



Research article

Containment of rumor spread by selecting immune nodes in social networks

Hong Wu¹, Zhijian Zhang^{2,*}, Yabo Fang^{2,3}, Shaotang Zhang¹, Zuo Jiang⁴, Jian Huang¹ and Ping Li¹

¹ School of Information Engineering, Qujing Normal University, Qujing 655000, China

² Faculty of Science, Kunming University of Science and Technology, Kunming 650031, China

³ School of Mathematics, Jilin University, Changchun 130012, China

⁴ School of Mathematics & Computer Science, Yunnan Minzu University, Kunming 650031, China

* **Correspondence:** Email: zhijian@kust.edu.cn.

Abstract: With the popularity of online social network these have become important platforms for the spread of information. This not only includes correct and useful information, but also false information, and even rumors which could result in panic. Therefore, the containment of rumor spread in social networks is important. In this paper, we propose an effective method that involves selecting a set of nodes in (k, η) -cores and immunize these nodes for rumor containment. First, we study rumor influence propagation in social networks under the extended Independent Cascade (EIC) model, an extension of the classic Independent Cascade (IC) model. Then, we decompose a social network into subgraphs via core decomposition of uncertain graphs and compute the number of immune nodes in each subgraph. Further we greedily select nodes with the Maximum Marginal Covering Neighbors Set as immune nodes. Finally, we conduct experiments using real-world datasets to evaluate our method. Experimental results show the effectiveness of our method.

Keywords: social networks; containment of rumor spread; (k, η) cores decomposition; extended independent cascade model; greedy algorithm

1. Introduction

With the popularity of online social networks, such as Facebook, Twitter and Wechat, more and more people are being involved. According to the statics, there have been 2.603 billion and 1.203

billion monthly active users on WeChat and Facebook, respectively [1,2]. All kinds of information disseminate through these social networks platforms. This information includes a large number of correct information and positive information, and this kind of information has great value to our life. For example, during the early state of the outbreak of 2019 novel coronavirus disease (COVID-19), the media reported that frequent hand washing, wearing masks and even quarantining themselves at home to avoid contacting with and being infected by others, and these information change people's behavior and mitigate the spread of COVID-19 [3]. Besides this correct and useful information, these platforms also spread misinformation even rumors. For example, "two explosions in the White House and Brack Obama is injured", and the news was confirmed to be fake by Associated Press. However, the fake news leads the temporary loss of market cap in S&B 500 to be reached \$136.5 billion [4]. Therefore, it is a crucial research issue to adopt effective strategies to contain the rumor spread in social networks.

Several recent approaches have studied the problem of containing the rumor spread from different perspectives. These mainly focus on extending the influence diffusion model to include rumor spread and contain the rumor spread under these diffusion models. Specifically, some studies proposed that adopt a positive cascade against the negative influence or rumor [5,6]. Another studies showed that blocking links could contain rumor spread [7,8]. Other studies showed that removing nodes was an effective strategy for blocking rumor spread [9]. (Refer to Section 2 for more details). These studies mainly employed a greedy strategy to select nodes or links to contain rumor spread.

However, it is time consuming to use a greedy strategy to select nodes or links to contain rumor spread even if in a small network. For example, He et al. have verified that the greedy algorithm takes more than 8 hours to finish for a network with 6.4k nodes under the competitive linear threshold model [5]. However, we need to swiftly contain rumor spread to avoid its harm before it spreads too far. For example, there is a common rumor that a traditional medicine can act against COVID-19, and so some people start to rush to purchase the medicine, however, this actually has no effect to COVID-19 [10]. If we can effectively contain the spread of this rumor, then the public can avoid buying invalid medicine. Therefore, in this paper, we mainly concentrate on efficiently containing the rumor spread in social networks.

Specifically, in this paper, we mainly focus on effectively selecting immune nodes to contain rumor spread under a spread model. Here, the immune nodes include two kinds: those nodes that are influenced by correct information; and those infected nodes which lose interest in a rumor with a certain probability over time. In reality, the susceptible people may be influenced by a rumor, and cheated by this rumor, then become the infective people. However, the infective people may be interested in other things and ignore this rumor with time going on, thus become immune people for this rumor. First, we propose the Extended Independent Cascade (EIC) model to reflect the characteristics, where the EIC model extended from the classic independent cascade (IC) model [11]. In the IC model, the node has two states: inactive and active, and each active node u tries to activate its inactive neighbor v with probability p_{uv} . In our EIC model, the node in the social network has three states: susceptible, infective and immune, where a susceptible node can transform its state into infective, and an infective node can transform the state into being immune. The description of the EIC model's state is similar to the Susceptible-Infected-Susceptible (SIS) model and the Susceptible-Infected-Recovered (SIR) model. However, the SIS model and SIR model assume that every individual directly contacts every other individual and spread the influence in a time unit. These two models are treated as continuous time modes. In our EIC model, the individual transfer its influence with a discrete time step. Further, to effectively select immune nodes, we employ the

ENHANCED- (k, η) -CORES (abbreviated to E- (k, η) -CORES) algorithm proposed by Bonchi et al. [12] to decompose the social network into dense subgraphs, where k is the core number of node v , η is a threshold and (k, η) -CORES is a subgraph G' . In (k, η) -CORES, the probability of the degree of node v that is greater or equal to k is no less than η . The subgraphs obtained by this algorithm can reflect well the characteristics of the degree of nodes and the influence spread probabilities among edges. Following this, we compute the number of immune nodes in each (k, η) -CORE based on the number of rumor spreaders and the value of k in each (k, η) -CORE, where the number of immune nodes is positively correlated with these two factors. Finally, we select the set of immune nodes S_k in each (k, η) -CORE based on its number of immune nodes. Evaluating the immune effectiveness in a subgraph of nodes is # p-hard under the EIC model even when this set is a singleton node [13]. Thus, it is necessary to employ a heuristic algorithm to approximately select immune nodes. In this paper, we propose the method of Marginal Covering Neighbors Set to approximately evaluate the immune effectiveness of susceptible nodes, and employ a greedy strategy that iteratively adds nodes based on the number of Maximum Marginal Covering Neighbors Set to the immune set, until $|S_k|$ nodes of each (k, η) -CORE are selected.

We compare our algorithms with other popular algorithms using two real-world social networks in the experimentation. Experimental results show that our algorithm is feasible and effective.

