Rooting our Rumor Sources in Online Social Networks: The Value of Diversity From Multiple Observations

Zhaoxu Wang, Wenxiang Dong, Wenyi Zhang, *Senior Member, IEEE*, and Chee Wei Tan, *Senior Member, IEEE*

***Abstract*—This paper addresses the problem of rumor source detection with multiple observations, from a statistical point of view of a spreading over a network, based on the susceptible-infectious model. For tree networks, multiple independent observations can dramatically improve the detection probability. For the case of a single rumor source, we propose a unified inference framework based on the joint rumor centrality, and provide explicit detection performance for degree-regular tree networks. Surprisingly, even with merely two observations, the detection probability at least doubles that of a single observation, and further approaches one, i.e., reliable detection, with increasing degree. This indicates that a richer diversity enhances detectability. Furthermore, we consider the case of multiple connected sources and investigate the effect of diversity. For general graphs, a detection algorithm using a breadth-first search strategy is also proposed and evaluated. Besides rumor source detection, our results can be used in network forensics to combat recurring epidemic-like information spreading such as online anomaly and fraudulent email spams.**

***Index Terms*—Graph networks, inference algorithms, maximum likelihood detection, multiple observations, rumor spreading.**

I. Introduction

IDENTIFICATION of malicious information sources in a network, be it in the case of an online spam spreading in the Internet or a misinformation or rumor propagating in an online social network, allows timely quarantine of the epidemic-like spreading to limit the damage caused. For example, law enforcement agencies may be interested in identifying the perpetrators of false information used to manipulate the market prices of certain stocks. It is challenging to understand how the source of the spreading can be identified in a reliable manner from a number of snapshot observations of the spread conditions, e.g., infected nodes in virus propagation or users possessing the rumor, and the underlying connectivity of the network. This can lead to a fundamental understanding of the role of network in aiding or constraining epidemic-like spreading [1], [2].

In this paper, we consider the issue of reliably detecting the rumor sources with multiple observations from a statistical point of view of a spreading. Our goal is to find the rumor sources in order to control and prevent network risks based on limited information about the network structure and multiple snapshot observations.

A. Related Works

This network inference problem is a combinatorial problem that is generally hard to solve, and has remained surprisingly unexplored until late. In the seminal work [3], [4], Shah and Zaman studied the single rumor source detection problem, and conducted asymptotic performance analysis of the Maximum Likelihood (ML) detector for a single rumor source using the susceptible-infectious (SI) model. Then, this work was extended in [5] for random graphs. The authors in [6] studied the Maximum A Posteriori (MAP) detection based on various suspect set characteristics. In [7], the authors investigated the detection of multiple rumor sources in the SI model. In [8], the authors considered the case where nodes randomly report their infection state in SI spreading, which means we only get an incomplete picture of the infection state. Besides, other types of models have also been considered in [9]–[12]. In [9], the authors analyzed the detection under susceptible-infected-recovered (SIR) spreading, and proposed a probabilistic method based on sample path estimation. In [10], the authors proposed an approach based on message passing algorithms to detect a single rumor source in the SIR model. The authors in [11] studied the detection in susceptible-infected-susceptible (SIS) spreading of regular tree networks. In [12], the authors presented an exercise in modeling social networks as sensors and used the Apollo tool to identify the provenance of the reported observations.

Several other source detecting algorithms have also been proposed recently, including an eigenvalue-based estimator [13], a fast Monte Carlo algorithm [14], and an algorithm utilizing a number of specialized "monitor-nodes" [15].

In the related work, multiple observations have been used to get good performance in network inference, e.g., in order to defend against attacks on the Internet, Hidden Markov Models (HMMs) have been used as an effective detector in sensor network [16]. The cases where multiple observers (ISPs) all see the same and different distributions of observations were raised

Manuscript received July 03, 2014; revised November 06, 2014; accepted December 21, 2014. Date of publication January 06, 2015; date of current version May 12, 2015. This work was supported in part by the Doctoral Program of Higher Education (SRFDP) and Research Grants Council Earmarked Research Grants (RGC ERG) Joint Research Scheme through Specialized Research Fund 20133402140001, in part by the National Natural Science Foundation of Chinaunder Grant 61379003, and in part by the Research Grants Council of Hong Kong under Project M-CityU107/13. A preliminary version of this work was presented in part at ACM SIGMETRICS 2014, Austin, TX, USA, June 2014. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Georgios Giannakis.

Z. Wang, W. Dong, and W. Zhang are with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China (e-mail: wzxboa@mail.ustc.edu.cn; javin@mail.ustc.edu.cn; wenyizha@ustc.edu.cn).

C. W. Tan is with the Department of Computer Science, City University of Hong Kong, Hong Kong SAR, China (e-mail: cheewtan@cityu.edu.hk).

Digital Object Identifier 10.1109/JSTSP.2015.2389191

in [17] and [18], respectively. Further in some other networks, e.g., speech processing systems, the author presented multiple new voice activity detectors (VAD) for improving speech detection robustness in noisy environments and the performance of speech recognition systems [19].

B. Our Contributions

One key question in rumor source detection is to determine the right number of observations or adaptive measurements such that the sources can be correctly detected. The existing works, however, rely critically on the assumption that only a single snapshot observation is made in the detection. In practice, there can possibly be multiple observations of a *recurrent* online anomaly, e.g., fraudulent email spams and recurring malcode, which are usually originated from a common culprit. Obviously, the interdependency between observations and network structure affect detectability. Single observation detection does not leverage all degrees of freedom and can perform poorly. Indeed, the multiple observations add an interesting dimension to detecting the source reliably. Can we do more with less by having as few observations as possible and yet enough to provide detectability guarantees? This paper answers this question. To the best of our knowledge, there has been no work done on rumor source detection with multiple observations. The work of identifying the single rumor source with multiple observations has been presented partly in SIGMETRICS, see [20].

In summary, the contributions of the paper are as follows:

1) First, we study the case of a single rumor source. For multiple *independent* observations in a tree network, we propose a unified detection framework based on a graph topological quantity called joint rumor centrality and design an efficient algorithm to identify the joint rumor center. For degree-regular trees, we characterize and evaluate the exact and asymptotic detection probability.
2) Then, we consider the case of multiple connected sources and show an explicit improvement in detectability for general trees.
3) For general graphs, we extend our detection framework by applying a breadth-first search (BFS) strategy to obtain induced trees from multiple observations. Our numerical results show that the algorithm performs more effectively than that for a single observation in small-world, scale-free and Newman's scientific collaboration networks.

C. Organization

The structure of this paper is as follows. We introduce the system model in Section II. In Section III, we study the single rumor source problem with multiple independent observations, and propose a detection framework for tree networks. We analytically characterize the detection performance for regular tree networks. In Section IV, we investigate the effect of diversity for multiple connected sources. In Section V, for tree-type networks, we propose algorithms to calculate the exact correct detection probability and identify the joint rumor center under multiple independent observations. We extend our framework to the general network case in Section V-D. In Section VI, we numerically evaluate the performance of our algorithms. Section VII provides proofs of the analytical results. We conclude the paper in Section VIII.

TABLE I
KEY TERMS AND SYMBOLS

Symbol	Definition
δ	node degree
s^*	original rumor source
$\widehat{s}$	detected rumor source
$T_u^{s^*}$	subtree rooted at node u, with node s^* as source
G_{n_j} $(1 \leq j \leq k)$	a snapshot observation of n_j infected nodes for the jth rumor spreading
$\{\cap G_{1\to k}\}$	the intersection of observations, i.e., $\{G_{n_1} \cap \ldots \cap G_{n_k}\}$
$\{\cup G_{1\to k}\}$	the joint of observations, i.e., $\{G_{n_1} \cup \ldots \cup G_{n_k}\}$
$R(u, G_n)$	rumor centrality of node u in G_n
$R_k(u, G_{n_1} \ldots G_{n_k})$	joint rumor centrality of node u in $\{\cap G_{1\to k}\}$
$\mathbf{P}_G(\cdot)$	probability distribution of an infection sample in a rumor spreading process, or of an infection sample equivalently constructed using the Pólya's urn model
$P_{\rm c}(\cdot)$	correct detection probability
$P_{\rm e}(\cdot)$	error detection probability

II. Model and Preliminaries

In this section, we focus on the single rumor source detection problem and introduce the SI rumor spreading model. The case of multiple connected sources is considered in Section IV. Then, we describe the ML detector for the rumor source in regular trees, general trees and graphs, respectively. We list the key parameters used in the paper in Table I.

A. Rumor Spreading Model

We consider a network of nodes modeled as an undirected graph $G=(V,E)$, where the set of vertices V represents the nodes in the network, and the set of edges E represents the links between the network nodes. We assume V is countably infinite in order to avoid boundary effects and consider the case where initially only one node $s^* \in V$ is the rumor source.

We use a variant of the SIR model for the rumor spreading known as the SI model, where a node that is infected with the rumor retains it forever. A rumor is spread from node i to node j if and only if there is an edge between them, i.e., if $(i,j) \in E$. Let τ_{ij} be the spreading time from i to j for all $(i,j) \in E$, which are mutually independent and have exponential distribution with parameter λ. Without loss of generality, take $\lambda=1$.

B. ML Rumor Source Detector

1) ML Detector: Suppose that a rumor originates from a node $s^* \in V$. We observe the network G at some time and find n infected nodes, which are collectively denoted by G_n. Our goal is to construct a detector to detect a node $\hat{s}$ as the rumor source s^*. The ML detector of s^* given G_n maximizes the correct detection probability and is given by

$$\widehat{s} \in \arg \max_{s^* \in G_n} \mathbf{P}_G \left(G_n \mid s^* \right), \tag{1}$$

where $\mathbf{P}_G(G_n|s^*)$ is the probability of observing G_n assuming s^* to be the rumor source.

2) Exact ML Detector for Regular Trees: In general, the evaluation of $\mathbf{P}_G(G_n|s^*)$ is computationally prohibitive since it is related to counting the number of linear extensions of a partially ordered set. By leveraging the concept of rumor centrality, first introduced in [4], the exact ML detector for a regular tree is given by

$$\widehat{s} \in \arg\max_{s^* \in G_n} \mathbf{P}_G\left(G_n \mid s^*\right) = \arg\max_{s^* \in G_n} R(s^*, G_n), \tag{2}$$

where $R(s^*, G_n)$ is the rumor centrality of node s^* in G_n and can be evaluated by

$$R(s^*, G_n) = (n-1)! \prod_{u \in \mathrm{child}(s^*)} \frac{R(s^*, Z_u^{s^*})}{T_u^{s^*}!}, \tag{3}$$

where $Z_u^{s^*}$ denotes the subtree rooted at node u, with node s^* as the source, and $T_u^{s^*}$ denotes the number of nodes in the subtree $Z_u^{s^*}$.

3) Approximate Detector: For general cases, intuitively, a rumor tends to travel from the source to each infected node along a minimum-distance path [4], and this serves, along with the rumor centrality, as a reasonable and useful heuristic for constructing approximate detectors. For general trees, an approximate detector is thus given by [4]

$$\widehat{s} \in \arg\max_{s^* \in G_n} \mathbf{P}_G(\sigma_s^* \mid s^*) \cdot R(s^*, G_n), \tag{4}$$

where $\mathbf{P}_G(\sigma_s^* \mid s^*)$ the probability of observing σ_s^* assuming s^* to be the rumor source and σ_s^* represents the breadth-first search (BFS) ordering of nodes in the tree, i.e., a permitted permutation assuming s^* to be the rumor source, defined in [4].

For general graphs, we approximate the diffusion tree by a BFS tree; that is, we assume that if node s^* was the source, then the rumor spreads along a BFS tree rooted at s^*, denoted by $T_b(s^*)$. So an approximate detector is given by

$$\widehat{s} \in \arg\max_{s^* \in G_n} \mathbf{P}_G(\sigma_s^* \mid s^*) \cdot R\left(s^*, T_b(s^*)\right), \tag{5}$$

where σ_s^* represents the BFS ordering of nodes in the BFS tree $T_b(s^*)$.

Besides, a message-passing algorithm has been proposed in [4] to compute the rumor centralities for all the nodes in a general tree with n nodes using only $O(n)$ computation steps.

III. Independent Observations

In this section, we describe the ML rumor source detector under multiple independent observations, which will be shown to be equivalent to a topological quantity that we call the joint rumor centrality, for regular tree graphs. Then we leverage the concept of joint rumor centrality to develop analytical performance results of the ML detector for regular tree-type networks.

A. ML Detection

Suppose k different rumors originate from a common node in the network, regarded as k times independent rumor spreading with the same rumor source. For the jth $(1 \le j \le k)$ rumor spreading, at some time, we observe a snapshot of n_j infected nodes carrying the rumor, which are collectively denoted by G_{n_j}. For arbitrary k independent observations, due to the SI model, each G_{n_j} $(1 \le j \le k)$ must form a connected subgraph and the rumor source must belong to $\{G_{n_1} \cap \ldots G_{n_k}\}$. We assume a uniform prior probability of the source node among all the nodes in the network. This assumption is for tractability and is common in literature. Although it affects the magnitude, it does not qualitatively affect the results [4]. The ML detector under multiple independent observations $G_{n_1}, \ldots, G_{n_k}$ maximizes the correct detection probability, as given by

$$\begin{aligned} \widehat{s} \in & \arg\max_{s^* \in \{G_{n_1} \cap G_{n_2} \ldots \cap G_{n_k}\}} \mathbf{P}_G(G_{n_1}, G_{n_2}, \ldots, G_{n_k} \mid s^*) \\ = & \arg\max_{s^* \in \{G_{n_1} \cap G_{n_2} \ldots \cap G_{n_k}\}} \mathbf{P}_G(G_{n_1}|s^*) \cdots \mathbf{P}_G(G_{n_k}|s^*). \end{aligned} \tag{6}$$

For a δ-regular tree, since all nodes have the same degree, every permitted permutation has the same probability that is independent of the source [4]. Hence, we have

$$\mathbf{P}_G\left(G_n \mid s^*\right) = \sum_{\sigma \in \Omega(s^*, G_n)} \mathbf{P}_G(\sigma|s^*) = R(s^*, G_n)\mathbf{P}_G(\sigma|s^*), \tag{7}$$

where $\Omega(s^*, G_n)$ is the set of all permitted permutations starting with s^* and resulting in G_n, and $\mathbf{P}_G(\sigma|s^*) = \prod_{i=1}^{n-1} \frac{1}{\delta - 2(i-1)} \equiv P(\delta, n)$.

