Inference of a Rumor's Source in the Independent Cascade Model

Petra Berenbrink, Max Hahn-Klimroth, Dominik Kaaser, Lena Krieg, Malin Rau

Introduction

We consider the so-called *Independent Cascade Model* for rumor spreading or epidemic processes popularized by Kempe et al. [2003]. In this model, a small subset of nodes from a network are the source of a rumor. In discrete time steps, each informed node "infects" each of its uninformed neighbors with probability p. While many facets of this process are studied in the literature, less is known about the inference problem: given a number of infected nodes in a network, can we learn the source of the rumor?

Independent Cascade Model

We observe the following model: Let $G = (V, E)$ be a graph. The model works in discrete steps. Every node has exactly one of the following attributes in every step:

- currently active
- formerly active
- inactive

I_t denotes the set of active nodes in step t. All active nodes I_t have a chance to infect each of their inactive neighbors with probability p. All the infected neighbors of the nodes ist set I_t form the set I_{t+1}.

ML-Estimator

We assume that we observe the network at an unknown time and can only see the set nodes ($\boldsymbol{X}^\star$) that are currently active. We also assume that the source of the rumor is a single node $I_0 = \{\boldsymbol{\omega}\}$. Note that $\boldsymbol{X}^\star$ can be empty.
Let

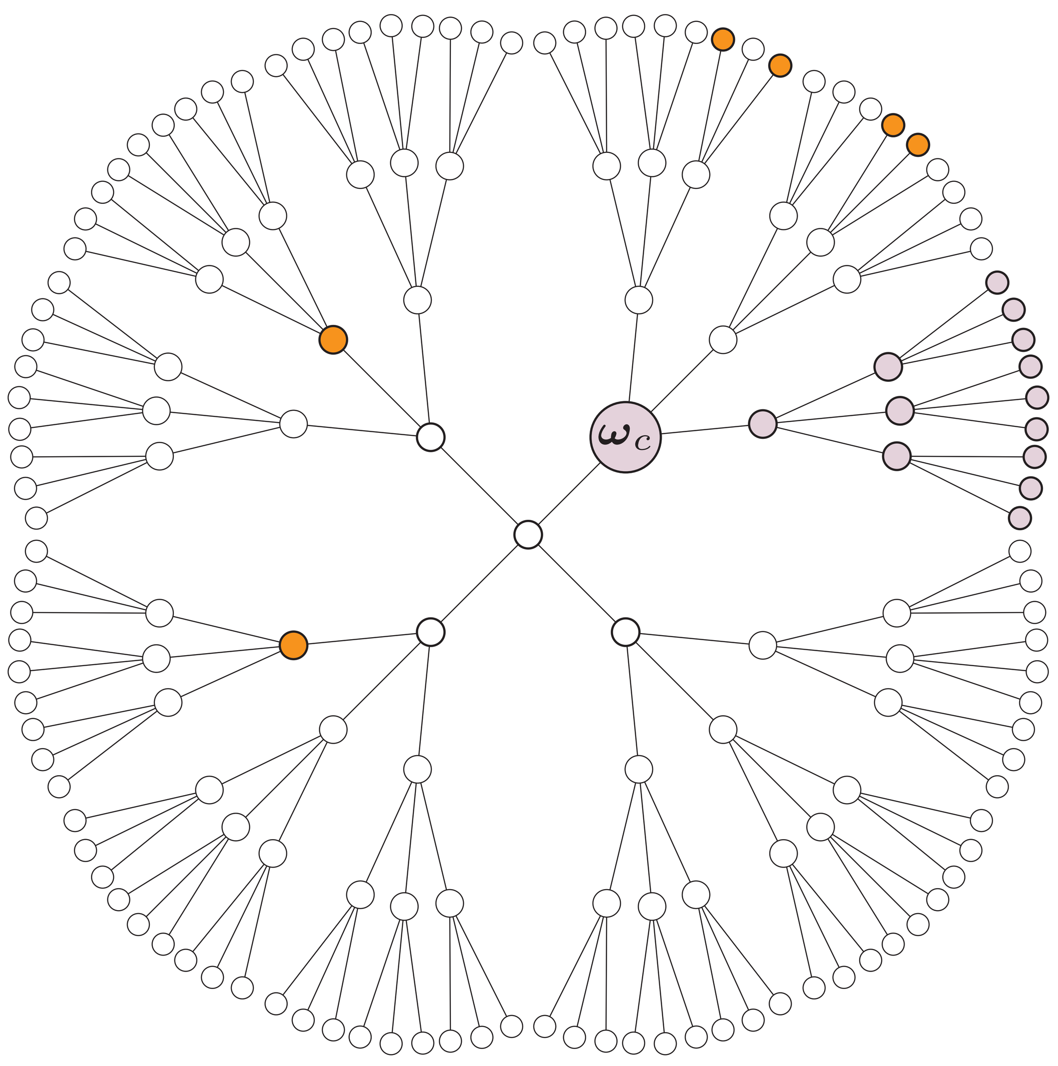
$$\mathcal{C} = \{v \in V \mid \exists t \in \mathbb{N} : \forall w \in \boldsymbol{X}^\star : \operatorname{dist}(v, w) = t\}$$