The remainder of this paper is organized as follows. In Section 2, we briefly survey the related work. In Section 3, we review the IC model and define the EIC model. In Section 4, we employ the E- (k, η) -CORES decomposition algorithm to obtain subgraphs and estimate the number of immune nodes in each (k, η) -CORE. In Section 5, we propose an improved greedy algorithm based on Maximum Marginal Covering Neighbors Set to select the immune set in each (k, η) -CORE. In Section 6, we verify the performance of the proposed algorithm via experimentation. In Section 7, we conclude and discuss the future work.

2. Related work

Kempe et al. first proposed the Linear Threshold (LT) model and IC model to describe the influence diffusion process and formulate the influence maximization problem under these two diffusion models as an optimization problem [11]. Then, a greedy algorithm was used to select seeds, where the greedy algorithm can achieve a $(1-1/e)$ approximation guarantee. The LT model and IC model only reflect one idea or opinion spread, and many variants of influence maximization have been proposed [7,14–16]. In this paper, we mainly focus on efficiently containing rumor spread in a social network. In the following, we summarize the related work of containing rumor spread as three categories.

2.1. Adopting a positive cascade against negative influences or rumors strategy

In this category, researchers have mainly focused on using positive influences to contain negative influences or rumors. He et al. [5] proposed a competitive linear threshold model to describe the process of competitive influence spread, and developed an efficient heuristic competitive local directed acyclic graph algorithm based on the Competitive linear threshold model for the influence blocking maximization problem. Liu et al. [6] considered that a participant uses their influence to contain an opponent's influence via the diffusion-containment model. Yang et al. [14] proposed a linear threshold model with one direction state transition for modeling competitive

influence spread, and proved that the problem of minimizing rumor spread in this model is NP-hard. They further presented three different approximation algorithms to solve the problem of minimizing rumor spreading and highlighted the proximity effect with the proposed algorithm. Guo et al. [16] focused on studying the Multi-Feature Rumor Blocking problem, and proposed the Revised-IMM (Influence Maximization via Martingales) algorithm to select a positive seed set to compete with the rumor spread under the Multi-Feature diffusion model. Zhu et al. studies the robust rumor blocking (RRB) problem while the exact positions of rumormongers are unknown. The Reverse Reachable Set was present to estimate the objective function of RRB, and a randomized greedy algorithm for RRB was designed to select k nodes as positive information spreader (called protector) to minimize rumor spread [17]. Compared with the following direct methods (Sections 2.2 and 2.3), this category is an indirect method, but it is also an effective method.

2.2. Removing links to block rumor spread

In this category, researchers have mainly focused on the removal or blockage of an edge set to minimize spread. Specifically, Kimura et al. [18] proposed a novel method, on the basis of a greedy algorithm and bond percolation, to minimize contamination by blocking links. Yao et al. proposed a greedy algorithm to select edges as blocking links to minimize the spread of negative influences [8]. Yan et al. [7] studied the problem of rumor spread minimization by removing an edge set from the network, i.e., finding an edge set $e=|k|$ from the candidate edge set E and removing this from the original graph G . They further proved that the objective function of Rumor Spread Minimization is not submodular, and a heuristic algorithm was used to select edges. Guo et al. presented the targeted protection maximization problem, which blocks the least edges to make the expected ration of nodes in the target set influenced by a rumor at most β . They employed the General-TIM algorithm based on Reverse Shortest Path and the Greedy algorithm were proposed to select edges [19].

2.3. Blocking nodes for minimizing negative influence spread

Some related research has been conducted on blocking nodes to minimize the influence of negative information. Newman et al. [9] experimentally verified that removing nodes in decreasing order of the out-degree about the 10% can immune to an epidemic of nodes in a social network. Wang et al. [20] employed a greedy algorithm to efficiently find k uninfected users in order to minimize negative information spread .

Comparing the above existing work with the study in this paper, the following aspects can be considered:

- Methods for influence maximization, that do not consider competitive situations, use the classic LT or IC model to model influence spread and use a greedy algorithm or other heuristic to select seeds.
- The models for influence containment only consider nodes influenced by positive and negative information. States are changed from inactive to positive or negative, and do not consider that a negative state can transform into a positive state, i.e., a susceptible node can switch its state to infective with probability p_{uv} at time t , and the infective node can't switch its state to the immune state at time step t_1 .
- Methods for selecting seeds to contain influence spread are always not feasible for containment, as it will take a long time to select an immune node when confronted with a large

number of nodes in a social network.

Therefore, in this paper, we propose the EIC model, which can well reflect the characters of rumor spread. The EIC model includes three states (i.e., susceptible, infective, and immune) and the process of state transformation. In addition, we discuss how to effectively select seeds in each (k, η) -CORE of a social network under the EIC model.

3. Influence spread model

Given an undirected graph $G = (V, E, p)$, where V denotes the node set, $E \subseteq V \times V$ denotes the edge set in a social network, and $p: E \rightarrow (0, 1)$ is a probability of existence of each edge, we extend the IC model to include rumor characteristics. First, we briefly review the IC model proposed by Kempe et al. and Chen et al. [11,21]. An uncertain graph can be defined as $G = (V, E, p)$ with S_0 , where S_0 denotes the node set. A node in G has two states: active and inactive. Let S_t denote the set of active nodes in step t . For every time step $t \geq 1$, each newly activated node $u \in S_t \setminus S_{t-1}$ tries to activate its inactive out-neighbor v with probability $p(u, v)$, and this attempt is independent from others. If u is successful, then v will switch its state from inactive to active; if not, then u will not try to activate v in subsequent steps. This diffusion process is repeated until there are no newly activated nodes, i.e., $S_t = S_{t-1}$.

Yao et al. addressed the minimization of negative influence via blocking links under the IC model [8]. In [7], under the same model, Yan et al. studied the Rumor Spread Minimization problem through link deletion. These studies suggest that the IC model can be used to model the spread of rumors. However, the IC model can only model a single opinion or idea. In reality, the correct or mistake information may spread in a social network simultaneously. In this paper, to better describe the characteristics of rumors, we extend the IC model to include three states: susceptible, infective, and immune, and the extended model is abbreviated to EIC.

We show the state transformation process of the EIC model in Figure 1. At step t , each node may infect its susceptible out-neighbors with probability $p(u, v)$, where this denotes the infective set at time step t . At step t_1 , the rumor disseminator v loses interest in spreading the rumor with probability α , and switches from being an infective state I_v to an immune state I'_v .

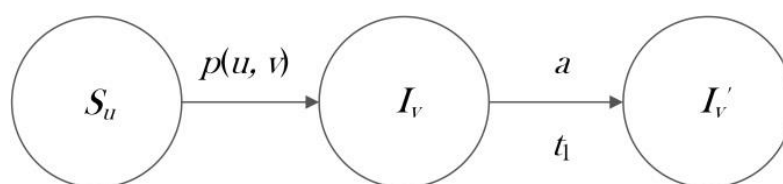


Figure 1. The state transformation process of influence spread.

