RUMORS IN A NETWORK: WHO'S THE CULPRIT? D.SHAH ,T.ZAMAN

Yujia Gu (yg2384)

Motivation

- Ubiquity of networks has made us vulnerable to new types of network risks. An isolated risk is amplified because it is spread by the network.

- Examples: worms of viruses in computer networks, cascading failure in large systems such as financial markets

- A Internet service provider, policy maker, would like to identify the source of the risk as quickly as possible.
- Thus, we study the question of finding the source of a rumor in a network.

Problem to Solve

- The goal is to find the source of the rumor in order to control and prevent network risks.
- However, information about network structure and the “rumor infected” nodes is limited.

- Need to cast a rumor spreading model.
- Need to find estimators to solve rumor source problem for networks of different structures
- Need to evaluate the performance of different estimators constructed.

Rumor Spreading Model

- Model a network of nodes as an undirected graph G(V,E), where V is a countably infinite set of nodes, E is the set of edges of the form (i,j) for some i and j in V.

- Use a variant of common SIR model, i.e. SI (susceptible-infected) model. This model does not allow for any nodes to recover. Once a node i has the rumor, it is able to spread it to another node j if and only if there is an edge between them.

- Let t_{ij} be the time it takes for node j to receive the rumor from node i once i has the rumor. In this model, tij are independent and have exponential distribution with parameter (rate) λ. Assume λ=1 .

Rumor Source Maximum Likelihood Estimator

- Suppose the rumor starting at a node v^* at time 0 has spread in the network G. We observe the network at some time and find N infected nodes. These N nodes form a connected subgraph of G, denote it by G_N.
- Assume a uniform prior probability of the source node among all nodes of G_N.
- Maximum likelihood estimator of v^* maximizes the correct detection probability.

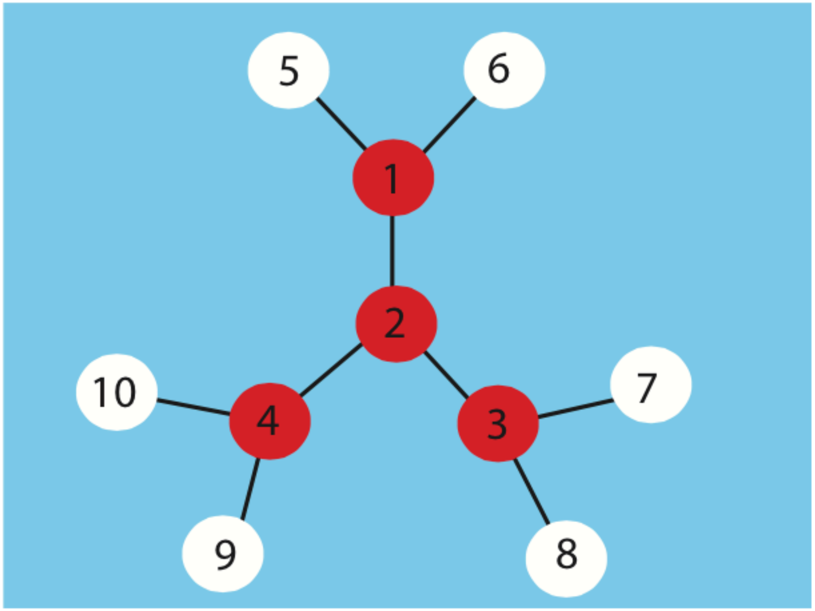
$$\hat{v} \in \arg\max_{v \in G_N} \mathbf{P}(G_N | v)$$

- $P(G_N|v)$ is the probability of observing G_N assuming v is the source.

Rumor Source Estimator for Regular Trees

- Definition: Given a connected tree G(V,E) and a source node $v \in V$
- Consider any permutation $\sigma: V \rightarrow \{1,2,\ldots, |V|\}$ of its nodes where $\sigma(u)$ denotes the position of node in the permutation.
- We call a permitted permutation for tree with source node if
- 1) $\sigma(v) = 1$
- 2) For any $(u, u') \in E$, if $d(v,u) < d(v,u')$, then $\sigma(u) < \sigma(u')$.

