Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Rumors clarification with minimum credibility in social networks

Xiaopeng Yao, Guangxian Liang, Chonglin Gu, Hejiao Huang *

Harbin Institute of Technology (Shenzhen), China

1. Introduction

The tremendous advance of the online social networks makes it more convenient and fast for the spread of information. In 2020, the information about COVID-19 is overwhelming, which is mixed with a lot of rumors. Rumor and truth can change people’s believes more than once, depending on who is more credible. Here we use credibility to measure the influence one person has on others. Considering costs, we often hope to find the people with the smallest credibility but can achieve the maximum influence. Therefore, we focus on how to use minimal credibility in a given amount of time to clarify rumors. Given the time $t$, the minimum credibility rumor clarifying (MCRC) problem aims to find a seed set with $k$ users such that the total credibility can be minimized when the total number of the users influenced by positive information reaches a given number at time $t$. In this paper, we propose a Longest-Effective-Hops algorithm called LEH to solve this problem that supposes each user can be influenced two or more times. The theoretical analysis proves that our algorithm is universal and effective. Extensive contrast experiments show that our algorithm is more efficient in both time and performance than the state-of-the-art methods.

In 2020, the information about Corona Virus Disease 2019 (COVID-19) is overwhelming, which is mixed with a lot of rumors. Rumor and truth can change people’s believes more than once, depending on who is more credible. Here we use credibility to measure the influence one person has on others. Considering costs, we often hope to find the people with the smallest credibility but can achieve the maximum influence. Therefore, we focus on how to use minimal credibility in a given amount of time to clarify rumors. Given the time $t$, the minimum credibility rumor clarifying (MCRC) problem aims to find a seed set with $k$ users such that the total credibility can be minimized when the total number of the users influenced by positive information reaches a given number at time $t$. In this paper, we propose a Longest-Effective-Hops algorithm called LEH to solve this problem that supposes each user can be influenced two or more times. The theoretical analysis proves that our algorithm is universal and effective. Extensive contrast experiments show that our algorithm is more efficient in both time and performance than the state-of-the-art methods.

In 2020, the information about Corona Virus Disease 2019 (COVID-19) is overwhelming, which is mixed with a lot of rumors. Rumor and truth can change people’s believes more than once, depending on who is more credible. Here we use credibility to measure the influence one person has on others. Considering costs, we often hope to find the people with the smallest credibility but can achieve the maximum influence. Therefore, we focus on how to use minimal credibility in a given amount of time to clarify rumors. Given the time $t$, the minimum credibility rumor clarifying (MCRC) problem aims to find a seed set with $k$ users such that the total credibility can be minimized when the total number of the users influenced by positive information reaches a given number at time $t$. In this paper, we propose a Longest-Effective-Hops algorithm called LEH to solve this problem that supposes each user can be influenced two or more times. The theoretical analysis proves that our algorithm is universal and effective. Extensive contrast experiments show that our algorithm is more efficient in both time and performance than the state-of-the-art methods.

In this work, we study the problem of minimizing credibility under the rumor-clarifying cascade, and call it Minimum Credibility Rumor Clarifying (MCRC) problem. Our goal is to find a seed set with $k$ nodes with the smallest credibility possible when the total number of clarification users reaches the given number at time $t$. Furthermore,
we also study the X-Minimum Credibility Rumor Clarifying (X-MCRC) problem under the x-rumor-clarifying cascade.

The main contributions of this paper are summarized as follows:

(1) We develop a rumor-clarifying cascade and propose the MCRC problem which has more practical and universal significance in real life. Furthermore, based on the original cascade, we put forward a x-rumor-clarifying cascade in which the state of each user can change two or more times.

(2) We design a Longest-Effective-Hops (LEH) algorithm which uses the longest distance, the number of neighbor nodes and the number of rumor nodes contained in out-neighbors to select the seed nodes. It improves the effectiveness and universality of the algorithm. The algorithm LEH can both solve MCRC problem and X-MCRC problem.

(3) We prove that the objective function of problem is submodular. Then we add the parameter of the increment of credibility each time to algorithm LEH to ensure that the algorithm can always find k seed nodes in the finite-time period while guaranteeing the approximate ratio.

(4) We analyze the properties of the algorithm and problems under different cascades. We evaluate our algorithm on three networks by selecting a number of different parameters and use three other methods for comparison. The results show that our algorithm outperforms these existing methods on different datasets.

The remaining part will be arranged as follows: In Section 2, we review the related work. In Section 3, the preliminaries and definitions are given and we conduct theoretical analysis on problems. In Section 4, we present our algorithm in detail and conduct theoretical analysis. Section 5 presents experiment results and analysis. Finally, we conclude the whole paper in Section 6.

2. Related work

In this section, we briefly discuss the prior works on rumor and controlling.

Many researchers [5–7] have already done a great number of research on the spread of rumors from different angles on Twitter or microblogging. In addition, V. Indu et al. [8] propose a novel nature-inspired algorithm that utilizes the identified prominent features to calculate the probability of a node to share a rumor and F. Chierichetti et al. [9] study the performance of rumor spreading in the classic preferential attachment model in social network. The ISS (Ignorant-Spreader-Stifler) rumor spreading model is studied by J. R. C. Piqueira [10]. Z. He et al. [11] propose a heterogeneous-network-based epidemic model that incorporates the two kinds of methods to describe rumor spreading in mobile social networks and S. Daum et al. [12] study rumor spreading with bounded in-degree by considering a restricted model where at each node only one incoming call can be answered in one time unit.

Controlling the spread of rumors has also been studied by many researchers. Generally speaking, there are three methods to control the spread of rumors. One is removing associations between users, such as [13,14] and the other one is blocking influential users, like [15,16]. The third one is spreading truth to clarify rumors, such as [17,18]. In addition, some researchers combine various methods to control rumors. For example, H. Tong et al. [19] associate the nodes removal method with the link removal method. However, the first two methods will significantly change the structure of the original network, so spreading truth to clarify rumors is widely used at present.

Many scholars [20–22] work to suppress rumors by spreading positive information in social networks. H. Zhang et al. [23] propose an effective algorithm, exploiting the critical nodes and using the greedy approach as well as applying the CELF heuristic to achieve the goal. S. Wen et al. [24] build a mathematical model to evaluate the efficiency of different rumor blocking methods and explore the strategies of different rumor blocking methods working together. Their analysis provides the exact numeric equivalence between the different strategies. B. Wang et al. [25] consider the use experiences into the problem and Z. He et al. [26] propose two cost-efficient strategies in mobile social networks for this problem. P. Zhang et al. [27] introduce a novel rumor control problem, called users’ Browsing based rUmor blocK (BUK) and try to find k nodes as protectors.

However, the problem of rumor blocking is NP-hard and the objective functions of these related problems are often very complicated to compute. Then W. Chen et al. [28] first show that computing the exact value of the expected influence is #P-hard. We can know that by using the greedy algorithm with Monte Carlo simulation, it often takes a lot of time on small networks. In order to solve these kind of theoretical problems, some researchers [29–31] design an improved switched fuzzy memory sampled-data control protocol and propose a dynamic event triggering controller including Markov switching topology in complex network to improve the effectiveness of the algorithm. Recently, Q. Fang et al. [32] propose an efficient random algorithm with unpredicted rumor seed set and it provides a (1 − 1/e − ε)-approximate solution with at least n/ε probability, where / is a probability indicator. G. Tong et al. [33] propose a randomized approximation algorithm which is provably superior to the state-of-the art methods with respect to running time.