Then, by substituting (7) into (6), the ML detector under arbitrary k observations for a regular tree becomes

$$\begin{aligned} \widehat{s} &= \arg\max_{s^* \in \{G_{n_1} \cap \ldots G_{n_k}\}} \left(R(s^*, G_{n_1}) \cdots R(s^*, G_{n_k})\right) \left(\mathbf{P}_G(\delta, n_1) \cdots \mathbf{P}_G(\delta, n_k)\right) \\ &= \arg\max_{s^* \in \{G_{n_1} \cap \cdots G_{n_k}\}} R(s^*, G_{n_1}) \ldots R(s^*, G_{n_k}). \end{aligned} \tag{8}$$

Note that in the model we do not require that all the rumors originate simultaneously from the original source, or that all the snapshots are observed simultaneously. In fact these k rumor spreadings can be completely asynchronous, as long as they are independent.

B. Joint Rumor Centrality for Regular Trees

We need to evaluate $R(s^*, G_{n_1}) \cdots R(s^*, G_{n_k})$ in (8) for multiple independent observations in a regular tree network. We call this quantity the joint rumor centrality, denoted by $R_k(s^*, G_{n_1} \ldots G_{n_k})$, which enables efficient implementation of the ML detector. Note that the rumor source $s^* \in \{\cap G_{1 \to k}\}$. If $R_k(s^*, G_{n_1} \ldots G_{n_k}) \ge R_k(u, G_{n_1} \ldots G_{n_k})$ for all $u \in \{\cap G_{1 \to k}\}$, then s^* is called a joint rumor center. For the joint rumor center, we have the following proposition.

Proposition 1: For a rumor source s^* with m $(0 \le m \le \delta)$ neighbors in $\{\cap G_{1 \to k}\}$, given multiple independent observations $G_{n_1}, \ldots, G_{n_k}$ of $n_1, \ldots, n_k$ nodes respectively, then:

1) When $m = 0$, there is only one node in $\{\cap G_{1 \to k}\}$, which is the rumor source s^*; when $1 \le m \le \delta$, node s^* is a joint rumor center, if and only if for any $u \in \{\cap G_{1 \to k} \setminus s^*\}$, we have

$$\frac{T_{u,G_{n_1}}^{s^*} \cdots T_{u,G_{n_k}}^{s^*}}{\left(n_1 - T_{u,G_{n_1}}^{s^*}\right) \cdots \left(n_k - T_{u,G_{n_k}}^{s^*}\right)} \le 1, \tag{9}$$

where $T_{u,G_{n_i}}^{s^*}$ denotes the number of nodes in subtree $Z_{u,G_{n_i}}^{s^*}$ of the observation G_{n_i} $(1 \le i \le k)$.

2) If there is a node s^* such that s^* has the maximum joint rumor centrality among all its neighbors in $\{\cap G_{1\to k}\setminus s^*\}$, then s^* is a joint rumor center.
3) Furthermore, the intersection $\{\cap G_{1\to k}\}$ under multiple independent observations has two joint rumor centers, say, s^* and s', if and only if

$$T^{s^*}_{s',G_{n_1}}\cdots T^{s^*}_{s',G_{n_k}}=\left(n_1-T^{s^*}_{s',G_{n_1}}\right)\cdots\left(n_k-T^{s^*}_{s',G_{n_k}}\right), \quad (10)$$

and these two nodes are neighbors. In addition, there can be at most two joint rumor centers in $\{\cap G_{1\to k}\}$.

Remark: The joint rumor center generalizes the rumor center introduced in [4]. Notably, the joint rumor center is a graph topological quantity that has a more restricted range $\{\cap G_{1\to k}\}$ than each observation G_{n_j} $(1\le j\le k)$. We present the proof in the following.

Proof:

a) Proof of Proposition 1-(1):

1) Here, we focus on the case that $1\le m\le\delta$. We relate a useful property about the joint rumor centrality that for any neighboring nodes u and s^* in $\{\cap G_{1\to k}\}$, we get

$$\frac{R_k\left(u,G_{n_1}\cdots G_{n_k}\right)}{R_k\left(s^*,G_{n_1}\cdots G_{n_k}\right)}=\frac{R\left(u,G_{n_1}\right)\cdots R\left(u,G_{n_k}\right)}{R\left(s^*,G_{n_1}\right)\cdots R\left(s^*,G_{n_k}\right)}$$
$$=\frac{T^{s^*}_{u,G_{n_1}}\cdots T^{s^*}_{u,G_{n_k}}}{\left(n_1-T^{s^*}_{u,G_{n_1}}\right)\cdots\left(n_k-T^{s^*}_{u,G_{n_k}}\right)}. \quad (11)$$

which can be derived from the rumor centrality in [4]. Suppose s^* is the rumor source that has the maximum joint rumor centrality, from (11), we get (9) can be achieved for any neighboring nodes u of s^*.

Since for any node v in subtrees $\{Z^{s^*}_{u,G_{n_1}}\cap\ldots Z^{s^*}_{u,G_{n_k}}\}$, we have $T^{s^*}_{v,G_{n_j}}\le T^{s^*}_{u,G_{n_j}}-1$ $(1\le j\le k)$. Hence, (9) will hold for any node $u\in\{\cap G_{1\to k}\setminus s^*\}$. Therefore, the first part of Proposition 1-i is proved.

2) When $1\le m\le\delta$, for any $u\in\{\cap G_{1\to k}\setminus s^*\}$, if (9), from (11), s^* with the maximum joint rumor centrality in $\{\cap G_{1\to k}\setminus s^*\}$, is a joint rumor center. Therefore, the second part of Proposition 1-i is proved.

b) Proof of Proposition 1-(2): To complete the proof, assuming s^* is the rumor source, we have the following important property about joint rumor centrality. Continuing carrying out the recursive relationship (11), we find that for any node $v\in\{\cap G_{1\to k}\setminus s^*\}$, we have

$$\frac{R_k(v,G_{n_1}\cdots G_{n_k})}{R_k(s^*,G_{n_1}\cdots G_{n_k})}=\prod_{i\in\mathbb{P}(s^*,v)}\frac{T^{s^*}_{i,G_{n_1}}\cdots T^{s^*}_{i,G_{n_k}}}{(n_1-T^{s^*}_{i,G_{n_1}})\cdots(n_k-T^{s^*}_{i,G_{n_k}})}, \quad (12)$$

where $\mathbb{P}(s^*,v)$ is the set of nodes in the path between s^* and v, not including s^*.

Since s^* has the maximum joint rumor centrality among all its neighbors in $\{\cap G_{1\to k}\setminus s^*\}$, combining with the fact that for any node v in subtrees $\{Z^{s^*}_{u,G_{n_1}}\cap\ldots Z^{s^*}_{u,G_{n_k}}\}$, then $T^{s^*}_{v,G_{n_j}}\le T^{s^*}_{u,G_{n_j}}-1$ $(1\le j\le k)$, we can prove that each term in the product on the right-hand side of (12) is no more than 1.

Hence, from (11), "(12) is no more than 1" will hold for any node $u\in\{\cap G_{1\to k}\setminus s^*\}$. Therefore, s^* with the maximum joint rumor centrality in $\{\cap G_{1\to k}\setminus s^*\}$, is a joint rumor center.

c) Proof of Proposition 1-(3): Here, rewriting the assumptions, we will prove that neighboring nodes s^* and s' with equal joint rumor centrality are two joint rumor centers. For this purpose we need the following lemma, which indicates that the error detection events are all disjoint.

Lemma 1: For any node s^* with m $(0\le m\le\delta)$ neighbors in $\{\cap G_{1\to k}\}$, i.e., $s_1^*,\ldots,s_m^*$, let random variable X_{i,n_j} be the number of nodes in subtree $Z^{s^*}_{s_i^*,G_{n_j}}$ of a snapshot observation G_{n_j} $(1\le i\le m, 1\le j\le k)$. Then, there is at most a neighbor of the node s^* satisfying

$$\prod_{j=1}^{k}\left(\frac{x_{i,n_j}}{n_j}\right)\ge\prod_{j=1}^{k}\left(1-\frac{x_{i,n_j}}{n_j}\right),\ i\in[1,m]. \quad (13)$$

Based on Lemma 1, since s' and s^* are neighbors that have the same joint rumor centrality, while s' is the only neighbor of s^* with a larger joint rumor centrality than all other neighbors, s^* also has a larger joint rumor centrality than its neighbors, thus leading to Proposition 2(3).

C. Detection Performance for Regular Trees

This section examines the performance of the ML detector leveraging the joint rumor center. We characterize analytically the performance for finite and asymptotically large number of nodes, for regular trees with an arbitrary degree δ. We present the proofs in Section VII.

The following two theorems study the special cases of node degree $\delta=2$ and 3.

Theorem 1: Suppose the rumor spreads on a regular tree with degree $\delta=2$, i.e., a linear network, given two independent observations G_{n_1} and G_{n_2}, then:

1) When $n_2=n_1=n\ge 1$, the correct detection probability P_c is given by

$$P_c=\binom{2n}{n}2^{-2n+1},\qquad n\ge 1. \quad (14)$$

As $n_1=n_2=n\to\infty$, P_c asymptotically scales like $\frac{2}{\sqrt{\pi n}}$.

2) When $n_1=2$, $n_2>n_1$, P_c is given by

$$P_c=\left\{\sum_{m=0}^{\lfloor\frac{n_2}{n_1}\rfloor}\binom{n_2-1}{m}-\mathbb{I}\binom{n_2}{n_1}\binom{n_2-1}{\lfloor\frac{n_2}{n_1}\rfloor}\right\}\cdot 2^{-n_2+1},$$

where

$$\mathbb{I}\binom{n_2}{n_1}=\begin{cases}1 & \frac{n_2}{n_1}\in\mathrm{Z}\\ 0 & \text{others}\end{cases}. \quad (15)$$

3) When $n_2>n_1\ge 3$, P_c is given by

$$P_c=\left\{\sum_{m=0}^{n_1-1}\left[\binom{n_2-1}{\lfloor m\frac{n_2}{n_1}\rfloor}\binom{n_1}{m}+S_{n_2-1}(m)\binom{n_1-1}{m}\right]\right.$$
$$\left.-2\binom{n_2-1}{\lfloor\frac{n_2}{n_1}\rfloor}-\mathbb{I}\binom{n_2}{n_1}\binom{n_2-1}{\lfloor\frac{n_2}{n_1}\rfloor}\right\}\cdot 2^{-(n_1+n_2)+2},$$

where $S_{n_2-1}(m)=\sum_{i=\lfloor m\frac{n_2}{n_1}\rfloor+1}^{\lfloor(m+1)\frac{n_2}{n_1}\rfloor-1}\binom{n_2-1}{i}$.

Remark: For linear networks, the above results provide exact correct detection probability under two independent observations. The performance does not improve much compared with that of a single observation in [4] and [6].

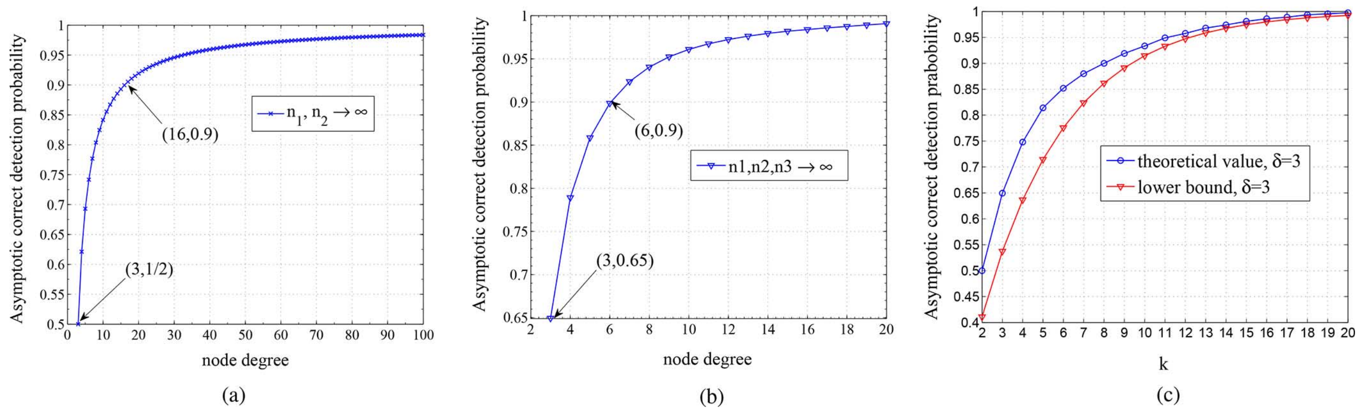


Fig. 1. Asymptotic correct detection probability under multiple independent observations. (a) Two independent observations, (b) Three independent observations, (c) $\phi_k(\delta)$ vs k, $\delta = 3$.

In the following, we see that when the degree is greater than two, multiple independent observations dramatically boost the performance.

Theorem 2: Suppose the rumor spreads on a regular tree with degree $\delta = 3$, given two independent observations G_{n_1} and G_{n_2}, then:

1) When $n_1 = n, n_2 = qn$ ($q \in \mathrm{Z}^+$), P_c is given by

$$P_\mathrm{c} = \frac{qn + q + 2}{2(qn + 1)}. \tag{16}$$

2) When $n_1 = n, n_2 = qn+1$ ($q \in \mathrm{Z}^+$), (16) still holds.
3) When $n_1 = n, n_2 = qn + t$ ($q \in \mathrm{Z}^+$), $t < n$, we have

$$P_\mathrm{c} = \frac{qn + q + 2}{2(qn + 1)} + \Delta P_c, \tag{17}$$

with $\Delta P_c < \frac{1}{2(qn+1)}$. This demonstrates that as $n \to \infty$, the asymptotic correct detection probability $\lim_{n\to\infty} P_\mathrm{c} = 1/2$.