.



Rumor Source Estimator for Regular Trees (cont.)

- Let $\Omega(v, G_N)$ be the set of all permitted permutations starting with node v and resulting in rumor graph G_N.
- Let $\sigma = \{v_1 = v, v_2, \ldots, v_N\}$, $G_k(\sigma)$ be the subgraph of G_N containing nodes $\{v_1 = v, v_2, \ldots, v_k\}$ for $1 \le k \le N$.
- Then, the probability $P(\sigma|v)$ for each $\sigma \in \Omega(v, G_N)$ is

$$\mathbf{P}(\sigma \mid v) = \prod_{k=2}^{N} \mathbf{P}\left(k^{\text{th}} \text{ infected node} = v_k \mid G_{k-1}(\sigma), v\right)$$

- Given $G_{k-1}(\sigma)$ (and source v), the next infected node could be any of the neighbors of nodes in $G_{k-1}(\sigma)$ which are not yet infected.
- Each of these nodes is equally likely to be the next infected node.
- If $G_{k-1}(\sigma)$ has $n_{k-1}(\sigma)$ uninfected neighboring nodes, then $\mathbf{P}(\sigma \mid v) = \prod_{k=2}^{N} \frac{1}{n_{k-1}(\sigma)}$.
- Suppose k-th added node to $G_{k-1}(\sigma)$ is $v_k(\sigma)$ with degree $d_k(\sigma)$, then

$$n_k(\sigma) = n_{k-1}(\sigma) + d_k(\sigma) - 2$$

- By induction, $$n_k(\sigma) = d_1(\sigma) + \sum_{i=2}^{k} (d_i(\sigma) - 2)$$

Rumor Source Estimator for Regular Trees (cont.)

- For a d regular tree, $\mathbf{P}(\sigma \mid v) = \prod_{k=1}^{N-1} \frac{1}{dk - 2(k-1)} \equiv p(d, N).$
- Denote number of distinct permitted permutations $|\Omega(v, G_N)|$ by $R(v, G_N)$.
- Definition: Given a graph G and vertex v of G, we define R(v,G) as the total number of distinct permitted permutations of nodes of G that begin with node $v \in G$ and respect the graph structure of G.
- Thus, the ML estimator for a regular tree becomes

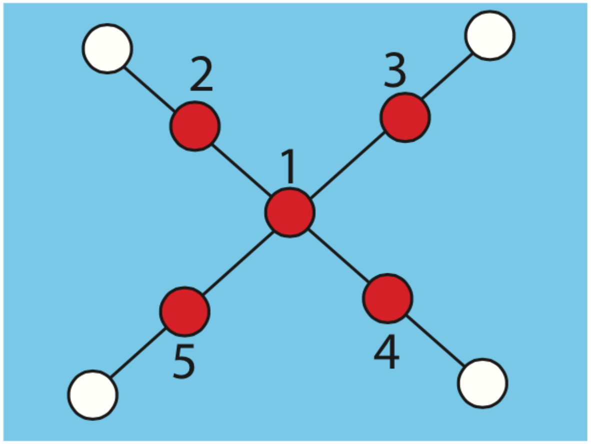
$$\begin{aligned}
\widehat{v} &\in \arg\max_{v \in G_N} \mathbf{P}\left(G_N \mid v\right) \\
&= \arg\max_{v \in G_N} \sum_{\sigma \in \Omega(v, G_N)} \mathbf{P}\left(\sigma \mid v\right) \\
&= \arg\max_{v \in G_N} R(v, G_N) p(d, N) \\
&= \arg\max_{v \in G_N} R(v, G_N)
\end{aligned}$$

