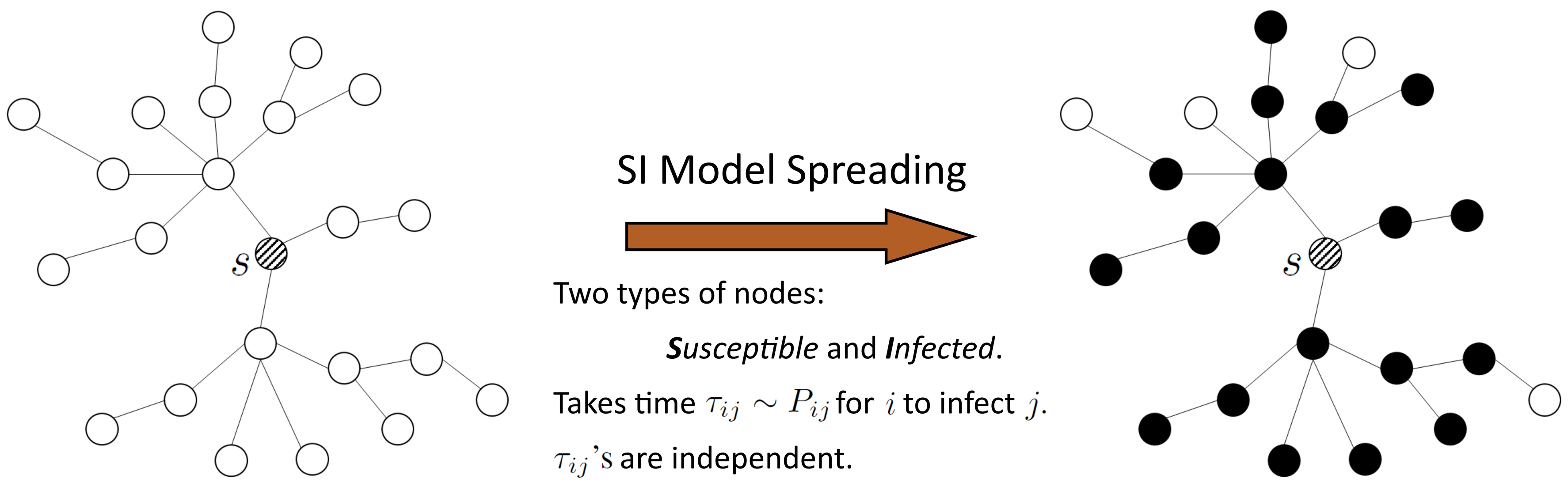
Rumor Source Detection: A Probabilistic Perspective

Ting-Han Fan and I-Hsiang Wang
National Taiwan University
{b02901062,ihwang}@ntu.edu.tw

Rumor Source Detection Problem

At time 0, start spreading. At unknown time t, observe G_I.



- Goal: Find s from G_I.
- It is natural to estimate s through the MLE:

$$\hat{s}_{\text{MLE}} = \arg\max_{v\in\mathcal{V}} L(v|G_I), \quad L(v|G_I) \triangleq \Pr(\exists\, t,\ \mathcal{G}_I^s(t) = G_I | s = v).$$

Challenge and Previous Works

✓ **Challenge:**

- MLE's computation time is **exponential** in # of infected nodes [1] .
- Previous works resort to approximations.

✓ **Previous Approximation Schemes:**

- [1] Rumor Center $\hat{s}_R$:

$R(v, G_I)$: # of ways infection starting from v leading to G_I.

$$\hat{s}_R = \arg\max_{v\in\mathcal{V}_I} R(v, G_I).$$

- [2] Jordan Center $\hat{s}_J$:

$d(u,v)$: minimum # of hops from node u to node v.

$$\hat{s}_J = \arg\min_{u\in\mathcal{V}_I} \max_{v\in\mathcal{V}_I} d(u,v).$$

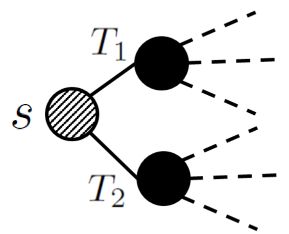
- [3] Dynamic Message Passing $\hat{s}_D$:

$\tilde{\mathbb{P}}_v(i,t)$: the estimate of the probability that when node v is the source, node i is infected at time t.

$$\hat{s}_D = \arg\max_{v\in\mathcal{V}_I} \max_{t\in\mathbb{R}_+} \prod_{i\in\mathcal{V}_I} \tilde{\mathbb{P}}_v(i,t) \prod_{j\notin\mathcal{V}_I} (1 - \tilde{\mathbb{P}}_v(i,t)).$$

Key Ideas

- If G_I is composed of two trees, $\mathscr{T}_1$ and $\mathscr{T}_2$, then the joint likelihood is factorized:



$$\mathbb{P}(G|s,t) = \mathbb{P}(\mathscr{T}_1|s,t)\mathbb{P}(\mathscr{T}_2|s,t).$$