4. Decomposition of the social network

Kitsak et al. [22] employed a k -shell decomposition method to divide a social network into a series of subgraphs, and found that the most efficient spreaders are the nodes which located within the core of the network. Moreover, Bonchi et al. proposed the (k, η) -CORES algorithm to decompose an uncertain graph [12]. Compared with k -shell decomposition, the (k, η) -CORES algorithm not only considers the degree of nodes but also takes into account the probability of edges. In order to improve the efficiency of decomposing (k, η) -CORES, we employ the E - (k, η) -CORES algorithm (Algorithm 1) to decompose a social network, proposed by Bonchi et al. [12].

 Algorithm 1. E - (k, η) -CORES of a social network

Input:

An uncertain graph $G = (V, E, p)$ and a threshold $\eta \in [0, 1]$

Output:

The (k, η) value $c[v]$ of each node v

```

1: compute  $\eta$ -deg( $v$ ) of each node  $v$ 
2:  $C \leftarrow [\phi, \dots, \phi]$ 
3: for all  $v \in V$ 
4:   put the node with same  $\eta$ -deg( $v$ ) into set  $C(\eta$ -deg( $v$ ))
5:   for  $i=1$  to  $\max(\eta$ -deg( $v$ ))
6:     select node  $v$  from  $C[i]$ ,  $c(v) \leftarrow i$ 
7:     remove edge  $e(u, v)$ 
8:     if  $\eta$ -deg( $u$ )  $> i$ 
9:       update  $\eta$ -deg( $u$ ), i.e., compute  $\eta$ -deg( $u$ ) where the edge  $e(u, v)$  is removed
10:    else
11:      do not update  $\eta$ -deg( $u$ )
12:    end for
13: remove  $v$  from  $C(\eta$ -deg( $v$ ))
14: end for
  
```

In Algorithm 1, we employ Eq (1) to compute η -deg(v) for each node v .

$$\eta\text{-deg}(v) = \max\{k \in [0, d_v] \mid \Pr[\text{deg}(v) \geq k] \geq \eta\} \quad (1)$$

In Eq (1), η -deg(v) is equal to the maximum k value in all $\Pr[\text{deg}(v) \geq k]$. To obtain the value of η -deg(v), we next need to consider how to compute $\Pr[\text{deg}(v) \geq k]$, which is equal to the sum of the probabilities $\Pr[\text{deg}(v) = i]$ for all $i \in [k, d_v]$ as shown in Eq (2).

$$\Pr[\text{deg}(v) \geq k] = \sum_{i=k}^{d_v} \Pr[\text{deg}(v) = i] \quad (2)$$

where $\Pr[\text{deg}(v) = i]$ denotes the probability that the degree of node v is equal to i , that is,

$$\Pr[\text{deg}(v) = i] = \sum_{\substack{N_v' \subseteq N_v \\ |N_v'| = i}} \prod_{e \in N_v'} p_e \prod_{e \in N_v \setminus N_v'} (1 - p_e) \quad (3)$$

Here, N_v denotes the adjacent edge set of node v and $N_v' \subseteq N_v$. We respectively set $|N_v| = m'$ and $|N_v'| = i$. There are $C_{m'}^i$ cases when we select i edges from m' edges, and the computation of $C_{m'}^i$ is shown in Eq (4):

$$C_{m'}^i = \frac{i}{m'(m'-i)} \quad (4)$$

The computation of $\eta\text{-deg}(v)$ as described in Eq (1) is time consuming. To improve the speed of computation, we employ a lower bound on the η -degree, which is proposed by [12] and described as follows.

Weisstein et al. [23] defined the regularized beta function $I_z(a, b)$ as the ratio of the incomplete beta function $B(z; a, b)$ and the beta function $B(a, b)$:

$$I_z(a, b) = \frac{B(z; a, b)}{B(a, b)} = \sum_{i=1}^{a+b-1} \binom{a+b-1}{i} z^i (1-z)^{a+b-1-i} \quad (5)$$

Given a node v , the minimum probability among edges incident to v is found, and set $p_{\min}(v) = \min_{e \in N_v} p_e$. Based on Eqs (2), (3) and (5), Eq (6) holds, and the detail of the proof can be found in [12].

$$\Pr(\text{deg}(v) \geq k) \geq I_{p_{\min}(v)}(k, d_v - k + 1) \quad (6)$$

Adding Eq (5) into Eq (6), we obtain

$$\Pr(\text{deg}(v) \geq k) \geq \sum_{i=k}^{d_v} C_{d_v}^i (p_{\min}(v))^i (1 - p_{\min}(v))^{d_v-i} \quad (7)$$

Equation (7) can avoid the combinatorial computation of Eq (4). If the right part of inequality Eq (7) from k to d_v is greater than or equal to the given η then this satisfies $\Pr(\text{deg}(v) \geq k) \geq \eta$. Based on Eqs (1) and (7), we can obtain the initial η -degree of all nodes.

In Figure 2 we give an example to illustrate the execution of Algorithm 1. The probabilities on the edges are as follows: $p_{12} = 0.5$, $p_{14} = 0.6$, $p_{15} = 0.8$, $p_{25} = 0.6$, $p_{24} = 0.4$, $p_{45} = 0.5$, $p_{23} = 0.3$, $p_{46} = 0.3$, $p_{56} = 0.4$, $p_{67} = 0.9$, $p_{68} = 0.5$, and $p_{78} = 0.4$, and we suppose $\eta = 0.05$.

Take v_2 as an example, the minimum probability value of the edge adjacent to v_2 is $p_{\min}(v_1) = 0.3$. According to Eq (7), we have:

$$\Pr(\text{deg}(v_2) = 4) = C_4^4 \times 0.3^4 = 0.0081,$$

$$\Pr(\text{deg}(v_2) = 3) = C_4^3 \times 0.3^3 \times 0.7 = 0.0756,$$

$$\Pr(\text{deg}(v_2) \geq 3) = 0.0081 + 0.0756 = 0.837 > 0.05.$$

Thus, we obtain $\eta\text{-deg}(v_2) = 3$. Similarly, we can compute $\eta\text{-deg}(v)$ of another node v in this simple social network, as shown in Table 1.

