Immune Algorithm to Suppress Rumor Propagation Based on Influence Maximization

Jing Chen,1,2 Nana Wei,1,3 and Hongbo Yang1,4

1School of Information Science and Engineering, Yanshan University, Qinhuangdao, China
2College of Electronic and Information Engineering, Guangdong Ocean University, Zhanjiang, China
3Hebei Key Laboratory of Virtual Technology and System Integration, Qinhuangdao, China
4Hebei Key Laboratory of Software Engineering, Qinhuangdao, China

Correspondence should be addressed to Nana Wei; liuli_weina@163.com

Received 27 October 2021; Revised 31 January 2022; Accepted 7 February 2022; Published 22 April 2022

Copyright © 2022 Jing Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to effectively reduce the spread range of rumor information, an immune method based on two-stage influence maximization to suppress rumor propagation is proposed. Firstly, User and Clustering Influence Maximizing (UCIM) algorithm is proposed to obtain the most influence nodes set at the current moment in the initial stage of event evolution based on the network topology and user characteristics. Secondly, the node set is identified and classified based on the RNN rumordetection model, and Immunosuppression Strategy considering the Average Path Weights and User-Clustering (IS-APWUC) strategy is proposed for the identified rumor nodes. In this strategy, rumor node set is taken as root nodes, and neighbor nodes with weak influence are pruned to construct an effective rumor path tree. Thirdly, considering the total probability of being infected by the rumors and the comprehensive influence factors, the nodes with high influence in the propagation stage are calculated as immune nodes so as to block the spread of the rumor information. Finally, the proposed method is verified by experiments on four real-world datasets. The results show that the IS-APWUC method has a better rumor suppression effect than the similar correlation algorithms.

1. Introduction

The massive hot topics generated by social network platforms spread rapidly among user groups through various modes. Among them, the uncontrolled spread of rumors has seriously affected the stability of society and economic development. Studies have shown that, to some extent, rumor messages spread farther, faster, deeper, and wider than nonrumor messages [1].

Rumor can be defined as a kind of public opinion, and it lacks a factual basis and is difficult for the public to distinguish between truth and falsehood. Some users with an irresponsible mindset publish and spread unconfirmed news, causing social panic psychology about unexpected events and causing certain potential damage. Therefore, how to effectively suppress the spread of rumors has become a research hot spot, and the proposed method is of great significance to social stability and network security. Maximum influence in information dissemination and public opinion control has become the key to leading the direction of communication, enhancing the control of public opinion, and reducing negative impact.

At present, many scholars have conducted research on how to control the rumor spreading scope and proposed suppression strategies based on the structural characteristics of social networks. However, most studies of suppression strategies ignore the comprehensive influence characteristics of rumor nodes, and there are problems such as the construction of rumor paths which is too complicated. In their experiments, most of them just randomly generate or search for influential nodes for blocking without introducing rumor detection models to pinpoint rumor nodes. To solve the above problems, an immune suppression strategy for identifying and detecting influential rumor nodes in the
event evolution stage based on the topology and user characteristics of social networks is proposed in this paper. By accurately locating the influential rumor nodes and constructing a simplified rumor path tree, the proposed immunization strategy based on two-stage influence maximization can effectively suppress the spread of rumors. The main contributions can be summarized as follows:

1. In view of the characteristics of the emergency dataset, the User and Clustering Influence Maximizing (UCIM) algorithm is proposed based on the topology structure of social network and user characteristics. The algorithm combines the activity of nodes with their neighbors and self-clustering coefficients to calculate the comprehensive influence of nodes and obtain the set with the maximum influence at the current moment.

2. After the rumor detection model based on recurrent neural network RNN classifies the opinion leader nodes, rumor leader set as root to prune the weakly influential neighbor nodes, and then effective rumor path trees are constructed, so the rumor propagation paths are extracted.

3. Based on effective rumor path trees, Immunosuppression Strategy considering the Average Path Weights and User-Clustering (IS-APWUC) is proposed. Considering the two factors of the sum of average path high influence of current node being infected by rumors and the comprehensive influence, the most influential nodes screened in the dissemination process are immunized. Therefore, the scope of rumor propagation is effectively suppressed.

