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

Abstract

We consider the so-called *Independent Cascade Model* for rumor spreading or epidemic processes popularized by Kempe et al. (2003). In this model, a node of a network is the source of a rumor – it is *informed*. 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? In the context of epidemiology this problem is often referred to as *patient zero problem*. It belongs to a broader class of problems where the goal is to infer parameters of the underlying spreading model. In this work we present a maximum likelihood estimator for the rumor's source, given a snapshot of the process in terms of a set of active nodes X after t steps. Our results show that, for acyclic graphs, the likelihood estimator undergoes a phase transition as a function of t. We provide a rigorous analysis for two prominent classes of acyclic network, namely d-regular trees and Galton-Watson trees, and verify empirically that our heuristics work well in various general networks.

1 INTRODUCTION

In this paper, we consider a stochastic diffusion process which models the spread of information or influence in networks. Influence propagation is motivated by many applications from various fields: in marketing these processes are studied to maximize the adoption of a new product; these processes are used to study how social media influencers manipulate humans in social networks, in epidemiology they are used to study the spread of viruses or disease [Brauer, 2017, Kermack and McKendrick, 1927], or, in general as processes that spread information in networks [Becker et al., 2020, Kempe et al., 2015, Lerman and Ghosh, 2010, Nekovee et al., 2007, Sadilek et al., 2012].

On a very high level these processes work as follows. Initially, a small subset of the vertices are in a distinguished state (they might have a piece of information, or they are infected, depending on the application in mind). In this paper, we will call these vertices *informed*. Informed vertices can inform their neighbors and the rumor spreads as time passes by through the network. Most publications studying stochastic diffusion processes study these processes in a *forward direction*, i.e., they consider how information spreads in a network, how many nodes will become informed, or which nodes have the largest influence on the vertices in a social network. Those processes are well understood in simple networks [Britton et al., 2007, Neal, 2003]. In this paper we study the problem in a *backward direction*. Our goal is to detect the *source* of the rumor, thus we study the learning problem of inferring ω, a problem which received far less attention. In the disease spreading settings this problem is referred to as the *patient zero problem*. This inference problem was studied with respect to the SI model from epidemiology rigorously [Shah and Zaman, 2012]. Understanding inference problems of this kind better will help us to find the source of outbreaks of infectious diseases like COVID-19 or to find the source of rumors. The later might help to prevent that political elections are influenced from the outside world.

Model Overview. In this paper we employ the well-known *Independent Cascade Model* (ICM) [Kempe et al., 2015]. The process starts with an initial set of active nodes I_0 and it works in discrete steps. When node v first becomes active in step t, it is given a single chance (one shot) to activate each (currently inactive) neighbor with probability p. Whether or not v succeeds to activate any of its neighbors, it cannot make any further attempts in later rounds. It remains inactive for all steps $t' > t$. The process runs until

Submitted to the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023). **To be used for reviewing only**.

no more activations are possible. We assume that I_0 (the *rumor's source*) contains only a single node denoted by $\boldsymbol{\omega}$. We call all nodes that were activated at one point of time during the process *informed*. Note that this *one-shot* property is a very fitting model for rumor or disease spreading in social networks. Indeed, once a user hears about an article supporting her opinion, she will either ignore it or share it within her social contacts in the near future. Every of the possible recipients either ignores her opinion (does not get activated) or decides to share it again with its peers (gets activated). Furthermore, users are unlikely to share the same article twice. In the case of disease spreading the informed vertices model the persons which caught the illness and the active ones model the persons which are infectious at any point of time. There are two well-known variants of the ICM. We study the variant in which we cannot distinguish formerly informed vertices from never informed vertices. Thus, a snapshot consists only of currently active vertices and the network graph itself which reduces the available information dramatically.

