Information Source Detection with Limited Time Knowledge

Xuecheng Liu [1] Luoyi Fu [1] Bo Jiang [1] Xiaojun Lin [2] Xinbing Wang [1]

[1]Shanghai Jiao Tong University [2]Purdue University

Problem Setup

We are interested in the source inference problem in a network: given the diffusion network G, can we identify the source node using partial observations at time t efficiently and precisely?

This problem has been addressed under various diffusion models [1, 2], we adopt the discrete-time version system model:

- Let $G(V, E)$ be a possibly infinite connected graph. Let $v^* \in V$ be a rumor source from which a rumor starts spreading.
- During each time-slot, the infected node attempts to infect its neighbors with probability p. A node, once becoming infected, remains infected.
- A set of observers $\mathcal{S} \subset V$ are chosen in advance to monitor the network. Each observer s records its infection time t_s.

In the Bayesian setting, the optimal estimator (in the sense of maximizing the correct detection probability $\mathbb{P}(\hat{v} = v^*)$)

$$\text{MLE:} \quad \hat{v} = \arg\max_{v \in V} \mathbb{P}(\{(s, t_s) : s \in \mathcal{S}\} | v^* = v)$$

is computationally intractable due to the #P hardness of the combinatorial nature and the unknown diffusion starting time.

Approximation Algorithm Design

Our approximation idea is similar to the sample-path-based approach proposed in [3]. Observing that the diffusion process is essentially Markovian, it is straightforward to compute the probability of a single trajectory. The difficulty of the MLE comes from the (possibly) infinitely many trajectories which are consistent with the partial observations $\{(s, t_s) : s \in \mathcal{S}\}$. Summing over these trajectories is intractable in the absence of structures. Alternatively, we seek for the most likely trajectory and then view the source node of that trajectory as $\hat{v}$.

Source Detection in a Tree

Due to the acyclicity of diffusion network G, the trajectory can be compactly represented by the infection time of each node.

1. Compute the probability of the trajectory rooted at $v \in V$ via Integer Linear Programming (**ILP**)

$$\begin{aligned} \text{minimize (over } \mathbf{t}) \quad & \textstyle\sum_{(i,j) \in E(T_{\mathcal{S}}(v))} \mathbf{t}(j) - \mathbf{t}(i) \\ \text{subject to} \quad & \mathbf{t}(u) = t_u \quad \forall u \in \mathcal{S} \\ & \mathbf{t}(j) - \mathbf{t}(i) \geq 1 \quad \forall (i,j) \in E(T_{\mathcal{S}}(v)) \\ & \mathbf{t}(u) \in \mathbb{Z} \quad \forall u \in V(T_{\mathcal{S}}(v)) \end{aligned}$$

   where $T_{\mathcal{S}}(v)$ is a directed tree rooted at v explaining the "parent-child" relationship in the diffusion process. The **ILP** can be solved exactly in $O(|V|)$ time using message passing.
2. Find the most likely trajectory rooted at $\hat{v}$ from a reduced node space $\mathcal{V}$ via breadth-first search. The procedure is illustrated in Figure 1, where $\mathcal{R}_1 \subset \mathcal{V} \subset \mathcal{R}_1 \cup \mathcal{R}_2$.

Source Detection in a Cyclic Graph

For each node v, we find a BFS tree rooted at v spanning as many observers as possible. Then compute the probability of the most likely trajectory on the BFS tree rooted at v. Finally view the node with largest trajectory probability as $\hat{v}$.