The organization of this paper is as follows. In Section 2, the related work of rumor suppression field in social network is introduced. In Section 3, the influence maximization UCIM algorithm based on the influence of nodes and nearest neighbor nodes and clustering coefficients is proposed. In Section 4, the two-stage IS-APWUC is proposed and discussed. In Section 5, the proposed algorithms are experimented and analyzed with other algorithms on real datasets. Finally, the conclusions and future work are described in Section 6.

2. Related Works

In the process of studying rumor propagation patterns, many improved models have been proposed by combining infectious disease models with complex networks. Modeling rumor propagation on scale-free networks, it was found that scale-free networks are more robust and susceptible to attacks compared to random networks [2]. Currently, the influence maximization problem is widely used in the field of rumor control [3, 4]. The immunization strategy is to immunize some key nodes to achieve the effect of blocking the spread of rumors. Common immunization strategies include random immunization, targeted immunization, and acquaintance immunization. Since then, scholars have proposed improved immunization strategies [5–7].

In terms of rumor propagation models and suppression strategies, Gu and Xia [7] proposed the Susceptible-Exposed-Infected-Removed (SEIR) model and important acquaintance immunization strategy to suppress rumor propagation. Wang et al. [8] proposed a dynamic rumor impact minimization model based on user experience by blocking a specific subset of nodes to minimize the impact of rumors. Ye et al. [9] constructed a multilevel propagation model based on entropy power, using a rumor path tree and constructing a score function to select the top k scoring nodes as truth initiator nodes, which can effectively stop rumor propagation. Liu et al. [10] used the node degree to define a new game income and proposed a dynamic network evolution model based on game theory and two rumor suppression strategies. Han et al. [11] proposed the Susceptible-Infected-Dubious-Removed (SIDR) rumor propagation model, which describes the case where a saboteur is persuaded and becomes a spreader by adding an additional state. Dong and Huang [12] proposed a model of improved Susceptible-Infected-Susceptible (SIS) rumor propagation combining propagation dynamics and population dynamics and a strategy to suppress the spread of social network rumors.

For rumor propagation suppression algorithms in social networks, Budak et al. [13] used the restricted activity to suppress disinformation propagation and verified that, in most cases, degree centrality has the same impact as greedy algorithms. Yao et al. [14] proposed a Longest-Effective-Hops (LEH) algorithm, which considers the use of minimum credibility to clarify rumors in a given amount of time. Tong et al. [15] considered the rumor propagation process and disinformation node propagation process and proposed the Multi-cascade (MC) problem for multipart rumor propagation and disinformation node proliferation. Wei and Li [16] used mediated centrality index values and point degree centrality methods to identify key nodes for shielding, which can effectively reduce the spread of rumor information. He [17] proposed an optimal suppression problem for the multiround propagation of rumors in social networks. Chen et al. [18] proposed a method to suppress the spreading of false messages based on influence maximization. By classifying the maximum influence node set using TextCNN and then filtering out a small number of false message nodes, the method is experimentally proven to be effective in suppressing the spread of false messages.

3. Influence Maximization Algorithm

3.1. Description and Definition of Influencing Factors. The weighted directed network graph can be formally defined as $G(V, E, w)$, where $V$ represents the user node set, $|V|$ is the number of nodes, $E$ represents the directed edge set, $|E|$ is the total number of edges, and $w$ represents the weighting function, which means that for each edge $(u, v)$ defines the probability that node $u$ sends a message to node $v$ and recognizes the message, in other words, the probability that node $u$ successfully activates $v$, represented by $w(u, v)$. All notations used in this paper are described in Table 1.
In most social networks, the search for the most influential nodes only considers the network topology and ignores users' own properties. Therefore, in order to better reflect the user dynamic characteristics in social networks, information related to users' own influence is considered based on the known topological structure, which is described as follows.