3. Problem model and theoretical analysis

In social networks, given a directed graph G = (V, E) where V denotes the set of nodes, E denotes the set of edges and each node represents a user and each directed edge represents that one user can influence another in a directed way. For each edge (i, w) ∈ E, let pr(Δ) ∈ [0, 1] denote the influence probability that the process spreads along edge v to node w. We define that an effective path between the two nodes is a one-way and connected path.

In what follows we provide the preliminaries to the rest of this paper. The important notations are listed in Table 1.

Table 1  NOTATIONS.

| Symbol | Definition |
|--------|------------|
| $p_{r,(w)}$ | The influence probability that the process spreads along edge v to node w |
| $N_{r}^<(v)$ | Node v’s in-neighbors |
| $N_{r}^>(v)$ | Node v’s out-neighbors |
| $N_{r}^{(i,j)}(v)$ | v’s j-hop in-neighbors(it have an effective path) |
| $N_{r}^{(i,j)}(v)$ | v’s j-hop out-neighbors(it have an effective path) |
| $C_{r}^<(v)$ | The maximum positive credibility in the node v’s in-neighbors |
| $C_{r}^>(v)$ | The maximum positive credibility just in the node v’s out-neighbors |
| $r_{i}$ | The negative credibility of each node in the set of the rumor cascade at the time t |
| $C_{r}$ | The positive credibility of each node v in the set of the clarifying cascade |
| $W_{r}((a))$ | The probability that v accepts the negative credibility or the positive credibility at v at time t |
| $S_{r}$ | The rumor seed nodes |
| $S_{c}$ | The clarifying seed nodes |
| $f_{r}((S_{c}))$ | The expected number of nodes which are in the clarifying cascade at time t when $S_{c}$ is selected as the seed set of the clarifying cascade |
| $C_{r}$ | Sum of the credibility of clarifying seed nodes |

Fig. 1 shows an example of different hops’ in-neighbors and out-neighbors of node $v_{i}$. As shown in Fig. 1, we can see that $N_{r}^{<}(v_{i}) = \{v_{1},v_{2}\}$ and $N_{r}^{>}(v_{i}) = \{v_{i},v_{0}\}$, $N_{r}^{<}(v_{0}) = \{v_{2},v_{0}\}$, $N_{r}^{>}(v_{0}) = \{v_{0}\}$ and $N_{r}^{<}(v_{2}) = \{v_{2}\}$.

In real life, users tend to be influenced two or more times. Therefore, we present the rumor-clarifying cascade model and x-rumor-clarifying cascade model in the following.
time each edge is $p$.

Fig. 2, there are 7 nodes. We assume that the influence probability of

time, the node $v$ credibility, then

probability of $t$ of each node in

rumor and the clarifying cascade, respectively. The negative credibility

of the rumor-clarifying cascade unfolds in discrete, as follows.

by clarifying cascade for once. Specially, by default, the states of nodes
twice, but it can be only changed by rumor cascade for once or changed
at the current moment, the larger $R$ is.

In this cascade model, the state of each node can be changed at most
twice, but it can be only changed by rumor cascade for once or changed
by clarifying cascade for once. Specially, by default, the states of nodes
which in $S_r$ and $S_c$ have changed for once at $t = 0$. The spread process
of the rumor-clarifying cascade unfolds in discrete, as follows.

(i) Initially all the nodes are inactive.

(ii) Secondly, at time $t = 0$, nodes in $S_r$ and $S_c$ are activated by the

rumor and the clarifying cascade, respectively. The negative credibility

of each node in $S_r$ is $R_i$ and the positive credibility of each node in $S_c$
is $C_i (i \in \{1, 2, 3, k\})$.

(iii) Thirdly, at time $t > 0$, each node $v$ which is influence at time

$t-1$ will influence each of its inactive neighbors $u$ with a success
probability of $\rho_{u,v}$. If node $v$ receives two opposing information at
the same time and the positive credibility is equal to the negative
credibility, then $v$ will be influenced by rumor. If node $v$ is influenced
by more than one node which are in the clarifying cascade at the same
time, the node $v$’s credibility is equal to the largest credibility.

(iv) Finally, if no node can be further influenced by any cascade,

the spread is over.

Fig. 2 shows an example of rumor-clarifying cascade. As shown in

Fig. 2, there are 7 nodes. We assume that the influence probability of
each edge is $p_{u,v}=1$ and $R_i=2$, $C_i=1$ and $C_i=4$. In Fig. 2(a), at
time $t=0$, $v_1$, $v_3$, $v_4$ are activated by rumor cascade and $v_2$, $v_7$ are activated
by clarifying cascade. Other nodes are inactive. In Fig. 2(b), at time

t=1, $v_1$ will be influenced as a rumor node and $v_1$, $v_6$ will be influenced
as clarification nodes. Although both node $v_2$ and $v_3$ try to influence
node $v_4$, node $v_4$ has larger positive credibility than node $v_3$. Then
$R_1=3$ and $C_{v_1}=C_{v_4}=4$. In Fig. 2(c), at time $t=2$, because $v_6$ has
already influenced by clarifying cascade, $v_6$ cannot influence $v_6$ again.
Thus, only $v_5$ may influence $v_6$. Because $C_{v_5}=4 > R_1=3$, the state of
$v_6$ stays the same. In addition, node $v_1$ will also be finally influenced
by $v_1$ and $C_{v_1}=4$ and $R_2=2$. Finally, no nodes that can be influenced
by any cascade.

In this example, under the rumor-clarifying cascade, node $v_1$ is
influenced first by rumors and then finally by positive information.
Its state has been changed twice and its credibility value has also
been changed twice. Therefore, our algorithm LEH uses the longest
distance, the number of neighbor nodes and the number of rumor nodes
contained in out-neighbors to select the seed nodes. Compared with
the classic greedy algorithm, this selection mechanism makes it faster
to find seed nodes in the rumor-clarifying cascade. Meanwhile,
the algorithm LEH also uses the number of rumor nodes contained in
the neighbors of the seed node as the initial credibility, which makes the
total credibility less than the total credibility of random algorithm
and greedy algorithm. The more details and properties of the algorithm LEH
will be introduced in Section 4.

3.2. X-rumor-clarifying cascade

In this cascade model, each node $v$ is initially inactive, the state
of each node $v$ can be changed at most $x (2 < x < +\infty)$ times, but
it can only be influenced by alternating between rumor and positive
information. If the rumor cascade influence the node $v$ at time $t$, then
the rumor cascade cannot influence the node $v$ at time $t+1$ unless the
node $v$ is influenced again by the clarifying cascade.