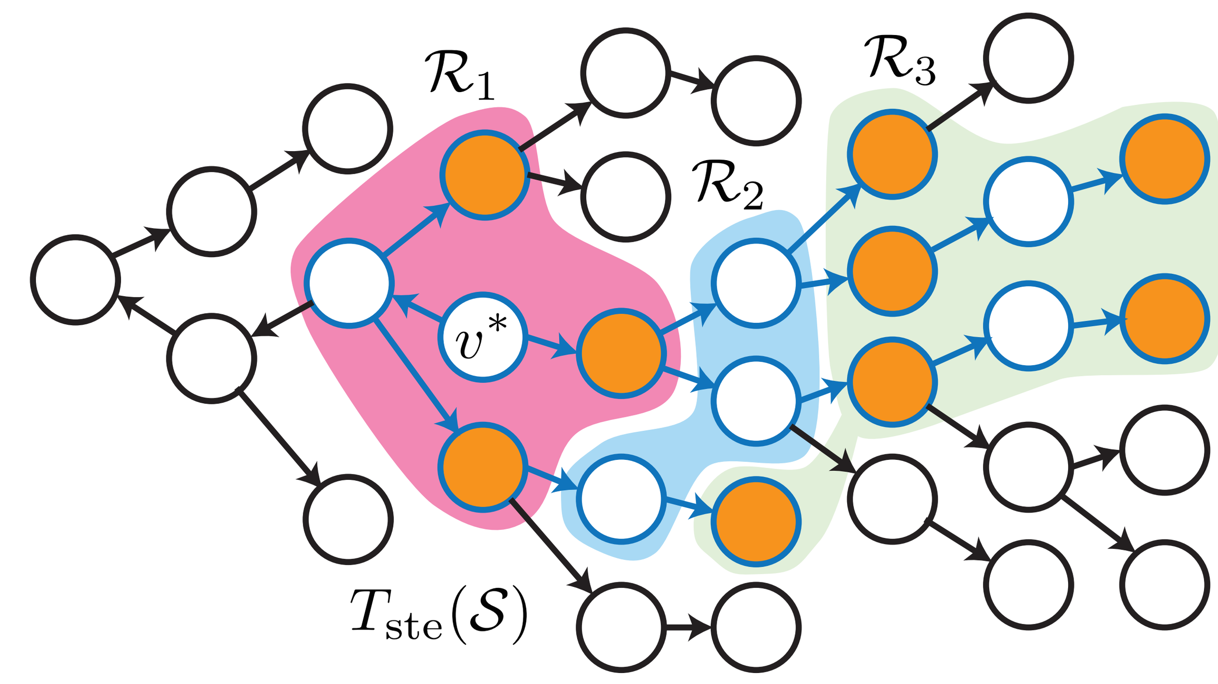


Figure 1. Illustration of the Most Likely Trajectory (MLT) estimator in a tree.

Analytic Results on Error Distance

Assume that the observers $\mathcal{S}$ are chosen uniformly at random w.p. q. By means of the *candidate path* [4], we obtain

- If G is an infinite regular tree with degree $g = 2$,

$$\mathbb{P}(\hat{v} = v^*) = q + \frac{(1-q)pq(pq + 3 - 3p)}{(pq + 2 - 2p)(pq + 1 - p)}$$