Now, we store the node v with $\eta\text{-deg}(v) = i$ in set $C[i]$, i.e., $C[1] = \{v_3\}$, $C[2] = \{v_7, v_8\}$, and $C[3] = \{v_1, v_2, v_4, v_5, v_6\}$. We further put these sets $C[1]$, $C[2]$ and $C[3]$ into set C , i.e., $C = \{C[1], C[2], C[3]\}$. After initialization, we compute the (k, η) -CORES of each node v in the social network. First, we cyclically select node v from $C[i]$ in order of i from small to large, and assign the (k, η) value of node v as $\eta\text{-deg}(v)$. Then, we compare $\eta\text{-deg}(u)$ with $\eta\text{-deg}(v)$, where u is the adjacent node of v . If $\eta\text{-deg}(u) \leq \eta\text{-deg}(v)$, then we do not need to recompute $\eta\text{-deg}_G(u)$. Otherwise, we need

to further calculate $\eta\text{-deg}_G(u)$ value after removing the edge $e(u, v)$. If Eq (1) holds, then we do not update $\eta\text{-deg}_G(u)$. Otherwise, $\eta\text{-deg}_G(u) = \eta\text{-deg}_G(u) - 1$.

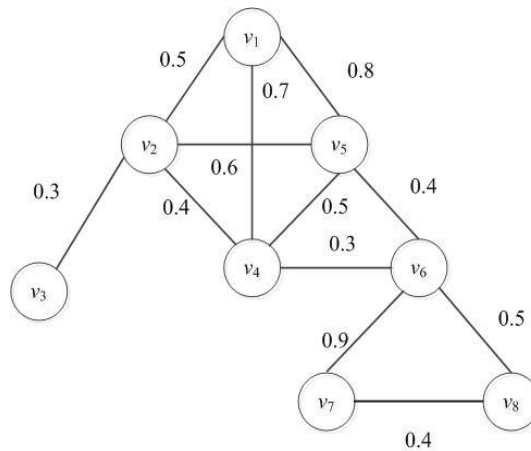


Figure 2. A simple social network G .

In Figure 2, we first select v_3 from $C[1]$ and assign its (k, η) -CORE value as 1. Since $\eta\text{-deg}(v_2) = 3 > \eta\text{-deg}(v_3) = 1$, we further employ Eq (7) to compute the value of $Pr(\text{deg}_G(v_2) \geq 3)$ after removing edge $e(v_1, v_2)$, and we have $Pr(\text{deg}_G(v_2) \geq 3) = 0.064 > 0.05$, therefore, $\eta\text{-deg}_G(v_2) = 3$. Similarly, we can obtain the (k, η) -CORE values of other nodes in Figure 2 as shown in Table 2.

Table 1. Initial $\eta\text{-deg}(v)$ of social network G in Figure 2.

Node	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
$\eta\text{-deg}(v)$	3	3	1	3	3	3	2	2

Table 2. The (k, η) -CORE of nodes in Figure 2.

Node	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
(k, η) -CORE	3	3	1	3	3	2	2	2

Based on the (k, η) -CORE value of each node v , we can obtain the (k, η) -CORES decomposition of the graph G , i.e., $(1, 0.05) = \{v_3\}$, $(2, 0.05) = \{v_6, v_7, v_8\}$ and $(3, 0.05) = \{v_1, v_2, v_4, v_5\}$. We next study how to select the set of immune nodes in (k, η) -CORE.

5. Selecting the set of immune nodes for each (k, η) -CORE

Selecting a set of immune nodes in social networks is time consuming. Thus, how to effectively select a set of immune nodes is an important research issue in this domain. In Section 4, the social network was decomposed into a serious of subgraphs, i.e., (k, η) -CORES, and we will focus on selecting the set of immune nodes of each (k, η) -CORE in this Section. Before selecting the set of immune nodes in each (k, η) -CORE, we first need to compute the number of immune nodes of each (k, η) -CORE. Then, we iteratively select the immune node of each (k, η) -CORE based on the number of immune nodes until $|S_k|$ nodes are selected.

5.1. Computation of the number of immune nodes in each (k, η) -CORE

The number of immune nodes is influenced by two factors: the number of rumor nodes set and the value of k in each (k, η) -CORE. Obviously, the larger the number of rumor nodes set in (k, η) -CORE, the more the number of immune nodes. Following this, we first give Eq (8) to describe the ratio of rumor nodes set in (k, η) -CORE:

$$\alpha_k^B = \frac{N(S_k^B)}{N(G_k)} \quad (8)$$

In Eq (8), $N(S_k^B)$, $N(G_k)$, and α_k^B denote the number of rumor nodes set, the number of nodes, and the ratio of rumor nodes in (k, η) -CORE, respectively.

The (k, η) -CORE of a node v can reflect its ability to spread its influence. Therefore, the larger the (k, η) -CORE value, the more easily this node will spread its influence. Accordingly, the larger the (k, η) -CORE value, the more immune nodes will be selected. Following this, we give Eq (9) to compute the ratio of rumor nodes based on the value of (k, η) and the ratio of the rumor node set α_k^B .

$$\beta_k = \frac{k \times \alpha_k^B}{\sum_{k=1}^{\max c_i} k \times \alpha_k^B} \quad (9)$$

Finally, we obtain the number of immune nodes in each (k, η) -CORE based on Eq (9), that is

$$|I_A^k| = |I_A| \times \beta_k \quad (10)$$

In Eq (10), $|I_A|$ and $|I_A^k|$ denote the number of immune nodes of social network G and (k, η) -CORE respectively.

We next study how to select the set of immune nodes of each (k, η) -CORE based on the number of immune nodes.

5.2. Selection of the set of immune nodes in each (k, η) -CORE

In Section 3, we have described the EIC model for the rumor propagation process. In this Section, we mainly focus on selecting the set of immune nodes of each (k, η) -CORE to maximally contain a rumor under the EIC model.

Kitsak et al. [22] found that the most efficient spreaders are located in the larger k -shell, and pointed out that nodes in the same core affect the nodes in the similar areas. Similarly, a node in the same (k, η) -CORE will also produce similar spread areas, i.e., v_1, v_2, v_4 , and v_5 of $(3, 0.05)$ -core in Figure 2 have similar spread areas. Cao et al. [24] proposed the Core Covering Algorithm based on k -shell decomposition to maximize the influence spread. In contrast to their work, the objective of this paper is to maximally contain rumor spread. Thus, we need to consider the characteristics of rumor spread and cannot directly select the nodes of the larger (k, η) -CORE or employ the Core Covering Algorithm to select the set of immune nodes.

Considering the characteristics of rumor spread, we have proposed an EIC model to describe the rumor spread process previously in this paper. Next, we consider how to select the set of immune nodes in each (k, η) -CORE in the EIC model. With this view, we can make use of the Degree-Discount algorithm proposed by Chen as this algorithm nearly matches the greedy algorithm for the IC model [13].