**Definition 1.** The number of user's followers: this attribute is highly relevant in terms of user influence. It represents popularity; the more followers the user has, the more the tweets that he or she will post will be followed by a larger number of followers and the more the potential users will be affected accordingly. The weight of followers of node $u$ is expressed as follows:

$$F(u, t) = \frac{\text{Num}[\text{Foll}(u, t)]}{\max_{u \in U} (\text{Num}[\text{Foll}(u, t)])}$$  

(1)

where $F(u, t)$ represents the proportion of the number of node $u$ followers to the maximum number of followers in a particular social network dataset at the current moment $t$, and its value range is $[0, 1]$. $\text{Num}[\text{Foll}(u, t)]$ represents the number of followers of node $u$ at time $t$. $U$ represents the set of all user nodes.

**Definition 2.** Behavioral characteristics of users: in social networks, behavioral characteristics of user participation in topics, which is the activity degree, play an influential role. It is divided into three categories: original, retweeting, and commenting. Original signifies that the user node is interested in a topic and publishes content to influence the surrounding users. Retweeting signifies that the user node sees interest in the information published by following users and retweets it to influence users around. Commenting signifies that the user node comments on the tweets published by other user nodes. Generally, the influence effect of users posting original tweets and retweeting tweets to influence surrounding users is equivalent, both expressing user emotions and influencing users around them to understand their own situation; users participate in the topic to comment to express their opinions, which may indirectly influence other user attitudes towards the tweet, thus causing retweeting and other behaviors. The behavioral characteristics of users are expressed as follows:

$$A(u, t) = a_0 \cdot \text{Num}[\text{Orig}(u, t)] + a_1 \cdot \text{Num}[\text{Retw}(u, t)] + a_2 \cdot \text{Num}[\text{Comm}(u, t)]$$

$$\text{avgA}(u, t) = \frac{A(u, t)}{\text{Num}[\text{Acti}(u, t)]}$$

(2)

where $\text{Num}[\text{Acti}(u, t)]$ represents the total behavior number of node $u$ at moment $t$ under a certain topic, $A(u, t)$ represents the behavior influence of node $u$ at moment $t$, and $\text{avgA}(u, t)$ normalizes $A(u, t)$, which takes values in the range $[0, 1]$. Due to the different roles of influence caused by different behaviors, coefficients are defined before counting the number of originals, retweets, and comments, which is $a_0$, $a_1$, and $a_2$, respectively, where $0 < a_2 < a_1 < a_0 \leq 1$. $\text{Num}[\text{Orig}(u, t)]$, $\text{Num}[\text{Retw}(u, t)]$, and $\text{Num}[\text{Comm}(u, t)]$ represent the number of original, retweeted, and commented behaviors of statistical node $u$ participating in the topic at moment $t$, respectively.

### 3.2. Influence Maximization UCIM Algorithm

In order to detect nodes with high influence during the rumor emergence time and take measures against them to suppress rumor propagation, it is necessary to identify the key nodes that cause public opinion. An influence maximization algorithm UCIM is proposed based on the influence of nodes and nearest neighbor nodes and clustering coefficients.

