IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 17, 2019, accepted March 30, 2019, date of publication April 10, 2019, date of current version April 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2909552

Rumor Source Detection in Networks Based on the SEIR Model

YOUSHENG ZHOU[1,2], CHUJUN WU[1], QINGYI ZHU[2,3], (Member, IEEE), YONG XIANG[3], (Senior Member, IEEE), AND SENG W. LOKE[3]

[1]College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2]School of Cyber Security and Information Law, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[3]School of Information Technology, Deakin University, Geelong, VIC 3127, Australia

Corresponding author: Qingyi Zhu (zhuqy@cqupt.edu.cn)

This work was supported by the Venture and Innovation Support Program for Chongqing Overseas Returnees under Grant CX2018122.

ABSTRACT Online social networks have become extremely important in daily life and can be used to influence lives in dramatic ways. Two issues are the veracity and provenance of posted information, including rumors. There is a need for methods for tracing rumors (or any piece of information) to their most likely source in such networks. We consider the detection problem of single rumor source based on observed snapshots based on the susceptible-exposed-infected-recovered (SEIR) model. According to the SEIR model, all nodes in the network are formulated into four possible states: susceptible (S), exposed (E), infected (I), and recovered (R). Given an observed snapshot in the network, from which we can know the relevant graph topology and all infected nodes, but where nodes in susceptible, exposed, or recovered status cannot be distinguished, the purpose of our research is to identify the rumor source based on the observed snapshot and graph topology. We propose the concept of the optimal infection process and derive an estimator for the rumor source based on this optimal infection process. Subsequently, we prove that this estimator matches the rumor source with a high probability. The effectiveness of the proposed scheme is validated using experiments based on regular tree networks with different degrees. We further evaluate the performance of our scheme on two well-known synthetic complex networks and four real-world networks; the results suggest that our proposed scheme outperforms the traditional rumor centrality heuristics. The performance analysis on computational complexity demonstrates that our scheme has advantages in efficiency compared with other rumor centrality heuristics used in rumor detection methods.

INDEX TERMS Rumor source detection, optimal infection process, susceptible-exposed-infected-recovered (SEIR) model, information security.

I. INTRODUCTION

The online network has facilitated our daily life but it also makes people vulnerable to risks [1], [2]. For instance, rumors can propagate rapidly across networks due to increases in network connectivity. These misleading information would undermine the stability of networks and even lead to pernicious influence on socity [3]. The intrinsic reason is the fact that anyone can release (false) information. Therefore, identifying the source is significant to possibly reduce the damage caused by the rumor [4]. Conventional techniques, such as stepping-stone detection [5] and IP traceback [6], are not efficient to find the actual rumor source since the source of these received packets is one of the propagation participants, not the origin of the rumor [7]. In addition, network topological structure and node properties should affect the propagation of rumors [8]–[11]. Therefore, it is necessary to find more practical means to detect information sources from the logical structure of the network.

In this paper, we use the susceptible-exposed-infected-recovered (SEIR) model to simulate the state transition of all nodes in networks. Then we show how to locate the actual rumor source and ultimately achieve the purpose of controlling the risks based on a known network snapshot and the graph topology observed at a certain moment.

A. RELATED WORK

In recent years, a series of epidemic model based methods to detect propagation sources have been proposed by researchers [12]. For example, traditional susceptible-infected

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

(SI) model is usually used to study the spread of information in a tree-like network with complete observations [13]–[15]. In these papers, nodes in networks were modeled in two possible states: susceptible state and infected state. Rumor centrality was proposed by Shah and Zaman [14] to identify the rumor source in these references. Shah *et al.* claimed that the node with maximum closeness centrality is the rumor source. Subsequently, some other epidemic models have been adopted to solve the problem of source detection, such as the susceptible-infected-recovery (SIR) model [16], [17] and the susceptible-infected-susceptible (SIS) model [18], [19], in which infected nodes may recover and will no longer be infected, or susceptible. Chen *et al.* [20] and Luo and Tay [21] extended the method of the rumor centrality heuristic from a single source problem to a multiple source problem. Wang *et al.* [22] extended the method of rumor centrality heuristic from complete observation to multiple observation through a breadth-first-search (BFS) technique on a tree-like topology. However, simulations of the previous work show that the rumor centrality heuristic has a low detection rate for identifying the rumor source [23].

Then, many methods for identifying the rumor source have been proposed for tree-like networks with partial observation. A sample path based estimator was proposed to estimate the rumor source by Zhu and Ying [17] and Luo and Tay [18]. They claimed that the optimal sample path has the highest probability of leading to the observed infection topology and the root of the optimal sample path, that is, the Jordan Center is the rumor source. Subsequently, Chen *et al.* [24] extended the rumor source detection from a single source to multiple sources through the Jordan Center technique. In addition, other observation methods such as sensor observation were proposed by Pinto *et al.* [25]. They proposed the central limit theorem on differences of infection times of sensors. Meanwhile, some researchers extended simulation experiments from tree-like networks to general networks. With respect to network partial observation, the Bayesian belief modeling was proposed by Altarelli *et al.* [26] to identify rumor source in general networks. As to sensors observation, Agaskar and Lu [27] leverage the Monte Carlo algorithm to detect rumor source in general networks.

In fact, the above work were based on a model with three states. The models with three states proposed previously are too simple to fully simulate the state change of nodes in the network. For example, a computer carrying a virus may be infected by this virus, or may be recovered due to the different anti-virus software installed. In another example, the users on the microblog after receiving a confusing message may hesitate to believe this message, and then forward this message to their friends or delete this message. Therefore, researchers believed that the node that received a rumor might believe this rumor and forward it, or it may not believe this rumor and drop it. In order to solve the problem of detecting the rumor source in such scenarios, an intermediate state between susceptible and infected is considered: exposed state [28]–[30]. A rumor model with the exposed state was discussed by Xia *et al.* [31] and used to simulate rumor propagation in complex social networks. However, the fact that the exposed nodes can change to the recovered nodes with a certain probability was not considered by the authors. Dong *et al.* [32] and Ran and Ling-Ling [33] adopted SEIR model to consider rumor spreading in online social networks. Liu and Sun [34] used SEIR to investigate the influence of heterogeneity of the underlying complex networks and control mechanisms on rumor propagation. However, the accuracy rates of the SEIR propagation model in the literature [31]–[34] used to identify rumor sources are not desirable. Motivated by [31]–[34], we extend SEIR propagation model to investigate the rumor detection problem in networks.

B. OUR CONTRIBUTIONS

- In this paper, we extend the well-known susceptible-exposed-infected-recovery (SEIR) model to study the detection problem of single rumor source according to an observed snapshot. Traditional SEIR model applied in the works [31]–[34] cannot accurately detect the rumor propagation in real networks, because the authors in [31]–[34] supposed that each infected node propagated the rumor to only one neighbor whose state is susceptible (S) at each time slot. If the chosen neighbor is infected (I) or recovered (R), the infected node that originally propagated the rumor would turn to the recovered state with probability 1. In addition, the authors assumed that the infected nodes are still likely to become the hesitant state (the E-state) and believe the same rumor again in the previous SEIR model hypothesis. Obviously, it is slightly inconsistent with the rumor propagation in the scenarios stated above. Therefore, in our approach, instead of assuming the rumor can be transmitted to only one neighbor, we consider the fact that each infected node can infect all S-state neighbors at a rate in each time slot. Meanwhile, nodes are assumed not to be infected again by the same rumor, i.e., the state transition I$\rightarrow$E$\rightarrow$I is not considered. An example as shown in Fig. 1 is given to intuitively illustrate the rumor propagation process based on the SEIR model in networks.
- We propose the concept of optimal infection process for identifying the actual rumor source, which represents a possible infection process with the maximum probability of generating the final observed graph topology, and the source of this optimal infection process is called the estimator of the rumor source. Subsequently, we show that this estimator of rumor source is equal to the Jordan infection center, which indicates that a node has the smallest infection eccentricity, where the infection eccentricity of a node is defined to be the maximum distance from the node to the infected nodes. Therefore, the problem of identifying the rumor source can be simplified to the problem of calculating the Jordan infection center in the final observed graph topology.

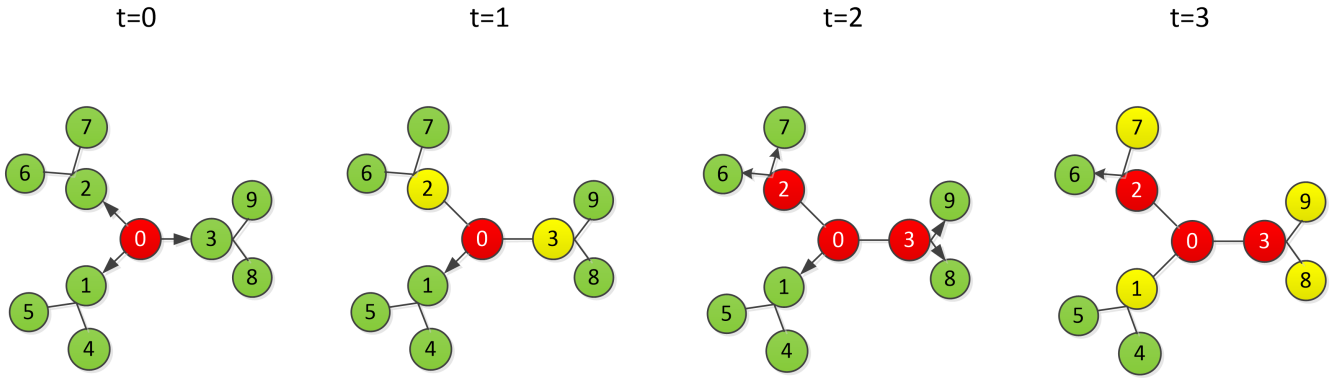

FIGURE 1. An example of infection process for the SEIR model. Suppose node 0 is the rumor source and the arrows indicate the direction of rumor propagation. At time $t = 1$, nodes 2 and 3 change to E-state after receiving the rumor from the node 0, but the node 1 maintains S-state at a given rate. At time $t = 2$, after believing this rumor, the node 2 and 3 change to I-state and forward the rumor to state S neighbors, respectively. At time $t = 3$, The neighbor node 7 changes to the state E after receiving the rumor from node 2. However, node 6 maintains the state S. Similarly, nodes 1, 8 and 9 change to state E after receiving the rumor from I-state neighbors.

- We evaluate the time complexity of our proposed scheme. According to the comparison between our proposed scheme and the traditional algorithm of rumor centrality heuristics used to identify the rumor source, our proposed scheme has the advantage of lower computational complexity. In addition, we evaluate the performance of our scheme based on regular tree networks with different degrees. Furthermore, we conduct extensive simulation experiments via various networks, including synthetic complex networks and real-world networks. The simulation results suggest that our scheme is more efficient to identify the rumor source than the traditional closeness centrality heuristic and betweenness centrality heuristic. The average detection rate of our scheme is about 55% under regular tree networks, which is superior to other centrality heuristics based methods (about 30%). In addition, our scheme outperforms other centrality heuristics in terms of the error distances in estimation under general networks as well.

C. ORGANIZATION

The rest of the paper is organized as follows. In Sect. II, the concept of the SEIR propagation model and the possible infection process are presented. In Sect. III, we describe in detail the optimal infection process for rumor propagation and derive an estimator for rumor source based on this optimal infection process. Then we demonstrate that this estimator of rumor source is equal to the Jordan infection center through detailed proofs. Subsequently, Sect. IV analyzes the time complexity of our scheme, and compares it with the traditional algorithm of rumor centrality heuristics used in rumor source detection. The effectiveness of the proposed scheme is evaluated using experiments based on various networks in Sect. V. Finally, we discuss the directions for future work in Sect. VI and conclude our paper in Sect. VII. In addition, the Appendix shows detailed proofs of our scheme.

II. PROBLEM FORMULATION

A. RUMOR PROPAGATION MODEL

In this section, we model the nodes and the edges in a network as an undirected graph $G = \{V, E\}$. Here, V is defined as a countably infinite set of nodes, E is defined as a set of edges for $(u, v) \in E$, $u, v \in V$. Under the SEIR rumor propagation model, a node in the network can be in any one of the four possible states: susceptible state (S), exposed state (E), infected state (I), recovered state (R). Under the discrete time slot propagation model, we assume that all nodes may change their own states at each discrete time slot based on their own states at the previous time slot [40].

At time slot $t = 0$, we suppose all nodes are in S-state but only one node s^* is in I-state, which we call this node as the rumor source. At the beginning of each time-slot, all I-state nodes would transmit the rumor to their S-state neighbors and those S-state nodes who have received the rumor would change to E-state with probability p_1. If some E-state nodes believe in this rumor message, then they should forward it to their neighbors with probability p_2, and thus change to the state I. However, after receiving this message, some E-state nodes may not believe this rumor message and drop it with probability r_2, then we consider them in R-state. If E-state nodes neither forward it nor drop it after receiving the rumor, this means these E-state nodes maintain their state stably with probability $1 - p_2 - r_2$. In addition, based on the previous state, the I-state nodes can recover and change to the state R with probability r_1 and R-state nodes would not be infected again (R-state nodes will no longer receive this rumor message). In addition, in reality, if a node is infected, the probability of recovery is smaller than the probability that a node receives this rumor but drops it, i.e., $r_1 < r_2$. Note that S-state nodes become E-state nodes depending on I-state neighbors, and the state transition of E→I, E→R and I→R only depend on their own previous states.

The state transition diagram is shown in Fig.2.

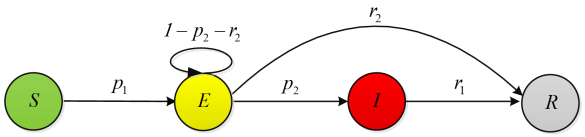

FIGURE 2. The state transition of SEIR model. S-state nodes who have received the rumor would change to the state E with probability p_1. If believing and forwarding this rumor to S-state neighbors with probability p_2, then E-state nodes would change to state I. However, if E-state nodes do not believe and drop it with probability r_2, then we consider them changing to state R. If E-state nodes neither forward it nor drop it after receiving this rumor, this means these E-state nodes maintain their state stably with probability $1-p_2-r_2$. In addition, the I-state nodes can recover and change to the state R with probability r_1 and R-state nodes would not be infected again.