Remark: Note that in [4], as $\delta \to \infty$, P_c asymptotically approaches 0.307 under a single observation, while here even with only two independent observations, we achieve $\lim_{n\to\infty} P_\mathrm{c} = 1/2$ for $\delta = 3$.

For general k, the following theorem describes certain basic properties of the correct detection probability pertaining to its monotonicity.

Theorem 3: Suppose the rumor spreads on a δ-regular tree, for k independent observations $G_{n_1}, \ldots, G_{n_k}$, then:

1) The correct detection probability P_c is increasing with δ. As δ grows sufficiently large, P_c approaches 1.
2) When fixing $n_1, \ldots, n_{k-1}$, P_c is non-increasing with the number of nodes of the kth observation G_{n_k}, i.e., n_k.

Remark: Theorem 3 reveals that, detection under multiple observations exhibits definite monotonicity behaviors with node degree and snapshot sizes. Such properties shed insights into understanding how the network structure and observations affect the ML detector. In addition, We derived corollary 2 as a complement to Theorem 3.

For any finite $n_1, \ldots, n_k$, we can use Algorithm 1 in Section V-A to calculate the exact value of the correct detection probability P_c. Furthermore, the following theorem precisely quantifies the asymptotic behaviors of P_c, for large snapshot sizes, degree δ, and number of observations k.

Theorem 4: Suppose the rumor spreads on a δ-regular tree, given k independent observations $G_{n_1}, \ldots, G_{n_k}$, then:

1) When $\delta \geq 3$, we have

$$\lim_{n_1,\ldots,n_k\to\infty} P_\mathrm{c} = \phi_k(\delta) := 1 - \delta\left(1 - \varphi_k\left(\frac{1}{\delta-2}, \frac{\delta-1}{\delta-2}\right)\right), \tag{18}$$

where

$$\varphi_k(\alpha, \beta) = \int\cdots\int_{\prod_{j=1}^{k} \frac{x_j}{1-x_j} \leq 1} \frac{\Gamma(\alpha+\beta)^k}{\Gamma(\alpha)^k \Gamma(\beta)^k} \prod_{j=1}^{k} \left(x_j^{\alpha-1}(1-x_j)^{\beta-1}\right)$$

$$\mathrm{d}x_1 \ldots \mathrm{d}x_k, \alpha = \frac{1}{\delta-2}, \beta = \frac{\delta-1}{\delta-2}.$$

As δ grows sufficiently large, $\phi_k(\delta) \to 1$; also, as k grows sufficiently large, $\phi_k(\delta) \to 1$.

2) In particular, when $\delta = 3$, $\phi_k(3)$ can be written as

$$\phi_k(3) = 1 - 3 \cdot 2^{k-2} \int_0^1 \cdots \int_0^1 \frac{\prod_{j=1}^{k-1} (x_j(1-x_j))}{\prod_{j=1}^{k-1} x_j + \prod_{j=1}^{k-1} (1-x_j)} \mathrm{d}x_1 \ldots \mathrm{d}x_{k-1}, \tag{19}$$

which is increasing with k and is bounded as follows:

$$1 - \frac{3}{4}\left(\frac{\pi}{4}\right)^{k-1} < \lim_{n_1,\ldots,n_k\to\infty} P_\mathrm{c} < 1, k \in \mathrm{Z}^+; \tag{20}$$

that is, as k grows sufficiently large, $\lim_{n_1,\ldots,n_k\to\infty} P_\mathrm{c} \to 1$ exponentially.

Remark: For general k observations, we mainly provide the results in the asymptotic regime as $n \to \infty$. The lower bound in case of $\delta = 3$ provides an intuitive insight that $\phi_k(3)$ increases with the number of observations k, and further indicates the convergence to be exponentially fast; see Fig. 1(c). According to Theorem 3 that P_c increases with δ and decreases with n, we have as $\delta > 2,\ k > 1,\ P_\mathrm{c} \geq P_\mathrm{c}(\delta = 3) > \phi_k(3) \geq \phi_2(3) = 1/2$. Indeed P_c may by far exceed 1/2, if $\delta > 2,\ k > 1$; and as any of k and δ grows sufficiently large, P_c converges to 1. Therefore, Theorem 4 reveals that multiple independent observations significantly improve the detection performance.

Corollary 1: Suppose the rumor spreads on a δ-regular tree, given two independent observations G_{n_1} and G_{n_2}, then:

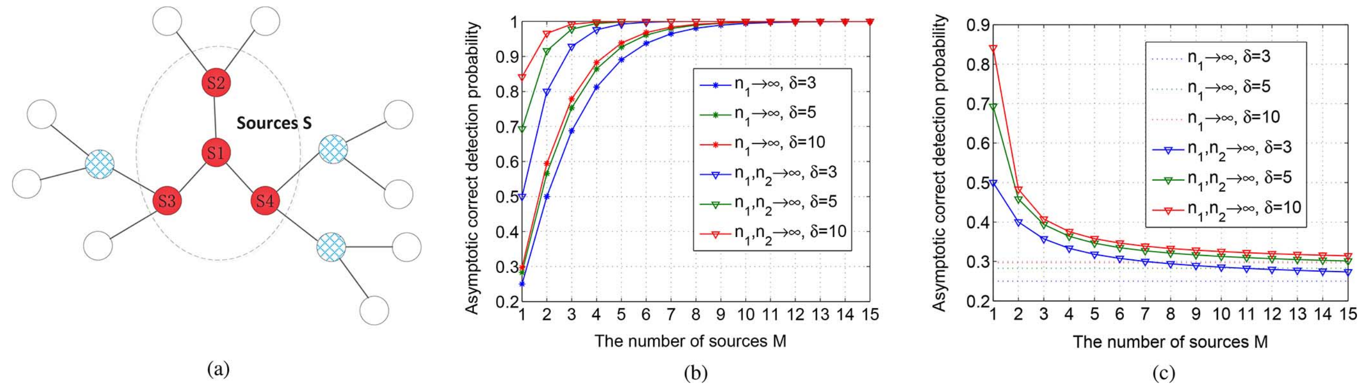


Fig. 2. Asymptotic correct detection probability for regular trees under multiple connected sources. (a) Illustration of multiple connected sources. (b) $\lim_{n\to\infty} P_{\mathrm{c}}(S)$ vs m. (c) $\lim_{n\to\infty} P_{\mathrm{c}}(s^*)$ vs m.

1) When $\delta \geq 3$, we have

$$\lim_{n_1,n_2\to\infty} P_{\mathrm{c}} = \phi_2(\delta) = 1-\delta\left(1-\varphi_2\left(\frac{1}{\delta-2},\frac{\delta-1}{\delta-2}\right)\right),$$

where $\varphi_2(\alpha,\beta) = \int\int_{x_1+x_2\leq 1} \frac{\Gamma(\alpha+\beta)^2}{\Gamma(\alpha)^2\Gamma(\beta)^2} x_1^{\alpha-1}x_2^{\alpha-1}(1-x_1)^{\beta-1}(1-x_2)^{\beta-1}\mathrm{d}x_1\mathrm{d}x_2$, $\alpha = \frac{1}{\delta-2}$, $\beta = \frac{\delta-1}{\delta-2}$. As δ grows sufficiently large, $\phi_2(\delta) \to 1$.

2) In particular, when $\delta = 3$, the asymptotic correct detection probability equals to $1/2$, i.e.,

$$\lim_{n_1\to\infty,n_2\to\infty} P_{\mathrm{c}} = \frac{1}{2}; \tag{21}$$

when $\delta = 4$, the asymptotic correct detection probability is

$$\lim_{n_1\to\infty,n_2\to\infty} P_{\mathrm{c}} = \frac{16}{\pi^2} - 1 \approx 0.621. \tag{22}$$

Remark: Corollary 1 is a special case of Theorem 4 for $k = 2$, illustrating that the asymptotic correct detection probability under merely two independent observations already significantly exceeds $1/2$; see Fig. 1(a). Besides, Fig. 1(b) displays an example of Theorem 4 in the case of three independent observations. As observed, when $\delta = 3$ or 6, the asymptotic correct detection probability equals to 0.65 or 0.9, in sharp contrast to that of 0.25 or 0.307 for $\delta = 3$ or $\delta \to \infty$ under a single observation in [4] and [6].

IV. Detection for Multiple Sources

A. Multiple Connected Sources

In many applications, there may be more than one rumor source in the network. For example, many fraudulent email spams, recurring malcode and cyberspace attacks are usually premeditated and organized by criminal gangs. Thus most of the victims are in fact innocent and only a few of suspicious nodes need to be examined first, which may have nontrivial connection patterns in the network. If we can identify one of them with high and reliable correct detection probability, we can utilize other messages from this node to further identify all the rumor sources. Our method provides a direction of main attack on this. Hence in this section, we consider the detection of multiple sources. Assume that there are m connected rumor sources, i.e., $S = \{s_1, s_2, \ldots, s_m\}$ $(m \geq 1)$. We wish to detect any of them (in other words, detecting any one of them is equally successful), and the correct detection probability of which is denoted by $P_{\mathrm{c}}(S)$. An illustration is a rumor source set $S = \{s_1, s_2, s_3, s_4\}$ with multiple connected nodes on a regular tree with node degree $\delta = 3$ is shown in Fig. 2(a). The sources in S form a connected subgraph of G.

The following theorem precisely quantifies the asymptotic behaviors of $P_{\mathrm{c}}(S)$, for large snapshot sizes, degree δ, and number of observations k $(k \geq 1)$. We present the proof in Section VII.

Theorem 5: Suppose the rumor spreads on a δ-regular tree with m connected rumor sources, i.e., $S = \{s_1, s_2, \ldots, s_m\}$ $(m \geq 1)$, given k independent observations $G_{n_1}, \ldots, G_{n_k}$ $(k \geq 1)$, when $\delta \geq 3$, we have

$$\begin{aligned}\lim_{n_1,\ldots,n_k\to\infty} & P_{\mathrm{c}}(S) = \phi_k(\delta,m)\\ &= 1-((\delta-2)m+2)\left(1-\varphi_k\left(\frac{1}{\delta-2},\frac{(\delta-2)m+1}{\delta-2}\right)\right),\end{aligned} \tag{23}$$

where $\varphi_k(\alpha,\beta) = \int\cdots\int_{\prod_{j=1}^{k}\frac{x_j}{1-x_j}\leq 1} \frac{\Gamma(\alpha+\beta)^k}{\Gamma(\alpha)^k\Gamma(\beta)^k}\prod_{j=1}^{k}(x_j^{\alpha-1}(1-x_j)^{\beta-1})\mathrm{d}x_1\ldots\mathrm{d}x_k$, $\alpha = \frac{1}{\delta-2}$, $\beta = \frac{(\delta-2)m+1}{\delta-2}$.

As $\delta \geq 3$, m grows sufficiently large, $\phi_k(\delta,m) \to 1$; $\delta \geq 3$, k grows sufficiently large, $\phi_k(\delta,m) \to 1$; also, as $k \geq 2$, δ grows sufficiently large, $\phi_k(\delta,m) \to 1$.

Remark: For multiples sources, we mainly provide the results in the asymptotic regime for large snapshot sizes. Fig. 2(b) is a special case of Theorem 5 for $k = 2$, showing that $P_{\mathrm{c}}(S)$ increases with m or δ and $P_{\mathrm{c}}(S)$ under merely two independent observations has already obviously improved compared with a single observation.

B. A Primary Rumor Source Case

In various practical scenarios, the outbreak of rumors always arises from a node cluster which formed by multiple rumor sources. The primary rumor source node is the node who have a great influence or the abundant connectivity, because it is the head of the crime gangs or has the abundant connectivity. That means the multiple source set $S = \{s_1, s_2, \ldots, s_m\}$ $(m \geq 1)$ is infected by a primary rumor source node s^* $(s^* \subseteq S)$ first.

Then they form the initial multiple source set S and collude in spreading a rumor. Based on multiple observations, our aim is to infer the primary rumor source node s^*, e.g., the head node of the source node cluster, and the correct detection probability of which is denoted by $P_{\mathrm{c}}(s^*)$.

The following theorem precisely quantifies the asymptotic behaviors of $P_{\mathrm{c}}(s^*)$, for large snapshot sizes, degree δ, and number of observations k ($k \geq 1$). We present the proof in Section VII.

Theorem 6: Suppose the rumor spreads on a δ-regular tree with m connected rumor sources, i.e., $S = \{s_1, s_2, \ldots, s_m\}$ ($m \geq 1$), given k independent observations $G_{n_1}, \ldots, G_{n_k}$ ($k \geq 1$), $\delta \geq 3$, then:

1) When $k = 1$, we have

$$\lim_{n\to\infty} P_{\mathrm{c}}(s^*) = \phi_1^*(\delta,m) = \phi_1^*(\delta,1) = 1-\delta\cdot\left(1-\varphi_0(\alpha_0,\beta_0)\right), \tag{24}$$

2) when $k \geq 2$, we have

$$\lim_{n_1,\ldots,n_k\to\infty} P_{\mathrm{c}}(s^*) = \phi_k^*(\delta,m)$$
$$= 1-\delta\cdot\sum_{i=0}^{m-1}\left(\mathbf{P}_{G_m}(X=i)\cdot\left(1-\varphi_k(\alpha_i,\beta_i)\right)\right), \tag{25}$$

where $\varphi_k(\alpha_i,\beta_i) = \int\cdots\int_{\prod_{j=1}^{k}\frac{x_j}{1-x_j}\leq 1} \frac{\Gamma(\alpha+\beta)^k}{\Gamma(\alpha)^k\Gamma(\beta)^k}\prod_{j=1}^{k}\left(x_j^{\alpha-1}(1-x_j)^{\beta-1}\right)\mathrm{d}x_1\ldots\mathrm{d}x_k$, $\alpha_i = \frac{i(\delta-2)+1}{\delta-2}$, $\beta_i = \frac{(\delta-2)(m-i)+1}{\delta-2}$, and $\mathbf{P}_{G_m}(X = x_1) = \binom{m-1}{x_1}\frac{\prod_{i=1}^{2} b_i(b_i+\delta-2)\cdots(b_i+(x_i-1)(\delta-2))}{\delta(\delta+\delta-2)\cdots(\delta+(m-2)(\delta-2))}$, where $x_2 = m-1-x_1$, $b_1 = 1$ and $b_2 = \delta-1$.