### Table 1: Notations.

<table>
<thead>
<tr>
<th>Notations</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F(u, t)$</td>
<td>The followers influence of node $u$ at $t$ moment</td>
</tr>
<tr>
<td>$\text{Num}[\text{Foll}(u, t)]$</td>
<td>The followers number of $u$ at $t$ moment</td>
</tr>
<tr>
<td>$U$</td>
<td>The user node set in current network</td>
</tr>
<tr>
<td>$A(u, t)$</td>
<td>The behavior influence of node $u$ at $t$ moment</td>
</tr>
<tr>
<td>$a_0$</td>
<td>The coefficient of original</td>
</tr>
<tr>
<td>$a_1$</td>
<td>The coefficient of retweeted</td>
</tr>
<tr>
<td>$a_2$</td>
<td>The coefficient of commented</td>
</tr>
<tr>
<td>$\text{Num}[\text{Orig}(u, t)]$</td>
<td>The original number of $u$ at $t$ moment</td>
</tr>
<tr>
<td>$\text{Num}[\text{Retw}(u, t)]$</td>
<td>The retweeted number of $u$ at $t$ moment</td>
</tr>
<tr>
<td>$\text{Num}[\text{Comm}(u, t)]$</td>
<td>The commented number of $u$ at $t$ moment</td>
</tr>
<tr>
<td>$\text{Num}[\text{Acti}(u, t)]$</td>
<td>The total behavior number of $u$ at $t$ moment</td>
</tr>
<tr>
<td>$\text{avgA}(u, t)$</td>
<td>The normalized with respect to $A(u, t)$</td>
</tr>
<tr>
<td>$\text{UInf}(u, t)$</td>
<td>The own influence of node $u$ at $t$ moment</td>
</tr>
<tr>
<td>$c(u)$</td>
<td>The clustering coefficient of node $u$</td>
</tr>
<tr>
<td>$\text{Nei}(u)$</td>
<td>The neighbor set of node $u$</td>
</tr>
<tr>
<td>$\text{IM}(u, t)$</td>
<td>The comprehensive influence of node $u$ at $t$ moment</td>
</tr>
<tr>
<td>$R$</td>
<td>The maximizes influence rumor set</td>
</tr>
<tr>
<td>$\text{Af f}(u, t)$</td>
<td>The total probability of node $u$ being infected by $R$ at $t$ moment</td>
</tr>
<tr>
<td>$\text{UIm}(u, t)$</td>
<td>The being infected influence of node $u$ by rumors $R$ at $t$ moment</td>
</tr>
</tbody>
</table>
The number of followers and the behavioral characteristics of participation in the topic are considered in the user’s own influence, namely, the node activity, expressed as follows:

\[ U_{Inf}(u, t) = (1 + F(u, t)) \ast (1 + \text{avg}A(u, t)), \]

where \( U_{Inf}(u, t) \) represents the own influence of node \( u \) at moment \( t \). In the process of calculating \( U_{Inf}(u, t) \), two cases may be encountered: (1) potential users do not participate in the topic at moment \( t \) and before moment \( t \), corresponding to \( \text{avg}A(u, t) \) that is 0; (2) users may participate in the topic, but their number of followers \( F(u, t) \) is 0. Therefore, in order to weigh the influence of users, the variable is tempered by adding 1 before it.

In addition to considering the user nodes themselves, topological attributes are also used as a judging indicator of node influence. Different topological attribute values represent different aspects of the topological properties of the node, such as degree centrality, mediator centrality, and kernel value. The comprehensive influence is about to combine the node activity with the clustering coefficient, expressed as follows:

\[ IM(u, t) = (1 + c(u)) \ast \left[ U_{Inf}(u, t) \ast \sum_{v \in \text{Nei}(u)} U_{Inf}(v, t) \right], \]

where \( IM(u, t) \) represents the comprehensive influence of node \( u \) at moment \( t \), \( c(u) \) represents the clustering coefficient of node \( u \), \( \text{Nei}(u) \) represents the neighbor set of node \( u \), and node \( v \) represents belonging to \( \text{Nei}(u) \).

The larger \( IM \) value of the node indicates that it is more active in the transmission process and more important in its topology location. It has a greater influence on information transmission. The algorithm introduces user characteristics based on the topology of social networks, which better reflects the dynamics of information dissemination by network users. In order to better suppress the spread of rumors, it is necessary to preferentially select the node set with a larger comprehensive influence \( IM \) value and then adopt relevant strategies for them.

### 3.3. UCIM Algorithm Pseudocode

A social network is denoted by \( G(V, E, w) \), \( k \) is the required number of nodes, and \( S \) is the node set with maximum influence. Algorithm 1 gives the execution process of the UCIM algorithm.

### 4. IS-APWUC Immunosuppression Strategy

#### 4.1. Problem Overview

The problem of suppressing the spread of rumors is considered in two stages in terms of maximizing influence. The first stage: in the initial stage of breaking news evolution, exploring opinion leader nodes plays a key role; the second stage: rumor nodes are identified by existing rumor detection methods on leader nodes, and then based on the influence maximizing nodes screened in the rumor propagation process, immunization strategy is proposed to reduce rumor propagation scope. The overall framework of the algorithm flow chart is presented in this paper, as shown in Figure 1.