B. THE POSSIBLE INFECTION PROCESS

1) THE SETTING OF THE OBSERVED GRAPH TOPOLOGY

We use the θ_v to indicate the state of the node v in the observed snapshot G at the observed time-slot. Note that this observed time-slot is an unknown value, and we refer to this value to help study the infection process of each node.

$$\theta_v = \begin{cases} 1, & v \text{ is in state I} \\ 0, & v \text{ is in state S, E, or, R} \end{cases}$$

Here, the reason why $\theta_v = 0$ when v is in the state S, E, or, R is that we can only know all infected nodes, and cannot distinguish other states in the actual observed graph topology.

In addition, we define

$$\Theta = \{\theta_v | \theta_{s^*}, \theta_{v_1}, \theta_{v_2}, \ldots, \theta_{v_i}, v_i \in V\} \tag{1}$$

as a set of the states of all nodes in the final observed graph topology. For instance, if all nodes in the final observed graph topology are in the non-infected state except for the rumor source, we have $\Theta = \{\theta_v | 1, 0, 0, \ldots, 0, v \in V\}$.

2) THE POSSIBLE INFECTION PROCESS FOR THE STATE OF NODES

In order to facilitate the study of the states transition of nodes and the process of rumor propagation in networks, we define the state of the node v at the time-slot t as follows.

$$\omega_v(t) = \begin{cases} S, & v \text{ is in state S} \\ E, & v \text{ is in state E} \\ I, & v \text{ is in state I} \\ R, & v \text{ is in state R} \end{cases}$$

and $\Omega(t) = \{\omega_v(t) | \omega_{s^*}(t), \omega_{v_1}(t), \omega_{v_2}(t), \ldots, \omega_{v_i}(t), v_i \in V\}$ denotes by the set of the state of all nodes. For example, at the time-slot $t = 0$, we have $\Omega(0) = \{I, S, S, S, \ldots, S\}$. Here, $\Omega_v(0) = Sfor(v \in V,\ v \neq s^*)$, $\Omega_{s^*}(0) = I$.

Next, we define the collection of the infection process based on the states of all nodes.

Definition 1: For each $v \in V$, an infection process starting from the node v from the time-slot 0 to T in the network is defined as follows:

$$span_{(0\to T)}\Omega(t)_v = \{\Omega(t)_v | \Omega(0), \Omega(1), \Omega(2), \ldots, \Omega(T), 0 \le t \le T\} \tag{2}$$

In order to map the state of the nodes in infection processes to the state of the nodes in the final observed graph topology, we define the following function *func*().

$$func(\omega_v(t)) = \begin{cases} 1, & \omega_v(t) \text{ is } I \\ 0, & \omega_v(t) \text{ is } S, E, or R \end{cases}$$

If the state of all nodes in the final observed snapshot (1) (Assume that the final observed graph topology is at time-slot t) coincides with all nodes' state of an infection process (2), i.e., $func(\Omega(t)) = \Theta$, we claim this infection process is a possible infection process. In order to understand the above propagation processes, a simple example is shown in Fig.3.

The left picture shows the network observed graph topology Θ observed at a certain moment. It is clear to see that the red nodes represent the infected nodes, i.e., $\theta_1 = \theta_3 = \theta_4 = 1$, and the white nodes represent the nodes in susceptible, exposed or recovered state since we can only identify infected nodes in the network observed graph topology. i.e., $\theta_2 = \theta_5 = 0$. Therefore, the observed graph topology is $\Theta = \{1, 0, 1, 1, 0\}$.

The picture on the right depicts two possible infection processes, where the state of the nodes in these two infection processes coincide with the state of the nodes in the observed graph topology Θ.

The state of the nodes in the first possible infection process (A) at time-slot $t = 4$ is $\Omega(4) = \{I, R, I, I, S\}$, and $func(\Omega(4)) = \Theta = \{1, 0, 1, 1, 0\}$. The possible infection process (A) is

$$\begin{aligned} span_{(0\to 4)}\Omega(4) &= \{\Omega(0), \Omega(1), \Omega(2), \Omega(3), \Omega(4)\} \\ &= \{\{I,S,S,S,S\},\{I,E,E,S,S\},\{I,I,I,S,S\}, \\ &\qquad \{I,R,I,E,S\}, \{I,R,I,I,S\}\}. \end{aligned}$$

In addition, the possible infection process (B) is expressed as follows:

$$\begin{aligned} &span_{(0\to 5)}\Omega(5) \\ &\quad = \{\Omega(0), \Omega(1), \Omega(2), \Omega(3), \Omega(4), \Omega(5)\} \\ &\quad = \{\{I,S,S,S,S\}, \{I,S,E,S,S\}, \{I,E,I,S,S\}, \\ &\qquad \{I,I,I,S,S\}, \{I,R,I,E,E\}, \{I,R,I,I,E\}\}. \end{aligned}$$

and it is easy to get $\Omega(5) = \{I, R, I, I, E\}$ and $func(\Omega(5)) = \Theta = \{1, 0, 1, 1, 0\}$.

3) THE POSSIBLE INFECTION PROCESS FOR ML ESTIMATOR

We denote our estimator for rumor source as $\hat{s}$ and assume each node in the network may be the rumor source. Suppose that in the network, there is only one rumor source s^* that has begun to propagate the rumor to surrounding neighbors when the time-slot $t = 0$. Subsequently, we observed a snapshot G and N infected nodes at a certain time slot. Based on the knowledge of the network infection snapshot G and the observed topology graph Θ, we need to derive an estimator $\hat{s}$ for the rumor source s^*.

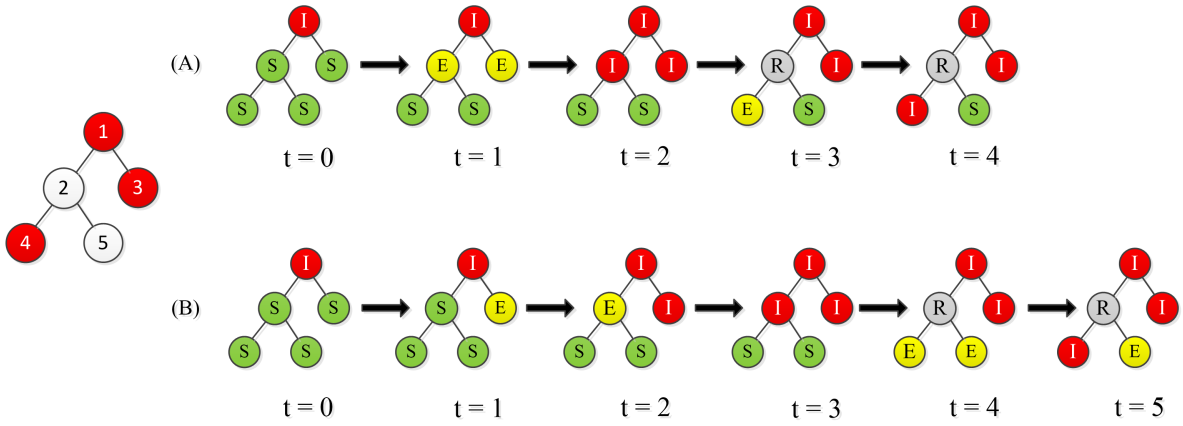

FIGURE 3. An example for rumor propagation. The left picture shows an observed graph topology $\Theta = \{1, 0, 1, 1, 0\}$ observed at a certain moment. Note, only infected nodes can be identified, i.e., $\theta_1 = \theta_3 = \theta_4 = 1, \theta_2 = \theta_5 = 0$. The right picture depicts two possible infection processes, where the state of the nodes in these two infection processes coincide with the state of the nodes in the observed graph topology Θ. The infection process (A) is $span_{(0\to 4)}\Omega(4) = \{\{I, S, S, S, S\}, \{I, E, E, S, S\}, \{I, I, I, S, S\}, \{I, R, I, E, S\}, \{I, R, I, I, S\}\}$, and the infection process (B) is $span_{(0\to 5)}\Omega(5) = \{\{I, S, S, S, S\}, \{I, S, E, S, S\}, \{I, E, I, S, S\}, \{I, I, I, S, S\}, \{I, R, I, E, E\}, \{I, R, I, I, E\}\}$.

TABLE 1. The table of notations.

Notation	Definitions
s^*	The actual source of the rumor in the network
$\hat{s}$	The estimator identified the rumor source in the network
J	The Jordan infection center
v_{root}	The root node in the tree network
p_1	The probability of a susceptible node receiving rumor from its I-state neighbors
p_2	The probability of an exposed node being infected
r_1	The probability of an infected node recovers
r_2	The probability of an exposed nodes dropping the rumor information
S_u^{-v}	The subtree rooted at node u which does not contain the branch from node v
t_u^E, t_u^I, t_u^R	The exposed, infected and recovered time of node u
T_u^*	The optimal time duration of the optimal infection process starting from the node u
$l(u, v)$	The shortest distance between node u and node v
$e\check{c}c(u)$	The infection eccentricity of node u
$\Omega(t)$	The states of all nodes at the time-slot t
$\omega_u(t)$	The state of node u at the time-slot t
$span_{(0\to T)}\Omega(t)_u$	An infection process from 0 to T starting from node u
$span_{(0\to T)\Omega(t)_u}, S_v^{-u}$	An infection process from time-slot 0 to T in the subtree S_v^{-u}
$C(u)$	The set of children of node u
θ_u	The state of the node u in the observed graph topology at the observed time-slot
Θ	The set of the states of all nodes in the graph topology based on the observed snapshot G
G	An observed snapshot in the network
I	The set of all infected nodes
E	The set of all edges

According to the above setup, we consider the maximum likelihood problem (ML) to identify the rumor source s^*.

$$\hat{s} = arg\max_{v\in V} \sum_{span_{(0\to T)}\Omega(t)_v : func(\Omega(t))=\Theta} Pr(span_{(0\to t)}\Omega(t)_v | s^* = v) \quad (3)$$

Here, $span_{(0\to T)}\Omega(t)_v : func(\Omega(t)) = \Theta$ is defined as all possible infection processes starting from the node v coinciding with the observed topology graph Θ.

$Pr(span_{(0\to T)}\Omega(t)_v | s^* = v)$ is the probability of the rumor spreads along the infection process $span_{(0\to T)}\Omega(t)_v$ starting from the rumor source v.

Assuming that the observed time slot t is known, we can get at least t^n possible infection processes based on the total number of nodes n in the network. Moreover, we need to set the time of state transition for each node during the infection process due to each node having four possible states. Similar to other models [14]–[18], solving the maximum likelihood problem requires an exponential number of calculations. To solve the problem of identifying the rumor source, we propose the concept of optimal infection process to optimize the above ML problem (3).

The notations used throughout this paper are summarized in Table 1.

III. THE OPTIMAL INFECTION PROCESS

A. THE OPTIMAL INFECTION PROCESS FOR RUMOR SOURCE

As stated above, the detection problem of the rumor source is to identify the source s^* in the graph topology Θ based

on an observed snapshot G. In order to identify the source, we propose two concepts: the optimal infection process $\tilde{span}_{(0\to T_v^*)}\Omega(t)_v$ and the optimal time duration T_v^* respectively as follows.

Definition 2: The optimal infection process is defined as the infection process that most likely led to the observed graph topology Θ in all the possible infection processes $span_{(0\to T)}\Omega(t)_v : func(\Omega(t)) = \Theta$.

$$\tilde{span}_{(0\to T_v^*)}\Omega(t)_v = arg \max_{v\in V, span_{(0\to T)}\Omega(t)_v : func(\Omega(t))=\Theta} Pr(span_{(0\to T)}\Omega(t)_v) \tag{4}$$

Accordingly, T_v^* is the optimal time duration of the optimal infection process starting from the node v. Therefore, we consider the source associated with the optimal infection process as a candidate for the rumor source.

B. THE OPTIMAL INFECTION PROCESS FOR REGULAR TREES

In this section, we construct our model based on a regular tree network with infinite levels in which each node has the same degree. Since there is no loop and each node has the same degree in the regular tree network, we first study the structure characteristics of the optimal infection process.

Our task is to identify the rumor source under the assumption that there is initially only one rumor source in the tree network. According to the literature [17], we adopt the same definition of infection eccentricity $\tilde{ecc}(v)$ and the Jordan infection center as follows:

Definition 3: For nodes $v, v' \in V$, we define $l(v, v')$ as the distance of the shortest path between v and the infected node v' given the observed graph topology Θ. Hence, the infection eccentricity $\tilde{ecc}(v)$ is defined as the maximum distance from v to any infected nodes.

$$\tilde{ecc}(v) = \max_{v\in V, v'\in \mathrm{I}} l(v, v') \tag{5}$$

Similarly, the Jordan infection center denotes the node with the minimum infection eccentricity, i.e.,

$$J = arg \min_{v\in V} \tilde{ecc}(v) \tag{6}$$

Take the network of Fig.3 as an example, where the graph on the left is an observed graph topology with 5 nodes. The maximum distance from node 1 to other infected nodes is 2, i.e., $\tilde{ecc}(1) = 2$. Similarly, it is easy to get that $\tilde{ecc}(2) = 2$, $\tilde{ecc}(3) = 3$, $\tilde{ecc}(4) = 3$, $\tilde{ecc}(5) = 3$. Therefore, we conclude that the Jordan infection centers are node 1 and node 2 in this observed graph topology.

Next, we will demonstrate the source related with the optimal infection process is the candidate for the rumor source and the Jordan infection center is the most likely to be the rumor source among all the candidates. This conclusion can be deduced from the following 3 lemmas.