As long as $\boldsymbol{X}^\star$ is not empty, our estimator is defined as the *closest candidate*:

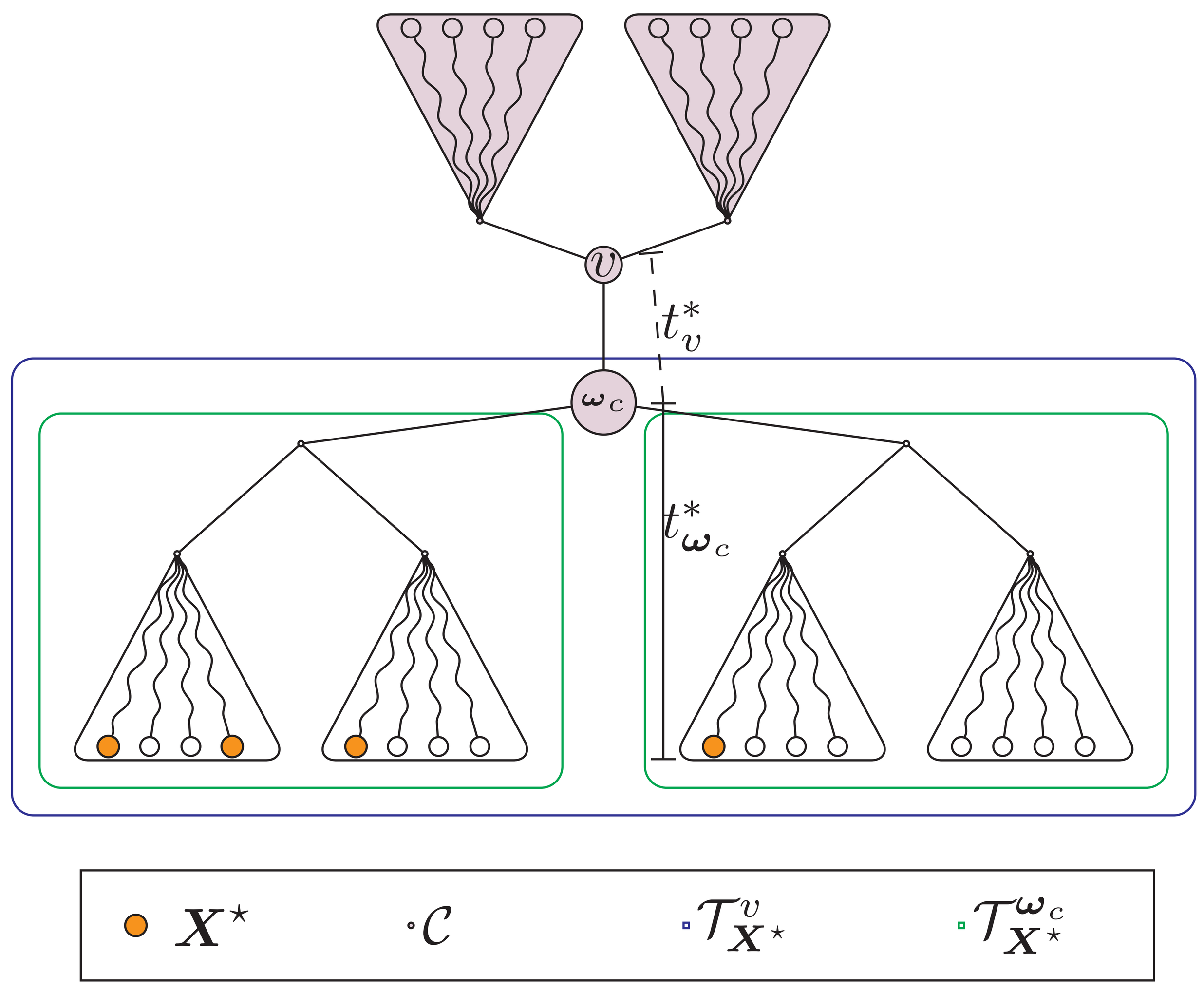
$$\boldsymbol{\omega}_c = \arg\min_{v \in \mathcal{C}} \{dist(v, w) \mid w \in \boldsymbol{X}^\star\}$$

If $|\mathcal{C}| < 2$, then the estimation fails.