As m grows large, $\phi_k^*(\delta,m) \to P_{\mathrm{c}}(G_m,\delta) \geq \phi_1^*(\delta,m)$, where $P_{\mathrm{c}}(G_m)$ is the correct detection probability in the finite regime consisting of m nodes for regular trees.

Remark: For the case of $k = 1$, i.e., a single observation, the number of sources m does not affect $\phi_1^*(\delta,m)$, while for the case of $k \geq 2$, $\phi_k^*(\delta,m)$ decreases with m and converges to $P_{\mathrm{c}}(G_m,\delta)$. Fig. 2(c) is a special case of Theorem 5 for $k = 2$, showing that $P_{\mathrm{c}}(S)$ increases with δ and decreases with m. From Fig. 2(c), the impossibility of performance improvement of identifying the primary rumor source for large m demonstrates the boost of detection by diversity is offset by the number of secondary rumor sources.

V. Algorithms for Trees

A. Detection Probability for Regular Trees

In this section, we present Algorithm 1 to calculate the exact correct detection probability under multiple independent observations $G_{n_1}, \ldots, G_{n_k}$ in the finite regime for regular trees, based on the Lemma 3 established in Section VII.

For a rumor source s^* with m neighbors in $\{\cap G_{1\to k}\}$, under multiple independent observations, we propose Algorithm 1 by exploiting Lemma 3 instead of Lemma 2 to calculate the exact detection probability that greatly reduces the number of iterations and the computation complexity.

Algorithm 1 Correct Detection Probability Calculation

Input: δ, k, G_{n_j} ($j = 1, \ldots, k$)
Initialize $P_e = 0$
for $x_{1,n_k} = 1 \to n_k - 1$ **do**
  ......
  for $x_{1,n_1} = 1 \to n_1 - 1$ **do**
    if $\prod_{j=1}^{k} x_{1,n_j} > \prod_{j=1}^{k}(n_j - x_{1,n_j})$ **then**
      $P_e = P_e + \prod_{j=1}^{k}\mathbf{P}_{G_{n_j}}(X_{1,n_j} = x_{1,n_j})$
    end if
    if $\prod_{j=1}^{k} x_{1,n_j} = \prod_{j=1}^{k}(n_j - x_{1,n_j})$ **then**
      $P_e = P_e + \frac{1}{2}\cdot\prod_{j=1}^{k}\mathbf{P}_{G_{n_j}}(X_{1,n_j} = x_{1,n_j})$
    end if
  end for
end for
Output: $P_c = 1 - \delta\cdot P_e$

Algorithm 2 Joint Rumor Centrality Calculation

Input: Graphs $G_{n_j}, j \in [1,k]$
Choose an arbitrary node $v \in \cap G_{1\to k}$ as a root.
for each $u \in \cup G_{1\to k}$ **do**
  if u is a leaf **then**
    $t^{\mathrm{up}}_{u\to\mathrm{parent}(u,\cup G_{1\to k})}: t^{\mathrm{up}}_{u,G_{n_j}\to\mathrm{parent}(u,G_{n_j})} = 1$, $j \in \{j : u, \mathrm{parent}(u) \in G_{n_j}, j \in [1,k]\}$
    $t^{\mathrm{up}*}_{u\to\mathrm{parent}(u,\cup G_{1\to k})} = 1$, $p^{\mathrm{up}}_{u\to\mathrm{parent}(u,\cup G_{1\to k})} = 1$
  else if u is root v **then**
    $v', v \in \cap G_{1\to k}, \forall v' \in \mathrm{child}(v,\cap G_{1\to k}), r^{\mathrm{down}}_{v\to v'} = \frac{\prod_{j=1}^{k}(n_j-1)}{\prod_{i\in\mathrm{child}(v,\cup G_{1\to k})} p^{\mathrm{up}}_{i\to v}}$
  else
    $t^{\mathrm{up}}_{u\to\mathrm{parent}(u,\cup G_{1\to k})}: t^{\mathrm{up}}_{u\to\mathrm{parent}(u,G_{n_j})} = \sum_{i\in\mathrm{child}(u,G_{n_j})} t^{\mathrm{up}}_{i\to u} + 1$, $j \in \{j : u, \mathrm{parent}(u) \in G_{n_j}, j \in [1,k]\}$
    $t^{\mathrm{up}*}_{u\to\mathrm{parent}(u,\cup G_{1\to k})} = \prod_{\{j:u,\mathrm{parent}(u)\in G_{n_j}, j\in[1,k]\}} t^{\mathrm{up}}_{u\to\mathrm{parent}(u,G_{n_j})}$
    if $u \in \cap G_{1\to k}$ **then**
      $t^{\mathrm{up}'}_{u\to\mathrm{parent}(u,\cup G_{1\to k})} = \prod_{\{j:u,\mathrm{parent}(u)\in G_{n_j}, j\in[1,k]\}}(n_j - t^{\mathrm{up}}_{u\to\mathrm{parent}(u,G_{n_j})})$
      $\forall u' \in \mathrm{child}(u,\cap G_{1\to k})$, $r^{\mathrm{down}}_{u\to u'} = r^{\mathrm{down}}_{\mathrm{parent}(u,\cup G_{1\to k})\to u}\cdot\frac{t^{\mathrm{up}*}_{u\to\mathrm{parent}(u,\cup G_{1\to k})}}{t^{\mathrm{up}'}_{u\to\mathrm{parent}(u,\cup G_{1\to k})}}$
    end if
    $p^{\mathrm{up}}_{u\to\mathrm{parent}(u,\cup G_{1\to k})} = t^{\mathrm{up}*}_{u\to\mathrm{parent}(u,\cup G_{1\to k})}\cdot\prod_{i\in\mathrm{child}(u,\cup G_{1\to k})} p^{\mathrm{up}}_{i\to u}$
  end if
end for
Output: Joint rumor centralities of all the nodes in $\{\cap G_{1\to k}\}$.

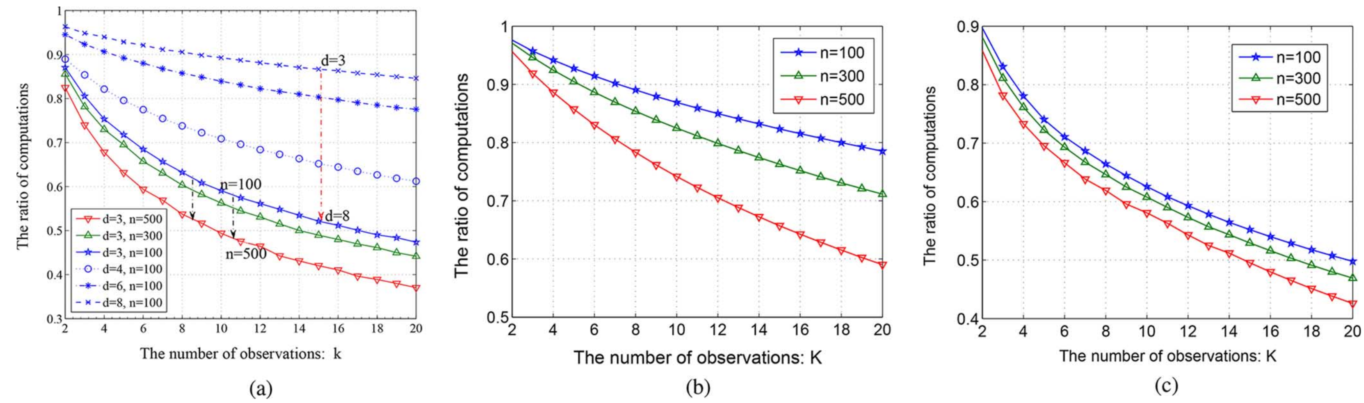


Fig. 3. Ratio of computations between Algorithm 2 and the direct implementation of multiple instances of single observation message-passing algorithm. (a) 3-regular tree, (b) Small-World networks, (c) Scale-free networks.

B. Joint Rumor Centrality Calculation

For the single rumor source detection problem, to find the joint rumor center from k different independent observations $G_{n_1}, \dots, G_{n_k}$, we need to evaluate the maximum joint rumor centrality in $\{\cap G_{1\to k}\}$. In [4], a message-passing algorithm is proposed to compute the rumor centralities for one time observation in a general tree G_n with n nodes, using $\mathcal{O}(n)$ computations. In this section, we propose an algorithm for calculating the joint rumor centralities among all infected nodes based on message-passing, using $\mathcal{O}(|G_{n_1} \cup \dots G_{n_k}|)$ computation steps. A baseline algorithm for comparison is to naively execute k instances of the single observation message-passing algorithm in [4] to find the joint rumor center, using $\mathcal{O}(\sum_{j=1}^{k} n_j)$ computations. Since $|G_{n_1} \cup \dots G_{n_k}| < \sum_{j=1}^{k} n_j$, our algorithm is more computationally efficient.

According to (3) and (8), the joint rumor centrality of node s $(s \in \{\cap G_{1\to k}\})$ can be written as

$$R_k(s,G_{n_1}\dots G_{n_k}) = R_k(s^*,G_{n_1}\cdots G_{n_k}) \prod_{i\in\mathbb{P}(s^*,s)} \frac{T^{s^*}_{i,G_{n_1}}\cdots T^{s^*}_{i,G_{n_k}}}{(n_1-T^{s^*}_{i,G_{n_1}})\cdots(n_k-T^{s^*}_{i,G_{n_k}})}$$
$$= \frac{(n_1-1)!\cdots(n_k-1)!}{\prod_{u\in\{G_{n_1}\setminus s\}} T^{s}_{u,G_{n_1}}\cdots \prod_{u\in\{G_{n_k}\setminus s\}} T^{s}_{u,G_{n_k}}}, \qquad (26)$$

where $\mathbb{P}(s^*, s)$ is the set of nodes in the path between the source s^* and s, not including s^*. This is the key to our algorithm for calculating the joint rumor centralities for all the nodes in $\{\cap G_{1\to k}\}$. Let us briefly describe the idea of our algorithm. We consider the extended infected tree-type graph formed by the k observations, $\{\cup G_{1\to k}\}$. We first select an arbitrary node $v \in \{\cap G_{1\to k}\}$ as the rumor source and calculate the size of all of its subtrees and its joint rumor centrality. This is accomplished by having each node u pass three messages up to its parent in $\cup G_{1\to k}$. The first message, denoted by $t^{\text{up}}_{u\to\text{parent}(u,\cup G_{1\to k})}$, contains up to k variables, each of which is denoted by $t^{\text{up}}_{u,G_{n_j}\to\text{parent}(u,G_{n_j})}$, representing the number of nodes in u's subtree of the observation G_{n_j}, as u and its parent both belong to G_{n_j}. The second message contains two variables, one of which is the cumulative product of the variables in the first message, which we call $t^{\text{up}*}_{u\to\text{parent}(u,\cup G_{1\to k})}$, and the other of which is an optional variable, denoted by $t^{\text{up}'}_{u\to\text{parent}(u,\cup G_{1\to k})}$, when both $u, v \in \cap G_{1\to k}$, thus assisting in calculating the joint rumor centralities of neighboring nodes. The third message is the cumulative product of the sizes of the subtrees of all nodes in u's subtree of each G_{n_j}, which we call $p^{\text{up}}_{u\to\text{parent}(u,\cup G_{1\to k})}$. These messages are passed upward until the source node receives its messages. By multiplying the cumulative subtree products of the children of the source v, we obtain the joint rumor centrality $R_k(v, G_{n_1} \dots G_{n_k})$.

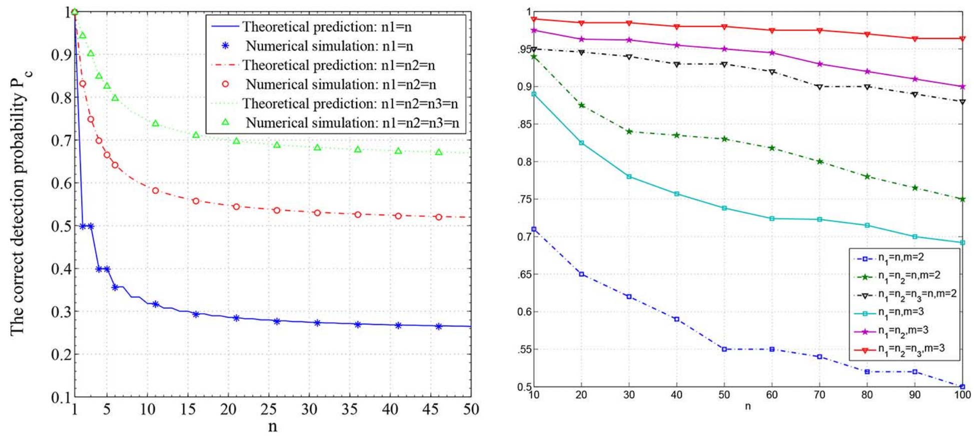
Based on the joint rumor centrality of node v, we evaluate the joint rumor centralities of the children of v using (26). Each node in $\cap G_{1\to k}$ passes its joint rumor centrality to its children in a message which we denote by $r^{\text{down}}_{u\to u'}$, for $\forall u' \in \text{child}(u, \cap G_{1\to k})$. Each node calculates its joint rumor centrality using its parent's joint rumor centrality and its messages $t^{\text{up}*}_{u\to\text{parent}(u,\cup G_{1\to k})}$ and $t^{\text{up}'}_{u\to\text{parent}(u,\cup G_{1\to k})}$. Therefore, our joint rumor centrality calculation algorithm calculates the joint rumor centralities of all infected nodes by message-passing, using $\mathcal{O}(|G_{n_1} \cup \dots G_{n_k}|)$ computation steps.