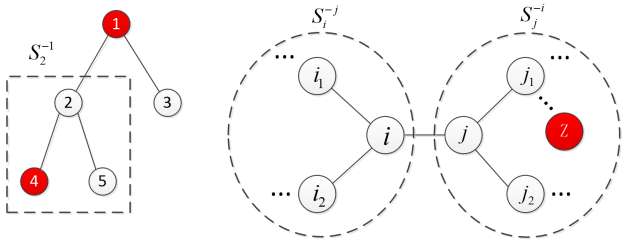

FIGURE 4. Illustration for Lemma1 and Lemma2. The left picture shows an observed graph topology Θ , where we assume that the node 1 is the rumor source and $\tilde{ecc}(1) = 2$. If the rumor propagates from the node 1 to the node 4, the time duration of infection would be at least 4. The right picture describes the observed graph topology Θ composed of two parts: the subtree S_i^{-j} and S_j^{-i}.

1) THE INFECTION ECCENTRICITY AND THE OPTIMAL TIME DURATION

Lemma 1: Assume the case that the underlying graph G is a regular tree network with infinite levels, in which the root node v_{root} is the rumor source, i.e., $\omega(v_{root}) = I$. In addition, a graph topology Θ at a certain time-slot has been observed, where exists at least one infected node. We can derive the following conclusion.

- The time duration $t_{v_{root}}$ of all the possible infection processes $span_{(0\to T)}\Omega(t)_{v_{root}} : func(\Omega(t)) = \Theta$ starting from node v_{root} is

$$t_{v_{root}} \in [2 \cdot \tilde{ecc}(v_{root}), \infty) \tag{7}$$

  where $\tilde{ecc}(v_{root})$ is the infection eccentricity of node v_{root}, where the infection eccentricity represents the maximum distance of one node to the other infected nodes (Definition 3).
- The probability $Pr(\tilde{span}_{(0\to T^*_{v_{root}})}\Omega(t)_{v_{root}})$ is a monotonically decreasing function for $t_{v_{root}} \in [2 \cdot \tilde{ecc}(v_{root}), \infty)$. e.g., for the infected time duration $2 \cdot \tilde{ecc}(v_{root}) \le T_1 < T_2$, we have

$$Pr(\tilde{span}_{(0\to T_1)}\Omega(t)_{v_{root}}) > Pr(\tilde{span}_{(0\to T_2)}\Omega(t)_{v_{root}}) \tag{8}$$

  where the probability $Pr(\tilde{span}_{(0\to T_v^*)}\Omega(t)_v)$ is the likelihood of the optimal infection process $\tilde{span}_{(0\to T_v^*)}\Omega(t)_v$ starting from v.
- For the graph topology Θ, we get the optimal time duration as follows:

$$T^*_{v_{root}} = 2 \cdot \tilde{ecc}(v_{root}) \tag{9}$$

  where $T^*_{v_{root}}$ is the optimal time duration in the optimal infection process $\tilde{span}_{(0\to T^*_{v_{root}})}\Omega(t)_{v_{root}}$ starting from node v_{root}.

Intuition 1: First, Eq.(7) is easily understood based on the observed graph topology and the definition of infection eccentricity. Under the discrete time slot model, the rumor can only spread at most one hop from the rumor source at each time-slot. Therefore, if the rumor source intends to infect the target node, the time duration is at least equal to $2 \cdot \tilde{ecc}(v)$. An example is shown in the left picture of Fig.4, it is an

observed graph topology Θ, where we assume that node 1 is the rumor source. It is easy to know $e\tilde{c}c(1) = 2$ according to the concept of infection eccentricity. If the rumor propagates from node 1 to node 4, the time duration of infection would be at least 4. i.e., the time duration $t_1 \geq 2 \cdot e\tilde{c}c(1) = 4$.

Next, we use the inductive hypothesis to prove the inequality (8). We construct two optimal infection processes that originated from the same rumor source v_{root}, where the time duration is different. First, suppose that the maximum distance from the rumor source to other infected nodes is 0 in the observed graph topology Θ , i.e., $\theta(v) = 0(v \in V,\ v \neq v_{root})$, $\theta(v_{root}) = 1$, we demonstrate that this inequality monotonically decreases with the time duration t. Next, we claim that the inequality (8) is still true when this maximum distance is equal to n. Finally, we prove that the inequality (8) still holds in the case of $n + 1$ based on the induction hypothesis.

According to the inequality (8), the optimal infection process $Pr(s\tilde{p}an_{(0\to T^*_{v_{root}})}\Omega(t)_{v_{root}})$ has the highest probability of leading to the observed graph topology when the time duration is $2 \cdot e\tilde{c}c(v_{root})$. Therefore, the optimal duration $T^*_{v_{root}}$ is equal to $2 \cdot e\tilde{c}c(v_{root})$ and Eqs.(9) holds. The detailed proof is shown in Appendix A.

2) THE OPTIMAL INFECTION PROCESS

According to Lemma 1, we can see that there is a unique optimal time duration T^*_v for each node $v \in V$. Next, we give out the lemma 2 as follows:

Lemma 2: Similar to Lemma 1, we assume the regular tree network with multiple levels, in which the root node v_{root} is the rumor source, and the observed graph topology Θ contains at least one infected node. For any pair of neighbor nodes i, j, if $T^*_i > T^*_j$, we can get the following conclusion.

$$Pr(s\tilde{p}an_{(0\to T^*_i)}\Omega(t)_i) < Pr(s\tilde{p}an_{(0\to T^*_j)}\Omega(t)_j) \tag{10}$$

We need the following steps to get the above conclusion.

- In the case of $T^*_i > T^*_j$, the subtree S_j^{-i} must contain at least one infected node, i.e., $S_j^{-i} \bigcap I \neq \emptyset$, where the S_j^{-i} represents the subtree rooted at node j which does not contain the branch from node i. Thus, we have the following conclusion based on the definition of infection eccentricity.

$$e\tilde{c}c(i) = e\tilde{c}c(j) + 1 \tag{11}$$

- We will show the exposed time of j is 1 and the infected time of j is 2 on the optimal infection process $s\tilde{p}an_{(0\to T^*_i)}\Omega(t)_i$ starting from node i, i.e., $\omega_j(1) = E$, $\omega_j(2) = I$, $t^E_j = 1, t^I_j = 2$.
- Given an optimal infection process $s\tilde{p}an_{(0\to T^*_i)}\Omega(t)_i$, we can always construct an optimal infection process $s\tilde{p}an_{(0\to T^*_j)}\Omega(t)_j$, $T^*_i > T^*_j$, whose probability of $s\tilde{p}an_{(0\to T^*_j)}\Omega(t)_j$ is higher than the probability of $s\tilde{p}an_{(0\to T^*_i)}\Omega(t)_i$.

Intuition 2: To prove the inequality (10), we consider the network graph topology Θ to be composed of two parts: the subtree S_i^{-j} and S_j^{-i}, as shown in the right picture of Fig.4. For the Eq.(11), it is easy to understand that the subtree S_j^{-i} contains at least one infected node in the observed graph topology Θ. If the subtree S_j^{-i} does not contain at least one infected node, it means that all the infected nodes are in the subtree S_i^{-j}. According to the definition of infection eccentricity, we get $e\tilde{c}c(i) + 1 = e\tilde{c}c(j)$ due to the fact that node i, j are neighbors. Thus, we have $T^*_i < T^*_j$ according to Eq.(9), which contradicts the assumption that $T^*_i > T^*_j$.

Next, we can easily see that the furthest infected node (we denote that this node as z) is on the optimal infection process $s\tilde{p}an_{(0\to T^*_i)}\Omega(t)_i$ from the node i must be in the subtree S_j^{-i} (If z is in the S_i^{-j}, it will contradict the fact that $e\tilde{c}c(i) = e\tilde{c}c(j) + 1$). Assuming that $t^E_j = 1, t^I_j > 2$ is on this optimal infection process $s\tilde{p}an_{(0\to T^*_i)}\Omega(t)_i$, we have $T^*_i - t^I_j = T^*_j + 2 - t^I_j < T^*_j$ based on $e\tilde{c}c(i) = e\tilde{c}c(j) + 1$ ($T^*_i = T^*_j + 2$). In other words, this means that the rumor staring from i cannot reach z in the subtree S_j^{-i} on this optimal infection process. Similarly, it does not hold when $t^E_v > 1, t^I_v > 2$. Therefore, we get the conclusion that $t^E_j = 1, t^I_j = 2$ on the $s\tilde{p}an_{(0\to T^*_i)}\Omega(t)_i$.

Finally, given an optimal infection process $s\tilde{p}an_{(0\to T^*_i)}\Omega(t)_i$, we can always construct an optimal infection process $s\tilde{p}an_{(0\to T^*_j)}\Omega(t)_j$, $T^*_i > T^*_j$, whose probability of $s\tilde{p}an_{(0\to T^*_j)}\Omega(t)_j$ is higher than the probability of $s\tilde{p}an_{(0\to T^*_i)}\Omega(t)_i$. The detailed proof for Lemma 2 is shown in Appendix B.

3) THE JORDAN INFECTION CENTER

According to Lemma 1 and Lemma 2, we can derive Lemma 3.

Lemma 3: Consider the situation that the underlying snapshot G is a regular tree network with infinite levels. A graph topology Θ observed at a certain time-slot contains at least one infected node. The Jordan infection center is the most likely to be the rumor source among all the possible candidates, which are the source of all optimal infection processes.

$$\hat{s} = arg \min_{v \in V} e\tilde{c}c(v) \tag{12}$$

Intuition 3: For any pair of non-adjacent nodes i, j and the infection eccentricity of i is larger than the infection eccentricity of j. i.e., $e\tilde{c}c(i) > e\tilde{c}c(j)$, we will prove that there is a path from the node i to the node j, in which the infection eccentricity of all nodes on this path monotonically decrease. By applying Lemma 2 repeatedly, we can deduce the conclusion that the optimal infection process starting from the node j is more likely to happen than the optimal infection process starting from the node i. Thus, the optimal infection process starting from the Jordan infection center is most likely to result in the observed graph topology Θ and the Jordan infection center is the most likely to be the real rumor source. The detailed proof is shown in Appendix C.

In addition, it is obvious that a tree network exists at most two Jordan infection centers; however, if the network contains

two Jordan infection centers, they must be neighbors. The detailed proof is presented in Appendix D.

IV. COMPUTATIONAL COMPLEXITY ANALYSIS

In this section, we analyze the computational complexity of our proposed scheme. According to Lemma 3, we need to firstly calculate the infection eccentricity of all nodes in networks in order to identify the Jordan infection center. According to the RI algorithm [17], we number all infected nodes based on the infection snapshot and let them broadcast their IDs to their neighbors. Every node checks the received IDs of the infected nodes. If an ID has not been received, the node records the ID and the time t, then forwards it to neighbors. The above procedure continues until a node receives all IDs of infected nodes, then this node is considered as the Jordan infection center. It is easy to verify that the duration of the above procedure is equal to the minimum infection eccentricity, which is the Jordan infection center. For the above description, the maximum number of IDs transmitted for each edge in networks is $|I|$. Therefore, it is obvious that in the worst-case scenario, the time complexity of identifying the Jordan infection center is $O(|I| \cdot |E|)$.

Here, we analyze other centrality heuristics traditionally applied to detect rumor source. In graph theory and network analysis, centrality heuristics is used to identify the critical vertices in networks. Closeness centrality and Betweenness centrality are traditional measurement methods used to estimate rumor source and these concepts are defined as follows:

(1) The closeness centrality of a node x is the average distance of the shortest path between x and all other reachable nodes in the graph.

$$Centr_C(x) = \frac{n-1}{\sum\limits_{x,y \in V} d(x,y)}$$

where $d(x, y)$ is the shortest distance between node x and node y. Thus, the computational complexity for the closeness centrality is O($|I|\cdot|E|$) [40].

(2) The Betweenness centrality quantifies the number of times a node x acts as a bridge along the shortest path between two other nodes i, j.

$$Centr_B(x) = \sum_{x \neq i \neq j \in V} \frac{\sigma_{i,j}(x)}{\sigma_{i,j}}$$

where $\sigma_{i,j}$ is total number of shortest paths from the node i to the node j, $\sigma_{i,j}(x)$ is the number of those paths that pass through x. Note that $\arg\max\limits_{x \in V} Centr_C(x)$ and $\arg\max\limits_{x \in V} Centr_B(x)$ indicate that the estimator of the rumor source is node x in the above case. Similarly, the time complexity of the betweenness centrality is O($|I|\cdot|E|$) [40].

The comparison of computational complexity for all the above heuristics is presented in Table 2. Note that the time complexity of the above showed centrality heuristics are the same; however in practice the efficiency of our scheme is better than other centrality heuristics since the time complexity of our scheme reaches O($|I|\cdot|E|$) in the worse-case.

TABLE 2. The comparison of computational complexity.

No	Heuristics	Computational complexity
1	Closeness centrality	O($\|I\|\cdot\|E\|$)
2	Betweenness centrality	O($\|I\|\cdot\|E\|$)
3	Ours	O($\|I\|\cdot\|E\|$)

V. EXPERIMENTS

We compared the performance of the optimal infection processes based estimator (OP) with the following centrality heuristics based estimators: Closeness centrality (CCE) and Betweenness centrality (BCE), and performed experiments over various networks including tree networks with different degrees, two well-known synthetic complex networks and four real-world networks. The description of the network datasets is shown in Table 3.

A. EXPERIMENTS ON REGULAR TREE NETWORKS

This section provides the simulation results on regular trees with different degrees to compare the performance of OP, CCE and BCE.

We first generated a graph of regular tree networks based on specified degree, which ranges from 2 to 6. In regular tree networks, a node was chosen randomly as the source of rumor and let this node spread following the model SEIR, in which the probability of infection processes was chosen uniformly. i.e., $q_1 \in (0, 1)$, $q_2 \in (0, 1)$, $r_2 \in (0, min(q_2, 1 - q_2))$, $r_1 \in (0, r_2)$. Since there are 4 states in the processes of infection, we set a smaller probability of recovery to ensure that there are sufficient infected nodes for our analysis. For each specified degree, we constructed multiple regular tree networks containing 200-1000 nodes and repeated the simulation experiment 1000 times, and then evaluated the average of the rumor detection probability. In order to observe the effect of different numbers of infected nodes, the duration t of infection processes was chosen uniformly. Considering the size of the network, the duration t of infection processes was chosen uniformly from [3,20] to ensure the full spread of infection processes.