The idea behind the Degree-Discount algorithm is as follows: Let node v be a neighbor of node u . If node u has been selected as a seed, then the degree of v should be discounted. Combining the idea of Cao et al. [24] and Chen et al. [13], we propose a method that uses Maximum Marginal Covering Neighbors Set as the seed selection rule. Specifically, let S_k and I_B be the set of immune nodes and rumor nodes, respectively; we can describe the marginal gain of v newly added to the set S_k as shown in Figure 3.

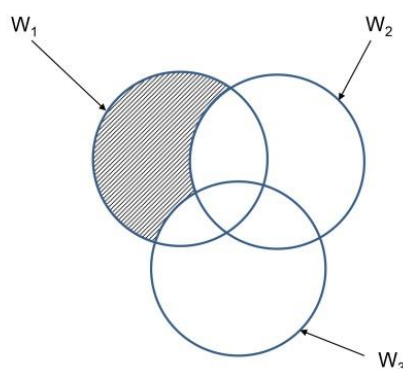


Figure 3. The marginal gain of newly added node v .

In Figure 3, let W_1 , W_2 and W_3 be the t -step neighbor set of $S_k \cup \{v\}$, S_k and I_B . The marginal gain of immune node v is the diagonal part, which can be formally described as in Eq (11).

$$\sigma(S_k \cup \{v\}, I_B) - \sigma(S_k, I_B) = |W_1| - |W_1 \cap W_2| - |W_1 \cap W_3| + |W_1 \cap W_2 \cap W_3| \quad (11)$$

The set of immune nodes S_k and rumor nodes I_B may influence the same neighbors, thus the influence will counteract. At the same time, the immune node set S_k and $S_k \cup \{v\}$ will also influence the same neighbors, and thus the influence will overlap. Therefore, we introduce the discount factor p , and let the counteracted part and overlap be multiplied by this discount factor p , that is

$$\sigma(S_k \cup \{v\}, I_B) - \sigma(S_k, I_B) = |W_1| - (|W_1 \cap W_2| - |W_1 \cap W_3| + |W_1 \cap W_2 \cap W_3|) \times p \quad (12)$$

Selecting the optimal set of immune nodes is NP -hard. The objective function $\sigma(S_k, I_B)$ is monotone and submodular in the EIC model, and the details of the proof are omitted here. Therefore, we employ a greedy algorithm to approximately select the set of immune nodes based on the theory in [25]. We next provide Algorithm 2 to describe how to select the set of immune nodes for each (k, η) -CORE. In this algorithm, we iteratively select the current best immune node u from $V - I_B$ in each (k, η) -CORE based on Maximum Marginal Covering Neighbors Set, which can maximally contain the rumor spread of (k, η) -CORE added into the immune set S_k starting from an empty set, until $|S_k|$ seeds are selected.

According to the conclusion in [25], Algorithm 2 can approximate the optimal result with $(1-1/e)$ for the problem of containing rumor spread. Step 2 is a critical step of Algorithm 2, and the time complexity of Algorithm 2 is $O(|S_k| \times |V_k| \times t)$, where $|S_k|$, $|V_k|$, and t are the number of immune nodes, the number of nodes, and the time step, respectively.

Algorithm 2: Greedy algorithm for selecting the set of immune nodes S_k of (k, η) -CORE

Input:

$G_k = (V_k, E_k, p)$, an undirected graph, which is obtained by Algorithm 1
 I_B , the set of rumor disseminators
 $|S_k|$, the number of immune nodes in G_k
 α , the probability of a rumor node transforming into an immune node
 t_1 , the time step of a rumor node transforming into an immune node

Output:

The set of immune nodes S_k of G_k

Variable:

t , the time step

Steps:**Step 1: Initialization**

- 1: $S_k \leftarrow \emptyset$
- 2: From time 1 to t , calculate t time step neighbors of S_k , $S_k \cup \{v\}$ and I_B , respectively
- 3: $W_1 \leftarrow \sigma(S_k \cup \{v\})$
- 4: $W_2 \leftarrow \sigma(S_k)$
- 5: $W_3 \leftarrow \sigma(I_B)$

Step 2: Selecting the set of immune nodes S_k

- 6: **For** $i = 1$ to $|S_k|$ **do**
 - 7: Select $u = \operatorname{argmax}(\sigma(S_k \cup \{v\}, I_B) - \sigma(S_k, I_B))$ // Executing Eq (12) for each node
 - 8: $S_k \leftarrow S_k \cup \{u\}$
 - 9: **End For**
 - 10: **Return** S_k
 - 11: Evaluate the immune effectiveness Q
-

5.3. Evaluating the immune effectiveness of S_k

We have obtained the set of immune nodes S_k in Section 5.2. Following this, we evaluate rumor spread value Q of *immune node set* S_k in the EIC model in T steps, i.e., we calculate the influence effectiveness of I_B when the set of immune nodes S_k exists. We have:

$$Q = \sum_{v \in V \setminus S_k \cup I_B} AP_T(v) \quad (13)$$

For this purpose, we need to calculate the value of $AP_T(v)$, i.e., the node v (i.e., $v \in G_k \setminus S_k \cup I_B$, $v \in G_k \setminus S_k \cup I_B$) influenced by I_B in T steps when the immune node set S_k exists. This can be defined as

$$AP_T(v) = \sum_{t=1}^T \eta_t ap_t(v) \quad (14)$$

In Eq (14), η_t denotes the immune coefficient. In the EIC model, the rumor disseminator will become an immune node with probability α in time step t_1 . Therefore, we have:

$$\eta_t = \begin{cases} 1 - \alpha & 1 \leq t \leq t_1 \\ 1 & t_1 \leq t \leq T \end{cases} \quad (15)$$

In Eq (14), $ap_t(v)$ denotes the influence effectiveness of I_B at time step t when the immune node set S_k exists, and can be described as in Eq (16).

$$ap_t(v) = 1 - \prod_{w \in N^m(v)} (1 - ap_{t-1}(w) \times p(w, v)) \quad (16)$$

6. Experiments

In this section, we conduct experiments using two real-world datasets to test and evaluate the performance of the algorithms. We first give a description of each dataset and the experimental setup. Then, we analyze and discuss the experimental results.

6.1. Experimental setup

Gnutella peer-to-peer network: This dataset is a sequence of snapshots of the Gnutella peer-to-peer file sharing network from August 2002. In this network, nodes represent hosts in the Gnutella network topology and edges represent connections between the Gnutella hosts. There are 10,876 nodes and 39,994 edges [26].