C. Computational Efficiency

We perform simulations with different node degrees and infected set sizes to empirically evaluate the computational efficiency. We define r_a as the ratio of computations between Algorithm 2 and the direct execution of multiple instances of single observation message-passing algorithm in [4]. For each simulation experiment we conduct the Monte-Carlo simulation 10000 times. As shown in Fig. 3(a), We get r_a decreases with the number of observations, the infected set size and node degree. These imply that Algorithm 2 is advantageous in cases with large node degrees and large number of observations. Then, we perform simulations on small-world networks [21] and scale-free networks [22] and an overview of these networks will be stated in the following Section VI-B. As shown in Fig. 3(b) and (c), when the number of observations is 20, and 500 infected nodes are observed in each observation, Algorithm 2 almost reduces half of the computations as compared to running multiple single observation message-passing algorithms. Furthermore, we can also set up a similar algorithm to realize the case of multiple sources.

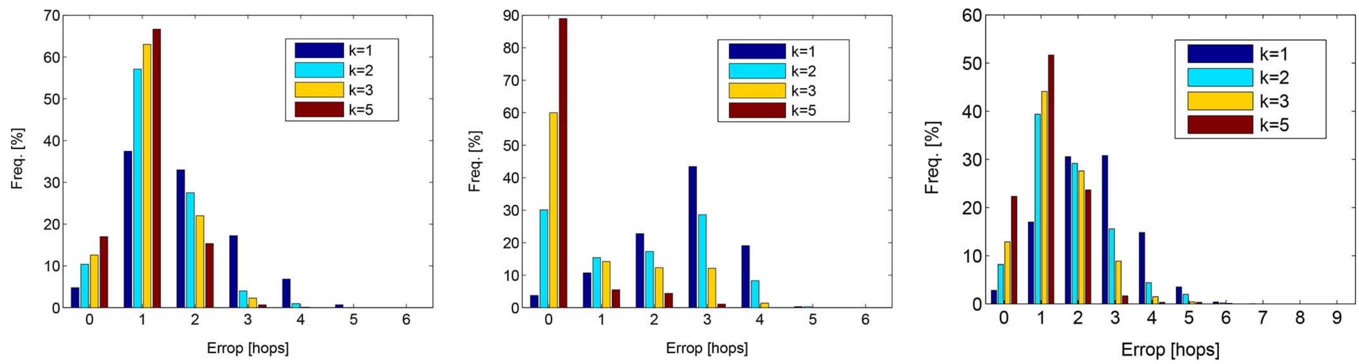
D. Detector for General Graphs

For general graphs, owing to the lack of knowledge of the spanning tree, we use the BFS heuristic to obtain an induced tree network from each observation. We assume that if node $s^* \in G_{n_j}$



(a) (b)

Fig. 4. Simulation results for regular trees. (a) P_c vs n,$\delta = 3$, (b) P_c vs n,$\delta = 3$, multiples sources.



(a) (b) (c)

Fig. 5. Simulation results for general graph networks. (a) Small-World networks, (b) Scale-free networks, (c) Newman's scientific collaboration network.

$(1\leq j\leq k)$ is the source, then the rumor spreads along a BFS tree rooted at s^*, denoted by $T_b(s^*, G_{n_j})$. Therefore, based on Section II-B(3), we obtain the following rumor source detector:

$$\widehat{s} = \arg\max_{s^*\in\{\cap G_{1\rightarrow k}\}} \prod_{j=1}^{k} P(\sigma_s^* \mid s^*, G_{n_j}) \cdot R_k(s^*, G_{n_1} \ldots G_{n_k}), \tag{27}$$

where σ_s^* represents the BFS ordering of nodes in $T_b(s^*, G_{n_j})$. In Section VI, we will show with simulations the performance of this detector for different networks.

VI. Numerical Simulations

In this section, we evaluate the performance of the proposed rumor source detector on different networks.

A. Tree Networks

Here we provide simulation results for regular trees in order to corroborate the analysis in Sections III and IV. For each configuration, we conduct the Monte-Carlo simulation 10000 times to assess the correct detection probability.

In Fig. 4(a), we show that when $\delta = 3$, P_c increases with the number of observations and decreases with the number of infected nodes. As shown, multiple independent observations significantly improve the performance, compared with a single observation.

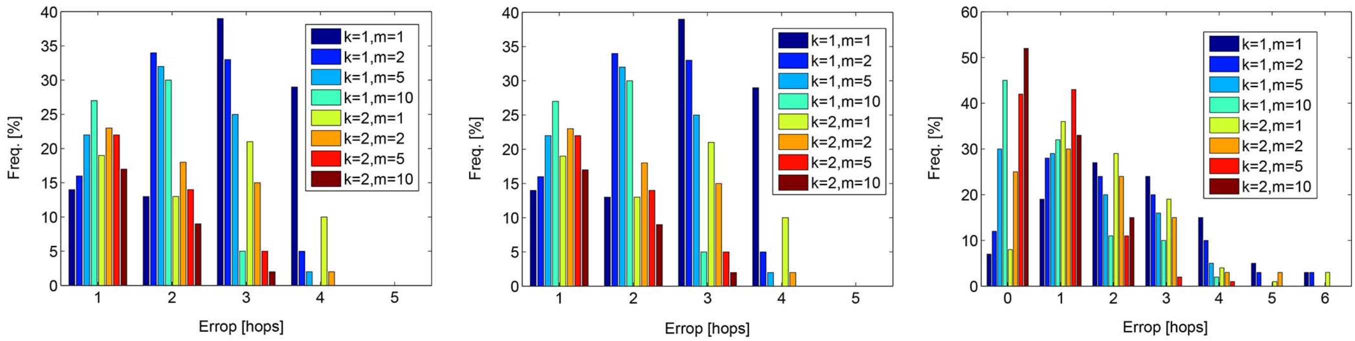
In Fig. 4(b), for $\delta = 3$, with multiple sources $m = 2$ or 3, the simulation results show multiple observations dramatically boost the performance, which coincides with the theoretical values as predicted by Theorem 5.

B. General Graph Networks

1) A Single Rumor Source: We perform simulations on small-world networks [21] and scale-free networks [22]. There are two very popular models for networks, so we would like our rumor source estimator to perform well on these topologies. Watts and Strogatz (WS) introduced the concept of small-world network [21]. An interesting experience of the small world is that, after meeting a stranger, one is surprised to find that they have a common friend in between. Small-world model is that the connectivity distribution of a network peaks at an average value and decays exponentially. Another significant recent discovery in the field of complex networks is the observation that many large-scale networks are scale-free [22], that is, their connectivity distributions are in a power-law form that is independent of the network scale. That means most nodes have very few link connections and yet a few nodes have many connections.

For both topologies, each underlying graph contains 10000 nodes and we let the rumor spread to infect up to 400 nodes in each observation.

We investigate the error (i.e., the number of hops between the detected source and the actual source) on the two networks, respectively. Comparing two independent observations with a single observation, we have a better performance in both networks, especially the scale-free network. As can be seen



(a) (b) (c)

Fig. 6. Simulation results for general graph networks under multiple connected sources. (a) Small-World networks, (b) Scale-free networks, (c) Newman's scientific collaboration network.

from Fig. 5(b), we have a fairly large correct detection probability under multiple observations,—nearly 90 percent when k is 5. The reason for this may be that scale-free networks have a highly heterogeneous degree distribution and thus contain many high-degree hubs.

Furthermore, we perform simulations on Newmans scientific collaboration network[1] which contained 13861 nodes. From Fig. 5(c), we show that multiple observations dramatically boost the performance.

2) Multiple Connected Sources: We evaluate the performance under multiple connected sources on the above general graphs. The number of multiple sources are randomly chosen to be 1, 2, 5 and 10. The simulation results are shown in Fig. 6. As can be seen, comparing two independent observations with a single observation, we have a better performance in both three topologies.

VII. Proofs

A. Lemmas

We need the following technical lemmas, whose proofs (along with that of Lemma 1) are in Appendix; see more details in [20]. For a rumor source s^* with m $(1 \le m \le \delta)$ neighbors in $\{\cap G_{1\to k}\}$, i.e., $s_1^*, \ldots, s_m^*$, let random variable X_{i,n_j} be the number of nodes in subtree $Z^{s^*}_{s_i^*, G_{n_j}}$ of observation G_{n_j} $(1 \le i \le m, 1 \le j \le k)$.

We have stated Lemma 1 in Section III-B.

Lemma 2: For the source s^* and the inferred $\hat{s}$, introduce

$$\begin{cases} P_1 := P\left(\hat{s} = s^* \middle| \forall i, 1 \le i \le m, \prod_{j=1}^k x_{i,n_j} < \prod_{j=1}^k (n_j - x_{i,n_j})\right) = 1, \\ P_{\frac{1}{2}} := P\left(\hat{s} = s^* \middle| \exists i, 1 \le i \le m, \prod_{j=1}^k x_{i,n_j} = \prod_{j=1}^k (n_j - x_{i,n_j})\right) = \frac{1}{2}, \\ P_0 := P\left(\hat{s} = s^* \middle| \exists i, 1 \le i \le m, \prod_{j=1}^k x_{i,n_j} > \prod_{j=1}^k (n_j - x_{i,n_j})\right) = 0. \end{cases}$$

The correct detection probability P_{c} is given by

$$\begin{aligned} P_{\mathrm{c}} = {} & P_{\frac{1}{2}} \cdot \sum_{\prod_{j=1}^k x_{i,n_j} = \prod_{j=1}^k (n_j - x_{i,n_j})}^{1\le i\le m, 1\le m\le\delta} \prod_{j=1}^k \mathbf{P}_{G_{n_j}}\left(\bigcap_{i=1}^{\delta}\{X_{i,n_j} = x_{i,n_j}\}\right) \\ & + P_1 \cdot \sum_{\prod_{j=1}^k x_{i,n_j} < \prod_{j=1}^k (n_j - x_{i,n_j})}^{1\le i\le m, 1\le m\le\delta} \prod_{j=1}^k \mathbf{P}_{G_{n_j}}\left(\bigcap_{i=1}^{\delta}\{X_{i,n_j} = x_{i,n_j}\}\right). \end{aligned} \tag{28}$$

Remark: Lemma 2 is deduced from Proposition 1. In order to prove Theorem 1 and Theorem 2, we should find the conditions that the estimator in (8) can correctly identify s^* as the source, i.e., s^* is the joint rumor center.

Lemma 3: The correct detection probability P_{c} in Lemma 2 can be rewritten as

$$\begin{aligned} P_{\mathrm{c}} = 1 - \delta \Bigg\{ & \frac{1}{2} \sum_{\prod_{j=1}^k x_{i,n_j} = \prod_{j=1}^k (n_j - x_{i,n_j})} \prod_{j=1}^k \mathbf{P}_{G_{n_j}}\left(\bigcap_{i=1}^{\delta}\{X_{i,n_j} = x_{i,n_j}\}\right) \\ & + \sum_{\prod_{j=1}^k x_{i,n_j} > \prod_{j=1}^k (n_j - x_{i,n_j})} \prod_{j=1}^k \mathbf{P}_{G_{n_j}}\left(\bigcap_{i=1}^{\delta}\{X_{i,n_j} = x_{i,n_j}\}\right) \Bigg\}. \end{aligned} \tag{29}$$

Remark: Lemma 3 uses the fact that the error detection events are all disjoint, established in Proposition 1, Lemma 1 and Lemma 5. Using this Lemma instead of Lemma 2, we design Algorithm 1 to calculate the exact correct detection probability, in Section V-A. This greatly decreases the number of iterations and the calculation time.

Lemma 4: Suppose a rumor source s^* spreads on a δ-regular tree, resulting in an observation G_n. As $n \to \infty$, the probability that a subtree $Z^{s^*}_{s_i^*, G_n}$ $(1 \le i \le \delta)$ is empty asymptotically vanishes:

$$\lim_{n\to\infty} \mathbf{P}_{G_n}(X_i = 0) = 0, \qquad 1 \le i \le \delta. \tag{30}$$

Remark: Lemma 4 shows that as the spreading infects a sufficiently large number of nodes, the rumor source tends to be surrounded, which is very meaningful for establishing our asymptotical analysis in detection probability under multiple observations; see its proof in Appendix.

Lemma 5: Suppose a rumor source s^* spreads on a δ-regular tree, with multiple independent observations $G_{n_1}, \ldots, G_{n_k}$. For $1 \le i \le \delta$, define

$$E_i = \left\{ \frac{x_{i,n_1}}{n_1} \cdots \frac{x_{i,n_k}}{n_k} < \left(1 - \frac{x_{i,n_1}}{n_1}\right) \cdots \left(1 - \frac{x_{i,n_k}}{n_k}\right) \right\},$$

$$D_i = \left\{ \frac{x_{i,n_1}}{n_1} \cdots \frac{x_{i,n_k}}{n_k} = \left(1 - \frac{x_{i,n_1}}{n_1}\right) \cdots \left(1 - \frac{x_{i,n_k}}{n_k}\right) \right\},$$

$$F_i = \left\{ \frac{x_{i,n_1}}{n_1} \cdots \frac{x_{i,n_k}}{n_k} \le \left(1 - \frac{x_{i,n_1}}{n_1}\right) \cdots \left(1 - \frac{x_{i,n_k}}{n_k}\right) \right\}.$$

[1]Source: http://konect.uni-koblenz.de/networks/opsahl-collaboration

As $n_1,\ldots,n_k\to\infty$, we have

$$\lim_{n_1,\ldots,n_k\to\infty} P_{\rm c}=1-\delta\cdot\lim_{n_1,\ldots,n_k\to\infty}\mathbf{P}_G(E_1^c)=1-\delta\cdot\lim_{n_1,\ldots,n_k\to\infty}\mathbf{P}_G(F_1^c). \quad (31)$$

Remark: Lemma 5 is deduced from Lemma 1 and Lemma 4. E_i and D_i denote the events of error detection and half correct detection probability from Proposition 1, respectively, and F_i denotes the events containing both E_i and D_i, for $1\le i\le\delta$. Interestingly, to evaluate the asymptotical correct detection probability, we only focus on any a subtree of the rumor source. We show that the lower and upper bounds in (31) asymptotically coincide to prove this lemma; see its proof in Appendix.