We defined the detection probability to be the fraction of experiments in which the estimator coincides with the actual source. In addition, the distance between the estimator and the source is defined as the error distance. The detection probability of the rumor source under the regular tree networks with different degrees is shown in Fig. 5. The histogram of average error distance as measured in number of hops under the regular tree networks of different degrees is shown in Fig. 6.

From Fig. 5, we can see that the average value of rumor detection probability of OP is 55% and the rumor detection rate for CCE and BCE is about 25%. OP is higher than CCE and BCE by approximately 30%. From Fig. 6, the estimators

TABLE 3. The description of the network datasets.

Networks	V	E	E/V	Description
Scale-free	3000	11000	3.5	Scale-free networks [35]
Small-world	3000	12000	4	Small-world networks [36]
Power grid	4941	6594	1.3	The network representing the topology of the Western States Power Grid of the United States [36].
Facebook	4039	88234	21.8	The network dataset consists of 'circles' (or 'friends lists') from Facebook and was collected from survey participants using the Facebook app [37].
Autonomous systems	10670	22002	2.1	The network for Autonomous Systems (AS) peering information inferred from Oregon route-views [38].
Wikipedia vote network	7115	103689	14.5	The network contains all the Wikipedia voting data from the inception of Wikipedia [39].

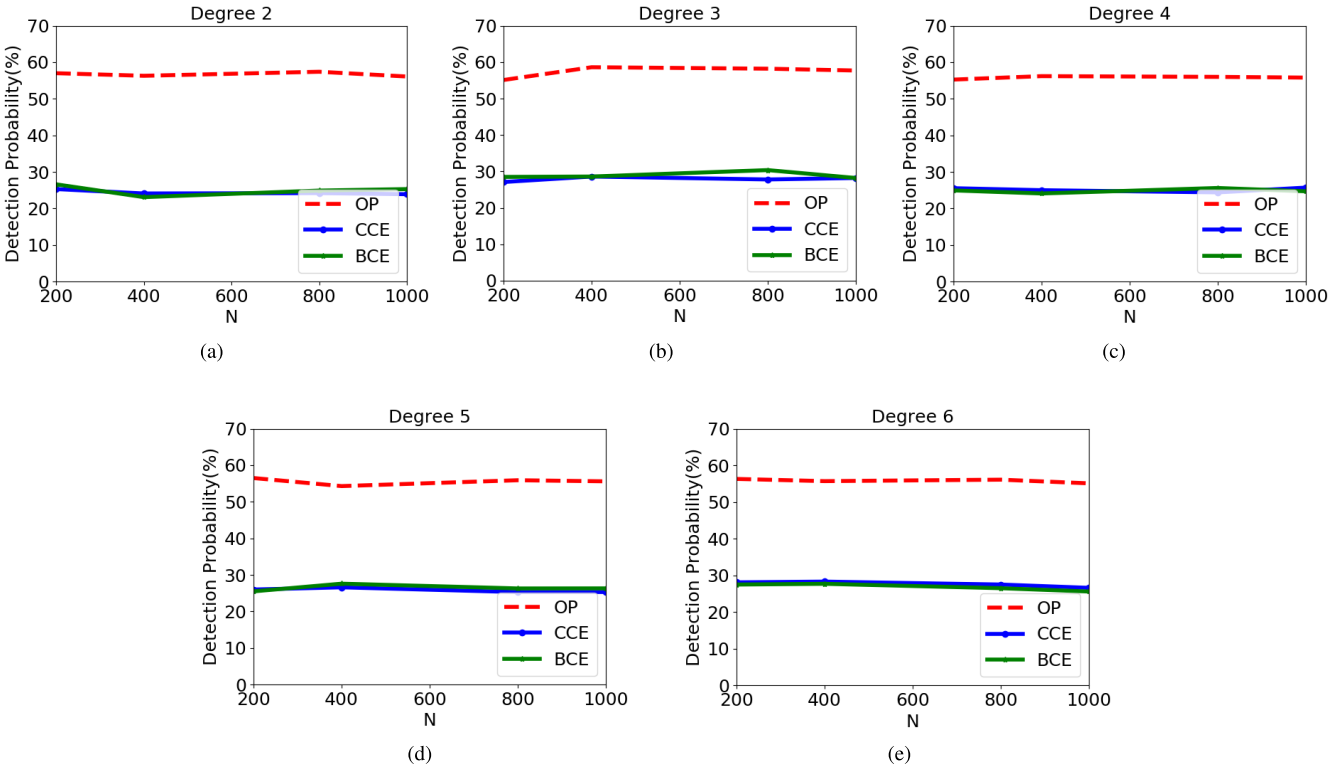

FIGURE 5. Rumor detection probability of OP, CCE, and BCE utilized on regular tree networks with different degrees. (a) Regular tree network with the degree 2. (b) Regular tree network with the degree 3. (c) Regular tree network with the degree 4. (d) Regular tree network with the degree 5. (e) Regular tree network with the degree 6.

of rumor source identified by OP, CCE and BCE are concentrated within 0-2 hops away from the rumor source due to the relatively small diameter of regular tree networks. Although CCE and BCE have higher rumor detection probability at 1-2 hops than OP, OP is more efficient to identify the rumor source (0 error distance for 55% versus 30%).

B. EXPERIMENTS ON SYNTHETIC COMPLEX NETWORKS

Next, we performed simulations under scale-free network and small-world network. These well-known synthetic networks were proposed in [35] and [36], where the scale-free networks contain 3000 nodes, 11000 edges, and the small-world networks contain 3000 nodes, and 12000 edges, respectively. According to the definition of Jordan infection center, it is easy to find the Jordan infection centers given infection snapshots of these network topologies. Thus we can view Jordan infection centers as the possible candidates of the rumor source. We compared the performance of the OP with CCE and BCE in these two synthetic networks, i.e., scale-free network and small-world network. Histograms were used to show the error distances (or hops) between the three estimators and the actual source. These experiment settings were the same as in the previous simulation. We randomly selected a node as the rumor source and let it spread in the network, in which the probability of the infection processes was chosen uniformly. i.e., $q_1 \in (0, 1)$, $q_2 \in (0, 1)$, $r_2 \in (0, min(q_2, 1-q_2))$, $r_1 \in (0, r_2)$. The infection duration t was chosen uniformly from [3, 100]. We chose a smaller recovery

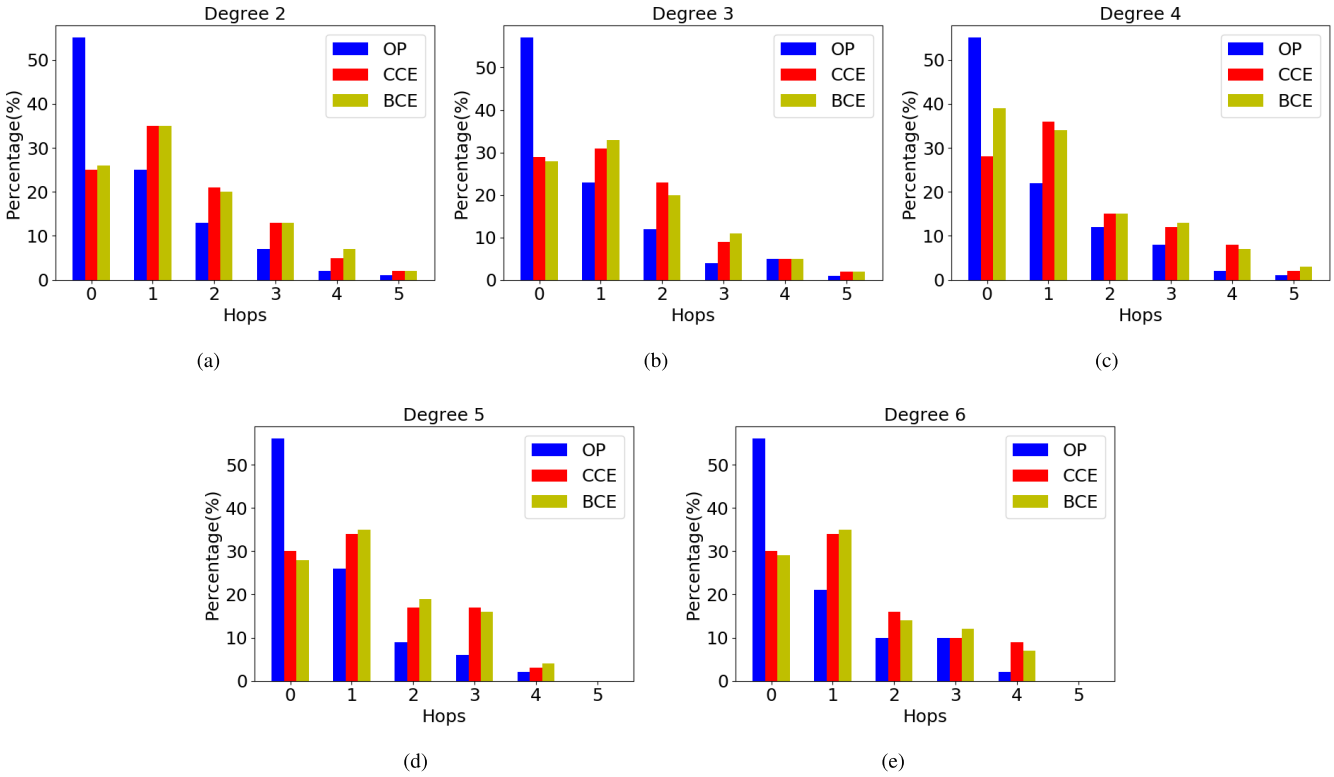

FIGURE 6. Histogram of average error distances as measured in number of hops of OP, CCE and DCE applied to regular tree networks with different degrees. (a) Regular tree network with degree 2. (b) Regular tree network with degree 3. (c) Regular tree network with degree 4. (d) Regular tree network with degree 5. (e) Regular tree network with degree 6.

probability and a longer time duration t in order to ensure that there were more infected nodes in the infection snapshot for analysis. Simulation experiments were repeated 2000 times and the average error distances were recorded.

Fig. 7(a) and Fig. 7(b) show experimental results of scale-free network and small-world network. For the scale-free network and the small-world we used, the average ratio of edges to nodes is 3.5 and 4, respectively. As it can be seen, the estimators of the rumor source identified by OP, CCE and BCE in this scale-free network are concentrated within 1 hop away from the actual source. However, the estimator of rumor source is within 0-5 hops away from the actual source in the small world network. Due to the power law degree distributions, the scale-free networks contain many nodes with high degrees; however, the degrees of nodes in the small-world network apt to be average.

From Fig. 7(a), we can see that CCE and BCE have higher detection probability at one hop than OP (80% for CCE, 82% for BCE versus 48% for OP) in scale-free network. However, OP is more effective to identify rumor source (0 error distance) than other heuristics(50% versus 20% for CCE, 14% for BCE). Similarly, we can see from Fig. 7(b) that estimators of rumor source of BCE and CCE are concentrated within 1-5 hops. Although CCE and BCE have a higher rumor detection probability at one hop (16% for CCE, 14% for BCE versus 9%), OP can effective identify the true rumor source (0 error distance 52% versus 4% for CCE, 5% for BBC).

C. EXPERIMENTS ON REAL NETWORKS

In this part, four real-world networks are simulated. We benchmark the performance of the OP with CCE and BCE on following real-networks—power grid network, Facebook, Autonomous systems and Wikipedia vote network, which are described as follows:

- The network topology of power grid is an undirected, unweighted network representing the topology of the Western States Power Grid of the United States. Simulation data was compiled by D. Watts and S. Strogatz, which includes are 4941 nodes and 6594 edges. The average ratio of edges to nodes in power grid is 1.3 [36].
- Facebook data was collected from survey participants using Facebook App, and it was anonymized by replacing the Facebook-internal ids for each user with a new value. It contains 4039 nodes and 88234 edges. The average ratio of edges and nodes is 21.8 [37].
- The network for Autonomous Systems (AS) peering information was obtained from Oregon route-views, it consists of 10670 nodes and 22002 edges. The average ratio of edges and nodes is 2.1 [38].
- Wikipedia vote network contains all Wikipedia voting data from inception, which contains 7115 nodes and 103689 edges, and the average ratio of edges and nodes is 14.5 [39].

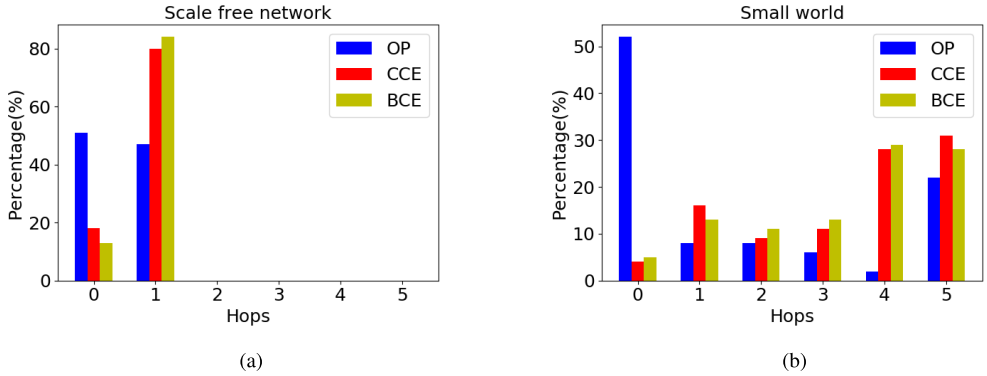

FIGURE 7. Histogram of error distances as measured in number of hops of OP, CCE and DCE applied on synthetic complex networks. (a) Scale-free network, (b) small world network.