Arxiv HEP-TH (High Energy Physics-Theory) collaboration network: This dataset is from the e-print arXiv and covers scientific collaborations between authors's papers submitted to the high energy physics theory category. There are 9877 nodes and 25,998 edges. Each node u in the network represents an author, and each edge $e(u, v)$ denotes that an author u coauthored a paper with author v [27].

The experiments were conducted on a machine with an Intel Core i7-8550 CPU and 8 GB main memory running the Windows 10 operating system. The code was written in Python.

In all experiments, we adopt the EIC model as the influence propagation model. In addition, since all datasets lack a propagation probability, we uniformly assign the propagation probability of each edge in the range [0.1, 0.9] at random.

6.2. Experimental results

In the experiments, we compare the value of the rumor spread value Q obtained with the greedy algorithm based on the Maximum Marginal Covering Neighbors Set (i.e., the greedy algorithm), the Degree Discount algorithm (the degree discount heuristic algorithm), MaxDegree algorithm (selecting seeds with the largest degree) and the Random algorithm (selecting seeds randomly). Next, we show relationships between Algorithm 2 and parameters. Further, we give the running times of the different algorithm. Finally, we compare the rumor spread value with Greedy algorithm (i.e., Algorithm2) under k -core decomposition and (k, η) decomposition methods.

6.2.1 Rumor spread value vs. four different algorithms

For the Gnutella peer-to-peer network and the Arxiv HEP-TH collaboration network, we first recorded the rumor spread value Q when selecting the set of immune nodes with Greedy (Algorithm2), DegreeDiscount, MaxDegree, and Random, respectively, with the increase of immune

seeds, where we assigned $\eta = 0.01$, $t = 3$, $t_1 = 2$, $\alpha = 0.7$, and $p = 0.5$. The number of rumor disseminators $|I_B|$ in Figure 4(a),(b) are 37 and 20, respectively. The results of comparisons are shown in Figure 4. It can be seen that with more immune seeds, the effectiveness improves. The rumor spread value Q for Greedy is better than for the Maxdegree, DegreeDiscount and Random algorithms for most immune seeds. This result means that Greedy algorithm proposed in this paper is effective.

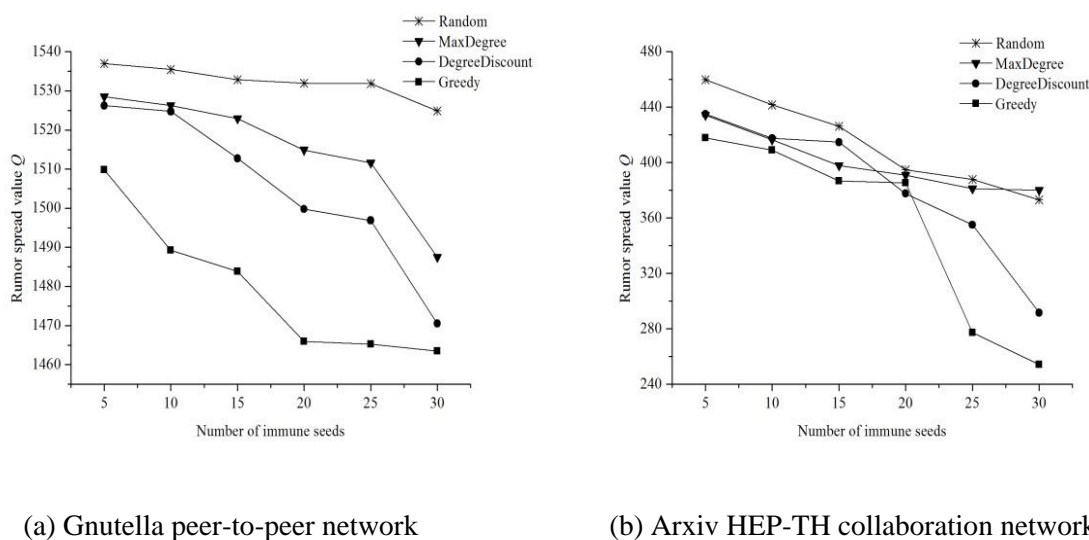


Figure 4. Rumor spread value vs. increasing of seeds on Gnutella peer-to-peer and Arxiv HEP-TH networks with the increase of immune seeds.

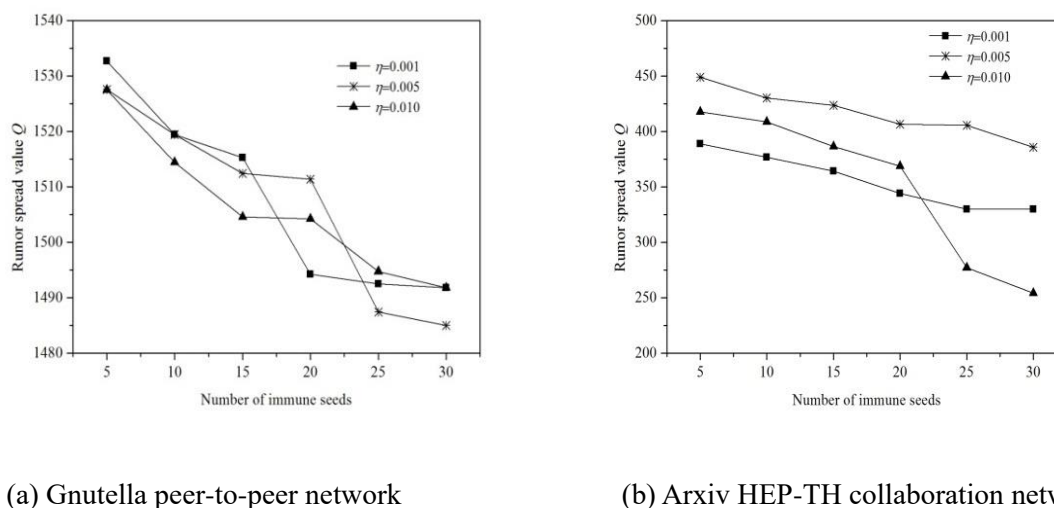
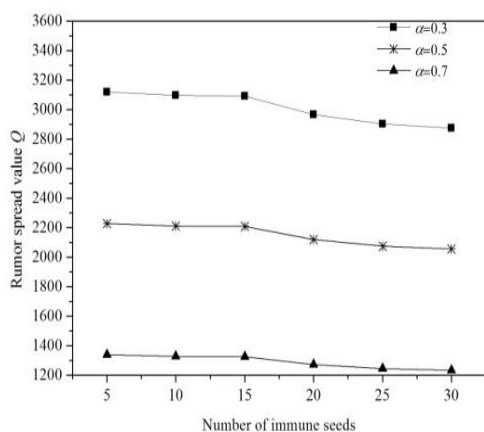


Figure 5. Rumor spread value Q with the increase of immune seeds on the networks with various η .