Visualization



(a) Visualization of a possible snapshot of the spreading process in a 4-regular tree. The orange nodes are the active elements. The candidate set $\mathcal{C}$ of possible rumor's sources consists of all vertices in the purple sub-tree.



(b) Here, $\boldsymbol{\omega}_c$ spawned three sub-trees out of which two contain active elements of $\boldsymbol{X}^\star$ (orange) and one does not contain active elements (purple). Thus, $\mathcal{C}$ consists of all vertices in the purple sub-tree rooted at $\boldsymbol{\omega}_c$.

D-regular trees

Theorem 1 (d-regular trees). *Let p be the spreading parameter of the Independant Cascade Model and $t = \omega(1)$ the amount of steps. Then, the following phase-transitions occur.*

- *If $(d-1) \cdot p \le 1$, any estimator fails at weak detection with probability $1 - o_t(1)$.*[a]
- *If $1 < (d-1) \cdot p = \Theta(1)$ then the closest candidate $\boldsymbol{\omega}_c$ is the source of the rumor $\boldsymbol{\omega}$ with constant probability (weak detection). Furthermore, the probability that $\operatorname{dist}(\boldsymbol{\omega}_c, \boldsymbol{\omega}) > k$ is at most $\exp(-\Omega(k))$.*
- *If $(d-1) \cdot p = \omega(1)$ then closest candidate $\boldsymbol{\omega}_c$ is the source of the rumor $\boldsymbol{\omega}$ with probability $1 - o_d(1)$ (strong detection).*

[a] We denote by $o_t(1)$ a quantity that tends to zero with $t \to \infty$.

Po(λ)-Galton-Watson trees

Theorem 2 (Galton-Watson processes). *Let p be the spreading parameter of the Independant Cascade Model and $t = \omega(1)$ the amount of steps. Then, the following phase-transitions occur.*

- *If $\lambda p \le 1$, any estimator fails at weak detection with probability $1 - o_t(1)$.*
- *If $1 < \lambda p = \Theta(1)$, then the closest candidate $\boldsymbol{\omega}_c$ is the source of the rumor $\boldsymbol{\omega}$ with positive probability (weak detection). Furthermore, the probability that $\operatorname{dist}(\boldsymbol{\omega}_c, \boldsymbol{\omega}) > k$ is at most $\exp(-\Omega(k))$.*
- *If $\lambda p = \omega(1)$, then closest candidate $\boldsymbol{\omega}_c$ is the source of the rumor $\boldsymbol{\omega}$ with probability $1 - o_\lambda(1)$ (strong detection).*

Experiments

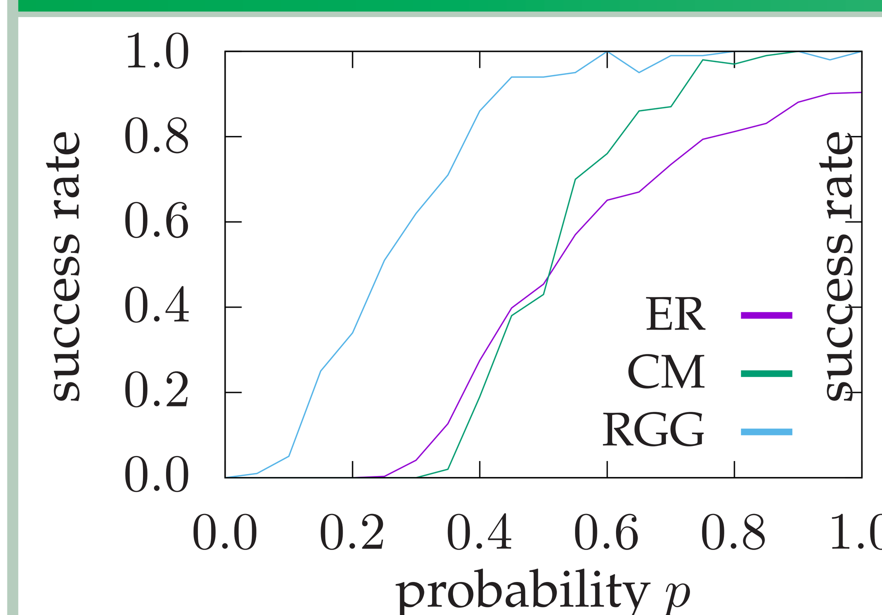


Figure 2: Success rates for Erdős-Rényi graphs, random regular graphs, and random geometric graphs.