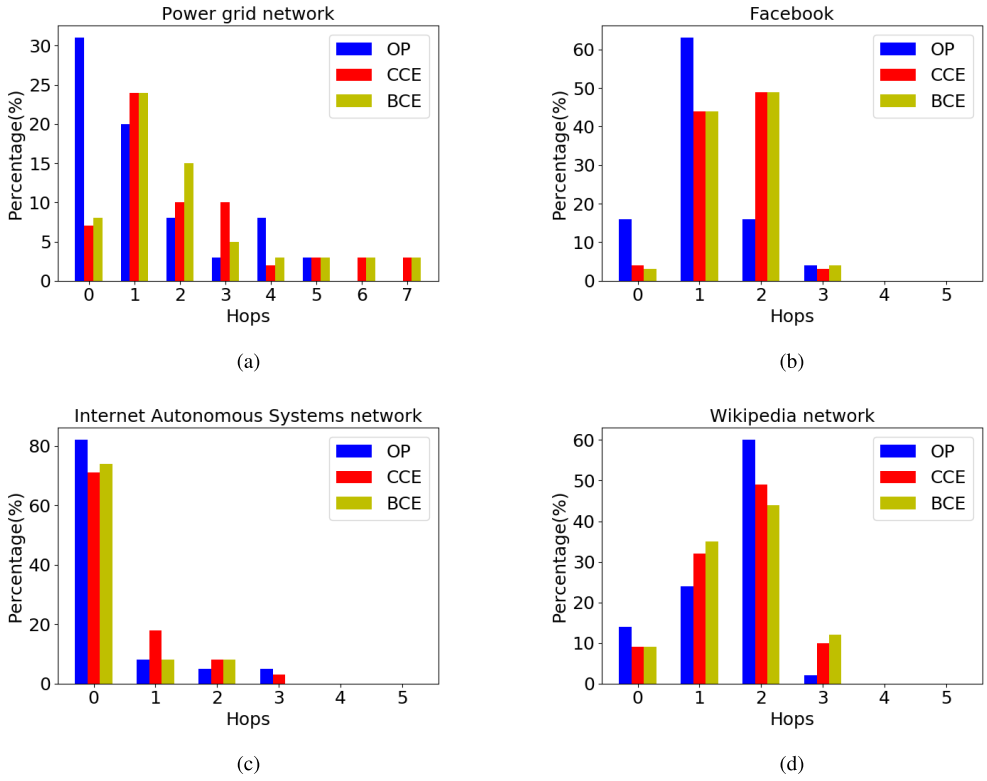

FIGURE 8. Histogram of error distances as measured in number of hops of OP, CCE and DCE applied on real-world networks. (a) Power grid network. (b) Facebook. (c) Internet Autonomous Systems network. (d) Wikipedia vote network.

These experiment settings were the same as in the previous simulation, in which the probabilities of infection and recovery were chosen uniformly. i.e., $q_1 \in (0, 1)$, $q_2 \in (0, 1)$, $r_2 \in (0, min(q_2, 1 - q_2))$, $r_1 \in (0, r_2)$. A node was randomly selected as the rumor source and the rumor spreads in the network. The infection duration t was chosen uniformly from [3, 100]. Simulation experiments were repeated 2000 times and the average error distances were recorded.

Fig. 8(a) and Fig. 8(b) show experimental results of the power grid network and the Facebook social network, respectively. As we can see, for the power grid network, the estimators of the rumor source identified by OP, CCE and BCE are within 7 hops from the actual rumor source. However, for the Facebook, the estimators of rumor source is within 3 hops from the actual rumor, which is on the account of the fact that the degrees of nodes in the power grid network coinciding with small world networks are relatively average, and the Facebook social network coinciding with scale-free networks have many nodes with high degrees.

From Fig. 8(a), we can see that CCE and BCE in the power grid network achieve higher detection probability at one hop than OP (23.5% for CCE, 24% for BCE versus 20% for OP). However, the estimators of rumor source for OP is more accurate (0 error distance, 32% versus 7% for

CCE and 8% for BCE). From Fig. 8(b), we can see that the detection probability of the rumor source drops due to the many 88234 edges in Facebook. The ratio of edges to nodes in this network (21.8) is larger than the ratio of edges to nodes in other networks. But we can clearly see that the estimators of the rumor source identified by OP is higher than CCE and BCE within 1 hop away from the actual rumor source. OP has a higher rumor detection probability at 1 hop (61% versus 43% for CCE, 42% for BCE). Moreover, OP can effectively identify the actual rumor source(0 error distance, 17% versus 4% for CCE, 3% for BCE).

Fig. 8(c) and Fig. 8(d) describe results of rumor source identification on Internet Autonomous Systems network and Wikipedia network, which shows the histogram of error distances for three different rumor source estimators: OP, CCE and BCE. From Fig. 8(c), our heuristic OP correctly identifies the rumor source than other centrality heuristics(81% versus 70% for CCE, and 74% for BCE, respectively). For the Internet Autonomous Systems network, similar to the power grid network and the small world network, the average ratio of edges to nodes is 2.1 and the average degree of nodes is small. Thus it can be explained why OP has a higher rumor identification than other centrality heuristics. For the Wikipedia network, it is obvious that OP is more capable of detecting rumor source than CCE and BCE (0 error distance, 14% versus 9% for CCE and BCE, respectively). The reason can be attributed to the fact that the Wikipedia network consists of 7115 nodes and 103689 edges, and many high degree hubs, which is similar to the Facebook and scale free networks.

VI. DISCUSSION

In this paper, we developed the optimal infection process to solve the rumor source identification issue under the SEIR model, where the optimal infection process represents a possible infection process with the maximum probability of leading to the observed graph topology. We compute an estimator for the rumor source according to the optimal infection process. Subsequently, we proved that this estimator of rumor source is equal to the Jordan infection center through detailed proofs. Therefore, rumor source identification can be simplified to the problem of identifying the Jordan infection center based on the final observed graph topology. The time complexity of our approach is $O(|I|\text{ąd}|E|)$, which is equal to closeness centrality heuristic and betweenness centrality heuristic used widely in rumor source detection in the worse-case, thus our approach performs better than other centrality heuristics. Next, simulations are performed on various networks, including regular networks with different degrees, two well-known synthetic complex networks, and real networks. Experimental results suggest that our proposed scheme improves the estimation accuracy for rumor source compared to other centrality heuristics. The average value of rumor detection probability for OP is higher than CCE and BCE by approximately 25% under regular tree networks. For the synthetic complex networks and real-world networks, more than 80% of the estimators of the rumor source for OP are within two hops away from the actual source. Moreover, simulations suggest that our OP heuristic performs well for rumor identification probability on synthetic complex networks and real-world networks compared with the previous centrality heuristics.

However, it can be seen that in the scenario with a large number of nodes, such as Facebook and Wikipedia networks, the estimation accuracy of our heuristic still needs to be improved. Due to the large average degree of nodes in these networks, and there are many high degree hubs in them, the spread of rumors in these networks is more rapid, and so, it is more difficult to track the actual source. One possible practical means of addressing this problem is effectively filtering the unrelated nodes in networks with many high degree nodes. In addition, MCMC methods [41], [42] should be considered to fully explore the uncertainty of rumor source location for improving the accuracy in our future work. In addition, identification of multiple rumor sources remains an attractive future direction. Assume that there are n rumor sources in the network, and each rumor source infects all S-state neighbors at the same time. Our goal is to infer the most likely n estimators of rumor sources based on one observed infection snapshot. We can firstly divide all nodes in the observed snapshot into n subsets by using the method from the literature [43], and then we can derive an estimator of rumor source in each subset based on the proposed optimal infection process. Finally, n estimators for the actual rumor source could be derived.

VII. CONCLUSION

In this paper, we addressed the detection problem for one single rumor source based on the observed snapshot under the SEIR model. We obtained an estimator of the rumor source in online networks through our proposed optimal infection processes (OP), and demonstrated that the estimator of the rumor source equals to the Jordan's infection center using induction hypotheses. Subsequently, we evaluated the performance of rumor source detection of OP over various networks, including regular tree networks, two well-known synthetic complex networks and four real-world networks. Based on the simulation results, the rumor identification probability of our heuristic is higher than closeness centrality heuristic (CCE) and betweenness centrality heuristic (BCE) traditionally adopted in rumor source detection. Through performance analysis of computational complexity, we showed that our heuristic has advantages in efficiency compared with the previous rumor centrality heuristics. We believe our work forms an important theoretical basis towards more effective ways of detecting rumor sources, which is crucial given the impact that false or misleading piece of information (e.g., rumors) can have in society, especially with large and growing social media networks today and in the future.

APPENDIX

This section shows all detailed proofs for the Lemma 1-4.

PROOF OF THE LEMMA 1

For Lemma 1, we will show the probability $Pr(\tilde{span}_{(0\to T^*_{v_{root}})}\Omega(t)_{v_{root}})$ of the optimal infection process is the monotonically decreasing function for $t_{v_{root}} \in [2\cdot \tilde{ecc}(v_{root}), \infty)$ in the graph topology Θ. In the observed graph topology Θ, It is obvious that the rumor that propagates from the source v_{root} to the farthest infected node requires at least $2\cdot\tilde{e}(v_{root})$ time-slots due to the rumor can only spread at most one hop at each time-slot. Therefore, the infection duration $t_{v_{root}}$ is at least $2\cdot\tilde{e}(v_{root})$.

Firstly, we assume there are two optimal infection processes with one time difference in the observed topology Θ: $\tilde{span}_{(0\to T^*_{v_{root}})}\Omega(t)_{v_{root}}$ and $\tilde{span}_{(0\to T^*_{v_{root}}+1)}\Omega(t)_{v_{root}}$, we will show

$$Pr(\tilde{span}_{(0\to T^*_{v_{root}})}\Omega(t)_{v_{root}}) > Pr(\tilde{span}_{(0\to T^*_{v_{root}}+1)}\Omega(t)_{v_{root}}) \tag{13}$$

According to definition of the optimal infection process (Definition 2), the inequality (13) can be expressed as follows:

$$\max_{span_{(0\to T)}\Omega(t)_{v_{root}}:func(\Omega(t))=\Theta} Pr(span_{(0\to T)}\Omega(t)_{v_{root}}) > \max_{span_{(0\to T+1)}\Omega(t)_{v_{root}}:func(\Omega(t))=\Theta} Pr(span_{(0\to T+1)}\Omega(t)_{v_{root}}) \tag{14}$$

Next we use the induction hypothesis to prove above inequality (14). First, we define $\{\Theta^k\}$ as the collection of the possible observed topology Θ, where k is the largest distance from root v_{root} to an infected node. For example, $\{\Theta^0\}$ is the collection of a possible observed topology Θ, in which this observed topology Θ only contains one infected node when we are observing. This infected node is the rumor source due to $k=0$.

A.

when $k=0$, v_{root} is the only infected node and all the possible infection processes follow observed graph topologies $\Theta\in\{\Theta^0\}$.

$$\begin{aligned} &Pr(span_{(0\to T)}\Omega(t)_{v_{root}}) \\ &= Pr(\omega(t)_{v_{root}} = I, 0\le t\le T) \\ &\quad\cdot\prod_{u\in C(v_{root})} Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})|\omega(t)_{v_{root}} = I) \end{aligned} \tag{15}$$

$$\begin{aligned} &= (1-r_1)^T \\ &\quad\cdot\prod_{u\in C(v_{root})} Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})|\omega(t)_{v_{root}} = I) \end{aligned} \tag{16}$$

Here, $(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})$ represents the infection process $span_{(0\to T)}\Omega(t)_{v_{root}}$ from time-slot 0 to T in the subtree $S_u^{-v_{root}}$. In addition, $C(v_{root})$ represents a collection of children of node v_{root}. $(1-r_1)^T$ means v_{root} maintains its own state I throughout the infection process.

Next, $u\in C(v_{root})$ has three possible states : S, E or R. Based on (16), we obtain the following analyses.

(1). If u is susceptible and $\omega_u(T)=$ S: u maintains the susceptible state throughout the T time-slots and v_{root} cannot infect u. we obtain

$$Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})|\omega(t)_{v_{root}} = I) = (1-q_1)^T \tag{17}$$

(2). If u is in the exposed state and $\omega_u(T)=$ E: u was exposed and turned to E(S→E) in a certain time-slot but it can maintain its own state and was not infected until time-slot T. We denote by t_u^E its exposed times, so we have

$$\begin{aligned} &Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})|\omega(t)_{v_{root}} = I) \\ &= (1-q_1)^{t_u^E-1}q_1(1-q_2-r_2)^{T-t_u^E} \\ &= \frac{q_1}{1-q_1}\left(\frac{1-q_1}{1-q_2-r_2}\right)^{t_u^E}(1-q_2-r_2)^T \end{aligned} \tag{18}$$

where the probability $(1-q_1)^{t_u^E-1}q_1(1-q_2-r_2)^{T-t_u^E}$ indicates that the node u changes to the E state at time-slot t_u^E and maintains its own state E until the T time-slots. Next, if $\frac{1-q_1}{1-q_2-r_2} > 1$, this probability (18) is maximum when t_u^E equals to T, we have

$$\max(Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})|\omega(t)_{v_{root}} = I)) = (1-q_1)^{T-1}q_1$$

or if $\frac{1-q_1}{1-q_2-r_2} < 1$, then when $t_u^E = 1$(u was exposed at time-slot 1), this probability (18) is maximum

$$\max(Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})|\omega(t)_{v_{root}} = I)) = q_1(1-q_2-r_2)^{T-1}$$

(3). If u is in the recovered state, $\omega_u(T)=$ R. Due to the infection processes of the node u have two possibilities: u was exposed to the state E from the state S and has not been infected, finally it recovered within T time-slots (S→E→R); or u was exposed to the state E and was infected at some time-slots, it recovered at another time-slots (S→E→I→R). The infected time, exposed time and recovered time are defined as t_u^E,t_u^I and t_u^R respectively.

(3.1). The processes of states transition of u: S→E→R. The probability (16) can be expressed as follows.

$$\begin{aligned} &Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})|\omega(t)_{v_{root}} = I) \\ &= (1-q_1)^{t_u^E-1}q_1(1-q_2-r_2)^{t_u^R-t_u^E-1}r_2 \\ &\le q_1\cdot r_2 \end{aligned} \tag{19}$$

Here, $t_u^E-1\ge 0$ and $t_u^R-t_u^E-1\ge 0$, the maximum value of (19) can be achieved when $t_u^E=1$, $t_u^R=2$, i.e., node u was exposed at the first time-slot and recovered at the next time-slot, Hence, $\omega_u(T)=$ R.