Figure 2 shows the whole process of suppressing rumor propagation in social networks from the initial \( t_0 \) to \( t_T \), where nodes represent users who post or comment on messages, and edges between nodes represent connections between users. The bolded nodes and edges are used to indicate the range that has been spread at the current time, and the unbudded nodes and edges indicate the range where the information has not yet been spread. As time grows, users keep publishing messages, propagating the information to spread, and expanding the range of influence. In this paper, the set \( S \) with the greatest influence at \( t_n \) is calculated by the UCIM algorithm and then classified by rumor detection methods. As shown in Figure 2, black nodes are rumor nodes, and white nodes are nonrumor nodes. The proposed strategy is applied to suppress the rumors at \( t_{n+1} \), which takes the rumor nodes as the source and uses the IS-APWUC algorithm to calculate the high influence nodes on the rumor path for immunization. Spot-filled nodes are immune nodes, thus achieving the rumor suppression effect.

#### 4.2. IS-APWUC Immunization Algorithm

In this paper, the RNN rumor detection model that represents deep features of the text is mined by learning training, where the network configuration uses Long Short-Term Memory (LSTM) [19]. This model is used to identify and classify the set \( S \) selected by the UCIM algorithm, and the recognized rumor set is defined as \( R \); that is, \( R = \{R_1, R_2, \ldots, R_S\} \). With the set \( R \) as root nodes, the immune nodes with high influence are found in two steps; the first step is to prune and construct an effective path tree to analyze the rumor propagation path; the second step proposes the IS-APWUC based on the total probability of being infected by the rumors and the comprehensive influence, and the high influence nodes are screened for protection or immunity using this algorithm. In this way, the rumor minimizes the information spread, meaning that the negative influence is minimized. The Susceptible-Infected-Removed (SIR) model is chosen as a simulation platform to analyze the effect of IS-APWUC on rumor suppression propagation [20].

The first step constructs a simplified rumor path tree. In order to analyze rumor propagation paths well, the initial weighted directed graph \( G \) is simplified, and a directed acyclic graph \( G' \) is constructed as follows:

1. Prune the weakly influential neighbor nodes of the set \( R \). Weak influence nodes are influenced by rumors but unable to expand the role of information dissemination. One-step breadth-first traversal algorithm Breadth First Search (BFS) is performed sequentially on the set \( R \) to remove the weakly influential neighbor nodes [21]. The node set involved in the BFS algorithm forms \( R_{adj} \), and then \( R_{adj} \) is judged. If the out-degree is equal to 0, the node is removed from the network. And if out-degree is not 0 and its neighbor nodes are all rumor nodes, it should also be removed. Finally, the network \( G1 \) is obtained. In Step 1 of Figure 3, some neighbor nodes
of set $R$ are $U_{10}$, $U_{11}$, and $U_{12}$, where $U_{10}$ has an out-degree of 0, and neighbor nodes of $U_{11}$ and $U_{12}$ are in $R$ set. It can be found that $U_{10}$, $U_{11}$, and $U_{12}$ are unable to expand information transmission; that is, it is meaningless if these nodes are selected as immune nodes. Therefore, weak
influence nodes are removed from \( G \), as shown in Figure 3.

(2) Obtain the surrounding node set of rumor nodes. In network \( G_1 \), the Depth First Search (DFS) algorithm is performed for each node in set \( R \), and subgraphs are constructed [21]. The subgraphs involved in the algorithm are formed into set sequence \( T \); that is, \( T = \{T_1, T_2, \ldots, T_S\} \).

(3) Get the maximum directed graph \( G' \) containing the rumor propagation paths. A similar acyclic algorithm is adopted for the sequence of set \( T \) to find the maximum Directed Acyclic Graph (DAG) initiated by \( R \) [22]. Firstly, on the basis of \( T_1 \), the remaining untraversed edge \( E' \) is traversed to judge whether a loop is formed. If no loop is formed, the edge is added to \( T_1 \) until a maximum directed acyclic graph is formed, such as Step 3 of Figure 3. Then, sequentially, \( T_2, \ldots, T_S \) are constructed. Finally, the network \( G' \) constructed from the set \( T \) is obtained. The purpose is to simplify the complex network diagram and facilitate the extraction of rumor propagation paths.