with ties broken uniformly at random

Rumor Source Estimator for General Trees

- The likelihood of a node is a sum of the probability of every permitted permutation for which it is the source. In general, these will have different values, but it may be that a majority of them have a common value.
- We need to determine this value of the probability of the common permitted permutations.
- Assume the nodes receive the rumor in a breadth-first search (BFS) fashion.
- Define the BFS permitted permutation σ_v* with node v as the source, then the rumor source estimator becomes (ties broken uniformly at random)

$$\widehat{v} \in \arg\max_{v \in G_N} \mathbf{P}\left(\sigma_v^* \mid v\right) R(v, G_N).$$



$$\mathbf{P}(\sigma_1^*) R(1, G_N) = \left(\frac{1}{4}\right)^4 4! = 6\left(\frac{1}{4}\right)^3$$

$$\mathbf{P}(\sigma_2^*) R(2, G_N) = \frac{1}{2}\left(\frac{1}{4}\right)^3 3! = 3\left(\frac{1}{4}\right)^3.$$

Rumor Centrality

- Call R(v, G_N) the rumor centrality of the node v with respect to G_N .
- The node with maximum rumor centrality will be called the rumor center of the network.
- Define T_u^v as the number of nodes in the subtree rooted at node u, with node v as the source.
- Need to calculate permitted permutations of N nodes of G_N.

- The first of N slots in a given permitted permutation must be the source node v.
- Then, from the remaining N-1 slots, we must choose $T_{v_1}^v$ slots for the nodes in the subtree rooted at v_1. These nodes can be ordered in R(v_1, $T_{v_1}^v$) different ways.
- With the remaining N-1- $T_{v_1}^v$ nodes, we must choose $T_{v_2}^v$ nodes for the tree rooted at node v_2 and these can be ordered R(v_2, $T_{v_2}^v$) ways.
- Continue this way recursively to obtain

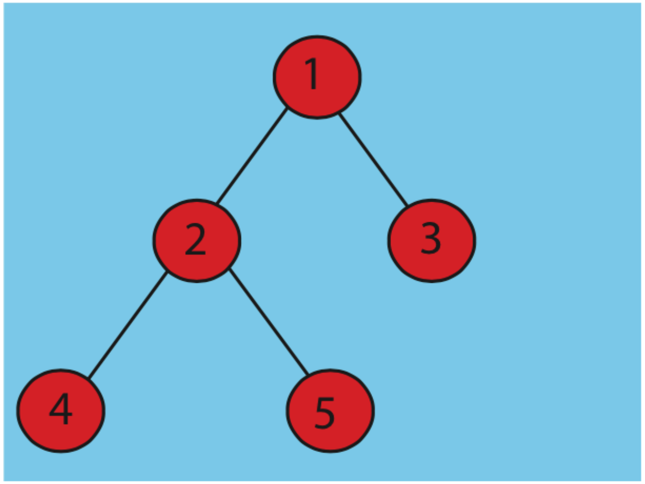
$$R(v, G_N) = \binom{N-1}{T_{v_1}^v}\binom{N-1-T_{v_1}^v}{T_{v_2}^v}\ldots\binom{N-1-\sum_{i=1}^{k-1}T_{v_i}^v}{T_{v_k}^v}\prod_{i=1}^{k}R(v_i, T_{v_i}^v)$$

$$= (N-1)!\prod_{i=1}^{k}\frac{R(v_i, T_{v_i}^v)}{T_{v_i}^v!}$$

Rumor Centrality (cont.)

- We continue this recursion until we reach the leaves of the tree.
- The leaf subtrees have 1 node and 1 permitted permutation.
- Therefore, the number of permitted permutations for a given tree G_N rooted at v is
-