(3.2). The processes of states transition of u: S→E→I→R.

$$\begin{aligned} &Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})|\omega(t)_{v_{root}} = I) \\ &= (1-q_1)^{t_u^E-1}q_1(1-q_2-r_2)^{t_u^I-t_u^E-1} \\ &\quad\cdot q_2(1-r_1)^{t_u^R-t_u^I-1}r_1 \end{aligned} \tag{20}$$

$$\cdot\prod_{w\in C(u)} Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_w^{-u})|t_u^E, t_u^I, t_u^R) \tag{21}$$

Here, the infinite subtree S_w^{-u} exists at least one node $\mu \in S_w^{-u}$ such that μ is either susceptible or exposed, and its parent node v is recovered. e.g., $\mu = w$ and parent node $v = u$. In addition, we denote the set of nodes $S_w^{-u} \setminus S_\mu^{-v}$ that nodes are in the subtree S_w^{-u} but not in the subtree S_μ^{-v}.

- If μ is susceptible, we denote by t_v^I, t_v^R node v infected and recovered time. The probability (21) can be expressed

$$\begin{aligned} &Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_w^{-u})|t_u^E, t_u^I, t_u^R) \quad (22)\\ &= Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_w^{-u}\setminus S_\mu^{-v})|t_u^E, t_u^I, t_u^R)\\ &\quad \cdot Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_\mu^{-v})|t_v^I, t_v^R)\\ &= Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_w^{-u}\setminus S_\mu^{-v})|t_u^E, t_u^I, t_u^R)\\ &\quad \cdot (1-q_1)^{t_v^R - t_v^I} \le (1-q_1)^{t_v^R - t_v^I} \le (1-q_1) \quad (23)\end{aligned}$$

  where the probability (23) is established due to $t_v^R - t_v^I \ge 1$.
- If μ is in exposed state, denoting by t_μ^E its own exposed time. We can get

$$\begin{aligned} &Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_w^{-u})|t_u^E, t_u^I, t_u^R)\\ &= Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_w^{-u}\setminus S_\mu^{-v})|t_u^E, t_u^I, t_u^R)\\ &\quad \cdot Pr(X([0,t], S_\mu^{-v})|t_\mu^E, t_v^I, t_v^R)\\ &= Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_w^{-u}\setminus S_\mu^{-v})|t_u^E, t_u^I, t_u^R)\\ &\quad \cdot (1-q_1)^{t_\mu^E - t_v^I - 1} q_1 (1-q_2-r_2)^{T-t_\mu^E} \quad (24)\\ &\quad \le q_1(1-q_2-r_2)^{T-t_\mu^E} \le q_1 \quad (25)\end{aligned}$$

  where the probability (24) is established due to $t_\mu^E - t_v^I - 1 \ge 0$. The maximum value of (24) can be achieved when $t_\mu^E = 3$, $t_v^I = 2$, i.e., node v was infected at the 2 time-slots and its child μ was exposed at the next time-slots.

Hence, the probability (21) can be expressed as follows

$$\begin{aligned} &\prod_{w\in C(u)} Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_w^{-u})|t_u^E, t_u^I, t_u^R)\\ &\qquad \le \max\{(1-q_1)^{|C(u)|}, q_1^{|C(u)|}\} \quad (26)\end{aligned}$$

According to the above formulation, the probability (16) can be expressed as

$$\begin{aligned} &Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})|\omega(t)_{v_{root}} = I)\\ &= (1-q_1)^{t_u^E - 1} q_1 (1-q_2-r_2)^{t_u^I - t_u^E - 1} q_2 (1-r_1)^{t_u^R - t_u^I - 1} r_1\\ &\quad \cdot \prod_{w\in C(u)} Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_w^{-u})|t_u^E, t_u^I, t_u^R)\\ &\quad \le q_1 \cdot q_2 \cdot r_1 \cdot \max\{(1-q_1)^{|C(u)|}, q_1^{|C(u)|}\} \quad (27)\end{aligned}$$

This probability (27) is maximized when $t_u^E = 1$, $t_u^I = 2$, $t_u^R = 3$, i.e.,u was infected to E at the first time-slot and turn to I at the next time-slot, then recovered in the third time-slot.

In summary, according to the definition of the optimal infection process below,

$$\begin{aligned} &Pr(\tilde{span}_{(0\to T^*_{v_{root}})}\Omega(t)_{v_{root}})\\ &\quad = \max_{span_{(0\to T)}\Omega(t)_{v_{root}}: func(\Omega(t))=\Theta} Pr(span_{(0\to T)}\Omega(t)_{v_{root}}) \end{aligned} \quad (28)$$

The probability (15) can be expressed as follows.

$$\begin{aligned} &\text{If } T = 0, \quad Pr(\tilde{span}_{(0\to T)}\Omega(t)_{v_{root}}) = 1\\ &\text{If } T = 1, \quad Pr(\tilde{span}_{(0\to T)}\Omega(t)_{v_{root}})\\ &\qquad = (1-r_1)\prod_{u\in C(v_{root})} \max Pr\{(1-q_1), q_1\}\\ &\text{If } T = 2, \quad Pr(\tilde{span}_{(0\to T)}\Omega(t)_{v_{root}})\\ &\qquad = (1-r_1)^2 \prod_{u\in C(v_{root})} \max Pr\{(1-q_1)^2, (1-q_1)q_1,\\ &\qquad\quad q_1(1-q_2-r_2), q_1\cdot r_2\}\end{aligned}$$

If $T \ge 3$, we have

$$\begin{aligned} &Pr(\tilde{span}_{(0\to T)}\Omega(t)_{v_{root}})\\ &= (1-r_1)^T \prod_{u\in C(v_{root})} \max Pr\{(1-q_1)^T, (1-q_1)^{T-1}q_1,\\ &\quad q_1(1-q_2-r_2)^{T-1}, q_1\cdot r_2,\\ &\quad q_1q_2r_1\cdot\max\{(1-q_1)^{|C(u)|}, q_1^{|C(u)|}\}\}. \end{aligned} \quad (29)$$

Here, if $\frac{1-q_1}{1-q_2-r_2} > 1$, it is easy to see that $(1-q_1)^{T-1}q_1 > q_1(1-q_2-r_2)^{T-1}$, if $\frac{1-q_1}{1-q_2-r_2} < 1$, we obtain $(1-q_1)^{T-1}q_1 < q_1(1-q_2-r_2)^{T-1}$. Note that $t = T$ is fixed and this probability (29) is monotonically decreasing of t.

For example, the optimal infection process (28) when $1-q_1 > q_1$ and $\frac{1-q_1}{1-q_2-r_2} > 1$ can be expressed as follows.

$$\begin{aligned} &Pr(\tilde{span}_{(0\to 1)}\Omega(t)_{v_{root}}) = (1-r_1)(1-q_1)^{|C(v_{root})|}\\ &Pr(\tilde{span}_{(0\to 2)}\Omega(t)_{v_{root}})\\ &\quad = (1-r_1)^2 \prod_{u\in C(v_{root})} \max P\{(1-q_1)^2, (1-q_1)q_1, q_1r_2\}\\ &Pr(\tilde{span}_{(0\to 3)}\Omega(t)_{v_{root}})\\ &\quad = (1-r_1)^3 \prod_{u\in C(v_{root})} \max P\{(1-q_1)^3, (1-q_1)^2q_1,\\ &\qquad\qquad q_1q_2r_1(1-q_1)^{|C(u)|}, q_1r_2\}\end{aligned}$$

Since $|C(u)| \ge 1$ and $r_2 > r_1$, we are easy to get $Pr(\tilde{span}_{(0\to 1)}\Omega(t)_{v_{root}}) > Pr(\tilde{span}_{(0\to 2)}\Omega(t)_{v_{root}}) > Pr(\tilde{span}_{(0\to 3)}\Omega(t)_{v_{root}})$. In a similar way, the same procedure can be easily adapted to obtain that $Pr(\tilde{span}_{(0\to 1)}\Omega(t)_{v_{root}}) > Pr(\tilde{span}_{(0\to 2)}\Omega(t)_{v_{root}}) > Pr(\tilde{span}_{(0\to 3)}\Omega(t)_{v_{root}})$. In summary, $Pr(\tilde{span}_{(0\to T)}\Omega(t)_{v_{root}})$ is a monotonically decreasing function for t when $k = 0$.

B.

Assume this inequality (13) holds for $k \le n$, then we consider $k = n+1$, i.e., the distance from the v_{root} to the farthest infected node is $n+1$ in the observed

graph topology Θ. Therefore, all the possible infection processes $span_{(0\to T)}\Omega(t)_{v_{root}}$ follow observed graph topologies $func(\Omega(t)_{root}) = \Theta \in \{\Theta^{n+1}\}$.

It is obvious that time duration t needs to satisfy $t \geq 2(n+1) \geq 1$ for each infection process $span_{(0\to T)}\Omega(t)_{v_{root}}$ in order to infect the farthest node.

First, we divide the set of subtrees $S = \{S_u^{-v_{root}} | u \in C(v_{root})\}$ into two subsets:

$$S^j = \{S_u^{-v_{root}} | u \in C(v_{root}), \quad \Theta(S_u^{-v_{root}}) \cap I = \emptyset\} \tag{30}$$

$$S^i = S \backslash S^j \tag{31}$$

where S^j is a set of subtree that does not contain infected nodes, and S^i is a subset of S that excludes the set S^j. Note that infection processes are mutually independent.

(1). Considering the set of $S^j \subset S_u^{-v_{root}}$, we can follow the conclusion (29) for $k = 0$.

- If $T \geq t_{v_{root}}^R \geq 1$, we have

$$\begin{aligned} &Pr(\tilde{span}_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}} | t_{v_{root}}^R) \\ &= \max Pr\{(1-q_1)^{t_{v_{root}}^R}, q_1(1-q_2-r_2)^{T-1} \\ &\quad (1-q_1)^{t_{v_{root}}^R - 1} q_1 (1-q_2-r_2)^{T-t_{v_{root}}^R}, q_1 \cdot r_2, \\ &\quad q_1 q_2 r_1 \cdot \max\{(1-q_1)^{|C(u)|}, q_1^{|C(u)|}\}\}. \end{aligned} \tag{32}$$

  where $t_{v_{root}}^R$ is the recovered time of node v_{root}.

- If $T < t_{v_{root}}^R$, then we have

$$\begin{aligned} &Pr(\tilde{span}_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}} | T_{v_{root}}^R) \\ &= \max Pr\{(1-q_1)^T, q_1(1-q_2-r_2)^{T-1}, q_1 r_2 \\ &\quad (1-q_1)^{T-1} q_1, q_1 q_2 r_1 \cdot \max\{(1-q_1)^{|C(u)|}, q_1^{|C(u)|}\}\}. \end{aligned} \tag{33}$$

  In short, we can get the conclusion that $Pr(\tilde{span}_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}} | t_{v_{root}}^R)$ in the subtree S^j is monotonically decreasing in t given any $t_{v_r}^R$.

(2). For the set of $S^i \subset S_u^{-v_{root}}$, given a optimal infection process $(\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})$, an infection process $(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})$ with a higher probability can be constructed. We represent t_u^E as the time at which the state of u changes to E in the infection process $(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})$, and similarly $\tilde{t}_u^E$ is the time at which the state of u changes to E in the optimal infection process $(\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})$.

- If $\tilde{t}_u^E > 1$, we let $\tilde{t}_u^E = t_u^E + 1$, i.e., in the optimal infection process $\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}$, u changed to E was one time-slot later than in the process $span_{(0\to T)}\Omega(t)_{v_{root}}$. Furthermore, we assume that these two infection processes are the same after node u receives the rumor. So we can get

$$\begin{aligned} &Pr(\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}, S_u^{-v_{root}}) \\ &= (1-q_1)^{\tilde{t}_u^E - 1} q_1 \cdot Pr(\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}, S_u^{-v_{root}} | \tilde{t}_u^E) \end{aligned} \tag{34}$$

and

$$\begin{aligned} &Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}}) \\ &= (1-q_1)^{t_u^E - 1} q_1 \cdot Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}} | t_u^E) \end{aligned} \tag{35}$$

where $Pr(\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})$ and $Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})$ are probabilities of node u receiving the rumor in the infection process $\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}$ and the infection process $span_{(0\to T)}\Omega(t)_{v_{root}}$.

Since we assume that the two infection processes of node u after receiving the rumor are the same, i.e.,

$$\begin{aligned} &Pr(\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}, S_u^{-v_{root}} | \tilde{t}_u^E) \\ &\quad = Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}} | t_u^E) \end{aligned} \tag{36}$$

Therefore, according to (34) and (35), we obtain

$$\begin{aligned} &Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}}) \\ &\quad > Pr(\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}, S_u^{-v_{root}}) \end{aligned} \tag{37}$$

where $\tilde{t}_u^E = t_u^E + 1$.

- If $\tilde{t}_u^E = 1$, we let $\tilde{t}_u^E = t_u^E = 1$, we have

$$\begin{aligned} &Pr(\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}, S_u^{-v_{root}}) \\ &\quad = q_1(1-q_2-r_2)^{\tilde{t}_u^I - 1 - 1} q_2 \\ &\quad\quad \cdot \prod_{w \in C(u)} Pr(\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}, S_w^{-u})) \end{aligned} \tag{38}$$

and

$$\begin{aligned} &Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}}) \\ &\quad = q_1(1-q_2-r_2)^{t_u^I - 1 - 1} q_2 \\ &\quad\quad \cdot \prod_{w \in C(u)} Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_w^{-u})) \end{aligned} \tag{39}$$

Here, since $S^i \subset S_u^{-v_{root}}$ and the subtree $S_u^{-v_{root}}$ must contain infected nodes, the node u will be infected at a certain time-slot. Respectively, we denote the infected time of node u by $\tilde{t}_u^I$ in $\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}$ and t_u^I in $span_{(0\to T)}\Omega(t)_{v_{root}}$. In addition, the node w is the child node of u.