Teacher Student Model. Our problem fits well into the so-called *teacher-student model*, introduced by Gardner and Derrida [1989] in the context of studying fundamental properties of the Perceptron, a fairly simple binary classifier. This framework is frequently used to model (machine) learning tasks [Zdeborová and Krzakala, 2016, Coja-Oghlan et al., 2020]. Suppose a *teacher* samples a *ground-truth* $\boldsymbol{\sigma}$ from a distribution μ_P called *teacher's prior*. Rather than directly revealing this ground-truth to a student, the teacher creates a condensed version $\hat{\boldsymbol{\sigma}}$ of the ground-truth by means of a *teacher's model* μ_M. Now, in the so-called Bayes optimal case, the teacher conveys $\hat{\boldsymbol{\sigma}}, \mu_P, \mu_M$ to her student. The student's task is to infer a non-trivial guess of $\boldsymbol{\sigma}$ from the observed data $(\hat{\boldsymbol{\sigma}}, \mu_P, \mu_M)$. In the context of our contribution, the teacher's prior is the uniform distribution on all nodes of a given network G. The ground-truth is a sample $\boldsymbol{\omega}$ from this distribution. The vertex $\boldsymbol{\omega}$ is the *rumor's source*. The teacher's model is the distribution of the outcome of the t-step forward process of the ICM on G starting with the source $\boldsymbol{\omega}$. The condensed information $\hat{\boldsymbol{\sigma}}$ that the student receives consists of the network G and a set $\boldsymbol{X}^\star$ of active nodes at the same time step. Note that the student does not receive any information about the time at which the set is drawn. The student's task now is to infer $\boldsymbol{\omega}$ from G and $\boldsymbol{X}^\star$.

Results in a Nutshell. In our first result, we prove that our model is Bayes optimal (or, in terms of statistical physics, the *Nishimori property* holds). Secondly, with respect to $d-$regular trees, we show that for a small spreading parameter $p(d-1) < 1$ it is not possible to infer the source of the rumor. For $p(d-1) \gg 1$ we show that we can detect the source node with a very large probability. For $1 < p(d-1) = \Theta(1)$ we show that the source of the rumor can be inferred with a constant probability. Furthermore, we bound the probability that the real rumor source $\boldsymbol{\omega}$ is far away from our algorithm's output. Finally, we establish a similar phase transitions with respect to Galton-Watson processes with spreading parameter $\mathrm{Po}(\lambda)$.

Related Work. Forward propagation processes, like the epidemic models [Brauer, 2017, Keeling and Rohani, 2008, Kermack and McKendrick, 1927, Yorke and Hethcote, 1984], rumor spreading [Goffman and Newill, 1964, Lerman and Ghosh, 2010, Nekovee et al., 2007, Sadilek et al., 2012], information cascades [Kempe et al., 2015, Zhao et al., 2012], blog propagation models [Leskovec et al., 2007], and marketing strategies [Becker et al., 2020] have been studied extensively and for a long time within different research communities and we refrain from discussing the extensive literature here. On the contrary, rigorous contributions on the corresponding inference problem, the source detection task, are scarce.

To the best of our knowledge, there is only little rigorous knowledge on how to infer a rumor's source. With respect to the SI model, the rigorous contributions are by Kazemitabar and Amini [2019], Shah and Zaman [2010, 2012]. Shah and Zaman [2010, 2012] prove that on some infinite acyclic networks like d-regular trees, super-critical Galton-Watson processes, and geometric graphs, approximate inference of $\boldsymbol{\omega}$ is always possible in the SI model as long as the infinite tree satisfies certain expansion properties (for instance, in the d-regular case, this requires $d \geq 3$). Furthermore, they prove that the probability of declaring a *far-apart* node as the rumor's source is small. The results are proved by a fundamental connection between a generalized Polya's urn and the SI model on acyclic networks. Dawkins et al. [2021] introduce a statistical framework for the recovery of the diffusion source. Their framework covers the SI model but can be extended to other diffusion processes as well. Finally, there is also a line of work towards the more general discrete time SIR model [Zhu et al., 2017, Zhu and Ying, 2016] which includes the ICM as a special case (if individuals recover after one round). Zhu et al. [2017] study a Jordan centrality based inference algorithm for d-regular trees. Our Theorem 2 deepens some of their results with regard to the ICM. The exact connection will be discussed in Section 5.

Moreover, there are many well explained heuristics towards the source detection task on various network types which are supported by extensive simulation studies [Bindi et al., 2017, Jain et al., 2016, Ji and Tay, 2017, Prakash et al., 2014, Wang et al., 2017], in recent contributions also based on neural networks [Biazzo et al., 2021, Shah et al., 2020, Shu et al., 2021]. Finally, a closely related problem that attained attention recently, is not to infer $\boldsymbol{\omega}$ given $\boldsymbol{X}^\star$ but to infer the parameters of the underlying spreading model. Over the last years, learning strategies towards this problem were proposed and studied experimentally [Lokhov, 2016, Mastakouri and Schölkopf, 2020].

