Rooting out the Rumor Culprit from Suspects

Wenxiang Dong*, Wenyi Zhang* and Chee Wei Tan†
University of Science and Technology of China* City University of Hong Kong†

***Abstract*—In this work, we study the problem of rumor source locationing: supposing that a rumor, which is originated from a set of suspect nodes in a network, is spread across the network following the susceptible-infectious model, how to locate that rumor source? For regular tree type of networks, we establish exact and asymptotic results on the detection probability of the maximum-likelihood source locator under several scenarios. The development utilizes ideas from the Pólya's urn model [1], and sheds insight into the behavior of the source locationing algorithm.**

I. Rumor Spreading Model and ML Estimator

A undirected network graph $G = (V, E)$ consists of a set of nodes V and a set of edges E, where V is possibly countably infinite, and any pair of nodes may infect each other if and only if they are connected by an edge in E. The rumor spreading process obeys the susceptible-infectious (SI) model, where once a node gets infected it keeps the rumor forever. We consider the case where only one of the nodes in an *a priori* specified suspects set, $S \subseteq V$, can be the rumor source. The *a priori* distribution over S is assumed to be uniform; namely, $\mathbf{P_s}(s)$, the probability that s is the rumor source, is equal to $1/k$ for all $s \in S$.

The rumor spreading process unfolds as follows. Initially only a single node $s^* \in S$ possesses a rumor and thus is termed infected. An infected node may infect its neighbors. Let $\tau_{ij}, (i, j) \in E$ be the time for node j to be infected by its neighbor i, after i gets infected, and assume that τ_{ij}'s are mutually independent and all exponentially distributed with rate unity. At some point, we get to observe the network G and find n infected nodes, which must form a connected subgraph of G and thus are denoted as G_n. Apparently G_n must contain at least a node in S. The goal is to estimate the rumor source s^*, given G_n.

We focus on the maximum likelihood (ML) estimator,

$$\hat{s} = \arg\max_{s\in\{S \bigcap G_n\}} \mathbf{P_G}(G_n|s), \tag{1}$$

where $\mathbf{P_G}(G_n|s)$ is the probability of observing G_n, assuming s to be the rumor source. Note that unlike [2], here we restrict the rumor source to be within the suspect set $S \subseteq V$. Similar to that in [2], the ML estimator when G is a regular tree is

$$\hat{s} = \arg\max_{s\in\{S \bigcap G_n\}} \mathbf{P_G}(G_n|s) = \arg\max_{s\in\{S \bigcap G_n\}} R(s, G_n), \tag{2}$$

where $R(s, G_n)$ is called the "rumor centrality" and is defined in [2].

II. Detection Probability on Regular Trees

(1) Suspecting all nodes, $S = V$

In this extreme case, every node in the resulting observation G_n has the potential to be the rumor source, and the estimator in (1) reduces into that in [2]. Following a different approach from those in [2][3], we obtain both exact and asymptotic results for $\mathbf{P_c}(n)$, the correct detection probability with observing n infected nodes. For a regular tree with degree $\delta = 2$, $\mathbf{P_c}(n) = \frac{1}{2^{n-1}}\binom{n-1}{\lfloor (n-1)/2 \rfloor} \approx O(\frac{1}{\sqrt{n}})$; for a regular tree with degree $\delta = 3$, $\mathbf{P_c}(n) = \frac{1}{4} + \frac{3}{4}\frac{1}{2\lfloor n/2 \rfloor + 1}$; for a regular tree with degree $\delta > 3$, $\mathbf{P_c}(n) = 1 - \delta\left(1 - I_{1/2}\left(\frac{1}{\delta-2}, \frac{\delta-1}{\delta-2}\right)\right)$ when $n \to \infty$, where $I_x(\alpha, \beta)$ is the regularized incomplete Beta function with parameters α and β. The above asymptotic expression for $\lim_{n\to\infty} \mathbf{P_c}(n)$ is positive and approaches $1 - \ln 2 \approx 0.307$ as δ grows large.

(2) Connected suspects

Suppose that the suspect set S of cardinality k forms a connected subgraph of G. For a regular tree with degree $\delta = 2$, $\mathbf{P_c}(n) = \frac{1}{k}\left[1 + \frac{k-1}{2^{n-1}}\binom{n-1}{\lfloor (n-1)/2 \rfloor}\right] \approx \frac{1}{k} + \frac{k-1}{k}O(\frac{1}{\sqrt{n}})$; for a regular tree with degree $\delta = 3$, $\mathbf{P_c}(n) = \frac{k+1}{2k} + kO(\frac{1}{n})$ when $k \ll n$; for a regular tree with degree $\delta > 3$, $\mathbf{P_c}(n) = \frac{1}{k}\left[k - 2(k-1)\left(1 - I_{1/2}\left(\frac{1}{\delta-2}, \frac{\delta-1}{\delta-2}\right)\right)\right]$ as $n \to \infty$, which further approaches one as δ grows large. We hence see that $\mathbf{P_c}(n)$ significantly exceeds the *a priori* distribution $1/k$ if the graph is not linear, and indeed achieves reliable detection (i.e., $\lim_{n\to\infty} \mathbf{P_c}(n) \to 1$) as the node degree δ becomes sufficiently large.

(3) Two suspects

Suppose that the suspect set contains only two nodes, $S = \{s_1, s_2\}$, and denote by d the shortest path length between s_1 and s_2. Without loss of generality, consider $d < n$. For a regular tree with degree $\delta = 2$, $\mathbf{P_c}(n) = \frac{1}{2} + \frac{1}{2^n}\sum_{z_1=(n-d-1)/2}^{(n+d+1)/2}\binom{n-1}{z_1}$ if $n-d$ is odd, or $\mathbf{P_c}(n) = \frac{1}{2} + \frac{1}{2^n}\sum_{z_1=(n-d)/2}^{(n+d-2)/2}\binom{n-1}{z_1}$ if $n-d$ is even. For a regular tree with degree $\delta = 3$, $\mathbf{P_c}(n) = 0.75$ when $d = 1$ and $\mathbf{P_c}(n) = 0.886$ when $d = 2$, as $n \to \infty$. For general regular trees, we find that $\mathbf{P_c}(n)$ increases as d increases. In summary, we obtain an interesting result that $\mathbf{P_c}(n) > 1/2$ always holds for regular trees with degree $\delta = 3$, and furthermore it is increasing with respect to d for all regular trees.

Acknowledgement

The research of W. Dong and W. Zhang has been supported by National Basic Research Program of China (973 Program) through grant 2012CB316004, and by the 100 Talents Program of Chinese Academy of Sciences. The research of C. W. Tan has been supported by the Research Grants Council of Hong Kong under Project No. RGC CityU 125212.

References

[1] N. L. Johnson and S. Kotz, *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory*, John Wiley & Sons, 1977.
[2] D. Shah and T. Zaman, "Rumors in a network: Who's the culprit?" *IEEE Trans. Inform. Theory*, 57(8), 5163–5181, Aug. 2011.
[3] ——, "Rumor centrality: A universal source detector," in *Proc. ACM SIGMETRICS 2012.*