According to conclusion of the Step 1 and the induction hypothesis, the subtree S_w^{-u} satisfies the condition $k \leq n$, $\Theta(S_w^{-u}) \in \{\Theta^m\}$, $m \leq n$, so we have

$$\begin{aligned} &\max Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_w^{-u}) \\ &\quad > \max Pr(span_{(0\to T+1)}\Omega(t)_{v_{root}}, S_w^{-u}) \end{aligned}$$

where $(span_{(0\to T)}\Omega(t)_{v_{root}}, S_w^{-u}) : func(\Omega(t)_{root}, S_w^{-u}) = \Theta(S_w^{-u})$. Then, according to (38) and (39), we can obtain

$$\begin{aligned} &\max Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}}) \\ &\quad > \max Pr(\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}, S_u^{-v_{root}}) \end{aligned}$$

In summary, we can always construct an infection process $(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})$ whose probability

$Pr(span_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})$ is larger than the probability $Pr(\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})$ of a given optimal process $(\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})$.

Therefore, $Pr(\tilde{span}_{(0\to T)}\Omega(t)_{v_{root}}, S_u^{-v_{root}})$ is the monotonically decreasing function for t when $k = n + 1$.

C.

Consider the recovered time of v_{root}, we define $\tilde{t}^R_{v_{root}}$ as the recovered time of v_{root} in $\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}$, and define $t^R_{v_{root}}$ as the recovered time in a constructed infection process $span_{(0\to T)}\Omega(t)_{v_{root}}$.

Here, according to the $\tilde{t}^R_{v_{root}}$ in different situations, we construct the following different $t^R_{v_{root}}$.

- If $\tilde{t}^R_{v_{root}} > T + 1$, then we let $t^R_{v_{root}} > T$.
- If $\tilde{t}^R_{v_{root}} \le T$, we let $\tilde{t}^R_{v_{root}} = t^R_{v_{root}}$.
- If $\tilde{t}^R_{v_{root}} = T + 1$, we choose $t^R_{v_{root}} = T$.

According to step 2, it can easily see that the likelihood of the infection process $span_{(0\to T)}\Omega(t)_{v_{root}}$ is larger than the likelihood of $\tilde{span}_{(0\to T+1)}\Omega(t)_{v_{root}}$. Therefore, the original inequality (13) is established under the condition $k = n + 1$. Based on the inductive hypothesis, for any k the inequality (13) holds.

D.

To repeatedly apply the inequality (13), we can get the conclusion that the minimum time required is $T^*_{v_{root}}$ in the optimal infection process $Pr(\tilde{span}_{(0\to T)}\Omega(t)_{v_{root}})$ based on the observed topology Θ. Therefore, $t^*_{v_r}$ is twice the infection eccentricity starting from v_{root}, i.e.,

$$T^*_{v_{root}} = 2 \cdot \tilde{ecc}(v_{root})$$

Thus, Lemma 1 holds.

PROOF OF THE LEMMA 2

Supposed an infinite tree network with multiple levels, the root is the rumor source and Θ is the observed infection topology. In addition, this network topology exists at least one infected node. For the adjacent nodes i, j, i.e., $(i, j) \in E$, if $T^*_i > T^*_j$, we will show

$$Pr(\tilde{span}_{(0\to T^*_i)}\Omega(t)_i) < Pr(\tilde{span}_{(0\to T^*_j)}\Omega(t)_j) \tag{40}$$

Here, $\tilde{span}_{(0\to T^*_i)}\Omega(t)_i$ is the optimal infection process starting from i, and the optimal infection process $\tilde{span}_{(0\to T^*_j)}\Omega(t)_j$ starts from j.

Step 1: We will show the subtree S_j^{-i} must contain infected nodes in the observed topology Θ. i.e., $S_j^{-i} \bigcap I \neq \emptyset$. If $S_j^{-i} \bigcap I = \emptyset$, it means that all infected nodes are in the subtree S_i^{-j}. Since the node j can only infect nodes in S_i^{-j} through the edge (i, j), thus we have $\tilde{ecc}(j) = \tilde{ecc}(i) + 1$ and $\tilde{ecc}(j) > \tilde{ecc}(i)$, this contradicts the condition $T^*_i > T^*_j$ and $\tilde{ecc}(i) > \tilde{ecc}(j)$. Therefore, $S_j^{-i} \bigcap I \neq \emptyset$.

Step 2:

- If $S_i^{-j} \bigcap I \neq \emptyset$, $\forall a \in S_i^{-j} \bigcap I$

$$l(i, a) = l(j, a) - 1 \le \tilde{ecc}(j) - 1$$

$\forall b \in S_j^{-i} \bigcap I$, we have

$$l(i, b) = l(j, b) + 1 \le \tilde{ecc}(j) + 1$$

Therefore, we can obtain

$$\tilde{ecc}(i) = \max_{i' \in I} l(i, i') \le \tilde{ecc}(j) + 1$$

Note that $\tilde{ecc}(j) < \tilde{ecc}(i) \le \tilde{ecc}(j) + 1$. Hence, we can get

$$\tilde{ecc}(i) = \tilde{ecc}(j) + 1$$

- If $S_i^{-j} \bigcap I = \emptyset$, it means that all infected nodes are in the subtree S_j^{-i}, obviously $\tilde{ecc}(i) = \tilde{ecc}(j) + 1$, so we have $T^*_i = T^*_j + 2$.

Step 3: Nextly, we will show $t^E_j = 1$, $t^I_j = 2$ on the optimal infection process $\tilde{span}_{(0\to T^*_i)}\Omega(t)_i$. Firstly, according to $\tilde{ecc}(i) = \tilde{ecc}(j) + 1$, we can easily to know that the furthest infected node from the node i must be in S_j^{-i} in the observed topology. (If the furthest infected node is in S_i^{-j}, it will contradict $\tilde{ecc}(i) = \tilde{ecc}(j) + 1$), i.e., there is a node z in the subtree S_j^{-i}, $\exists z \in S_j^{-i}$, $l(i, z) = \tilde{ecc}(i) = \tilde{ecc}(j) + 1$, thus $l(v, z) = \tilde{ecc}(v)$.

Then, if $t^E_j = 1$, $t^I_j > 2$ on the optimal infection process $\tilde{span}_{(0\to T^*_i)}\Omega(t)_i$, we have

$$T^*_i - t^I_j = T^*_j + 2 - t^I_j < T^*_j \tag{41}$$

According to the definition of the optimal infection process, the node j needs at least time duration T^*_j to infect the furthest infected node z along optimal infection process $\tilde{span}_{(0\to T^*_i)}\Omega(t)_i$. Thus, according to the (41), node j cannot infect z within the time duration $T^*_i - t^I_j$.

Similarly, it does not hold when $t^E_j > 1$, $t^I_j > 2$. Therefore, $t^E_j = 1$, $t^I_j = 2$ on the optimal infection process $\tilde{span}_{(0\to T^*_i)}\Omega(t)_i$.

Step 4: Given the optimal infection process $\tilde{span}_{(0\to T^*_i)}\Omega(t)_i$, we can construct an infection process $\tilde{span}_{(0\to T^*_j)}\Omega(t)_j$ whose probability is higher than the probability of $\tilde{span}_{(0\to T^*_i)}\Omega(t)_i$.

Firstly, we divide the given infection process $\tilde{span}_{(0\to T^*_i)}\Omega(t)_i$ into two parts: S_i^{-j} and S_j^{-i}. According to $t^E_j = 1$, $t^I_j = 2$, we have

$$\begin{aligned} &Pr(\tilde{span}_{(0\to T^*_i)}\Omega(t)_i) \\ &\quad = q_1 q_2 \cdot Pr((\tilde{span}_{(0\to T^*_i)}\Omega(t)_i), S_j^{-i} | t^E_j = 1, t^I_j = 2) \\ &\qquad \cdot Pr((\tilde{span}_{(0\to T^*_i)}\Omega(t)_i), S_i^{-j}) \end{aligned} \tag{42}$$

where q_1 is the probability that j changed to exposed at the first time-slot, q_2 is the probability that j was infected at the next time-slot.

Next, we construct a possible infection process $\bar{span}_{(0\to T^*_j)}\Omega(t)_j$ starting from j, where we assume that i changes to the

exposed state at the first time-slot and is infected at the next time-slot in the subtree S_i^{-j}, i.e., $t_i^E = 1, t_i^I = 2$.

$$\begin{aligned} &Pr(\tilde{span}_{(0\to T_j^*)}\Omega(t)_j) \\ &\quad = q_1 q_2 \cdot Pr(((\tilde{span}_{(0\to T_j^*)}\Omega(t)_j), S_i^{-j})|t_i^E = 1, t_i^I = 2) \\ &\qquad \cdot Pr((\tilde{span}_{(0\to T_j^*)}\Omega(t)_j), S_j^{-i}) \end{aligned} \tag{43}$$

Now we analyze the properties of equation (42) (43).

(1). According to $t_j^E = 1, t_j^I = 2$ for the subtree S_j^{-i} in the given optimal infection process $(\tilde{span}_{(0\to T_i^*)}\Omega(t)_i)$, a partial infection process $(\tilde{span}_{(0\to T_j^*)}\Omega(t)_j, S_j^{-i})$ can be constructed, where this partial process equals to $(\tilde{span}_{(0\to T_i^*)}\Omega(t)_i, S_j^{-i})$ except for the first two time-slots. So we have

$$(\tilde{span}_{(0\to T_j^*)}\Omega(t)_j, S_j^{-i}) = (\tilde{span}_{(2\to T_i^*)}\Omega(t)_i, S_j^{-i})$$

then

$$\begin{aligned} &Pr(\tilde{span}_{(0\to T_j^*)}\Omega(t)_j, S_j^{-i}) \\ &\qquad = Pr((\tilde{span}_{(0\to T_i^*)}\Omega(t)_i, S_j^{-i})|t_v^E = 1, t_v^I = 2) \end{aligned} \tag{44}$$

Here, since $T_i^* = T_j^* + 2$ (Step 1), equation (44) holds.

(2). For the S_i^{-j}, we can construct the infection process $(\tilde{span}_{(0\to T_j^*)}\Omega(t)_j, S_i^{-j}|t_i^E = 1, t_i^I = 2)$ as follows.

$$\begin{aligned} &(\tilde{span}_{(0\to T_j^*)}\Omega(t)_j, S_i^{-j}|t_i^E = 1, t_i^I = 2) \\ &\qquad \in \arg\max Pr(span_{(0\to T_j^*)}\Omega(t)_j, S_i^{-j}|t_i^E = 1, t_i^I = 2) \end{aligned}$$

where $(span_{(0\to T_j^*)}\Omega(t)_j, S_i^{-j}) : func(\Omega(t)_j) = \Theta(S_i^{-j})$.

According to the Lemma 1, we have

$$\begin{aligned} &Pr(\tilde{span}_{(0\to T_j^*)}\Omega(t)_j, S_i^{-j}|t_i^E = 1, t_i^I = 2) \\ &\quad = \max Pr(span_{(0\to T_j^*)}\Omega(t)_j, S_i^{-j}|t_i^E = 1, t_i^I = 2) \\ &\quad = \max Pr(span_{(0\to T_i^*-2)}\Omega(t)_j, S_i^{-j}|t_i^E = 1, t_i^I = 2) \\ &\quad = \max Pr(span_{(0\to T_i^*)}\Omega(t)_j, S_i^{-j}) \\ &\quad = Pr(\tilde{span}_{(0\to T_i^*)}\Omega(t)_i), S_i^{-j}) \end{aligned} \tag{45}$$

Therefore, according to (44) (45), we have constructed an infection process $\tilde{span}_{(0\to T_j^*)}\Omega(t)_j$ rooted at j whose probability is higher than the probability of a given process $\tilde{span}_{(0\to T_i^*)}\Omega(t)_i$, thus the Lemma 3 holds.

PROOF OF THE LEMMA 3

Now we will show there will be infection process from a certain node to Jordan center, in which the infection eccentricity of nodes on this process will be monotonically decreased.

A.

First, assuming the tree network exists only one Jordan center u, we can find an infected node $\mu \in I$ as follows.

$$l(u, \mu) = e\tilde{c}c(u) = \eta$$

Step 1: we will show that $l(u, v) \geq l(u, \mu) - 1$, where $v = arg \max_{v\in I, v\neq\mu} l(u, v)$ and node v is not on the path from u to μ. Suppose $l(u, v) \leq l(u, \mu) - 2$, we have

$$\begin{aligned} e\tilde{c}c(k) &= \max(l(k, \mu), l(k, v)) \\ &= \max(l(u, \mu) - 1, l(u, v) + 1) \\ &= l(u, \mu) - 1 = \eta - 1 \\ &< e\tilde{c}c(u) \end{aligned}$$

where k is a neighbor of u on the path from u to μ. Since the assumption $l(u, v) \leq l(u, \mu) - 2$, we can find it contradicts the fact that u is the Jordan infection center with minimum infection eccentricity. Thus, $l(u, v) \geq l(u, \mu) - 1$ holds.

Step 2: Suppose a certain infected node u_i is not on the infection process from the Jordan infection center u to μ, and we assume $l(u, u_i) = \lambda_{u_i}$, it is easy to see

$$e\tilde{c}c(u_i) = e\tilde{c}c(u) + \lambda_{u_i} = \eta + \lambda_{u_i} > e\tilde{c}c(u)$$

When a certain node u_i is on the path from u to μ as $[u, u_i, u_{i+1}, u_{i+2}, u_{i+3}, \ldots, \mu]$, we will show $e\tilde{c}c(u_i) < e\tilde{c}c(u_{i+1})$ on this path. Suppose $e\tilde{c}c(u_i) \geq e\tilde{c}c(u_{i+1})$, we will have

$$\begin{aligned} e\tilde{c}c(u_{i+1}) \geq l(u_{i+1}, v) &= l(u, u_{i+1}) + l(u, v) \\ &= l(u, v) + \lambda_{u_{i+1}} \end{aligned}$$

From Step 1, we can get

$$\begin{aligned} e\tilde{c}c(u_{i+1}) &\geq l(u, v) + \lambda_{u_{i+1}} \\ &\geq l(u, \mu) - 1 + \lambda_{u_{i+1}} \\ &= l(u, \mu) + \lambda_{u_i} = \eta + \lambda_{u_i} \\ &\geq e\tilde{c}c(u_i) + \lambda_{u_i} \\ &> e\tilde{c}c(u_i) \end{aligned}$$

It is obvious that it contradicts the assumption $e\tilde{c}c(u_i) \geq e\tilde{c}c(u_{i+1})$. Hence, we can have $e\tilde{c}c(u) < e\tilde{c}c(u_i) < e\tilde{c}c(u_{i+1})$ on the infection process.