Fig. 3 shows an example of x-rumor-clarifying cascade when $x = 3$.
As shown in Fig. 3, there are 7 nodes. We assume that the influence
probability of each edge is $p_{u,v}=1$ and $R_i=2$, $C_{v_1}=5$, $C_{v_3}=1$
and $C_{v_1}=1$. In Fig. 3(a), at time $t = 0$, $v_1$, $v_4$ are activated by rumor
cascade and $v_2$, $v_5$, $v_7$ are activated by clarifying cascade. Other nodes
are inactivated. In Fig. 3(b), at time $t = 1$, $v_3$ will be influenced as a
rumor node and $v_4$, $v_6$ will be influenced as clarification nodes. Then
$R_1=3$, $C_{v_3}=5$ and $C_{v_6}=1$. In Fig. 3(c), at time $t = 2$, because
$C_{v_2}=1 < R_1=3 < C_{v_5}=5$, $v_2$ will be influenced by $v_3$ again as
a clarification node and $v_3$ will be influenced by $v_5$ as a clarification
node. In Fig. 3(d), at time $t = 3$, because $C_{v_1}=1 < R_3=3$, $v_1$
will be influenced by $v_3$ as a rumor node and $R_3=4$. Finally, the state of
$v_1$ has been changed 3 times and there are no more nodes that can be
influenced by any cascade.

In this example, under the x-rumor-clarifying cascade, the state of
node $v_1$ has been changed three times and the state of nodes $v_6$, $v_7$ have
been changed twice. In algorithm LEH, we propose a two-way increase
and decrease mechanism to ensure that the difference between the final
number of clarified nodes and the given value is as small as possible.
In addition, compared with other heuristic algorithms, algorithm
LEH can effectively guarantee the approximate ratio of the results by
adjusting the increment of credibility each time. The more properties
of algorithm LEH in solving X-MCRC problem will be discussed in
Section 4.

3.3. Problem definition

In this part, we will give the definition of two rumor clarifying prob-
lems based on the rumor-clarifying cascade and the x-rumor-clarifying
cascade, respectively.

Minimum Credibility Rumor Clarifying (MCRC) problem. In a
rumor-clarifying cascade directed social network $G = (V, E)$, it has
$n$ nodes and $m$ directed edges. Given a rumor seed set $S_r$, a positive
number $q \in (0, 1)$, positive integer $k$ and $t$. Let $f_t(S_r)$ be the expected
number of nodes which are in the clarifying cascade at time $t$ when

$X. Yao et al.
For MCRC problem, \( f_i(S_k) \) can be defined as follows:

\[
\begin{align*}
    f_i(S_k) &= \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(C_{N_t(i)}) \\
    &\quad - \sum_g \Pr[g] \sum_{t=1}^{g} W_{i,t-1}(R_{i-1}) \\
    &\quad + \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(C_{N_t(i)} - R_{i-1}) \\
    &\quad + \ldots \\
    &\quad + \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(C_{N_t(i)} - R_{i-1}) \\
    &= \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(C_{N_t(i)} - R_{i-1}) \\
    &\quad + \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(C_{N_t(i)} - R_{i-1}) \\
\end{align*}
\]

where \(|h_{i,0}| = k, |h_{i-1,0}| = r, |h_{0,0}| = 0, |h_{0,0}| = n - k - r.

Eq. (1) represents that \( f_i(S_k) \) is composed of 3 main parts:

(i) \( \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(C_{N_t(i)}) \) denotes the total number of nodes converted from rumor nodes to clarification nodes at each moment.

(ii) \( \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(R_{i-1}) \) denotes the total number of nodes converted from clarification nodes to rumor nodes at each moment.

(iii) \( \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(C_{N_t(i)} - R_{i-1}) \) denotes the total number of nodes converted from inactive nodes to clarification nodes at each moment.

For \( X - MCRC \) problem, \( f_i(S_k) \) can be defined as follows:

\[
\begin{align*}
    f_i(S_k) &= \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(C_{N_t(i)}) - \sum_g \Pr[g] \sum_{t=1}^{g} W_{i,t-1}(R_{i-1}) \\
    &\quad + \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(C_{N_t(i)}) \\
    &\quad - \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(R_{i-1}) \\
    &\quad + \ldots \\
    &\quad + \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(C_{N_t(i)}) - \sum_g \Pr[g] \sum_{t=1}^{g} W_{i,t-1}(R_{i-1}) \\
    &= \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(C_{N_t(i)} - R_{i-1}) \\
    &\quad + \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(C_{N_t(i)} - R_{i-1}) \\
\end{align*}
\]

Eq. (2) represents that \( f_i(S_k) \) is composed of 4 main parts:

(i) \( \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(C_{N_t(i)} - R_{i-1}) \) denotes the rumor nodes where the state has changed less than \( t \) times at time \( t - 1 \) that become clarification nodes at time \( t \).

(ii) \( \sum_g \Pr[g] \sum_{t=0}^{g} W_{i,t-1}(C_{N_t(i)} - R_{i-1}) \) denotes the total number of nodes converted from rumor nodes to clarification nodes at time \( t - 1 \).
negative credibility of the rumor is.

Suppose that $\sum_{t \geq 1} Pr[\exists \text{rumor nodes at time } t \geq 1 - 1]$. $\sum_{t \geq 1} Pr[\exists \text{clarifying cascade at time } t \geq 1 - 1]$. Let $f_t(A) = \text{expected number of nodes which are in the clarifying cascade at time } t \geq 1$ when the sum of the positive credibility of nodes in $A \subseteq V$. Therefore, $f_t(A) \leq f_t(B)$.

### Corollary 1. The function $f_t(A)$ of MRC problem is non-decreasing.

### Corollary 2. The function $f_t(A)$ of MRC problem has no monotonicity.

### Corollary 3. The function $f_t(A)$ of MRC problem is supermodular.

Proof. Suppose that $A \subseteq B \subseteq V$. According to each node in $A$ or $B$ has equal positive credibility, we can know that before the spread stops, the rumors can be clarified at time $t \geq 1$ only if $C_t > R_t(j \in 1, 2, \ldots, t)$. Then we can indicate that $t_A \leq t_B$, where $t_A$ denotes the expected time of spread when $A$ is selected as the seed set of the clarifying cascade and $t_B$ denotes the expected time of spread when $B$ is selected as the seed set of the clarifying cascade. Therefore, $f_t(A) \geq f_t(B) \geq 0$.

### Corollary 5. Independent variable $C_{\text{min}}$ is nonlinearly related to dependent variable $I_t(C_{\text{min}})$.

Proof. According to the rumor-clarifying cascade, when the positive credibility is greater than the negative credibility, node $v$ can be influenced by clarifying cascade. Then suppose there is a node $v$ with the negative credibility of 3 and one node $w \in S_t$. Moreover, assume that the positive credibility of node $u$ is 1, 2 and 4 in the case of different $C_{\text{min}}$. Then we have the following three cases:

- **Case 1.** The credibility of node $u$ is 1, then $C_{\text{min}} = 1$.
- **Case 2.** The credibility of node $u$ is 2, then $C_{\text{min}} = 2$.
- **Case 3.** The credibility of node $u$ is 4, then $C_{\text{min}} = 4$.

We can infer that $I_t(C_{\text{min}}) = I_t(C_{\text{min}}) < I_t(C_{\text{min}}')$. Thus, $(I_t(C_{\text{min}})/ I_t(C_{\text{min}}')#(I_t(C_{\text{min}})/ I_t(C_{\text{min}}'))$. Therefore, there is no linear correlation between dependent variable and independent variable.

### Theorem 3.4. Calculating the $f_t(S_t)$ is #P-hard.

Proof. As shown in [35] that $s - t$ connectedness problem is #P-complete. Based on the [36], this problem is equivalent to calculating the probability of connecting two nodes when each edge in $G$ is connected with a probability of $1/2$.