Corollary 2: Suppose the rumor spreads on a δ-regular tree. For two independent observations G_{n_1} and G_{n_2}, given the number of nodes of G_{n_1}, i.e., n_1, and considering $n_2=qn_1$ and $n_2=qn_1+1$, $q\in Z^+$, these two cases have the same correct detection probability, i.e.,

$$P_{\rm c}(n_2=qn_1)=P_{\rm c}(n_2=qn_1+1).$$

B. Theorem 1

Here we only present the proof of Theorem 1(1) for $n_1{=}n_2{=}n$; the other cases can be deduced similarly.

Since the rumor source s^* has m neighbors ($0\le m\le\delta$), we have $\delta{+}1$ cases depending on m. For clarity, we abbreviate x_{i,n_1} and x_{i,n_2} by x_i and y_i ($1\le i\le\delta$), respectively. Our proof proceeds in two steps. First, for each of the $\delta{+}1$ cases, we calculate $P_{\rm c}$ using Lemma 2. Second, we sum up all the cases.

When $\delta{=}2$, the joint distributions of $\{X_i, 1\le i\le 2\}$ and $\{Y_i, 1\le i\le 2\}$ are respectively (see [6])

$$\mathbf{P}_{G_{n_1}}\left(\bigcap_{i=1}^{2}\{X_i=x_i\}\right)=\frac{(n-1)!}{x_1!x_2!}\frac{1}{2^{n-1}},\mathbf{P}_{G_{n_2}}\left(\bigcap_{i=1}^{2}\{Y_i=y_i\}\right)=\frac{(n-1)!}{y_1!y_2!}\frac{1}{2^{n-1}}. \quad (32)$$

Consider the case of $m{=}2$; see Fig. 7. According to Proposition 1, (x_1,x_2) and (y_1,y_2) should satisfy

$$\begin{cases} x_1+y_1\le n,\\ x_2+y_2\le n,\\ x_1+x_2=n-1, \quad x_1\ge 1, x_2\ge 1,\\ y_1+y_2=n-1, \quad y_1\ge 1, y_2\ge 1. \end{cases} \quad (33)$$

Then when multiplied by the joint distributions $P_G(X)$ and $P_G(Y)$, from Lemma 2, the contribution to $P_{\rm c}$ is given by

$$\begin{aligned} P_{{\rm c},2}=&\left\{\binom{n-1}{1}\left(\binom{n-1}{1}+\frac{1}{2}\binom{n-1}{2}\right)+\sum_{x_1=2}^{n-3}\binom{n-1}{x_1}\left(\frac{1}{2}\binom{n-1}{x_1-1}+\right.\right.\\ &\left.\left.\binom{n-1}{x_1}+\frac{1}{2}\binom{n-1}{x_1+1}\right)+\binom{n-1}{n-2}\left(\binom{n-1}{n-2}+\frac{1}{2}\binom{n-1}{n-3}\right)\right\}\cdot 2^{-2(n-1)}\\ =&\left(\frac{1}{2}\binom{2n}{n}-2n\right)\cdot 2^{-2(n-1)}. \end{aligned} \quad (34)$$

Likewise, the contributions to $P_{\rm c}$ in the case of $m{=}1$ and $m{=}0$ are $P_{{\rm c},1}=2(n-1)\cdot 2^{-2(n-1)}$ and $P_{{\rm c},0}=2\cdot 2^{-2(n-1)}$, respectively. Summing up all the cases from $m{=}0$ to 2, Theorem 1(1) is proved.

As $n{\to}\infty$, by Stirling's formula, from (14), we have

$$P_{\rm c}=\frac{1}{2^{2n-1}}\cdot\frac{(2n)!}{n!n!}\approx\frac{1}{2^{2n-1}}\frac{\sqrt{2\pi n}(\frac{2n}{e})^{2n}}{(\sqrt{2\pi n}(\frac{2n}{e})^{n})^2}=\frac{2}{\sqrt{\pi n}}. \quad (35)$$

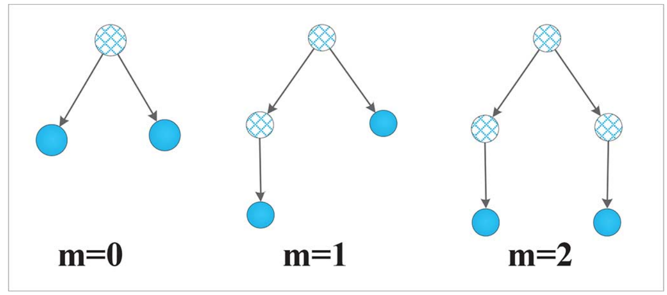


Fig. 7. Illustration of a rumor source s^* with m neighbors ($0\le m\le\delta$) in $\{\cap G_{1\to k}\}$, $\delta=2$. Latticed circles represent infected nodes in $\{\cap G_{1\to k}\}$.

C. Theorem 2

Here we only present the proof of Theorem 2(1); the other cases can be deduced similarly.

Our proof proceeds in two steps. First, we identify error detection events from Proposition 1 and Lemma 1. Second, we calculate $P_{\rm c}$ by Lemma 3.

Since $\delta{=}3$, the joint distributions of $\{X_i, 1\le i\le 3\}$ and $\{Y_i, 1\le i\le 3\}$ are $\mathbf{P}_{G_{n_1}}(\bigcap_{i=1}^{3}\{X_i=x_i\})=\frac{2}{n(n+1)}$, $\mathbf{P}_{G_{n_2}}(\bigcap_{i=1}^{3}\{Y_i=y_i\})=\frac{2}{qn(qn+1)}$, respectively. Hence, the marginal distributions of X_1 and Y_1 are

$$\mathbf{P}_{G_{n_1}}(X_1{=}x_1)=\frac{2(n-x_1)}{n(n+1)},\quad \mathbf{P}_{G_{n_2}}(Y_1{=}y_1)=\frac{2(n-y_1)}{qn(qn+1)}, \quad (36)$$

respectively.

From Proposition 1 and Lemma 1, when X_1 and Y_1 satisfy

$$\frac{x_1y_1}{(n-x_1)(qn-y_1)}\ge 1, \quad (37)$$

the error detection events happen.

Applying Lemma 3, when enumerating the error detection events,

$$\begin{aligned} P_{\rm e}=&\sum_{x_1=1}^{n-1}\frac{2(n-x_1)}{n(n+1)}\left(\sum_{y_1=qn-qx_1+1}^{qn-1}\frac{2(n-y_1)}{qn(qn+1)}\right.\\ &\left.+\frac{1}{2}\sum_{y_1=qn-qx_1}\frac{2(n-y_1)}{qn(qn+1)}\right)=\frac{qn-q}{6(qn+1)}. \end{aligned} \quad (38)$$

So $P_{\rm c}$ is

$$P_{\rm c}=1-3\cdot\frac{qn-q}{6(qn+1)}=\frac{qn+q+2}{2(qn+1)}. \quad (39)$$

Theorem 2(1) is thus proved.

D. Theorem 3

a) Theorem 3(1): From Lemma 3, in order to prove that $P_{\rm c}$ is increasing with δ, we just need to demonstrate that for any $(x_{1,n_1},\ldots,x_{1,n_k})$ satisfying $\prod_{j=1}^{k}x_{1,n_j}\ge\prod_{j=1}^{k}(n_j-x_{1,n_j})$, $\delta\prod_{j=1}^{k}\mathbf{P}_{G_{n_j}}(X_{1,n_j}{=}x_{1,n_j})$ is decreasing with δ. Note that the marginal distribution of X_1 is given by

$$\begin{aligned} &\mathbf{P}_G(X_1=x_1)\\ &=\binom{n-1}{x_1}\frac{\prod_{i=1}^{2}b_i(b_i+\delta-2)\cdots(b_i+(x_i-1)(\delta-2))}{\delta(\delta+\delta-2)\cdots(\delta+(n-2)(\delta-2))}, \end{aligned} \quad (40)$$

where $x_2 = n-1-x_1$, $b_1 = 1$ and $b_2 = \delta - 1$. By defining $f(a,b) = \frac{\frac{1}{\delta-2}+a}{\frac{\delta}{\delta-2}+b}$ and $F(\delta) = \delta \prod_{j=1}^{k} \mathbf{P}_{G_{n_j}}(X_{1,n_j} = x_{1,n_j})$, we have

$$F(\delta)=\delta\prod_{j=1}^{k}\binom{n_j-1}{x_{1,n_j}}\frac{\prod_{z_1=0}^{x_{1,n_j}-1}(1+z_1(\delta-2))\prod_{z_2=0}^{n_j-x_{1,n_j}-1}(\delta-1+z_2(\delta-2))}{\delta(\delta+\delta-2)\cdots(\delta+(n-2)(\delta-2))}$$
$$=\delta^{-(k-1)}\prod_{j=1}^{k}\big(f(1,1)f(2,2)\cdots f(x_{1,n_j}-1,x_{1,n_j}-1)$$
$$\cdot\, f(1,x_{1,n_j})\cdots f(n_j-2-x_{1,n_j},n_j-2)\big). \quad (41)$$

Rearranging $f(a,b)$, we have

$$f(a,b) = \frac{a\delta - a + 1}{(1+b)\delta - b} = \frac{b+1}{a} + \frac{\frac{b+1-a}{a}}{(1+b)\delta - b}.$$

Hence if $b \geq a$, $f(a,b)$ is decreasing with δ. So all factors of (41) are decreasing with δ, and so is $F(\delta)$. As δ grows large, $\lim_{\delta\to\infty} f(a,b) = \frac{b+1}{a}$ is a finite number. By applying (41), $\lim_{\delta\to\infty} F(\delta) = \lim_{\delta\to\infty} \delta^{-(k-1)} \cdot C = 0$, where C is a bounded number. Due to the fact that the total number of the error detection events is finite, as δ grows without bound, the correct detection probability P_c approaches 1 asymptotically.

b) Theorem 3(2): Here, we will prove that P_c is non-increasing as n_k increases to n_k+1, with given $G_{n_1},\ldots,G_{n_{k-1}}$ (i.e., fixing $n_1,\ldots,n_{k-1}$). We define events $D(\cdot)$ and $E(\cdot)$ as follows:

$$D(n_k)=\left\{\frac{x_{1,n_k}}{n_k} = \frac{\prod_{j=1}^{k-1}(n_j-x_{1,n_j})}{\prod_{j=1}^{k-1}(n_j-x_{1,n_j})+\prod_{j=1}^{k-1}x_{1,n_j}}\right\}, \quad (42)$$

$$E(n_k)=\left\{\frac{x_{1,n_k}}{n_k} > \frac{\prod_{j=1}^{k-1}(n_j-x_{1,n_j})}{\prod_{j=1}^{k-1}(n_j-x_{1,n_j})+\prod_{j=1}^{k-1}x_{1,n_j}}\right\}. \quad (43)$$

Thus $D(n_k)\cup E(n_k)$ is

$$D(n_k)\cup E(n_k) = \left\{\frac{\prod_{j=1}^{k}x_{1,n_j}}{\prod_{j=1}^{k}(n_j-x_{1,n_j})} \geq 1\right\}. \quad (44)$$

It can then be shown that the correct detection probability in Lemma 3 is equivalent to

$$P_c=1-\delta\left\{\frac{1}{2}\sum_{x_{1,n_k}=1}^{n_k-1}\mathbf{P}_{G_{n_k}}(X_{1,n_k}=x_{1,n_k})\sum_{D(n_k)}\prod_{j=1}^{k-1}\mathbf{P}_{G_{n_j}}(X_{1,n_j}=x_{1,n_j})\right.$$
$$\left.+\sum_{x_{1,n_k}=1}^{n_k-1}\mathbf{P}_{G_{n_k}}(X_{1,n_k}=x_{1,n_k})\sum_{E(n_k)}\prod_{j=1}^{k-1}\mathbf{P}_{G_{n_j}}(X_{1,n_j}=x_{1,n_j})\right\}. \quad (45)$$

We assume $|G_{n_j}|=n_j$ $(1\leq j\leq k)$ and $|G_{n_k}|=n_k'=n_k+1$, and evaluate the change of P_c as $|G_{n_k}|$ changes from n_k to n_k'.

Here we outline a sketch of the subsequent proof procedure. For any fixed $(x_{1,n_1},\ldots,x_{1,n_{k-1}})$, we can determine the solution set of x_{1,n_k} as well as that of $x_{1,n_k'}$, through (44). For the fixed $(x_{1,n_1},\ldots,x_{1,n_{k-1}})$, we can then evaluate the change in the correct detection probability. Therein, several possible cases need to be considered. Finally, based on (45), we conclude the proof by observing that for each fixed $(x_{1,n_1},\ldots,x_{1,n_{k-1}})$, the change of P_c from n_k to n_k' is bounded by zero from above.

In the following, we use X_j $(1\leq j\leq k-1)$, Y_k and Y_k' to denote X_{1,n_j} $(1\leq j\leq k-1)$, X_{1,n_k} and $X_{1,n_k'}$, respectively; furthermore, define

$$\bar{y}_k = \min\{y_k : (x_1,\ldots,x_{k-1},y_k)\in\{D(n_k)\cup E(n_k)\}\}, \quad (46)$$

$$\bar{y}_k' = \min\left\{y_k' : (x_1,\ldots,x_{k-1},y_k')\in\{D(n_k')\cup E(n_k')\}\right\}. \quad (47)$$

For arbitrary $(x_1^*,x_2^*,\ldots,x_{k-1}^*)$, take $x_k^* = \bar{y}_k$. If $x_k^*\in E(n_k)$, due to the fact that $\frac{x_k^*+1}{n_k+1} > \frac{x_k^*}{n_k} > \frac{x_k^*}{n_k+1}$, we have $x_k^*+1\in E(n_k')$; but x_k^* may not belong to either $E(n_k')$ or $D(n_k')$. Besides, note that x_k^*-1 cannot belong to $E(n_k')$ or $D(n_k')$, because x_k^*-1 does not belong to $E(n_k)$ and $\frac{x_k^*}{n_k} > \frac{x_{k-1}^*}{n_k} > \frac{x_k^*-1}{n_k+1}$. Hence, for given $(x_1^*,x_2^*,\ldots,x_{k-1}^*)$, when $x_k^* = \bar{y}_k\in E(n_k)$, we have $\bar{y}_k' = x_k^*+1\in E(n_k')$ or $\bar{y}_k' = x_k^*\in E(n_k')\cup D(n_k')$. Similarly, if $x_k^* = \bar{y}_k\in D(n_k)$, according to the definition of $D(n_k)$, we have $\bar{y}_k' = x_k^*+1\in E(n_k')$. In summary, for arbitrary $(x_1^*,x_2^*,\ldots,x_{k-1}^*)$, we have four possible cases:

1) $\bar{y}_k = x_k^*\in E(n_k)$, $\bar{y}_k' = x_k^*+1\in E(n_k')$ and $\bar{y}_k' = \bar{y}_k+1$;
2) $\bar{y}_k = x_k^*\in E(n_k)$, $\bar{y}_k' = x_k^*\in E(n_k')$ and $\bar{y}_k' = \bar{y}_k$;
3) $\bar{y}_k = x_k^*\in E(n_k)$, $\bar{y}_k' = x_k^*\in D(n_k')$ and $\bar{y}_k' = \bar{y}_k$;
4) $\bar{y}_k = x_k^*\in D(n_k)$, $\bar{y}_k' = x_k^*+1\in E(n_k')$ and $\bar{y}_k' = \bar{y}_k+1$.