$$\begin{aligned} R(v, G_N) &= (N-1)! \prod_{i=1}^{k} \frac{R(v_i, T_{v_i}^v)}{T_{v_i}^v!} \\ &= (N-1)! \prod_{i=1}^{k} \frac{(T_{v_i}^v - 1)!}{T_{v_i}^v!} \prod_{v_{ij} \in T_{v_i}^v} \frac{R(v_{ij}, T_{v_{ij}}^v)}{T_{v_{ij}}^v!} \\ &= (N-1)! \prod_{i=1}^{k} \frac{1}{T_{v_i}^v} \prod_{v_{ij} \in T_{v_i}^v} \frac{R(v_{ij}, T_{v_{ij}}^v)}{T_{v_{ij}}^v!} \\ &= N! \prod_{u \in G_N} \frac{1}{T_u^v} \end{aligned}$$



$$R(1, G) = \frac{5!}{5 * 3} = 8.$$

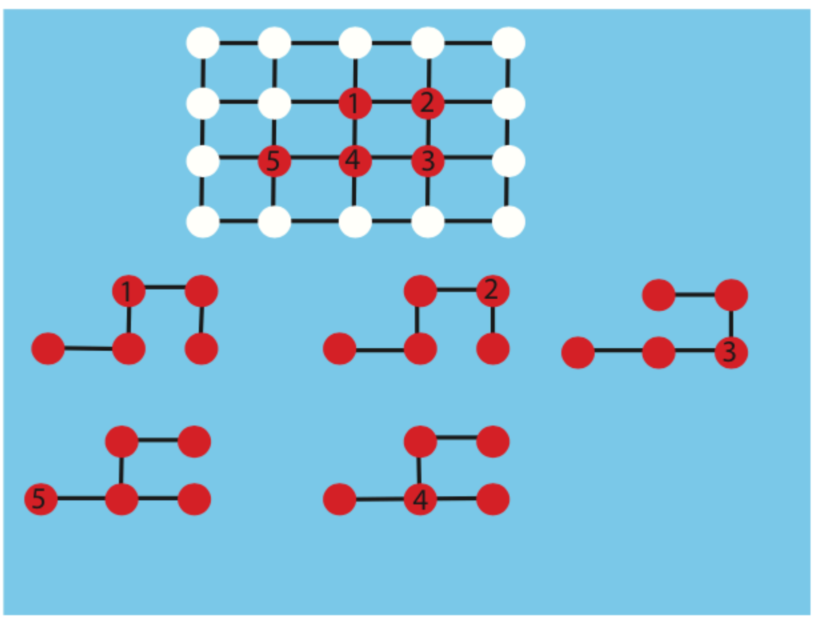
$$\{1,3,2,4,5\}, \{1,2,3,4,5\}, \{1,2,4,3,5\}, \{1,2,4,5,3\}$$
$$\{1,3,2,5,4\}, \{1,2,3,5,4\}, \{1,2,5,3,4\}, \{1,2,5,4,3\}$$

Rumor Source Estimator for General Graphs

- Even in a general graph the rumor spreads along a spanning tree of the observed graph corresponding to the *first time* each node receives the rumor.
- Assume if node v ∈ G_N was the source, then the rumor spreads along a breadth first search (BFS) tree $T_{bfs}(v)$ rooted at v.
- Then, the rumor source estimator for a general rumor graph G_N is

$$\widehat{v} \in \arg\max_{v \in G_N} \mathbf{P}(\sigma_v^* \mid v) R(v, T_{\mathrm{bfs}}(v)).$$

σ_v* represents the BFS ordering of nodes in the tree $T_{bfs}(v)$



$$\because P(\sigma|v) = \prod_{k=1}^{N-1} \frac{1}{dk-2(2k-1)}, d = 4$$

$$\therefore P(\sigma_1^*|1) = \prod_{k=1}^{5-1} \frac{1}{2k+2} = \frac{1}{4\times6\times8\times10}$$

$$\mathbf{P}(\sigma_1^* \mid 1) R(1, T_{\mathrm{bfs}}(1)) = \frac{1}{4*6*8*10} \frac{5!}{20}$$

$$\mathbf{P}(\sigma_2^* \mid 2) R(2, T_{\mathrm{bfs}}(2)) = \frac{1}{4*6*8*10} \frac{5!}{30}$$

$$\mathbf{P}(\sigma_3^* \mid 3) R(3, T_{\mathrm{bfs}}(3)) = \frac{1}{4*6*8*10} \frac{5!}{20}$$

$$\mathbf{P}(\sigma_4^* \mid 4) R(4, T_{\mathrm{bfs}}(4)) = \frac{1}{4*6*8*10} \frac{5!}{10}$$

$$\mathbf{P}(\sigma_5^* \mid 5) R(5, T_{\mathrm{bfs}}(5)) = \frac{1}{4*6*8*10} \frac{5!}{40}.$$