- Recursively factorizing leads to the Belief Propagation Algorithm.
- In Contrast, the likelihood of the node is not factorizable because

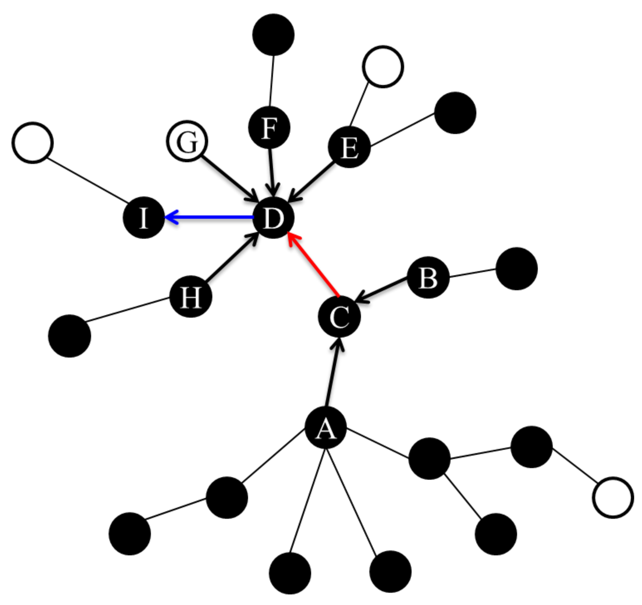
$$\mathbb{P}(G|s) = \int_t \mathbb{P}(\mathscr{T}_1|s,t)\mathbb{P}(\mathscr{T}_2|s,t)p(t)dt \neq \mathbb{P}(\mathscr{T}_1|s)\mathbb{P}(\mathscr{T}_2|s).$$

- Thus, we aim to find the *Joint Maximum Likelihood Estimator*:

$$(\hat{s}_{\text{JMLE}}, \hat{t}_{\text{JMLE}}) = \arg\max_{(v,t)\in\mathcal{V}\times\mathbb{R}_+} L(v,t|G_I),\ L(v,t|G_I) \triangleq \Pr(\mathcal{G}_I^s(t) = G_I | s = v, t).$$

Belief Propagation Algorithm

- $m_{i\to j}(t)$: probability that if j is the source, j causes the infection in the subtree rooted at i.
- $f_{ij}(t)$: PDF of the time it takes for node i to infect node j.



$$m_{C\to D} = (m_{A\to C}\, m_{B\to C}) * f_{DC}$$

$$m_{D\to I} = \Big(\prod_{k\in\{C,E,F,G,H\}} m_{k\to D}\Big) * f_{ID}$$

$*$ denotes convolution.

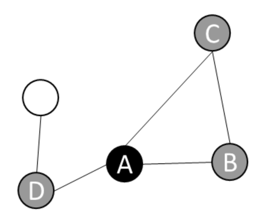
$$L(j,t|G_I) = \prod_{i\in\text{Nei}(j)} m_{i\to j}$$

- By Fast Fourier Transform, a convolution takes time $O(L\log L)$.
- Each edge needs to propagate two times (forward and backward).
- Since a tree has $n-1$ edges, the time complexity is $O(nL\log L)$.

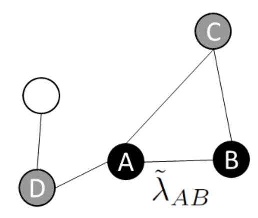
Extension to General Graph (GGT Heuristic)

Generate a weighted spanning tree. Then apply the BP algo.

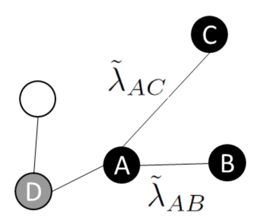
1. Given G_I and the rates on edges, Assume A is the source. $\mathbf{I} = \{A\}$



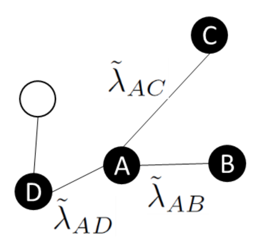
2. Choose the next infected node as
$$\arg\max_{v\in\mathcal{V}\setminus\mathbf{I}} \sum_{j\in\mathbf{I}\cap\text{Nei}(v)} \lambda_{jv}$$
Assume B is chosen. $\mathbf{I} = \{A, B\}$
3. Update B's parent and infection rate.
$$\text{Parent}_B = \arg\min_{i\in\mathbf{I}\setminus\{B\}} \mathbb{E}[\tilde{T}_i] = A$$
$$\tilde{\lambda}_{AB} = 1/\mathbb{E}[\tilde{T}_B - \tilde{T}_A | \tilde{T}_B > \tilde{T}_A]$$



4. Choose the next infected node again. Assume C is chosen. $\mathbf{I} = \{A, B, C\}$ Update C's parent and infection rate.