To proceed, we need the following fact from [6]:

$$\sum_{y_k'=x_k^*+1}^{n_k}\mathbf{P}_{G_{n_k+1}}\left(Y_k'=y_k'\right) = \frac{1+x_k^*(\delta-2)}{2+n_k(\delta-2)}\mathbf{P}_{G_{n_k}}(Y_k=x_k^*)$$
$$+\sum_{y_k=x_k^*+1}^{n_k-1}\mathbf{P}_{G_{n_k}}(Y_k=y_k). \quad (48)$$

We also notice a complementary relationship that, if some $\overrightarrow{x^*} = (x_1^*,\ldots,x_{k-1}^*)$ satisfies Case 1, then $\overleftarrow{x^*} = ((n_1-x_1^*),(n_2-x_2^*),\ldots,(n_{k-1}-x_{k-1}^*))$ has to satisfy Case 2. Hence, in the following we combine Case 1 and Case 2 to evaluate the change of P_c.

In Case 1, for $\overrightarrow{x^*}$, we have $\bar{y}_k = x_k^*\in E(n_k)$, $\bar{y}_k' = x_k^*+1\in E(n_k')$ and $\bar{y}_k' = \bar{y}_k+1$. Let $y_0 = n_k+1-x_k^*$. From (48), by letting $P_d(\overrightarrow{x^*}) = (\sum_{y_k'=y_0}^{n_k}\mathbf{P}_{G_{n_k+1}}(Y_k'=y_k') - \sum_{y_k=y_0}^{n_k-1}\mathbf{P}_{G_{n_k}}(Y_k=y_k))\cdot\prod_{j=1}^{k-1}\mathbf{P}_{G_{n_j}}(X_j=x_j^*)$, we have

$$P_d(\overrightarrow{x^*})=\left(\frac{1+x_k^*(\delta-2)}{2+n_k(\delta-2)}-1\right)\mathbf{P}_{G_{n_k}}(Y_k=x_k^*)\prod_{j=1}^{k-1}\mathbf{P}_{G_{n_j}}(X_j=x_j^*). \quad (49)$$

Correspondingly, in Case 2, by letting $P_d(\overleftarrow{x^*}) = (\sum_{y_k'=y_0}^{n_k}\mathbf{P}_{G_{n_k+1}}(Y_k' = y_k') - \sum_{y_k=y_0}^{n_k-1}\mathbf{P}_{G_{n_k}}(Y_k = y_k))\cdot\prod_{j=1}^{k-1}\mathbf{P}_{G_{n_j}}(X_j=x_j^*)$, we have

$$P_d(\overleftarrow{x^*})=\frac{1+(n_k-x_k^*)(\delta-2)}{2+n_k(\delta-2)}\mathbf{P}_{G_{n_k}}(Y_k=n_k-x_k^*)\prod_{j=1}^{k-1}\mathbf{P}_{G_{n_j}}(X_j=n_j-x_j^*). \quad (50)$$

By adding up (49) and (50), we can show that the total change in Case 1 and Case 2 as n_k increases by one is positive, i.e.,

$$P_d(\overrightarrow{x^*}) + P_d(\overleftarrow{x^*}) > 0,$$

in which we use the fact that $\frac{\mathbf{P}_{G_n}(X=n-x)}{\mathbf{P}_{G_n}(X=x)} = \frac{x}{n-x}$, which can be verified by (40).

Following similar way, for all the four cases we have that for any fixed $n_1, \ldots, n_{k-1}$, the change in the error detection probability, as n_k changes to $n_k + 1$, is lower bounded by zero, i.e., that the correct detection probability is monotonically non-increasing with n_k.

E. Theorem 4

a) Theorem 4(1): For the rumor source s^* with m $(1 \leq m \leq \delta)$ neighbors, i.e., $s_1^*, s_2^*, \ldots, s_m^*$, let random variable X_{i,n_j} be the number of nodes in subtree $Z^{s^*}_{s_i^*, G_{n_j}}$ of observation G_{n_j} $(1 \leq i \leq m, 1 \leq j \leq k)$. Note that as $n_j \to \infty$, the limiting marginal distribution of the ratio $X_{1,n_j}/n_j$ converges to a Beta distribution with a density function given by

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \tag{51}$$

where $\alpha = \frac{1}{\delta-2}$, $\beta = \frac{\delta-1}{\delta-2}$. So for k independent observations, $\mathbf{P}_G(E_1)$ is regarded as the cumulative distribution function of the ratios $\frac{X_{1,n_1}}{n_1}, \ldots, \frac{X_{1,n_k}}{n_k}$, given by

$$\lim_{n_1,\ldots n_k\to\infty} \mathbf{P}_G(E_1) = \int\!\cdot\!\cdot\!\int_{\prod_{j=1}^k \frac{x_j}{1-x_j} \leq 1} \frac{\Gamma(\alpha+\beta)^k}{\Gamma(\alpha)^k \Gamma(\beta)^k} \prod_{j=1}^k \left(x_j^{\alpha-1}(1-x_j)^{\beta-1}\right) \overline{d_k},$$

where $\overline{d_k} = \mathrm{d}x_1 \cdots \mathrm{d}x_k$. From Lemma 5, we obtain (18). From Theorem 3, $\phi_k(3) \leq \phi_k(\delta) < 1$ $(\delta \geq 3)$. Combining with the property that as $k \to \infty$, $\phi_k(3)$ goes to 1, which we will prove next, we have that as k grows without bound, $\phi_k(\delta)$ approaches one for any $\delta \geq 3$.

b) Theorem 4(2): According to (18), when $\delta = 3$, as $n_1, \ldots n_k \to \infty$,

$$\lim_{n_1,\ldots n_k\to\infty} P_{\mathrm{c}} = 1 - 3 \cdot (1 - \varphi_k(1,2)), \tag{52}$$

where

$$\begin{aligned}
&1 - \varphi_k(1,2) \\
&= 1 - \int\!\cdot\!\cdot\!\int_{\prod_{j=1}^k \frac{x_j}{1-x_j} \leq 1} 2^k \cdot \prod_{j=1}^k (1-x_j)\, \overline{d_k} \\
&= 2^k \int_0^1\!\cdot\!\cdot\!\int_0^1 \left(\int_{\frac{\prod_{j=1}^{k-1}(1-x_j)}{\prod_{j=1}^{k-1} x_j + \prod_{j=1}^{k-1}(1-x_j)}}^1 \prod_{j=1}^k (1-x_j) \mathrm{d}x_k \right) \overline{d_{k-1}} \\
&= 2^k \int_0^1\!\cdot\!\cdot\!\int_0^1 \frac{\prod_{j=1}^{k-1} x_j^2 \prod_{j=1}^{k-1}(1-x_j)}{2\left(\prod_{j=1}^{k-1}(1-x_j) + \prod_{j=1}^{k-1} x_j\right)^2} \overline{d_{k-1}} \\
&\overset{(a)}{=} 2^{k-2} \int_0^1\!\cdot\!\cdot\!\int_0^1 \frac{\prod_{j=1}^{k-1} x_j^2(1-x_j) + \prod_{j=1}^{k-1}(1-x_j)^2 x_j}{\left(\prod_{j=1}^{k-1}(1-x_j) + \prod_{j=1}^{k-1} x_j\right)^2} \overline{d_{k-1}} \\
&= 2^{k-2} \int_0^1\!\cdot\!\cdot\!\int_0^1 \frac{\prod_{j=1}^{k-1} x_j(1-x_j)}{\prod_{j=1}^{k-1}(1-x_j) + \prod_{j=1}^{k-1} x_j} \overline{d_{k-1}},
\end{aligned} \tag{53}$$

where (a) follows from substituting $x_j = 1 - x_j^{'}$, $1 \leq j \leq k - 1$. By applying (53) into (52), we get (19). By rearranging (19),

$$\begin{aligned}
&\lim_{n_1,\ldots n_k\to\infty} P_{\mathrm{c}} \\
&= 1 - 3 \cdot 2^{k-2} \int_0^1\!\cdot\!\cdot\!\int_0^1 \frac{\prod_{j=1}^{k-1}(x_j \cdot (1-x_j))}{\prod_{j=1}^{k-1} x_j + \prod_{j=1}^{k-1}(1-x_j)} \overline{d_{k-1}} \\
&\overset{(b)}{\geq} 1 - 3 \cdot 2^{k-2} \int_0^1\!\cdot\!\cdot\!\int_0^1 \frac{\prod_{j=1}^{k-1}(x_j \cdot (1-x_j))}{2\sqrt{\prod_{j=1}^{k-1}(x_j \cdot (1-x_j))}} \overline{d_{k-1}} \\
&\overset{(c)}{=} 1 - 3 \cdot 2^{k-3} \cdot \left(\frac{\pi}{8}\right)^{k-1} = 1 - \frac{3}{4} \cdot \left(\frac{\pi}{4}\right)^{k-1},
\end{aligned} \tag{54}$$

where (b) follows from the inequality of arithmetic and geometric means, and (c) follows from $\int_0^1 \sqrt{x(1-x)}\mathrm{d}x = \pi/8$. This leads to the desired result.

F. Theorem 5

Treat the connected graph formed by connected multiple sources as a super node s^T, with the tree G_n rooted at this super node. Hence the degree of the super node s^T is $(\delta - 2)m + 2$. According to Proposition 1 and Lemma 2,

$$\begin{aligned}
P_{\mathrm{c}}(S) &\geq \mathbf{P}_G\left(\bigcap_{j=1}^{\delta'} E_j\right) = 1 - \mathbf{P}_G\left(\bigcup_{j=1}^{\delta'} E_j^c\right) \\
&\overset{(a)}{\geq} 1 - \sum_{j=1}^{\delta'} \mathbf{P}_G\left(E_j^c\right) \overset{(b)}{=} 1 - \delta' \cdot \mathbf{P}_G\left(E_1^c\right),
\end{aligned} \tag{55}$$

where (a) is the joint bound, and (b) follows from symmetry, and $\delta' = (\delta-2)m+2$. Again using Proposition 1 and Lemma 2, we have $P_{\mathrm{c}}(S) \leq \mathbf{P}_G(\bigcap_{j=1}^{\delta'} F_j) = 1 - \delta' \cdot \mathbf{P}_G(F_1^c)$. Therefore, we have

$$\lim_{n_1,\ldots,n_k\to\infty} P_{\mathrm{c}}(S) = 1 - \delta' \cdot \lim_{n_1,\ldots,n_k\to\infty} \mathbf{P}_G(E_1^c) = 1 - \delta' \cdot \lim_{n_1,\ldots,n_k\to\infty} \mathbf{P}_G(F_1^c).$$

To evaluate E_1^c and F_1^c, we use a Polyas urn model in which we start with 1 ball of type 1 and $((\delta-2)m+1)$ balls of type 2. With these initial conditions, the limit law of fraction of balls of type 1 turns out to be (see [1] for details) a Beta distribution with parameters $\alpha = \frac{1}{\delta-2}$, $\beta = \frac{(\delta-2)m+1}{\delta-2}$. As $n_1, \ldots, n_k \to \infty$, we have (23).

Hence, Theorem 5 is proved.

G. Theorem 6

Similarly to Theorem 5, from Proposition 1 and Lemma 2, we can derive

$$\lim_{n_1,\ldots,n_k\to\infty} P_{\mathrm{c}}(s^*) = 1 - \delta \cdot \lim_{n_1,\ldots,n_k\to\infty} \mathbf{P}_G(E_1^c).$$

$\lim_{n_1,\ldots,n_k\to\infty} \mathbf{P}_G(E_1^c)$ can be evaluated by summing up all the products, i.e., the marginal distribution $\mathbf{P}_{G_m}(X=i)$ in one case of spreading graphs by m sources multiplies the limiting laws of the ratios of the sizes of rumor boundaries under this initial case. From the above discussion, we can prove Theorem 6.

VIII. Conclusion

We studied the rumor source detection problem for the SI model with multiple observations. By providing characterization through the interdependency of observations and network topology, we established a number of explicit analytical results for regular tree-type networks. For the case of the single rumor source, we showed that having multiple independent observations dramatically enhances detectability: even two observations can more than double the detection probability of a single observation. We may thus find the right number of observations needed to provide detectability guarantees. We also showed that the detection probability increases with the degree as well as the number of observations, i.e., richer connectivity and diversity both enhance detection. For the case of multiple sources, diversity is also a powerful performance booster. We provided a unified inference framework based on message passing for tree networks and leveraged it as an effective heuristic for general graphs. In the next step of work, it would be interesting to explore exact detectability results on general graphs. Other future topics also include developing efficient network forensics protocols, and addressing practical issues, e.g., non-unique initial sources, complicated spreading models, and outliers.

