Finding Rumor Sources on Random Graphs

Devavrat Shah*[1] and Tauhid Zaman†[2]

[1]Massachusetts Institute of Technology
[2]The Wharton School, University of Pennsylvania

We consider the problem of detecting the source of a rumor (information diffusion) in a network based on observations about which set of nodes possess the rumor. Finding rumor sources is a very general type of problem which arises in many different contexts. For example, the rumor could be a computer virus on the Internet, a contagious disease in a human population, or a trend in a social network. In each of these scenarios, detection of the source is of interest as this source may be a malicious agent, patient zero, or an influential person.

In a recent work [1], the authors proposed *rumor centrality* as an estimator for detecting the source. They established it to be the maximum likelihood (ML) estimator with respect to the popular Susceptible Infected (SI) model with exponential spreading times for d-regular trees. The key limitations of this prior work are: (i) the results do not quantify the exact detection probability, say α_d, for d-regular graphs, under the proposed maximum likelihood estimator other than $\alpha_2 = 0$, $\alpha_3 = 0.25$ and $0 < \alpha_d < 0.5$ for $d \geq 4$ for the SI model with exponential spreading times; (ii) the results do not quantify the magnitude of the error in the event of not being able to identify the source; and more generally, (iii) the results do not provide any insights into how the estimator behaves for generic heterogeneous tree (or tree-like) graphs under the SI model with a generic spreading time distribution.

The primary reason behind the limitations of the results in [1] is the fact that the analytic method employed there is quite specific to regular trees with homogeneous exponential spreading times. To overcome these limitations, as the main contribution of this work we introduce a novel analysis method that utilizes connections to the classical multi-class Markov branching process (equivalently, generalized Polyas urn). As a consequence of this, we are able to quantify the probability of the error event precisely and thus eliminate the shortcomings of the prior work. The following is a summary of the key results:

- **Regular tree, SI model with exponential spreading times**: We find the exact value for α_d, the detection probability for d-regular trees, for all d. We are able to show that $\lim_{d\to\infty} \alpha_d = 1 - \ln(2)$. Further, we show that the probability of rumor centrality estimating the k^{th} infected node as the source decays as $\exp(-\Theta(k))$.
- **Generic random tree, SI model with generic spreading times**: For generic expanding random trees with generic spreading times, we establish that the probability of correct source detection remains bounded away from 0.

The above results collectively establish that, even though, rumor centrality is an ML estimator only for regular trees and the SI model with exponential spreading times, it is universally effective with respect to heterogeneity in the tree structure and spreading time distribution. Its effectiveness for generic random trees immediately implies its utility for finding sources in sparse random graphs that are locally tree-like, such as Erdos-Renyi and random regular graphs.

References

[1] D. Shah and T. Zaman. Rumors in a network: Who's the culprit? *IEEE Transactions on Information Theory*, 57:5163–5181, 2011.

*devavrat@mit.edu
†tzaman@wharton.upenn.edu