Then we reduce this problem to the calculating the $f_t(S_t)$ as follows. Let $f_t(S, G)$ denote the expected number of nodes which are in the clarifying cascade at time $t \geq 1$ when $S_t$ is selected as the seed set of the clarifying cascade. Without loss of generality, we suppose that $S_t = \{v_1, v_2, \ldots, v_k\}$ and $p_{ij} = 0.5$ for all edges $e \in E$. Then we add nodes $u_i \in \{1, 2, 3, \ldots\}$ and the corresponding directed edges from $u_i \rightarrow v_j$ to the graph $G$ to make up the new graph $G'$. Let $p_{S_t, U, G'}$ denote the probability that $U$ is influenced by $S_t$ in $G$. We can indicate that $f_t(S_t, G') = f_t(S_t, G) + p_{S_t, U, G} \cdot p_{S_t, U, G'}$, where $U = \{u_j\} = \{1, 2, 3, \ldots\}$. Therefore, let $p_{S_t, U, G} = 1$, the probability that any node $u_i(i \in \{1, 2, \ldots, k\})$ connects to node $w_j$ in $G'$ is $p_{S_t, U, G'}$. Thus we converted the problem of calculating the $f_t(S_t)$ to the $s - t$ connectedness problem and it is #P-hard.

### Corollary 6. Calculating the $f_t(S_t)$ is #P-hard.

### Corollary 7. The MRC problem and $X - MRC$ problem are both N P-hard.

Lemma 3.5. For MRC problem, given the same $C_{\text{min}}$ when $f_t(S_t) = f_t(S_t')$, $t \leq t'$ where $t'$ is the spread time in the case where each node in the clarifying cascade seed set $S_t$ has equal positive credibility $C'_t$.

Proof. Assume that $S_t = \{v_1, v_2, \ldots, v_k\}$, $C_{t_1} \geq C_{t_2} \geq \cdots \geq C_{t_k}$, $\sum_{t_i} C_{t_i} = C_{\text{min}}$ and $S'_t = \{w_1, w_2, \ldots, w_k\}$, $C'_{t_1} \geq C'_{t_2} \geq \cdots \geq C'_{t_k} = C_{\text{min}}/k$. Then we divide possibilities into the following two cases:

- **Case 1.** $C_{t_1} = C_{t_2} = \cdots = C_{t_k}$
  It is obviously that when $C_{t_1} = C_{t_2} = \cdots = C_{t_k}$, $t = t'$.

- **Case 2.** $C_{t_1}, C_{t_2}, \ldots, C_{t_k}$ are not exactly the same
  Without loss of generality, suppose that $C_{t_1} > C_{t_2} \geq \cdots \geq C_{t_k}$. We can indicate that $C_{t_1} = C_{\text{min}}/k = C'_{t_1} = C'_{t_2} = \cdots = C'_{t_k} \geq \cdots \geq C_{t_k}$. The greater the value of credibility, the greater the number of points that can be clarified. Thus $t \leq t'$.

### Corollary 8. For $X - MRC$ problem, given the same $C_{\text{min}}$ when $f_t(S_t) = f_t(S_t')$, $t \leq t'$ where $t'$ is the spread time in the case where each node in the clarifying cascade seed set $S_t$ has equal positive credibility $C'_t$. 


Theorem 3.6. In a rumor-clarifying cascade directed social network \( G = (V,E) \) with \( n \) nodes, the spread time \( t \) is finite and \( t \leq n - 1 \).

Proof. In a rumor-clarifying cascade, the state of each node can be changed at most twice. Suppose that there is one rumor node \( v_1 \) and one clarification node \( v_2 \) at time \( t = 0 \). From time \( t = 0 \), both node \( v_1 \) and \( v_2 \) start spreading until all nodes are influenced twice. Because \( v_1 \) can influence at most \( n - 1 \) (except itself) nodes and \( v_2 \) can influence at most \( n - 1 \) (except itself) nodes. Thus, \( t \leq (n - 1) \).

Corollary 9. In a \( x \)-rumor-clarifying (\( 2 < x < +\infty \)) cascade directed social network \( G = (V,E) \) with \( n \) nodes, the spread time \( t \) is finite and \( t \leq (x - 1)n - 1 \).

Proof. In an \( x \)-rumor-clarifying cascade, the state of each node can be changed at most \( x \) times. Suppose that there is one rumor node \( v_1 \) and one clarification node \( v_2 \) at time \( t = 0 \). From time \( t = 1 \) until all \( n \) nodes' state are changed \( x \) times, then we can infer that the spread time \( t = (x - 1)n \) and the states of node \( v_1 \) and \( v_2 \) can no longer be changed. Thus, at most after time \( n - 1 \), the states of other \( n - 2 \) nodes will be changed again. Therefore, the spread time \( t \leq (x - 1)n - 1 \).

4. Algorithm and theoretical analysis

4.1. Algorithm LEH

We propose the Longest-Effective-Hops Algorithm (LEH) for solving MCRC problem and \( X - MCRC \) problem. Compared with the classic greedy algorithm, algorithm LEH uses the longest distance, the number of neighbor nodes and the number of rumor nodes contained in out-neighbors to select the seed nodes. We also propose a two-way increase and decrease mechanism to ensure that the difference between the final number of clarified nodes and the given value is as small as possible. Meanwhile, we add a parameter \( \epsilon \) to the algorithm LEH to represent the increment of credibility each time, which ensures our algorithm LEH obtain effective approximation ratio. The details of phase I and II of LEH are shown in Algorithm 1 and Algorithm 2, respectively.

In the proposed algorithm, each node \( v_i \) has a unique node id, \( id \in ID, ID = \{1, 2, 3, \ldots \} \). Let \( d(v, u) \) be the length of the shortest effective path from node \( v \) to node \( u \). Moreover, for a node set \( V' \), we define \( d(V', u) = \min\{d(v, u) : v \in V'\} \). Let \( l \) denote the length of the longest distance in the \( G = (V,E) \).

In Algorithm 1 (LEH phase I), we try to find clarification seed set \( S_i \) in a greedy manner. At the beginning, every node (except rumor seed nodes) is inactive. For each inactive node, the algorithm selects a node as a clarification seed node according to the following priorities: (1) \( \sum_{v_i} N(x) \cdot p^x_v \) is the largest. (2) \( \sum_{v_i} N(x) \) including the largest number of rumor nodes. (3) \( v \) has the smallest id number.

The initial positive credibility of each clarification seed node \( v^* \) is equal to the sum of \( d(v^*, S_i) \) and the number of rumor nodes in its all hops neighbors. The algorithm will terminate when \( f_i(S_i) \) is equal to the given number. If \( f_i(S_i) \) is more than the given number, the algorithm will reduce the positive credibility of the nodes in \( S_i \) in turn according to the degree of positive credibility. If \( f_i(S_i) \) is less than the given number, then do Algorithm 2 (LEH phase II).

In Algorithm 2 (LEH phase II), we try to increase the number of clarification nodes by increasing the positive credibility of the clarification seed nodes. At the beginning, the algorithm sorts the nodes in \( S_i \) according to the following priorities: (1) \( v^* \) is nearest to \( S_i \). (2) \( \sum_{v_i} N(x) \) including the largest number of rumor nodes. (3) \( \sum_{v_i} N(x) \cdot p^x_v \) is the largest. (4) \( v \) has the smallest id number.