Appendix A

Proof of Lemma 1: We illustrate the proof for ruling out the hypothetic case where node v has two neighbors satisfying the condition in Lemma 1. Denote the two neighbors by v_1 and v_2, which satisfy $x_{1,n_1}x_{1,n_2}\cdots x_{1,n_k} \geq (n_1 - x_{1,n_1})(n_2 - x_{1,n_2})\cdots(n_k - x_{1,n_k})$ and $x_{2,n_1}x_{2,n_2}\cdots x_{2,n_k} \geq (n_1 - x_{2,n_1})(n_2 - x_{2,n_2})\cdots(n_k - x_{2,n_k})$, respectively. Summing up the two inequalities, we have

$$\prod_{j=1}^{k} x_{1,n_j} + \prod_{j=1}^{k} x_{2,n_j} \geq \prod_{j=1}^{k}(n_j - x_{1,n_j}) + \prod_{j=1}^{k}(n_j - x_{2,n_j}). \quad (56)$$

Noting that v_1 and v_2 are neighbors of node v, we have $x_{1,n_j} + x_{2,n_j} \leq n_j - 1, \forall j \in [1,k]$. It follows that $x_{1,n_j} < n_j - x_{2,n_j}$ and $x_{2,n_j} < n_j - x_{1,n_j}, \forall j \in [1,k]$. Adopting into the left hand side of (56), we obtain

$$\prod_{j=1}^{k} x_{1,n_j} + \prod_{j=1}^{k} x_{2,n_j} < \prod_{j=1}^{k}(n_j - x_{1,n_j}) + \prod_{j=1}^{k}(n_j - x_{2,n_j}),$$

which contradicts (56). Hence it is impossible to have two neighbors of node v satisfying the condition in Lemma 1. Similarly, the possibility of having more than two neighbors can be ruled out, and hence there can be at most a neighbor of node v that satisfies the condition in Lemma 1.

Proof of Lemma 2: If for all $i \in [1,m]$, $\prod_{j=1}^{k} x_{i,n_j} < \prod_{j=1}^{k}(n_j - x_{i,n_j})$, according to Proposition 1, s^* is the unique joint rumor center. Therefore, we can make sure to correctly detect s^* as the rumor source. If for some $1 \leq i \leq m, \prod_{j=1}^{k} x_{i,n_j} = \prod_{j=1}^{k}(n_j - x_{i,n_j})$, according to Proposition 1 and Lemma 1, there are two joint rumor centers. Therefore, the probability to correctly detect s^* as the rumor source is 1/2. If for some $1 \leq i \leq m, \prod_{j=1}^{k} x_{i,n_j} > \prod_{j=1}^{k}(n_j - x_{i,n_j})$, then s^* does not have the maximum joint rumor centrality. Therefore s^* is not a joint rumor center, and we cannot detect it as the rumor source. Then Lemma 2 follows from summarizing all the correct detection events.

Proof of Lemma 3: Denote $P_{\rm e} = 1 - P_{\rm c}$ as the error detection probability. Define

$$D_j = \left\{\prod_{i=1}^{k} x_{j,n_i} = \prod_{i=1}^{k}(n_i - x_{j,n_i})\right\}, \text{ and}$$
$$F_j = \left\{\prod_{i=1}^{k} x_{j,n_i} > \prod_{i=1}^{k}(n_i - x_{j,n_i})\right\}, (1 \leq j \leq \delta).$$

According to Lemma 1, $D_j, F_j, 1 \leq j \leq \delta$ are all disjoint. Hence, based on Proposition 1, $P_{\rm e}$ can be evaluated by

$$P_{\rm e} = \frac{1}{2}\mathbf{P}_G\left(\bigcap_{j=1}^{\delta} D_j\right) + \mathbf{P}_G\left(\bigcap_{j=1}^{\delta} F_j\right) = \delta\left(\frac{1}{2}\mathbf{P}_G(D_1) + \mathbf{P}_G(F_1)\right),$$

where the second equality follows from symmetry.

Proof of Lemma 4: Note that the marginal distribution of X_1 is given by

$$\begin{aligned}&\mathbf{P}_G(X_1 = x_1)\\ &= \frac{(n-1)!}{x_1! x_2!} \cdot \frac{\prod_{i=1}^{2} b_i(b_i + \delta - 2)\cdots(b_i + (x_i - 1)(\delta - 2))}{\delta(\delta + \delta - 2)\cdots(\delta + (n-2)(\delta - 2))},\end{aligned} \quad (57)$$

where $x_2 = n - 1 - x_1$, $b_1 = 1$, $b_2 = \delta - 1$. As $n \to \infty$, we have

$$\begin{aligned}&\mathbf{P}_G(X_1 = 0)\\ &= \frac{(n-1)!}{(n-1)!} \frac{(\delta-1)\cdot(\delta-1+(\delta-2))\cdots(\delta-1+(n-2)(\delta-2))}{\delta\cdot(\delta+(\delta-2))\cdot(\delta+2(\delta-2))\cdots(\delta+(n-2)(\delta-2))}\\ &= \frac{\Gamma(\frac{\delta}{\delta-2})}{\Gamma(\frac{\delta-1}{\delta-2})} \cdot \frac{\Gamma(n + \frac{1}{\delta-2})}{\Gamma(n + \frac{2}{\delta-2})} \overset{(a)}{\sim} \frac{\Gamma(\frac{\delta}{\delta-2})}{\Gamma(\frac{\delta-1}{\delta-2})} \cdot \frac{\Gamma(n) n^{\frac{1}{\delta-2}}}{\Gamma(n) n^{\frac{2}{\delta-2}}} \to 0,\end{aligned}$$

where (a) follows from $\lim_{n\to\infty} \frac{\Gamma(n+\alpha)}{\Gamma(n)n^{\alpha}} = 1, \alpha \in \mathbb{R}$.

Proof of Lemma 5: First, according to Lemma 4, as $n_1, \ldots, n_k \to \infty$, the probability for the rumor source to have less than δ infected neighbors (i.e., empty subtrees) asymptotically vanishes. So we only need to consider the case where the rumor source has $m = \delta$ infected neighbors, and denote by $P_{\rm c,m=\delta}$ the correct detection probability in that case.

According to Proposition 1 and Lemma 2,

$$\begin{aligned}P_{\rm c,m=\delta} &\geq \mathbf{P}_G\left(\bigcap_{j=1}^{\delta} E_j\right) = 1 - \mathbf{P}_G\left(\bigcup_{j=1}^{\delta} E_j^c\right)\\ &\overset{(a)}{\geq} 1 - \sum_{j=1}^{\delta}\mathbf{P}_G\left(E_j^c\right) \overset{(b)}{=} 1 - \delta\cdot\mathbf{P}_G\left(E_1^c\right),\end{aligned} \quad (58)$$

where (a) is the joint bound, and (b) follows from symmetry.

Again using Proposition 1 and Lemma 2, we have

$$\begin{aligned}P_{\rm c,m=\delta} &\leq \mathbf{P}_G\left(\bigcap_{j=1}^{\delta} F_j\right) = 1 - \mathbf{P}_G\left(\bigcup_{j=1}^{\delta} F_j^c\right)\\ &\overset{(c)}{=} 1 - \sum_{j=1}^{\delta}\mathbf{P}_G\left(F_j^c\right) \overset{(d)}{=} 1 - \delta\cdot\mathbf{P}_G\left(F_1^c\right),\end{aligned} \quad (59)$$

where (c) follows from that $F_1^c, \ldots, F_\delta^c$ are disjoint (Lemma 1), and (d) follows from symmetry.

Therefore, we have

$$\begin{aligned}\lim_{n_1,\ldots,n_k\to\infty} P_{\mathrm{c}} &= \lim_{n_1,\ldots,n_k\to\infty} P_{\mathrm{c,m}=\delta}\\ &= 1-\delta\cdot \lim_{n_1,\ldots,n_k\to\infty} \mathbf{P}_G(E_1^c) = 1-\delta\cdot \lim_{n_1,\ldots,n_k\to\infty} \mathbf{P}_G(F_1^c).\end{aligned}$$

References

[1] A. Ganesh, L. Massoulie, and D. Towsley, “The effect of network topology on the spread of epidemics,” *Proc. IEEE INFOCOM*, pp. 1455–1466, 2005.

[2] D. Easley and J. Kleinberg*, Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[3] D. Shah and T. Zaman, “Detecting sources of computer viruses in networks: Theory and experiment,” in *Proc. ACM SIGMETRICS*, 2010, vol. 38, no. 1, pp. 203–214.

[4] 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.

[5] D. Shah and T. Zaman, “Rumor centrality: A universal source detector,” in *Proc. ACM SIGMETRICS*, 2012, vol. 40, no. 1, pp. 199–210.

[6] W. Dong, W. Zhang, and C. W. Tan, “Rooting out the rumor culprit from suspects,” in *Proc. IEEE ISIT*, 2013, pp. 2671–2675.

[7] W. Luo, W. P. Tay, and M. Leng, “Identifying infection sources and regions in large networks,” *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2850–2865, Nov. 2013.

[8] N. Karamchandani and M. Franceschetti, “Rumor source detection under probabilistic sample,” in *Proc. IEEE ISIT*, 2013, pp. 2184–2188.

[9] K. Zhu and L. Ying, “Information source detection in the SIR model: A sample path based approach,” *Proc. IEEE ITA*, pp. 1–9, 2013.

[10] A. Y. Lokhov, M. Mezard, H. Ohta, and L. Zdeborova, “Inferring the origin of an epidemy with dynamics message-passing algorithm,” *arXivv1303.5315*, 2013.

[11] W. Luo and W. P. Tay, “Finding an infection source under SIS model,” in *Proc. IEEE ICASSP*, 2013, pp. 2930–2934.

[12] D. Wang *et al.*, “Humans as sensors: An estimation theoretic perspective,” in *Proc. ACM/IEEE Inf. Process. Sens. Netw. (IPSN)*, 2014, pp. 35–46.

[13] Prakash, B. Aditya, Vreeken, Jilles, Faloutsos, and Christos, “Spotting culprits in epidemics: How many and which ones?,” in *Proc. IEEE ICDM*, 2012, pp. 11–20.

[14] A. Agaskar and Y. M. Lu, “A fast monte carlo algorithm for source localization on graphs,” in *Proc. SPIE Opt. Eng. Applicat.*, 2013.

[15] E. Seo, P. Mohapatra, and T. Abdelzaher, “Identifying rumors and their sources in social networks,” in *Proc. SPIE Defense, Security, Sensing*, 2012.

[16] D. Ariu, G. Giacinto, and R. Perdisci, “Sensing attacks in computers networks with hidden Markov models,” in *Proc. IEEE Mach. Learn. Data Mining Pattern Recogn. (MLDM)*, 2007, p. 449C463.

[17] H. X. Nguyen and M. Roughan, “Multi-observer privacy-preserving hidden markov models,” in *Proc. IEEE Netw. Operat. Manage. Symp. (NOMS)*, 2012, pp. 514–517.

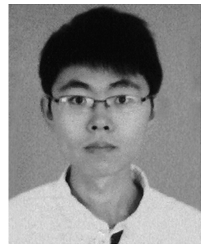
[18] H. X. Nguyen and M. Roughan, “Improving hidden Markov model inferences with private data from multiple observers,” *IEEE Signal Process. Lett.*, vol. 19, no. 10, pp. 696–699, Oct. 2012.

[19] J. C. S. J. Ramłrez *et al.*, “Statistical voice activity detection using a multiple observation likelihood ratio test,” *IEEE Signal Process. Lett.*, vol. 12, no. 10, pp. 689–692, Oct. 2005.

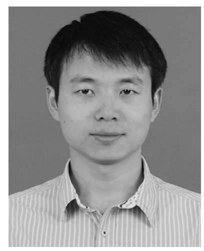
[20] Z. Wang, W. Dong, W. Zhang, and C. W. Tan, “Rumor source detection with multiple observations: Fundamental limits and algorithms,” in *Proc. ACM SIGMETRICS*, 2014.

[21] D. J. Watts and S. H. Strogatz, “Collective dynamics of smallworldnetworks,” *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[22] Barabási, Albert-László, and A. Réka, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

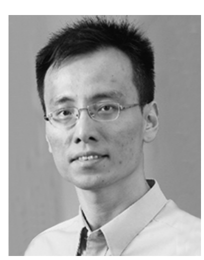


Zhaoxu Wang received his B.S. degree in electronic engineering from University of Science and Technology of China (USTC) in 2011. He is now a Ph.D. student in electronic engineering at USTC, Hefei, China. From July to November 2014, he worked as an intern in Qualcomm Incorporated, Beijing. His research interests include green (energy- efficient) communications, wireless heterogeneous networks, social network analysis, rumor source detection.



Wenxiang Dong received the B.S. degree in electronic and information engineering from University of Science and Technology of China (USTC), Hefei, China, in 2009. He is currently a Ph.D. candidate at USTC. From August to November 2012, he was a visiting student in Prof. Tan Chee Weis group at City University of Hong Kong. From August to November 2013, he was an intern in Microsoft Research Asia, Beijing. His research interests include social network analysis, mobile computing, and wireless network optimization.

Wenyi Zhang (S’00–M’07–SM’11) attended Tsinghua University and obtained his Bachelor’s degree in automation in 2001. He studied in the University of Notre Dame, Indiana, USA, and obtained his Master’s and Ph.D. degrees, both in electrical engineering, in 2003 and 2006, respectively. Prior to joining the faculty of Department of Electronic Engineering and Information Science, University of Science and Technology of China, he was affiliated with the Communication Science Institute, University of Southern California, as a Postdoctoral Research Associate, and with Qualcomm Incorporated, Corporate Research and Development. His research interests include wireless communications and networking, information theory, and statistical signal processing.



Chee Wei Tan (M’08–SM’12) received the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 2006 and 2008, respectively. He is an Assistant Professor at the City University of Hong Kong. Previously, he was a Postdoctoral Scholar at the California Institute of Technology (Caltech), Pasadena, CA. He was a Visiting Faculty at Qualcomm R&D, San Diego, CA, in 2011. His research interests are in networks, inference in online large data analytics, and optimization theory and its applications. Dr. Tan was the recipient of the 2008 Princeton University Wu Prize for Excellence and was awarded the 2011 IEEE Communications Society AP Outstanding Young Researcher Award. He was a selected participant at the U.S. National Academy of Engineering China-America Frontiers of Engineering Symposium in 2013. Dr. Tan currently serves as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and the Chair of the IEEE Information Theory Society Hong Kong Chapter.