The second step is to find immune nodes based on the construction of a valid rumor path tree. In \( G' \), the set \( Imm_S \) with high influence is sought for immunity, and the specific steps are as follows:

(1) Obtain the candidate node set as \( Cand_S \). The set \( R \) in graph \( G' \) is searched by \( n \)-step depth-first search, and the found nodes are added to the set \( Cand_S \), where \( 1 \leq m \leq n \).

(2) Calculate the probability that the candidate set \( Cand_S \) is infected by the set \( R \). The weights of all simple paths and edges between two points from node \( r \) in set \( R \) to node \( c \) in set \( Cand_S \) are obtained, and the average activation probability between two nodes is calculated as the sum of path weights divided by the number of paths. Then, the maximum value is picked as the probability that node \( c \) is infected by node \( r \). As shown in Part tree of Figure 3, there are multiple paths for \( R_1 \) to send messages to reach \( U_2 \), and the analysis of which path is efficient is calculated to give preference to the \( R_1 \rightarrow U_4 \rightarrow U_5 \rightarrow U_2 \) path. Node \( c \) may be infected by more than one rumor node, and the total probability of
4.3. IS-APWUC Algorithm Pseudocode. A social network is represented by $G(V, E, w)$, $k_1$ is the number of immune nodes, and $Imm_S$ is the immune node set with the highest influence. $R$ and $Cand_S$ are rumor set and candidate set, respectively. Algorithm 2 gives the execution process of the IS-APWUC algorithm.

In Algorithm 2, line 1 indicates initializing the sets of $R$, $Cand_S$, and $Imm_S$ as empty sets. Line 2 indicates that the rumor nodes identified using the RNN rumor detection model for the set $S$ are added to the set $R$. Lines 3 to 6 represent obtaining the set of nearest neighbor nodes $R_{adj}$ and judging and pruning operation is performed if the condition is satisfied. Lines 7 to 9 show that, after pruning, the effective directed acyclic graph $G'$ is further constructed for the set $R$. Lines 10 to 13 represent that DFS is performed on the set $R$ to obtain $Cand_S$ in the graph $G'$, and then $Aff$ is calculated using equation (5). The value of $IM$ is known in Algorithm 1, and finally, $Ulm$ is calculated using equation (6). Line 14 indicates that the first $k_1$ nodes with maximum influence are identified and added to the set $Imm_S$. Finally, line 15 represents returning the set $Imm_S$.

5. Experiments and Results Analysis

5.1. Twitter Dataset. The experiments are conducted using the PHEME dataset [23], and four datasets containing Twitter rumors and nonrumor collections posted during breaking news are selected. In order to facilitate the creation of directed graphs and analysis of user information data, user information table and relationship table are created for the raw data of each breaking news. The user information table is the record of the original, retweeting, commenting, time, number of followers, and other related pieces of information involved in the topic. The user relationship table, which is the connection between nodes, constructs a directed graph. If user $v$ retweets or comments behavior to user $u$, the directed edge is pointed from user $u$ to user $v$, indicating that user $u$ has an influence on user $v$.

The topological characteristics of the datasets are analyzed as shown in Table 2. The number of out-degrees is counted and visualized in a double logarithmic coordinate system for each of the four datasets, where $K$ is the number of out-degrees and $P(K)$ is the percentage of node out-degrees counted for the dataset.
degrees to the total number of out-degrees, as shown in Figure 4. The distribution of datasets conforms to the power-law distribution of complex networks, indicating that the dataset has scale-free properties, which further validates the data validity.

5.2. Impact Maximization Algorithm Comparison Experiment. The Weighted Independent Cascade Model (WICM) is used as the information dissemination model. $k$ is the size of the set of selected nodes, and the coefficients of the number of originals, retweets, and comments in the behavioral characteristics are $a_0 = a_1 = 1$ and $a_2 = 0.6$, respectively. The UCIM algorithm is replicated with other algorithms, and the comparison experiments are conducted on four real datasets. As shown in Figure 5, the proposed algorithm UCIM is experimentally compared with other algorithms to analyze the impact range of each algorithm.