First, the algorithm increases the positive credibility of the first seed node by \( \epsilon \), where \( \epsilon \) denotes the value of increased positive credibility at each time and it is artificially set by us (\( \epsilon \geq 1 \)). After the positive credibility of the node increases, if the difference between \( f_i(S_i) \) and the previous \( f_i(S_i) \) is greater than \( \epsilon \), the algorithm repeatedly increases the positive credibility of the same node by \( \epsilon \). If not, then the algorithm will increase the positive credibility of each sorted seed node by \( \epsilon \) in turn. Finally, the algorithm will terminate when \( f_i(S_i) \) is equal to or more than the given number.

4.2. Theoretical analysis of algorithm LEH

Lemma 4.1. For MCRC problem, according to the Algorithm LEH, when the spread time of rumor are same and \( f_i(S_i) = f_i'(S_i) \),

\[ C_{min} \leq kC' \]

where

\[ C_{min} = \sum_{i=1}^{k} C_{v_i} | C_{v_1} \geq C_{v_2} \geq \cdots \geq C_{v_k} \] and \( C' = (\sum_{i=1}^{k} C'_{v_i})/k \),

\( C'_{v_1} = C'_{v_2} = \cdots = C'_{v_k} \).

Proof. Assume that \( S_i = \{v_1, v_2, \ldots, v_k\} \), \( C_{v_1} \geq C_{v_2} \geq \cdots \geq C_{v_k} \) and \( S_i' = \{v_1', v_2', \ldots, v_k'\} \), \( C'_{v_1} = C'_{v_2} = \cdots = C'_{v_k} \). Let \( t' \) be the spread time in the case where each node in the clarifying cascade seed set \( S_i' \) has equal positive credibility \( C' \). We can know that when \( t = t' \), \( f_i(S_i) = f_i'(S_i') \). Then we divide possibilities into the following two cases:

- Case 1. \( C_{v_i} = C_{v_j} = \cdots = C_{v_k} \).
  It is obviously that when \( C_{v_i} = C_{v_j} (i \in \{1, 2, \ldots, k\}) \), \( C_{min} = kC = kC' \).

- Case 2. \( C_{v_1}, C_{v_2}, \ldots, C_{v_k} \) are not exactly the same
Algorithm 2: (LEH phase II). Input: $S_c$, $C_{\min}$, $k$, $C_A$, $t$, $p_{(i,j)}$, $\eta$ and $c$. Output: The $C_{\min}$, $S_c$.

1. Initialization: $i = 0$, $W = \emptyset$, $F = 0$. Let $M$ be a positive number and $M > C_{\min}$.
2. Let $A = \max\{\sum_{i \in S_c} |N^+(v_i)| \cdot p_{(i,j)}^l |v_i \in S_c\}$
3. $B = \max\{\sum_{i \in S_c} |N^+(v_i) \cap S_c| \cdot p_{(i,j)}^l |v_i \in S_c\}$

4. for $S_c \neq \emptyset$ do
   5. Find the nodes which are nearest to $S_c$
   6. if there is one node $v'$ which is nearest to $S_c$, then
      7. $W = W \cup \{v'\}$, $S_c = S_c \setminus v'$
   else
      9. Find the nodes which have the largest $B$
      10. if there is one node $v'$ which has the largest $B$ then
          11. $W = W \cup \{v'\}$, $S_c = S_c \setminus v'$
          else
                  13. Find the nodes which have the largest $A$
                  14. if there is one node $v'$ which has the largest $A$ then
                     15. $W = W \cup \{v'\}$, $S_c = S_c \setminus v'$
                     else
                     16. $W = W \cup \{v'\}$, $S_c = S_c \setminus v'$, where $v'$ has the smallest id

18. Renumber all the nodes in $W$ in order that $W = \{w_1, w_2, \ldots, w_k\}$.
19. while $f_i(S_c) < \eta n$ do
20.    while $M > c$ do
21.       $C_{w_i} = C_{w_i} + |v|$, $C_{\min} = |v| + C_{\min}$, $M = f_i(S_c) - F$, $F = f_i(S_c)$
22.       if $i = k$ then $i = 1$ else $i = i + 1$
23. end
24. end

We use contradiction to prove this case. Without loss of generality, suppose that $C_{\min} = kC'$ and $C_{w_i} > C_w \geq \cdots \geq C_{w_j}$. We can indicate that $C_{w_j} > kC' = C_{w_1} = \cdots = C_{w_j} \geq \cdots \geq C_{w_k}$. According to the algorithm LEH, for node $v_i$ and $w_i$, because $C_{w_i} > C_{w_j}$, then $d(S_c, v_i) < d(S_c, w_i)$ and $d(S_c, w_i) = d(S_c, w_j) \leq d(S_c, v_i) \leq \cdots \leq d(S_c, v_j) \in (2, 3, \ldots, k)$. Thus at the same time, nodes $v_i$ clarifies more nodes than nodes $w_i$. We can conclude that when $f_i(S_c) = f'_i(S_c)$, $t < t'$. This contradicts with the assumption. Therefore, $C_{\min} < kC'$.

Corollary 10. For $X \sim MCRC$ problem, according to the Algorithm LEH, when the speed time of rumor are same and $f'_i(S_c) = f'_i(S_c)$,

$C_{\min} \leq kC'$

where $C_{\min} = \sum_{i=1}^{k} C_{w_i}(C_{v_i} \geq C_{v_2} \geq \cdots \geq C_{v_k})$ and $C' = (\sum_{i=1}^{k} C_{w_i})/k$,

$C_{w_i} = C_{w_1} = \cdots = C_{w_k}$

Theorem 4.2. For $MCRC$ problem, the Algorithm LEH can find $C_{\min}$ in the finite-time period.

Proof. The proof process is divided into the following 2 parts:

Part I For Algorithm 1 (LEH phase I)

During each round of running of the algorithm, the algorithm picks only one node. Each clarification node can be select according to the number of neighbor nodes and rumor nodes in its neighbors and node id. After $k$ rounds, the terminal condition satisfied that $k$ clarification seed nodes are already selected.

Then the algorithm will calculate $f_i(S_c)$. If $f_i(S_c)$ is equal to the given number, the algorithm will terminate and output the $C_{\min}$. If $f_i(S_c)$ is less than the given number, the algorithm will do Algorithm 2 (LEH phase II). Otherwise, the algorithm will reduce the positive credibility of the nodes in $S_c$ in turn according to the degree of positive credibility and the algorithm will terminate when $f_i(S_c)$ is equal to the given number. We prove by contradiction that this terminal condition must be satisfied in finite time. Assume that when $C_{\min}$ goes down to 1, $f_i(S_c)$ is still more than the given number. Then we can infer that the positive credibility of each point in $S_c$ is less than 1. According to the rumor-clarifying cascade, clarification seed nodes cannot clarify any other node. This contradicts our assumption.

Part II For Algorithm 2 (LEH phase II)

During each round of running of the algorithm, the algorithm picks only one node to sort. After $k$ rounds, The terminal condition satisfied that $k$ clarification seed nodes are already sorted.

Then the algorithm will recalculate $f_i(S_c)$. If $f_i(S_c)$ is equal to or more than the given number, the algorithm terminate. We also use contradiction to prove that this terminal condition must be satisfied in finite time. Assume that when $C_{\min}$ goes up to $kn$, $f_i(S_c)$ is still less than the given number. We can infer that no matter how many rumor nodes there are, these rumor nodes can always be clarified. This contradicts our assumption.