B.

Next, we suppose that nodes u and v are two adjacent Jordan infection centers in the network, and assume that $e\tilde{c}c(u) = e\tilde{c}c(v) = \eta$.

Step 1: It is obvious that $\forall o \in S_u^{-v} \cap I$, we have

$$l(u, o) \leq \eta - 1 \tag{46}$$

If $l(u, o) > \eta - 1$, then $l(v, o) = l(u, o) + 1 > \eta$, it contradicts the fact that v is the Jordan infection center.

Now we will show that there exists an infected node $o' \in S_u^{-v} \cap I$, which makes $l(u, o') = \eta - 1$ holds.

We assume $l(u, \tau) \leq \eta - 2$ for all the infected nodes $\tau \in S_u^{-v} \cap I$, it implies that

$$l(v, \tau) = l(u, \tau) + 1 \leq \eta - 1 \tag{47}$$

Furthermore, we can find one infected node $\tilde{\tau} \in S_v^{-u} \cap I$ based on (46).

$$l(v, \tilde{\tau}) \leq \eta - 1 \tag{48}$$

According to (47)(48), $\forall \tau \in I$, we have

$$e\tilde{c}c(v) = \max l(v, \tau) = \eta - 1$$

It contradicts the fact that $e\tilde{c}c(v) = \eta$.

Therefore, the subtree S_u^{-v} contains an infected node o' and $l(u, o') = \eta - 1$. (Furthermore, we can derive a same conclusion that there exists a node $o'' \in S_v^{-u} \cap I$, such that $l(v, o') = \eta - 1$.)

Step 2: We can find a certain infected node $\alpha \in S_u^{-v}$ and assume $l(u, \alpha) = \lambda$, then we can have

$$l(\alpha, \beta) = l(\alpha, u) + l(u, v) + l(v, \beta) \leq \lambda + 1 + \eta - 1 = \lambda + \eta$$

where $\beta \in S_v^{-u} \cap I$.

Similarly, we can find an infected node $\gamma \in S_u^{-v} \cap I$

$$l(\alpha, \gamma) \leq l(\alpha, u) + l(u, \gamma) \leq \lambda + \eta - 1$$

Hence, we have

$$e\tilde{c}c(\alpha) = \lambda + \eta$$

Therefore, we get the conclusion that the infection eccentricity will decrease monotonically along the infection process from a certain infected node α to the Jordan infection center.

PROOF OF THE LEMMA 4

Firstly, we assume that the nodes i, j are two non-adjacent Jordan infection centers and $e\tilde{c}c(i) = e\tilde{c}c(j) = \eta$, the node k is a neighbor of i on the path from i to j, i.e., $l(i, k) = 1$.

If subtree $S_i^{-k} \bigcap \mathrm{I} = \emptyset$ and $\forall p \in I$, we have

$$l(k, p) = l(i, p) - 1 < l(i, p) \leq e\tilde{c}c(i)$$

It contradicts the fact that i is a Jordan infection center with the minimum infection eccentricity.

If subtree $S_i^{-k} \bigcap I \neq \emptyset$ and $\forall q \in S_i^{-k} \bigcap I$

$$l(k, q) = l(j, q) - l(j, k) < l(j, q) \leq e\tilde{c}c(j)$$

Therefore, $\forall o \in V \bigcap I$, $l(k, o) < \eta$, it is contrary to the fact that i, j are the Jordan infection centers with the minimum infection eccentricity. Therefore, we get the conclusion that all Jordan infection centers must be adjacent. Furthermore, if there are $n > 2$ Jordan infection centers in the tree network, these Jordan infection centers would form a clique. Therefore, this situation is impossible in a tree network, thereby Lemma 4 holds.

REFERENCES

[1] J.-B. Wang, L. Wang, and X. Li, "Identifying spatial invasion of pandemics on metapopulation networks via anatomizing arrival history," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2782–2795, Dec. 2016.

[2] J. B. Wang and X. Li, "Uncovering spatial invasion on metapopulation networks with SIR epidemics," *IEEE Trans. Netw. Sci. Eng.*, to be published.

[3] B. Doerr, M. Fouz, and T. Friedrich, "Why rumors spread so quickly in social networks," *Commun. ACM*, vol. 55, no. 6, pp. 70–75, Jun. 2012.

[4] Y. Wang, S. Wen, Y. Xiang, and W. Zhou, "Modeling the propagation of worms in networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 2, pp. 942–960, 2nd Quarter 2014.

[5] I. M. Parker *et al.*, "Impact: Toward a framework for understanding the ecological effects of invaders," *Biol. Invasions*, vol. 1, no. 1, pp. 3–19, Mar. 1999.

[6] S. Savage, D. Wetherall, A. Karlin, and T. Anderson, "Practical network support for IP traceback," *ACM Comput. Commun. Rev.*, vol. 30, no. 4, pp. 295–306, Aug. 2000.

[7] Y. Xie, V. Sekar, D. A. Maltz, M. K. Reiter, and H. Zhang, "Worm origin identification using random moonwalks," in *Proc. IEEE Symp. Secur. Privacy*, May 2005, pp. 242–256,

[8] Z. Bu, H-J. Li, J. Cao, Z. Wang, and G. Gao, "Dynamic cluster formation Game for attributed graph clustering," *IEEE Trans. Cybern.*, vol. 49, no. 1, pp. 328–341, Jan. 2019.

[9] H.-J. Li, Z. Bu, A. Li, Z. Liu, and Y. Shi, "Fast and accurate mining the community structure: Integrating center locating and membership optimization," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 9, pp. 2349–2362, Sep. 2016.

[10] X. L. Li *et al.*, "Punishment diminishes the benefits of network reciprocity in social dilemma experiments," *Proc. Natl. Acad. Sci. USA*, vol. 115, no. 1, pp. 30–35, Jan. 2018.

[11] H. J. Li and J. J. Daniels, "Social significance of community structure: Statistical view," *Phys. Rev. E.*, vol. 91, no. 1, 2015, Art. no. 012801.

[12] L. Wang and X. Li, "Spatial epidemiology of networked metapopulation: An overview," *Chin. Sci. Bull.*, vol. 59, no. 28, pp. 3511–3522, 2014.

[13] D. Shah and T. Zaman, "Detecting sources of computer viruses in networks: Theory and experiment," *SIGMETRICS Perform. Eval. Rev.*, vol. 38, no. 1, pp. 203–214, Jun. 2010.

[14] D. Shah and T. Zaman, "Rumors in a network: Who's the culprit?" *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5163–5181, Aug. 2011.

[15] D. Shah and T. Zaman, "Rumor centrality: A universal source detector," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 40, no. 1, pp. 199–210, Jun. 2012.

[16] N. T. J. Bailey, "The mathematical theory of infectious diseases and its applications," *Nat. Immunol.*, vol. 28, no. 2, pp. 479–480, 1977.

[17] K. Zhu and L. Ying, "Information source detection in the sir model: A sample path based approach," *IEEE Trans. Inf. Theory*, vol. 24, no. 1, pp. 408–421, Feb. 2016.

[18] W. Luo and W. P. Tay, "Finding an infection source under the SIS model," in *Proc. ICASSP*, Vancouver, BC, Canada, May 2013, pp. 2930–2934.

[19] H. Kang and X. Fu, "Epidemic spreading and global stability of an SIS model with an infective vector on complex networks," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 27, nos. 1–3, pp. 30–39, Oct. .2015.

[20] Z. Chen, K. Zhu, and L. Ying, "Detecting multiple information sources in networks under the SIR model," *IEEE Trans. Netw. Sci. Eng.*, vol. 3, no. 1, pp. 17–31, Jan. 2016.

[21] W. Luo and W. P. Tay, "Identifying multiple infection sources in a network," in *Proc. ASILOMAR*, Pacific Grove, CA, USA, Nov. 2012, pp. 1483–1489.

[22] Z. Wang *et al.*, "Rumor source detection with multiple observations:fundamental limits and algorithms," *ACM Int. Conf. Meas. Model. Comput. Syst.* , vol. 42, no. 1, Jun. 2014.

[23] J. Jiang, S. Wen, S. Yu, Y. Xiang, and W. Zhou, "K-center: An approach on the multi-source identification of information diffusion," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 12, pp. 2616–2626, Dec. 2015.

[24] Z. Chen, K. Zhu, and L. Ying, "Detecting multiple information sources in networks under the SIR model," *IEEE Trans. Netw. Sci. Eng.*, vol. 3, no. 1, pp. 17–31, Mar. 2014.

[25] P. C. Pinto, P. Thiran, and M. Vetterli, "Locating the source of diffusion in large-scale networks," *Phys. Rev. Lett.*, vol. 109, no. 6, 2012, Art. no. 068702.

[26] F. Altarelli, A. Braunstein, L. Dall'Asta, A. Lage-Castellanos, and R. Zecchina, "Bayesian inference of epidemics on networks via belief propagation," *Phys. Rev. Lett.*, vol. 112, no. 11, 2014, Art. no. 118701.

[27] A. Agaskar and Y. M. Lu, "A fast monte carlo algorithm for source localization on graphs," *Proc. SPIE*, vol. 26, p. 8858, Sep. 2013.

[28] Y. Yao, X. Luo, F. Gao, and S. Ai, "Research of a potential worm propagation model based on pure P2P principle," in *Proc. Int. Conf. Commun. Technol.*, Guilin, China, Nov. 2006, pp. 1–4.

[29] M. Y. Li and J. S. Muldowney, "Global stability for the SEIR model in epidemiology," *Math. Biosciences*, vol. 125, no. 2, pp. 155–164, Feb. 1995.

[30] M. Y. Li and J. S. Muldowney, *Global Stability in Some Seir Epidemic Models*. New York, NY, USA: Springer, 2002.

[31] L-L. Xia, G-P. Jiang, B. Song, and Y. R. Song, "Rumor spreading model considering hesitating mechanism in complex social networks," *Phys. A, Stat. Mech. Appl.*, vol. 437, pp. 295–303, Nov. 2015.

[32] S. Dong, Y-B. Deng, and Y-C. Huang, "SEIR model of rumor spreading in online social network with varying total population size," *Commun. Theor. Phys.*, vol. 68, pp. 545–552, Oct. 2017.

[33] G. Y. Ran and X. Ling-Ling, "The propagation and inhibition of rumors in online social network," *Acta Phys. Sinica*, vol. 61, p. 23, Jan. 2012.
[34] Q. Liu, T. Li, and M. Sun, "The analysis of an SEIR rumor propagation model on heterogeneous network," *Phys. A, Stat. Mech. Appl.*, vol. 469, pp. 372–380, Mar. 2017.
[35] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
[36] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
[37] J. McAuley and J. Leskovec, "Learning to discover social circles in ego networks," in *Proc. NIPS*, Aug. 2012, pp. 539–547.
[38] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: Densification laws, shrinking diameters and possible explanations," in *Proc. 11th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2005, pp. 177–187.
[39] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," in *Proc. 19th Int. Conf. World Wide Web*, Apr. 2010, pp. 641–650.
[40] A. Jain, V. Borkar, and D. Garg, "Fast rumor source identification via random walks," *Social Netw. Anal. Mining*, vol. 6, no. 1, p. 62, 2016.
[41] L. Wang and J. T. Wu, "Characterizing the dynamics underlying global spread of epidemics," *Nature Commun.*, vol. 9, no. 1, p. 218, 2018.
[42] X. Li and X. Li, "Reconstruction of stochastic temporal networks through diffusive arrival times," *Nat. Commun.*, vol. 8, no. 8, p. 15729, 2017.
[43] F. Ji and W. P. Tay, "An algorithmic framework for estimating rumor sources with different start times," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2517–2530, May 2017.

YOUSHENG ZHOU received the Ph.D. degree from the Beijing University of Posts and Telecommunications, in 2011. He is currently an Associate Professor with the Chongqing University of Posts and Telecommunications. He has published more than 20 academic papers in peer-reviewed international journals. He has also served as an Invited Reviewer for various international journals and conferences. His research interests include mobile security and cloud security.

CHUJUN WU is currently pursuing the master's degree with the Chongqing University of Posts and Telecommunications. His research interest includes privacy preserving of online social networks.

QINGYI ZHU received the B.E. and Ph.D. degrees from the College of Computer Science, Chongqing University, in 2009 and 2014, respectively. He is currently an Associate Professor with the Chongqing University of Posts and Telecommunications, China. He has published more than 20 academic papers in peer-reviewed international journals. He has also served as an Invited Reviewer for various international journals and conferences. His current research interests include cybersecurity dynamics, complex systems, and blockchain.

YONG XIANG (SM'12) received the B.E. and M.E. degrees from the University of Electronic Science and Technology of China, China, and the Ph.D. degree from The University of Melbourne, Australia. He is currently a Professor with the School of Information Technology, Deakin University, Australia. He is also the Associate Head of School (Research) and the Director of the Artificial Intelligence and Data Analytics Research Cluster. He has obtained many research grants, including five Discovery and Linkage Grants from the Australian Research Council. He has published two monographs, over 100 refereed journal articles, and numerous conference papers in these areas. He is the co-inventor of two U.S. patents and some of his research results have been commercialized. His current research interests include information security and privacy, signal and image processing, data analytics and machine intelligence, and the Internet of Things. He is the Associate Editor of the IEEE Signal Processing Letters and the IEEE Access. He has been invited to give keynote speeches and chair committees in a number of international conferences.

SENG W. LOKE received the Ph.D. degree from the University of Melbourne, Melbourne, VIC, Australia, in 1998. He is currently a Professor with the School of IT, Deakin University, Australia. He has coauthored more than 270 research papers, including over 58 journal papers, ten book chapters, and over 180 conference/workshop papers, with numerous work on context-aware computing and mobile and pervasive computing. His current research interests include the Internet-of-Things, intelligent transport systems, smart cities, mobile cloud computing, and crowd computing.