Degree algorithm is representative of heuristic algorithms, where the node with the largest degree is selected as the seed node until $k$ nodes are chosen [24].

DegreeDiscount algorithm is to select the node with the largest degree as the seed node and then discount the degree of the selected node neighbors until $k$ nodes are selected [25].

Core Covering Algorithm (CCA) is a heuristic algorithm based on network hierarchy and influence radius influence $d$, which is generally taken as $d = 1$ in experiments [26].

Louvain Clustered Local Degree Centrality (LCLD) algorithm is a node centrality method that combines the sum of nearest neighbor degrees and node clustering coefficients by the Louvain association division idea [18].
Random Maximum Degree algorithm (RMD) is a local information method for randomly selecting nodes and neighbors based on global node characteristics, set \( \alpha = 2 \) in the experiment [18].

In Figure 5, the effect of the influence range for the relevant algorithms in the four datasets is shown. The influence range called influence is the initial phase of the network by calculating the node set selected, for its propagation in the network, counting the total number of nodes finally affected. The wider the influence of the node set, the higher the accuracy of the algorithm. As can be seen in Figure 5, when \( k = 500 \), the UCIM algorithm has a 3.8%, 28.1%, 27%, 55.2%, and 70.2% higher impact range than Degree, DegreeDiscount, CCA, LCLD, and RMD, respectively, on the Charlie Hebdo dataset. On the Germanwings Crash dataset, the UCIM algorithm outperforms the Degree, DegreeDiscount, CCA, LCLD, and RMD algorithms, respectively, on the Ottawa Shooting dataset, the UCIM algorithm outperforms the Degree, DegreeDiscount, CCA, LCLD, and RMD in terms of impact range by 4.2%, 19.6%, 21.6%, 29.8%, and 50.7%, respectively. On the Ottawa Shooting dataset, the UCIM algorithm has a higher impact range than Degree, DegreeDiscount, CCA, LCLD, and RMD by 2.7%, 25.6%, 40.3%, 39%, and 71.2%, respectively. On the Sydney Siege dataset, the UCIM algorithm has a higher impact range than Degree, DegreeDiscount, CCA, LCLD, and RMD by 3.3%, 32.8%, 37.5%, 52.6%, and 73.4%, respectively. From the above, it can be seen that, on the four real datasets, the UCIM algorithm has the best effect on the range of influence, the Degree algorithm is second, the DegreeDiscount algorithm and CCA algorithm are better, the LCLD algorithm is slightly worse, and the RMD algorithm is the least effective. The UCIM algorithm combines user characteristics based on the network topology to demonstrate the dynamic characteristics of users in a realistic network. The reason for the UCIM algorithm to have the widest range of impact on different data compared to other algorithms is due to both critical position and node activity, which reflects the effectiveness of this algorithm.

5.3. Immunization Strategy Comparison Experiment. A simulation experiment is conducted using the SIR model to compare, analyze, and explore the effect of immune
nodes on the effectiveness of rumor resistance. In the experiment, $k_1$ is a value of 100, the evolutionary stages of breaking events Charlie Hebdo and Ottawa Shooting $t_a$ are the first 20% of the time cycle, the evolutionary stages of Germanwings Crash and Sydney Siege $t_b$ are the first 40% of the time cycle, and the network is constructed. The influential rumor set $R$ is obtained and classified by the UCIM algorithm and RNN rumor detection model. In network $G$, $R$ is used as the initial infectious node, unlike the previous randomly selected nodes as infectious nodes. The number of immune nodes all accounts for 8% of the total number of users in network $G'$. The IS-APWUC algorithm is compared with other algorithms on four datasets for five iterations of simulation experiments, and then the results of the five experiments are averaged and displayed in line graphs, as shown in Figures 6 and 7.

Targeted Immunization (TI) is to immunize the $k_1$ nodes that have a larger degree in the selected network [27].