Calculating Rumor Centrality: A Message Passing Algorithm

- In order to find the rumor center of an N node tree G_N , we need to first find the rumor centrality of every node in G_N. A naive algorithm can lead to $\Omega(N^2)$ operations.

- A special relation between subtrees rooted at two neighboring nodes u and v in G_N :

$$T_u^v = N - T_v^u.$$

- Rumor centralities of any two neighboring nodes (key to the algorithm):

$$R(u, G_N) = R(v, G_N)\frac{T_u^v}{N - T_u^v}$$

Calculating Rumor Centrality: A Message Passing Algorithm (cont.)

- Algorithm:
- Select any node v as the source node and calculate the size of all of its subtrees T_u^v and its rumor centrality R(v, G_N)
- Each node u pass two messages up to its parent
  1. The number of nodes in u's subtree, call it t_{up}
  2. The cumulative product of the size of the subtrees of all nodes in u's subtree, call it p_{up}
- The parent node adds the t_{up} messages together to obtain the size of its own subtree, and multiplies the p_{up} messages together to obtain its cumulative subtree product.
- These messages are then passed upward until the source node receives the messages. The source node will obtain its rumor centrality, R(v, G_N).
- Each node passes its rumor centrality to its children down in a message we define as r_{down}. Each node u can calculate its rumor centrality using its parent's rumor centrality and its own subtree size T_u^v

Results: Theory

- <u>Definition</u>*:*

Define C_t to be the event of correct rumor source detection using the rumor centrality based estimator after the rumor has spread for a time t on a graph G(V,E) .

Asymptotic detection probability:

1. For linear graphs it is 0
2. For trees which grow faster than a line it is strictly greater than 0

Results: Theory (Linear Graphs)

- (I). Linear Graphs: No Detection
- Consider a linear graph which is a regular tree of degree 2.
- Suppose the rumor starts spreading on a linear graph at time 0 as per the SI model.
- Then, $\mathbf{P}(\mathcal{C}_t) = O\left(\frac{1}{\sqrt{t}}\right)$
- Result: the linear graph detection probability goes to 0 as t goes to infinity.
- i.e. the estimator provides very little information because of the linear graph's trivial structure

Results: Theory (Regular Expander Trees)

- (II). Regular Expander Trees: Nontrivial Detection
- Suppose the rumor starts spreading on a regular expander tree with degree d>2 at time 0 as per the SI model. Then there exists a constant $\alpha_d > 0$ for all d>2 so that

$$0 < \alpha_d \leq \liminf_t \mathbf{P}(\mathcal{C}_t) \leq \limsup_t \mathbf{P}(\mathcal{C}_t) \leq \frac{1}{2}$$