Corollary 11. For $X \sim MCRC$ problem, the Algorithm LEH can find $C_{\min}$ in the finite-time period.

Theorem 4.3. The Algorithm LEH is correct.

Proof. For Algorithm 1 (LEH phase I), the proof process is divided into the following 2 parts:

Part I Step1–step17.

For the first iteration of the loop, there is only one node in set $S_c$ and it has the largest number of all hops neighbor nodes and it is sorted. It is indicated that the loop invariant ($i = 1$) holds for the first iteration of the loop.

For the next iterations of the for loop, each time, the nodes with the largest number of all hops neighbor nodes will be added to $S_c$ in order and there are sorted. Then each iteration can always maintain this invariant($i = 2, 3, \ldots, k$). In the meantime, the $S_c$ has been replaced by the current $S_c$.

Finally, when the for loop terminates with $i > k$, $|S_c| = k$. For each iteration of the loop, $i$ increases by 1, then it must have the case of that $i$ is equal to $k + 1$. If we replace $i = k + 1$ in the circular invariant, we cannot find any node that has more neighbor nodes than a node in $S_c$. Therefore, we can infer that there are already $k$ nodes in $S_c$ and the nodes in $S_c$ are already sorted by the number of all hops neighbor nodes. Therefore, the step1–step17 are correct.

Part II Step18–step25.

If $f_i(S_c) = \eta n$, the algorithm is terminated and output the $C_{\min}$. If $f_i(S_c) > \eta n$, the algorithm will reduce the positive credibility of the nodes in $S_c$.

First, before the first iteration of the loop ($f_i(S_c) > \eta n$), $C_{\min}$ has not changed. For the next iterations of the for loop ($f_i(S_c) > \eta n$), $C_{\min}$ is gradually decreasing. Then each iteration can always maintain this invariant true ($f_i(S_c) > \eta n$). Finally, when the for loop terminates with $f_i(S_c) \leq \eta n$. For each iteration of the loop, $f_i(S_c)$ decreases, then it must have the case of that $f_i(S_c) < \eta n$. If we replace ($f_i(S_c) \leq \eta n$) with ($f_i(S_c) < \eta n$) in the circular invariant, $C_{\min}$ stays the same. Therefore, the step18–step25 are correct.

For Algorithm 2 (LEH phase II), we divide the proof process into the following 2 main parts:

Part III Step1–step18.

For the first iteration of the loop, there are $k$ nodes in set $S_c$ and it is sorted, but there is only one node in set $W$ and it is sorted.

For the next iterations of the for loop ($|S_c| = k−1, k−2, \ldots, 1$), each time, the node nearest to $S_c$ in $S_c$ is added to $W$. The nodes in $W$ are sorted and the $W$ has been replaced by the current $W$. Therefore, each iteration can always maintain this invariant true ($|S_c| = k−1, k−2, \ldots, 1$).

Finally, when the for loop terminates with $|S_c| = 0$, $|W| > k$. For each iteration of the loop, $|S_c|$ decreases by 1, then it must have the
Fig. 4. Total positive credibility under Power5000 when $p(v, w) = 0.1$, $t = 5$, $\eta = 0.1$.

Fig. 5. Total positive credibility under Wiki-Vote when $p(v, w) = 0.1$, $t = 5$, $\eta = 0.1$.

Fig. 6. Total positive credibility under Soc-Slashdot0811 when $p(v, w) = 0.1$, $t = 10$, $\eta = 0.1$.

Fig. 7. Running time of different algorithms under different datasets when $p(v, w) = 0.1$.

case of that $|S_c|$ is equal to 0. If we replace $|S_c| = 1$ with $|S_c| = 0$ in the circular invariant, $W$ stays the same and $W$ is already all of the nodes which are all in order. Therefore, the step1–step18 are correct.

Part IV  Step19–step23.
Before the first iteration of the loop $(f_t(S_c) < \eta n)$, $C_{min}$ has not changed.

For the next iterations of the loop $(f_t(S_c) < \eta n)$, if $M > \epsilon$, $C_{min}$ is gradually increasing. Otherwise, the algorithm does the appropriate operations to satisfy $M > \epsilon$. Then each iteration can always maintain this invariant true $(f_t(S_c) < \eta n)$.

Finally, when the loop terminates with $f_t(S_c) \geq \eta n$. For each iteration of the loop, $(f_t(S_c))$ increases, then it must have the case of that $f_t(S_c) > \eta n$. If we replace $(f_t(S_c) \geq \eta n)$ with $(f_t(S_c) > \eta n)$ in the circular invariant, $C_{min}$ stays the same. Therefore, the step19–step23 are correct.

\textbf{Theorem 4.4.} For the MCR problem, the time complexity of the Algorithm LEH is $O(km(n - r)/|\epsilon|)$, where $k$ is the number of seed nodes, $m$ is the number of edges, $n$ is the total number of nodes and $r$ is the number of rumor nodes.

\textbf{Proof.} In Algorithm LEH, for $k$ clarifying seed nodes, each time when we find the clarifying seed node, we need to find it within the maximum hops and search from at most $(n - r)$ nodes. Then when we calculate $f_t(S_c)$, we loop $n$ times at most and each time the algorithm repeatedly increases the positive credibility of the node by $|\epsilon|$. Therefore, the time complexity is $O(km(n - r)/|\epsilon|)$.

\textbf{Corollary 12.} For the $X$–MCR problem, the time complexity of the Algorithm LEH is $O(km(n - r)(x - 1)/|\epsilon|)$, where $k$ is the number of nodes in the clarifying seed set, $m$ is the number of edges, $n$ is the total number of nodes and $r$ is the number of rumor nodes.

\textbf{Theorem 4.5.} For the MCR problem, the approximation of the Algorithm LEH is not less than $1/(\epsilon)(1/k)(1 - 1/\epsilon)$, where $\epsilon \geq 1$ and $k$ is the number of seed nodes.
1 We consider other three rumor blocking algorithms for comparison and these algorithms shown as follows:

(i) **Greedy**. This method uses Monte Carlo simulation.

(ii) **Random**. This method is to randomly select positive seed nodes.

(iii) **Proximity**. This method is to select the out-neighbors of the rumor seed nodes as the positive seed nodes.

In our experiment, all experiments are performed using codes written in C++ on an Intel(R) Xeon(R) Gold 5117 with 2.00 GHz CPU and 64 GB RAM, we set the probability on the edges to be either uniformly \( p_{(u,v)} = 0.1 \) or \( p_{(u,v)} = 1/d(u) \) [34], where \( d(u) \) is the in-degree of the node. The probability \( p_{(u,v)} \) is small, we set the value of \( \eta \) to be relatively small. Moreover, let \( R \) denote the number of rumor seed nodes and \( k \) denote the number of clarification seed nodes. The parameters of these parameters are shown in Table 3.

### 5.2. Results of MCRC problem

The analysis of the experimental results include the following three parts.

(i) **Comparisons under** \( p_{(u,v)} = 0.1 \) and \( \epsilon = 2 \)

#### Case 1. Comparisons of the total positive credibility

Fig. 4 and Fig. 5 show the total positive credibility of four algorithms under dataset Power5000 and Wiki-Vote when \( p_{(u,v)} = 0.1 \), \( t = 5 \), \( \eta = 0.1 \), respectively.