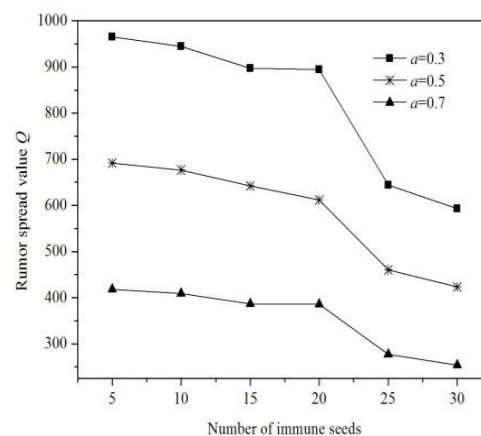
6.2.2 Rumor spread value with greedy algorithm vs. different parameters

Firstly, we tested the rumor spread value by Algorithm 2 with $\eta = 0.001$, $\eta = 0.005$, and $\eta = 0.01$, respectively. In this experiment, we set $t = 3$, $t_1 = 2$, $\alpha = 0.7$, and $p = 0.5$, and the number of rumors

spread $|I_B|$ in Figure 5(a), 5(b) are 37 and 20, respectively. The results are shown in Figure 5. It can be seen that different η will have a different immune effectiveness Q w.r.t. the same number of immune nodes for the Gnutella peer-to-peer network and the Arxiv HEP-TH collaboration network. This means that the immune effectiveness of Q is related to the value of η . This is because different η will generate different subgraphs. Accordingly, the set of immune nodes generated by these subgraphs will be different, and these immune nodes will lead to different values of immune effectiveness Q , which are consistent with intuition and real world situations.

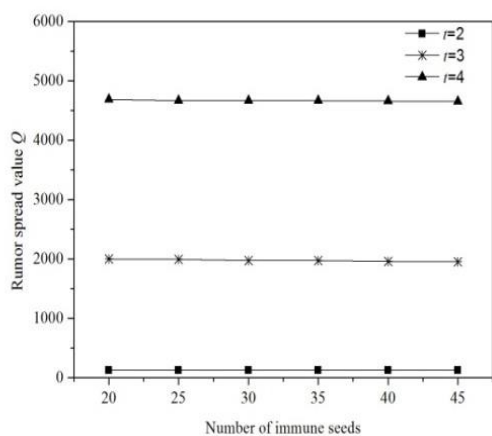


(a) Gnutella peer-to-peer network

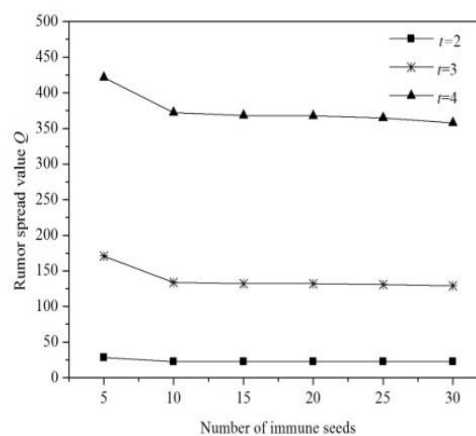


(b) Arxiv HEP-TH collaboration network

Figure 6. Rumor spread value Q with the increase of immune seeds on the networks with various α .



(a) Gnutella peer-to-peer network.



(b) Arxiv HEP-TH collaboration network

Figure 7. Rumor spread value Q with the increase of immune seeds on the networks with various time steps t .

Secondly, we compared the value of rumor spread value Q with the increase of immune seeds for the two networks with $\alpha = 0.3$, $\alpha = 0.5$, and $\alpha = 0.7$, respectively, where we set $\eta = 0.01$, $t = 3$, $t_1 = 2$, and $p = 0.5$. In addition, the number of rumor spreaders in the Gnutella peer-to-peer network and the Arxiv HEP-TH collaboration networks are 15 and 20, respectively. The comparisons are shown in Figure 6. It can be seen that the larger the transition probability α , the better the value of the immune effectiveness Q (i.e., the better containment effectiveness) will be. This is because a larger transition probability α will make more nodes change their state from disseminator to immune.

Finally, we test the value of immune effectiveness Q for time steps $t = 2$, $t = 3$, and $t = 4$ with the increase of immune seeds, where we assigned $\eta = 0.005$, $t_1 = 2$, and $p = 0.5$. In Figure 7(a), we assigned $|I_B| = 5$ and $\alpha = 0.2$, and in Figure 7(b), $|I_B| = 7$ and $\alpha = 0.7$. Figure 7 shows that the rumor spread value Q will increase as the time steps increase when the number of immune nodes is the same, and the trend of change will be stable. This is because the propagation of rumors will increase with the increase of time steps.

6.2.3 The running times with greedy algorithm vs. other algorithms

Table 3 shows the running times of the different algorithms for selecting $|S_k| = 30$ immune nodes in the Gnutella peer-to-peer network and Arxiv HEP-TH collaboration network. In these two networks, the Greedy algorithm proposed in this paper only takes 61.78 seconds and 4.48 seconds respectively. Compared with the original greedy algorithm, which takes several hours to select seeds in the 6.4k network [5], the greedy algorithm proposed in our paper is very effective.

Table 3. The running time under the *EIC* model when $k = 30$.

The running time under the EIC model with $k = 30$				
	Greedy	DegreeDiscount	MaxDegree	Random
Gnutella peer-to-peer network	61.78 s	1.16 s	0.67 s	0.71 s
Arxiv HEP-TH collaboration network	4.48 s	0.77 s	0.52 s	0.54 s

6.2.4 Rumor spread value with greedy algorithm vs. different methods of decomposition social network

For the Arxiv HEP-TH collaboration network, we compared the rumor spread value Q of Greedy algorithm with k -decomposition and (k, η) -decomposition, respectively, with the increase of immune seeds, where we assigned $\eta = 0.01$, $t = 3$, $t_1 = 2$, $\alpha = 0.7$ and $p = 0.5$. In addition, the number of rumor spreads is 15 and 20 respectively. The results of comparisons are shown in Table 4. It can be seen that the rumor spread value Q of with (k, η) -decomposition outperforms k -decomposition. This result means that (k, η) -decomposition employed in this paper is effective.