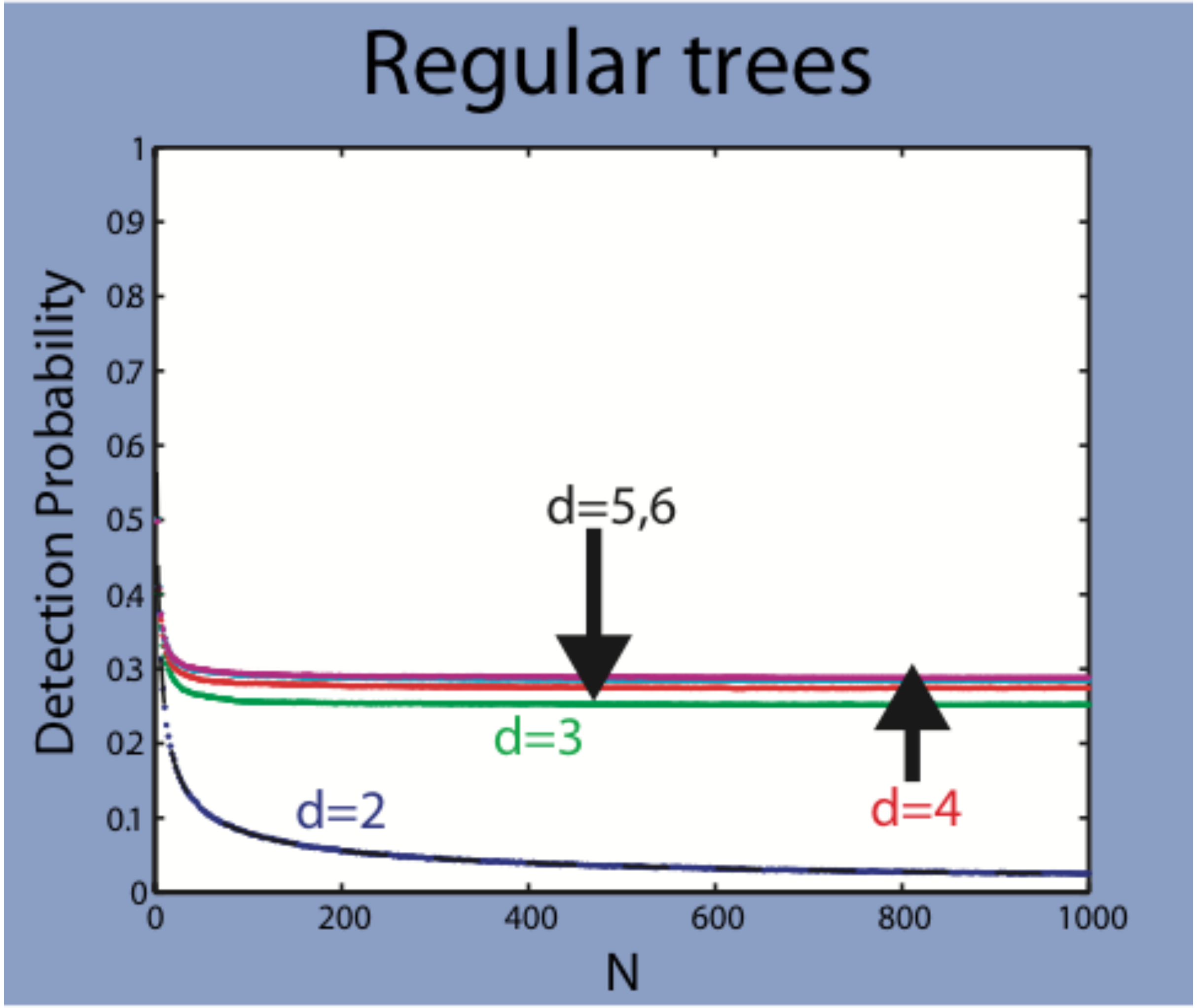
- Results:
- 1.The estimator performs detection of the rumor source with strictly positive probability irrespective of the size of the rumor network.
- 2.The detection probability is always upper bounded by 1/2 for any d > 2.