In Fig. 4, we can see that except the algorithm **Greedy**, the total positive credibility of other three algorithms increases with the number of clarification seed nodes when the \( R \) are the same. Fig. 4(a)–(d) show that the larger \( R \) is, the more total positive credibility is required. From Fig. 4, it can be inferred that the total positive credibility of algorithm LEH is slightly less than the other three algorithms’.

Fig. 5 shows similar trends as Fig. 4, we can see that except the algorithm **Greedy**, the total positive credibility of other three algorithms increases with the number of clarification seed nodes when the \( R \) are the same. Fig. 5(a)–(d) show that the larger \( R \) is, the more total positive credibility is required. From Fig. 5, it can be inferred that the total positive credibility of algorithm LEH is almost the same as the greedy algorithm's and it is less than the other two algorithms'.

Fig. 6 shows the total positive credibility of three algorithms (except algorithm **Greedy**) under dataset Soc-Slashdot0811 when \( p_{(u,v)} = 0.1 \), \( t = 10 \), \( \eta = 0.1 \). Because the running time of algorithm **Greedy** is more than \( 10^5 \) seconds, we do not show the results of the algorithm **Greedy**. From Fig. 6, it can be inferred that the total positive credibility of algorithm LEH is less than the other two algorithms’.

Overall, from Figs. 4–6, it can be inferred that the results our algorithm LEH are slightly better than the other three algorithms’ and

### 5. Experiment

#### 5.1. Datasets and parameters

In this section, we evaluate the performance of our algorithms with three datasets. We selected two real-world networks and one synthetic power-law network which has one of the most important features of social networks: a power-law distribution [37]. The details of these datasets are shown in Table 2, while the brief description of networks is given below.

**Power5000**: This dataset is artificially generated synthetic power-law network.

**Wiki-Vote**: This dataset is who-votes-on-whom network from Wikipedia and it is provided by the SNAP.1

**Soc-Slashdot0811**: This dataset is Slashdot social network from November 2008 and it is provided by the SNAP.1

1 https://snap.stanford.edu/
compare the result of the algorithm LEH with the \( \text{Greedy} \) on dataset Soc-Slashdot0811 to get valid results. Therefore, we only get valid results on dataset Power5000 and Wiki-Vote, but it needs to more than \( 3 \times 10^5 \) seconds. Furthermore, Algorithm \( \text{Proximity} \) fails to give a valid result no matter how much time it takes. Algorithm LEH can always get valid results in \( p \) different datasets when \( \eta \) is the same. Fig. 7(c) and (b) also show similar trends as Fig. 7(a). We do not show the results of the algorithm \( \text{Greedy} \) here, because the running time of it is more than \( 3 \times 10^5 \) seconds.

\( \eta \) denotes that the algorithm can run effectively and get a valid result in \( 3 \times 10^5 \) seconds for each set of experiments. \( \circ \) denotes that the algorithm cannot get the result in \( 3 \times 10^5 \) seconds for each set of experiments. \( \Box \) denotes that no matter how much time is given, the algorithm always fails to give a valid result.

Table 4 shows the availability of different algorithms under different datasets when \( p_{(v,w)} = 1/d(w) \).

| Algorithm | Availability |
|-----------|--------------|
| Power5000 | Wiki-Vote    | Soc-Slashdot0811 |
| \( t = 5 \) | 8            | 5               | 8               |
| \( \eta = 0.06 \) | 0.1          | 0.04            | 0.04            |

| Algorithm | Availability |
|-----------|--------------|
| LEH       | ✓            | ✓               | ✓               |
| \( \text{Greedy} \) | ✓            | ✓               | ✓               |
| \( \text{Random} \) | ✓            | ✓               | ✓               |
| \( \text{Proximity} \) | ✓            | ✓               | ✓               |

\( \Box \) denotes that each node can be influenced at most 2, 3 and 5 times for each set of experiments. \( \Box \) denotes that each node can be influenced at most 2, 3 and 5 times for each set of experiments. \( \Box \) denotes that each node can be influenced at most 2, 3 and 5 times for each set of experiments.

Case 2. Comparisons of the total positive credibility

Fig. 8 shows the total positive credibility of LEH and under different datasets and \( \eta \) when \( t = 8 \), \( p_{(v,w)} = 1/d(w) \). In Fig. 8, we can see that the total positive credibility of algorithm LEH is less than the \( \text{Greedy} \)'s on the same dataset and it also shows that the larger \( R \) is, the more total positive credibility of LEH is required.

Overall, from Table 3 and Fig. 8, it can be inferred that our algorithm LEH works better than the algorithm \( \text{Greedy} \) and it runs in less time.

5.3. Results of X-MCRC problem

In this experiment, we set \( \epsilon = 2 \), \( p_{(v,w)} = 0.1 \) and \( x = \{2, 3, 5\} \) on dataset Wiki-Vote. \( x = 2, x = 3 \) and \( x = 5 \) denote that each node can be influenced at most 2, 3 and 5 times, respectively. In Fig. 9, we can see that the total positive credibility of LEH is almost same when \( \epsilon = \{1, 2, 3, 5\} \). Fig. 10 shows that the running time of LEH decreases with the \( \epsilon \).

6. Conclusions

In this paper, we have studied the minimum credibility rumor clarifying(MCRC) problem for online social networks. We first develop a rumor-clarifying cascade which allows each user to be influenced at most twice. Then we present a Longest-Effective-Hops (LEH) algorithm for solving the problem. Furthermore, we propose a x-rumor-clarifying cascade and it is extended theoretically. The proposed algorithm LEH theoretically dominates the existing algorithms, and extensive experiments show that our algorithm is more efficient and effective. Our future work is to investigate the clarifying the rumor in the case of dynamic selection of seed nodes.
CRediT authorship contribution statement

Xiaopeng Yao: Investigation, Conceptualization, Methodology, Prove theories, Design algorithm, Writing draft. Guangxian Liang: Generate some figures, Design algorithm and programming, Optimization of the approximation ratio. Chonglin Gu: Algorithm improvement and optimization, Optimization of time complexity, Revise the paper. Hejiao Huang: Put forward ideas, Discussion of algorithms and methods, Revise the paper, Put forward some corollaries.

Declaration of competing interest

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

Acknowledgments

This work is financially supported by National key R&D program, China under Grant No. 2017YFB0803002 and No. 2016YFB08008-04, National Natural Science Foundation of China under Grant No. 61672195 and No. 61732022, China Postdoctoral Science Foundation, 04, National Natural Science Foundation of China under Grant No.