Table 4. the rumor spread value Q of Greedy algorithm with k -decomposition and (k, η) -decomposition.

Decomposition method	$ S_k = 40$	$ S_k = 45$	$ S_k = 50$	$ S_k = 55$	$ S_k = 60$	$ S_k = 65$
(k, η) -decomposition	271.56	269.52	259.09	254.51	250.80	237.48
k -decomposition	343.02	336.41	336.41	335.77	322.41	318.56

7. Conclusions and future work

With the goal of effectively selecting immune nodes to contain the spread of rumors, we employ the E -(k , η)-CORES algorithm to decompose a social network. Moreover, we extend the IC model to establish the EIC model by incorporating the specialties and characteristics of rumor spread. Further, we propose a greedy algorithm based on Maximum Marginal Covering Neighbors Set to select the set of immune nodes for containing rumor spread.

The E -(k , η)-CORE algorithm, EIC model, and the greedy algorithm strategy for containing rumor spread proposed in this paper can provide underlying techniques to handle rumor spread and negative information spread. However, containing rumor spread in large-scale social networks is still time consuming. For future work, we plan to implement these methods on a big data processing platform. Moreover, we still need to further study the relationships among relevant parameters.

Acknowledgments

This work was supported by the Yunnan Province Joint Youth Program Foundation of Local Colleges and Universities (No. 2017FH001-116), the Open Foundation of Key Laboratory for the Structure and Evolution of Celestial Objects, Chinese Academy of Sciences (No. OP201710), the School Level Excellent Course of Qujing Normal University (JPKC 2016005), the Research Foundation of Educational Department of Yunnan Province (No. 2017ZZX133), and the Project on the Humanities and Social Sciences Research of Ministry of Chinese Education (No. 18XJC880012)

Conflicts of interest

The authors declare no conflict of interest.

References

1. H. Tankovska, Number of monthly active Facebook users worldwide as of 1st quarter 2020 (in millions), 2021. Available from: <http://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwid/>.
2. L. L. Thomala, Number of monthly active WeChat users from 2nd quarter 2010 to 1st quarter 2020 (in millions), 2020. Available from: <http://www.statista.com/statistics/255778/number-of-active-wechat-messenger-accounts/>.
3. W. Zhou, A. Wang, F. Xia, Y. Xiao, S. Tang, Effects of media reporting on mitigating spread of COVID-19 in the early phase of the outbreak, *J. Math. Biosci. Eng.*, **17** (2020), 2693–2707.
4. P. Domm, False rumor of explosion at White House causes stocks to briefly plunge; AP confirms its Twitter feed was hacked, *CNBC. COM*, **23** (2013), 2062.
5. X. He, G. Song, W. Chen, Q. Jiang, Influence blocking maximization in social networks under the competitive linear threshold model, in *Proceedings of the 2012 siam international conference on data mining*, (2012), 463–474.
6. W. Liu, K. Yue, H. Wu, J. Li, D. Liu, D. Tang, Containment of competitive influence spread in social networks, *Knowl. Based Syst.*, **109** (2016), 266–275.

7. R. Yan, Y. Li, W. Wu, D. Li, Y. Wang, Rumor blocking through online link deletion on social networks, *ACM Trans. Knowl. Discovery Data*, **13** (2019), 1–26.
8. Q. Yao, C. Zhou, L. Xiang, Y. Cao, L. Guo, Minimizing the negative influence by blocking links in social networks, in *International conference on trustworthy computing and services*, Springer, Berlin, Heidelberg, (2014), 65–73.
9. M. E. J. Newman, S. Forrest, J. Balthrop, Email networks and the spread of computer viruses, *Phys. Rev. E*, **66** (2002), 035101.
10. A. Abdoli, Gossip, Rumors, and the COVID-19 Crisis, *Disaster Med. Public Health Preparedness*, **14** (2020), E29–30.
11. D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, (2003), 137–146.
12. F. Bonchi, F. Gullo, A. Kaltenbrunner, Y. Volkovich, Core decomposition of uncertain graphs, in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2014), 1316–1325.
13. W. Chen, Y. Wang, S. Yang, Efficient influence maximization in social networks, in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2009), 199–208.
14. L. Yang, Z. W. Li, A. Giua, Containment of rumor spread in complex social networks, *Inf. Sci.*, **506** (2020), 113–130.
15. W. Liu, X. Chen, B. Jeon, L. Chen, B. Chen, Influence maximization on signed networks under independent cascade model, *Appl. Intell.*, **49** (2019), 912–928.
16. J. X. Guo, T. T. Chen, W. L. Wu, A multi-feature diffusion model: Rumor blocking in social networks, *IEEE/ACM Trans. Networking*, 2020.
17. J. M. Zhu, S. Ghosh, W. L. Wu, Robust rumor blocking problem with uncertain rumor sources in social networks, *World Wide Web*, **24** (2021), 229–247.
18. M. Kimura, K. Saito, H. Motoda, Minimizing the spread of contamination by blocking links in a network, in *Aaai*, **8** (2008), 1175–1180.
19. J. X. Guo, Y. Li, W. L. Wu, Targeted protection maximization in social networks, *IEEE Trans. Network Sci. Eng.*, **7** (2020), 1645–1655.
20. S. Wang, X. Zhao, Y. Chen, Z. Li, K. Zhang, J. Xia, Negative influence minimizing by blocking nodes in social networks, in *Proceedings of the 17th AAAI Conference on Late-Breaking Developments in the Field of Artificial Intelligence*, (2013), 134–136.
21. W. Chen, L.V. Lakshmanan, C. Castillo, Information and influence propagation in social networks, *Synth. Lect. Data Manage.*, **5** (2013), 1–177.
22. M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, et al., Identification of influential spreaders in complex networks, *Nat. Phys.*, **6** (2010), 888–893.
23. E. W. Weisstein, *Binomial Distribution-from Wolfram Mathworld*, 2013. Available from: <http://mathworld.wolfram.com/BinomialDistribution.html>.
24. J. Cao, D. Dong, S. Xu, X. Zheng, B. Liu, J. Luo, A k-core based algorithm for influence maximization in social networks, *Chin. J. Comput.*, **38** (2015), 238–248.
25. S. Fujishige, *Submodular Functions And Optimization*, Elsevier, 2005.
26. SNAP Datasets: Stanford large network dataset collection, Available from: <http://snap.stanford.edu/data/p2p-Gnutella04.html>.

-
27. SNAP Datasets: Stanford large network dataset collection, Available from: <http://snap.stanford.edu/data/ca-HepTh.html>.



AIMS Press

©2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)