Results: Theory(Degree 3 Regular Expander Trees)

- (III). Degree 3 Regular Expander Trees: Exact Detection Probability
- For d=3, we are able to obtain the exact value asymptotic detection probability as t goes to infinity.

$$\lim_{t} \mathbf{P}(\mathcal{C}_t) = \frac{1}{4}$$

Simulation Results (Regular Trees)



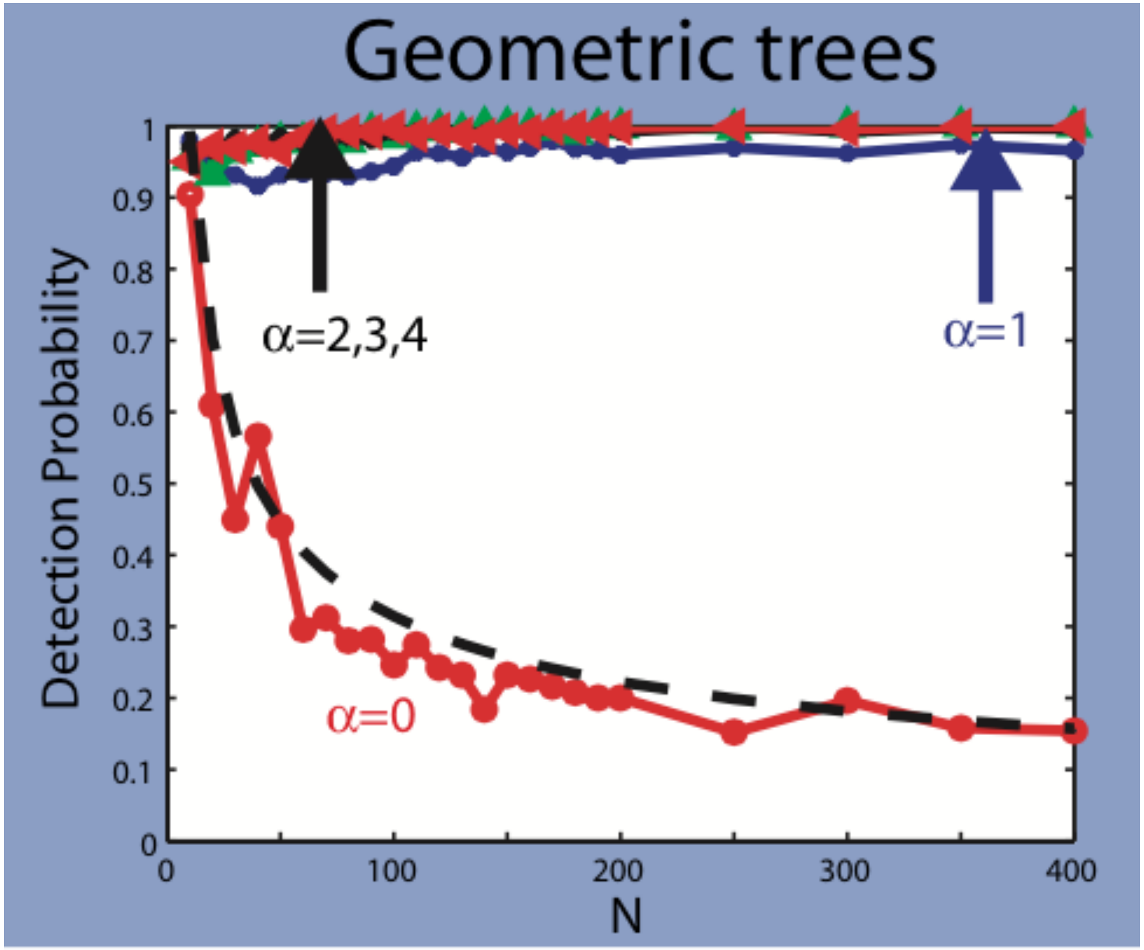
Results: Theory (*Geometric Trees*)

- IV. *Geometric Trees: Correct Detection*
- For cases of detection probability of the estimator in nonregular trees. Consider trees that grow polynomially.
- Suppose α>0, 0<b≤ c, fix a source node v*, d* be the degree of v*.
- Let v be any node in any subtree($T_1, T_2, \ldots T_{d^*}$) of v*,
- Let $n^i(v,r)$ be the number of nodes in Ti at distance exactly r from the node v.
- Require: for all 1≤ i ≤ d* and v∈Ti,