Self-Degree and Neighbor-Degree (SDND) immunization is to select $k_1$ nodes for immunization by using the product of node out-degree and maximum neighbor out-degree as node importance metrics [6].

Acquaintance Immunization Strategy Considering the Weights and Degrees (AI-CWD) immunization is the acquaintance immunization strategy with the largest product of node degree and neighbor node edge weights, and $k_1$ nodes are selected for immunization [5].

Local Shortest-Paths for Multiple Influencers (LSMI) is to evaluate the influence of each node using the shortest path
while selecting the high influence node as the truth node, and the node is the immune node in this experiment [28]. Figure 6 shows the effect of node infection range for the relevant algorithms in the four real datasets. The infection range called nums is the number of final infected nodes at each moment of time when the network is propagated in network G with known rumor nodes as the initial infected nodes, and \( t \) represents time. In Figure 6, overall, it can be visualized that IS-APWUC is effective in reducing the number of node infections compared to other methods. Figure 7 shows the range effect of nodes gaining immunity after being infected, where counts refers to the number of nodes immunized. The lower the number of nodes gaining immunity after being infected, the better the immunization effect of the algorithm. As seen in Figure 7, when \( t = 6 \), IS-APWUC is 2.1%, 4.5%, 3.4%, and 1.7% more effective than TI, SDND, AI-CWD, and LSMI on the Charlie Hebdo dataset. IS-APWUC is 0.4%, 1.6%, 1.7%, and 0.6% more effective than TI, SDND, AI-CWD, and LSMI on the Germanwings Crash dataset. On the Ottawa Shooting dataset, IS-APWUC is 1.5%, 2.7%, 1.7%, and 0.7% more effective than TI, SDND, AI-CWD, and LSMI. On the Sydney Siege dataset, IS-APWUC is 2.8%, 5.6%, 3.9%, and 1% more effective than TI, SDND, AI-CWD, and LSMI. From the above, it can be seen that the relevant algorithms in the four real datasets achieve suppression of rumor propagation to a certain extent, with IS-APWUC being the most effective, followed by TI and LSMI, SDND and AI-CWD being slightly worse. The reason is that the IS-APWUC algorithm considers the
comprehensive influence of the nodes in the effective rumor path tree so that the obtained immune nodes are mostly in the key position of susceptibility to infection and their own influence is larger. As the number of nodes in the dataset increases, the suppression effect of this method becomes better. Therefore, the algorithm has more advantages compared to other algorithms in suppressing the effect of the rumor spreading range.

6. Conclusion

In order to effectively suppress the rumor propagation influence range in social networks, an immunization method to suppress rumor propagation based on two-stage influence maximization is proposed. In the initial stage of dissemination, the UCIM algorithm is proposed to find the opinion leader nodes based on both network topology and user characteristics. Then, in the information dissemination stage, the leader nodes are classified based on the RNN rumor detection model. Finally, an effective rumor propagation tree is constructed for the identified rumors, and an IS-APWUC immune strategy is proposed so as to suppress the scope of rumor propagation. In the actual network, the influence range of the UCIM algorithm is much higher than that of Degree-Discount, CCA, LCLD, and RMD algorithms, while it is similar to that of the classical Degree algorithm, but the effect is relatively better. The IS-APWUC immunization has a better effect on rumor suppression compared with TI, SDND, AI-CWD, and LSMI algorithms.

In future work, the following research will be conducted: (1) the real online social network topology has community and self-similarity, and community structure characteristics are considered so that the rumors can be targeted to suppress the propagation; (2) in the construction of the rumor path tree, the process of adding new edges to construct the maximum directed acyclic graph has randomness and ignores the influence role of edge weights, which can be improved.

Data Availability

The data used to support the findings of this study are available from the corresponding authors upon request. The original dataset link is: https://figshare.com/articles/PHEME_dataset_of_rumours_and_non-rumours/4010619.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (62172352), the Natural Science Foundation of Hebei Province (F2019203157), and the Science and Technology Research Project of Hebei Higher Education Institutions (ZD2019004).

References