References

[1] D. Kempe, J. Kleinberg, E. Tardos, Maximizing the spread of influence through a social network, in: Proc. 9th SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2003, pp. 137–146.
[2] C. Budak, D. Agrawal, A. El Abbadi, Limiting the spread of misinformation in social networks, in: Proc. of the 20th WWW, ACM, 2011, pp. 665–674.
[3] X. He, G. Song, W. Chen, Q. Jiang, Influence blocking maximization in social networks under the competitive linear threshold model, in: SDM, SIAM, 2012, pp. 463–474.
[4] L. Pan, Z. Lu, W. Wu, B. Thuraisingham, H. Ma, Y. Bi, Least cost rumor blocking in social networks, in: Proc. of the 33rd ICDCS, IEEE, 2013, pp. 540–549.
[5] E. Serrano, C.A. Iglesias, M. Garijo, A novel agent-based rumor spreading model in twitter, in: Proc. 24th Int. Conf. World Wide Web, 2015, pp. 811–814.
[6] Y. Zhou, B. Zhang, X. Sun, Q. Zheng, T. Liu, Analyzing and modeling dynamics of information diffusion in microblogging social network, J. Netw. Comput. Appl. (2016) 92–102.
[7] Y. Ping, Z. Cao, H. Zha, Sybil-aware least cost rumor blocking in social networks, in: Proc. Globecom, IEEE, 2014, pp. 692–697.
[8] V. Indu, S.M. Thampi, A nature - inspired approach based on forest fire model for modeling rumor propagation in social networks, J. Netw. Comput. Appl. (2019) 28–41.
[9] F. Chierichetti, S. Lattanzi, A. Panconesi, Rumor spreading in social networks, Theoret. Comput. Sci. (2011) 2602–2610.
[10] J.R.C. Piqueira, Rumor propagation model: An equilibrium study, Math. Probl. Eng. (2010) 242–256.
[11] Z. He, Z. Cai, J. Yu, X. Wang, Y. Sun, Y. Li, Cost-efficient strategies for restraining rumor spreading in mobile social networks, IEEE Trans. Veh. Technol. (2017) 1–1.
[12] S. Daum, K. Fabian, M. Yannic, Rumor spreading with bounded in-degree, Theoret. Comput. Sci. (2018) 43–57.
[13] E. Khalil, B. Dilkina, L. Song, Scalable diffusion-aware optimization of network topology, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Datamining, 2014, pp. 1226–1235.
[14] Kimura, M., Motoda, H., Blocking links to minimize contamination spread in a social network, ACM Trans. Knowl. Discov. Data 3 (2) (2009) 9:1–9:23.
[15] H. Habiba, Y. Yu, T.Y. Berger-Wolf, J. Saia, Finding spread blockers in dynamic networks, in: Proc. SNACKD, 2010, pp. 55–76.
[16] R.M. Tripathy, A. Bagchi, S. Mehta, A study of rumor control strategies on social networks, in: Proc. CIKM, 2010, pp. 1817–1820.
[17] N.P. Nguyen, G. Van, M.T. Thai, S. Eidenbenz, Containment of misinformation spread in online social networks, in: Proc. WebSci, 2012, pp. 213–222.
[18] H. Tong, B.A. Prakash, T. Eliaši-Rad, M. Faloutsos, C. Faloutsos, Gelling, and melting, large graphs by edge manipulation, in: Proc. of the 21th ACM CIKM, 2012, pp. 245–254.
[19] L. Yang, Z.W. Li, A. Giusa, Containment of rumor spread in complex social networks, Inform. Sci. (2020) 113–130.
[20] Z. Tan, D. Wu, T. Gao, I. You, V. Sharma, AIM: Activation increment minimization strategy for preventing bad information diffusion in OSNs, Future Gener. Comput. Syst. (2018) 293–301.
[21] A.I.E. Honni, K. Li, S. Ahmad, Minimizing rumor influence in multiplex online social networks based on human individual and social behaviors, Inform. Sci. (2020) 1458–1480.
[22] H. Zhang, H. Zhang, X. Li, M.T. Thai, Limiting the spread of misinformation while effectively raising awareness in social networks, in: Proc. Int. Conf. Comput. Social Netw., 2015, pp. 35–47.
[23] S. Wen, J. Jiang, Y. Xiang, S. Yu, W. Zhou, W. Jia, To shut them up or to clarify: Restraining the spread of rumors in online social networks, IEEE Trans. Parallel Distrib. Syst 25 (12) (2014) 3306–3316.
[24] B. Wang, G. Chen, L. Fu, L. Song, X. Wang, DRIMUX: Dynamic rumor influence minimization with user experience in social networks, IEEE Trans. Knowl. Data Eng. 29 (10) (2017) 2168–2181.
[25] Z. He, Z. Cai, J. Yu, X. Wang, Y. Sun, Y. Li, Cost-efficient strategies for restraining rumor spreading in mobile social networks, IEEE Trans. Veh. Technol. 66 (3) (2017) 2789–2800.
[26] P. Zhang, Z. Bao, Y. Niu, Y. Zhang, S. Mo, F. Gong, Z. Peng, Prusive rumor control in online networks, World Wide Web 22 (4) (2019) 1799–1818.
[27] W. Chen, C. Wang, Y. Wang, Scalable influence maximization for prevalent viral marketing in large-scale social networks, in: Proc. 16th SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 1029–1038.
[28] C. Zhao, S. Zhong, Q. Zhong, K. Shi, Synchronization of Markovian complex networks with input mode delay and Markovian directed communication via distributed dynamic event-triggered control, Nonlinear Anal. Hybrid Syst. 36 (2020) 10883.
[29] C. Zhao, S. Zhong, X. Zhang, Q. Zhong, K. Shi, Novel results on nonfragile sampled-data exponential synchronization for delayed complex dynamical networks, Internat. J. Robust Nonlinear Control 30 (10) (2020) 4022–4042.
[30] K. Shi, J. Wang, S. Zhong, Y. Tang, J. Cheng, Non-fragile memory filtering of T–S fuzzy delayed neural networks based on switched fuzzy sampled-data control, Fuzzy Sets and Systems 394 (2020) 40–64.
[31] Q. Fang, X. Chen, Q. Nong, Z. Zhang, Y. Cao, Y. Feng, et al., General rumor blocking: An efficient random algorithm with martingale approach, Theoret. Comput. Sci. (2020) 82–93.
[32] G. Tong, W. Wu, L. Guo, D. Li, C. Liu, B. Liu, D. Du, An efficient randomized algorithm for rumor blocking in online social networks, IEEE Trans. Netw. Sci. Eng. 7 (2) (2020) 845–854.
[33] G. Tong, W. Wu, L. Guo, D. Li, C. Liu, B. Liu, D. Du, An efficient randomized algorithm for rumor blocking in online social networks, in: Proc. Annu. Joint Conf. IEEE Comput. Commun., 2017, pp. 1–9.
[34] L.G. Valiant, The complexity of enumeration and reliability problems, SIAM J. Comput. 8 (3) (1979) 410–421.
[35] W. Chen, C. Wang, Y. Wang, Scalable influence maximization for prevalent viral marketing in large-scale social networks, in: Proc. 16th SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 1029–1038.
[36] A. Clauset, C.R. Shalizi, M.E. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (4) (2009) 661–703.
Chonglin Gu received Ph.D. degree in computer science and technology from Harbin Institute of Technology (Shenzhen) in 2018. After that, he has been a postdoctoral fellow in the Chinese University of Hong Kong (Shenzhen). Now, he is an assistant professor in the school of computer science and technology in Harbin Institute of Technology (Shenzhen). His research interests are in the fields of cloud computing, cloud storage security, parallel and distributed computing, big data, especially algorithm design and system implementation.

Hejiao Huang graduated from the City University of Hong Kong and received the Ph.D. degree in computer science in 2004. She is currently a professor in Harbin Institute of Technology (Shenzhen), China, and previously was an invited professor at INRIA, France. Her research interests include cloud computing, network security, trustworthy computing, formal methods for system design and wireless networks.