEE Access*, 7, 13745–13757.

[46] Zhou, F., Jiao, J. R., & Lei, B. (2015). A linear threshold-hurdle model for product adoption prediction incorporating social network effects. *Information Sciences*, 307, 95–109.

[47] Zhang, H., Zhang, H., Li, X., & Thai, M. T. (2015). Limiting the spread of misinformation while effectively raising awareness in social networks. In *Proc. of the Int. Conf. on Computational Social Networks*, Beijing, China (pp. 35–47).

[48] Zhu, L., Zhao, H., & Wang, H. (2016). Complex dynamic behavior of a rumor propagation model with spatial-temporal diffusion terms. *Information Sciences*, 349, 119–136.

[49] Zhao, Y., Kou, G., Peng, Y., & Chen, Y. (2018). Understanding influence power of opinion leaders in e-commerce networks: An opinion dynamics theory perspective. *Information Sciences*, 426, 131–147.