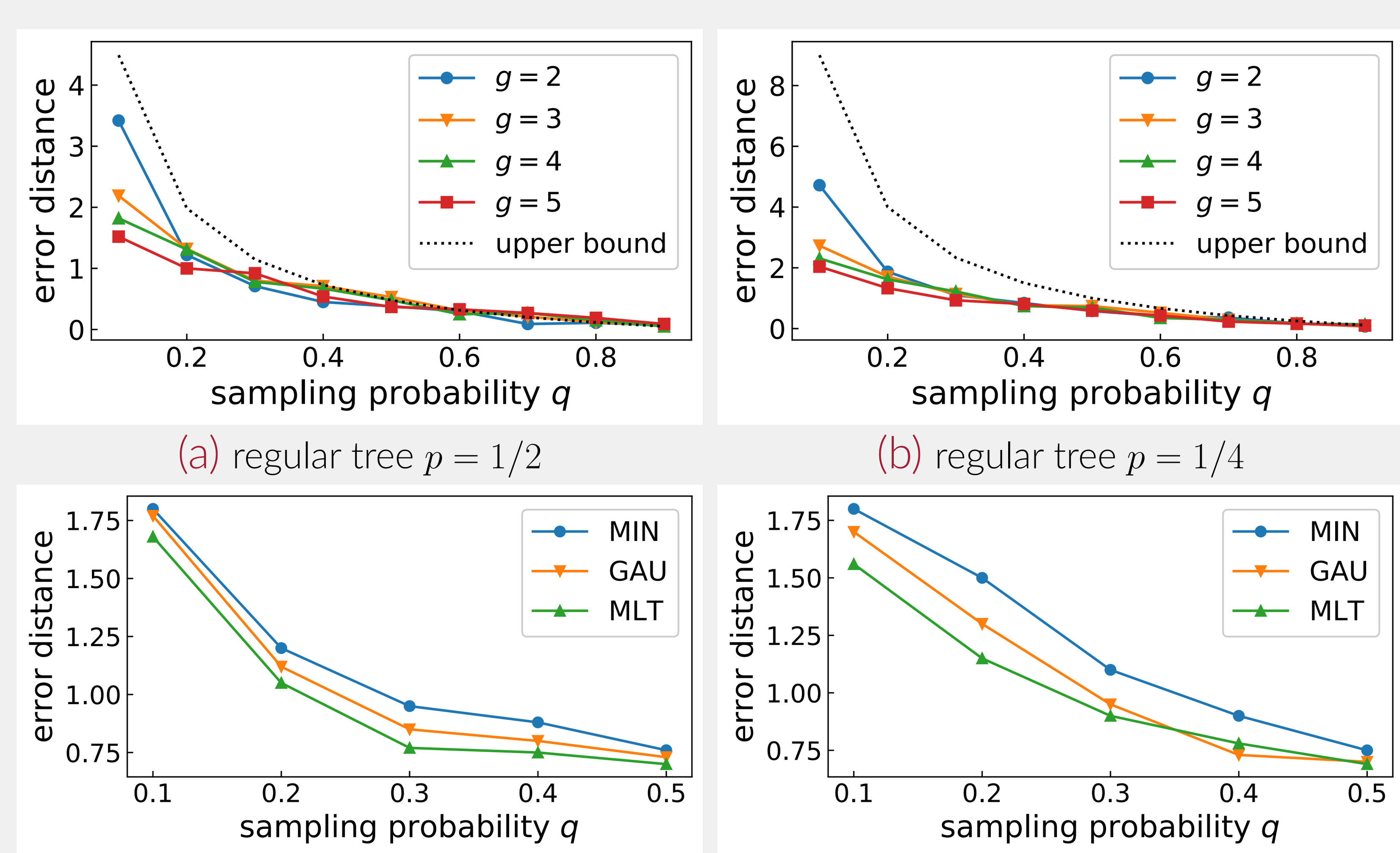
$$\mathbb{E}[d(\hat{v}, v^*)] \leq (1-q) \min \left\{ \frac{1}{q}, \frac{2(1 - p + pq)(1-p)^2}{pq(2 - 2p + pq)^2} \right\}$$

- If G is an infinite regular tree with degree $g \geq 3$,

$$\mathbb{P}(d(v^*, \hat{v}) \leq D) \geq 1 - (1-q)(1 - p + p(1-q)x_1)^g$$

  where x_1 is given by function iteration $x_D = 1$ and $x_i = h(x_{i+1}) = (1 - p + p(1-q)x_{i+1})^{g-1}$ for $i \in [D-1]$.
- The MLT estimator stochastically dominates the MIN estimator in terms of error distance.

Numerical Evaluations



(a) regular tree $p = 1/2$ (b) regular tree $p = 1/4$

(c) Erdös-Rényi (density 10.24) (d) Scale-free (density 3.98)

Figure 2. Simulation results on diffusion network G where $|V| = 1024$.

References

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

[2] P. C. Pinto, P. Thiran, and M. Vetterli, "Locating the source of diffusion in large-scale networks," *Phys. Rev. Lett.*, vol. 109, p. 068702, Aug 2012.

[3] K. Zhu and L. Ying, "Information source detection in the sir model: A sample-path-based approach," *IEEE/ACM Transactions on Networking*, vol. 24, pp. 408–421, Feb 2016.

[4] X. Liu, L. Fu, B. Jiang, X. Lin, and X. Wang, "Information source detection with limited time knowledge," *CoRR*, vol. abs/1905.12913, 2019.