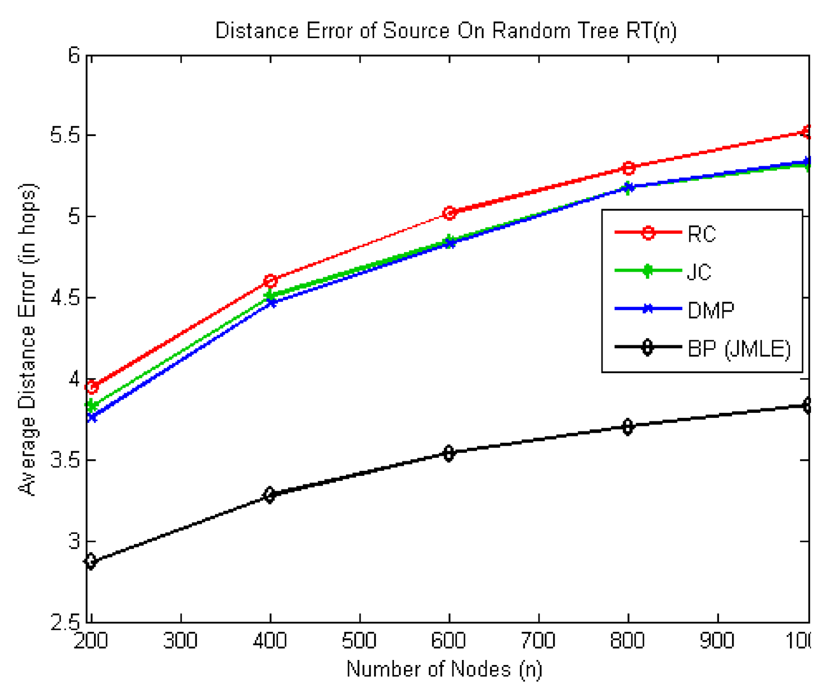
5. Do again for D.

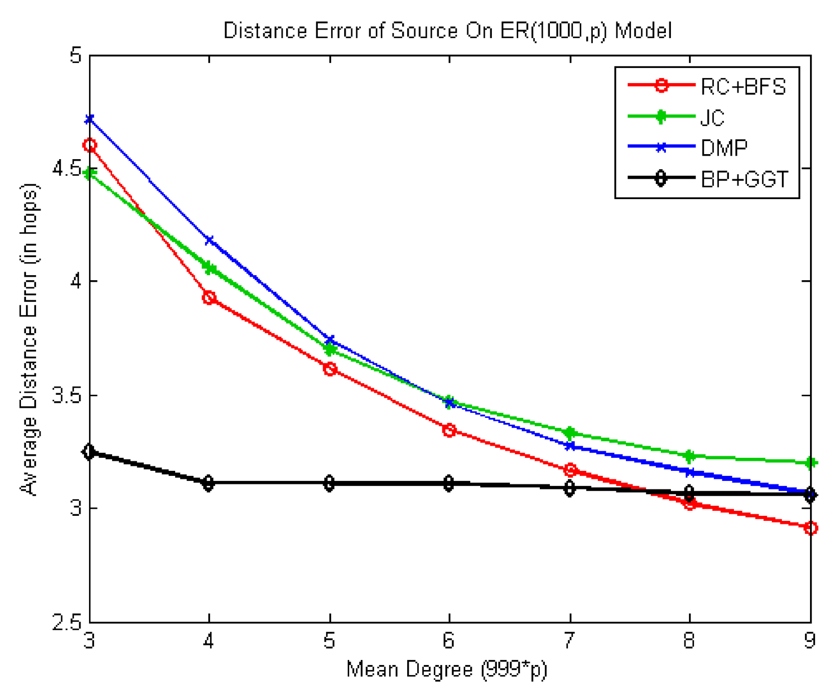


6. Compute $L(A, t|\text{GGT}_A)$ by BP algo. GGT_A is the weighted tree above.

Note:
$\tilde{T}_i$'s are computed and modeled by Gamma distribution.

Experimental Results





- All simulation is run with $\tau_{ij} \overset{\text{i.i.d.}}{\sim} \text{Exponential}(1)$.
- The more similar a graph is to a tree, the more powerful the BP algorithm will be.

Reference:
[1] D. Shah and T. Zaman, "Rumors in a network: Who's the culprit?" *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5163–5181, Aug 2011.
[2] W. Luo, W. P. Tay, M. Leng, and M. K. Guevara, "On the universality of the jordan center for estimating the rumor source in a social network," in *2015 IEEE International Conference on Digital Signal Processing (DSP)*, July 2015, pp. 760–764.
[3] A. Y. Lokhov, M. M´ezard, H. Ohta, and L. Zdeborov´a, "Inferring the origin of an epidemic with a dynamic message-passing algorithm," *Phys. Rev. E*, vol. 90, p.012801, Jul 2014.