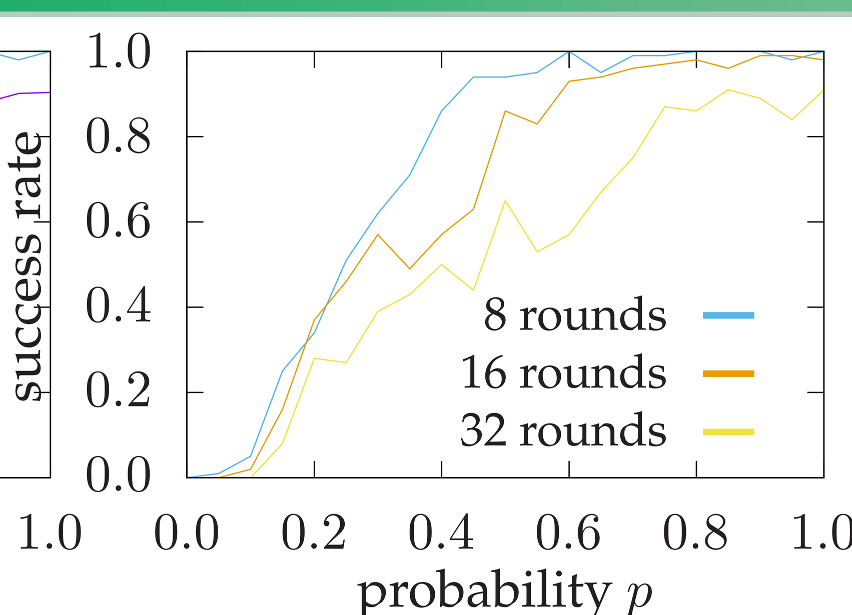


Figure 3: Success rates in a random geometric graph with expected node degree 16 after 8, 16, 32 steps

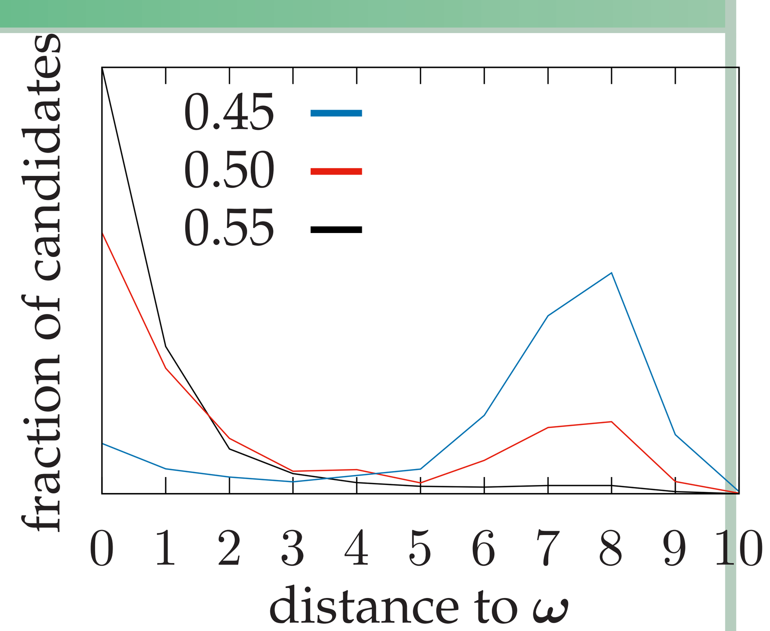


Figure 4: Histogram of the distribution of the distances of the candidates returned by our heuristic to $\boldsymbol{\omega}$ for $p = 0.45, 0.5, 0.55$.

References

[1] Petra Berenbrink, Max Hahn-Klimroth, Dominik Kaaser, Lena Krieg, Malin Rau: Inference of a Rumor's Source in the Independent Cascade Model. UAI 2023.

[2] David Kempe, Jon Kleinberg, Éva Tardos: Maximizing the spread of influence through a social network. SIGKDD 2003.