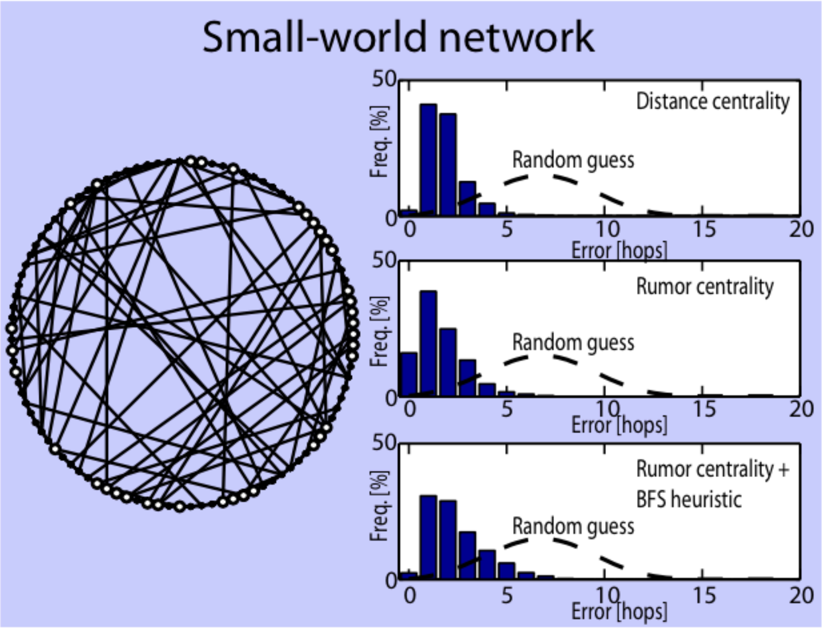
$$br^\alpha \leq n^i(v,r) \leq cr^\alpha$$

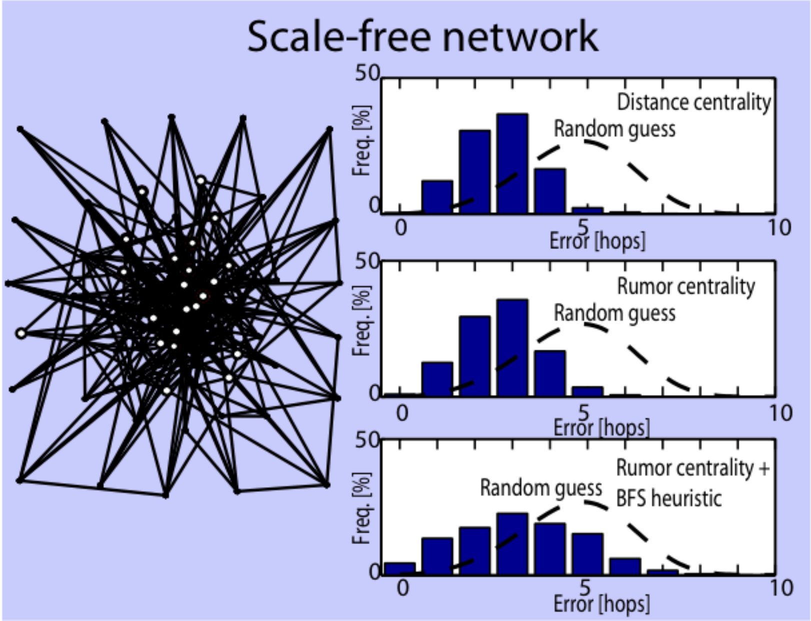
- Typically, with degree $d_{v^*} \geq 3$, let $d_{v^*} > \frac{c}{b} + 1$.
- Then, $\liminf_t \mathbf{P}(C_t) = 1$.
- Results: it is as good as the best possible estimator.

Simulation Results (*Geometric Trees*)



Simulation Results (*General Network*)





- The graphs show the rumor infected nodes in white.
- The histogram shows the error of the rumor source estimator for 400 node rumor graphs in each network.

Thank you!

Q&A