2 MODEL AND RESULTS

We are given a communication network $G = (V, E)$ with n vertices $\rho_1 \dots \rho_n$ and m edges. We assume $I_0 = \{\boldsymbol{\omega}\}$ and $\boldsymbol{\omega}$ is chosen uniformly at random from all vertices. Hence, the rumor originates at a single source. A *spreading parameter* $p \in (0, 1)$ determines the viciousness of the rumor. The diffusion process in the ICM runs in discrete, synchronous steps. Let I_t be the set of vertices which are active in step t. In step t we will call the vertices in $I_0 \cup I_1 \dots \cup I_t$ *informed*. In every step $t \geq 1$ every active node $\rho_i \in I_t$ activates any of its uninformed neighbors with probability p. All these newly informed vertices form the set I_{t+1}. Note that every node becomes active exactly once but vertices have potentially the chance to be activated by each of their neighbors. Note that it can happen that the process dies out at a step t. In this case for all $t' \geq t$ it holds that $I_{t'} = \emptyset$.

The *interference problem* is defined as follows. We observe the state of the network at an arbitrary time t. The task is to infer $\boldsymbol{\omega}$ given only (G, I_t) and the parameter p. In this paper we study the following variants of the problem.

- *Strong detection:* infer $\boldsymbol{\omega}$ with high probability,
- *Weak detection:* infer $\boldsymbol{\omega}$ with positive probability[1].

Our first result relates the probability that a given set X is active conditioned on some node v being the source to the probability that v is the source conditioned on X being active. This establishes the so-called *Nishimory property* (or Bayes optimality) of our inference problem. In terms of the teacher-student model, it states that the student has access to the teacher's prior and the teacher's model. Equivalently, in expectation there is no statistical difference between the ground-truth and a uniform sample from the posterior distribution [Zdeborová and Krzakala, 2016]. Theorem 1 applies to all types of networks as long as the rumor's source $\boldsymbol{\omega}$ is chosen uniformly at random.

Theorem 1. *Let $G = (V, E)$ be an arbitrary network and fix an arbitrary step t. Let $\boldsymbol{X}^\star$ be the set active vertices in step t. For any $X \subseteq V$*

$$\arg\max_{v \in V} \mathbb{P}(\boldsymbol{X}^\star = X \mid \boldsymbol{\omega} = v) = \arg\max_{v \in V} \mathbb{P}(\boldsymbol{\omega} = v \mid \boldsymbol{X}^\star = X).$$

The above theorem is used to show the main results of this paper. It allows us to calculate $\mathbb{P}(\boldsymbol{X}^\star = X \mid \boldsymbol{\omega} = v)$ instead of $\mathbb{P}(\boldsymbol{\omega} = v \mid \boldsymbol{X}^\star = X)$, which often is more accessible. Note that calculating the first probability is quite challenging in general networks G since the entropy of $\boldsymbol{X}^\star$ is very large.

For the remaining analytical part of the of the paper we consider acyclic networks, namely d-regular trees and Galton-Watson trees with offspring distribution $\mathrm{Po}(\lambda)$. The latter

[1] We say that a sequence of events $\mathcal{E}_1, \mathcal{E}_2, \dots$ takes place with positive probability if $\lim_{t\to\infty} \mathbb{P}(\mathcal{E}_t) > 0$ and with high probability (w.h.p.) if $\lim_{t\to\infty} \mathbb{P}(\mathcal{E}_t) = 1$.

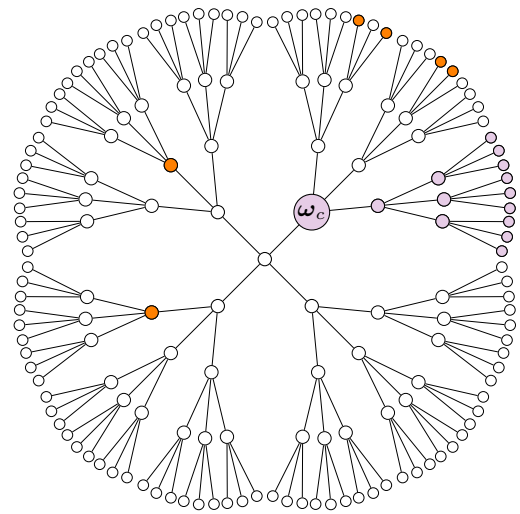

Figure 1: Visualization of a possible snapshot of the spreading process. Here, $\boldsymbol{\omega}_c$ spawned four sub-trees out of which three contain active elements of $\boldsymbol{X}^\star$ (orange nodes) and one does not contain active elements (purple). Thus, the candidate set $\mathcal{C}$ of possible rumor's sources consists of all vertices in the purple sub-tree.

Galton-Watson trees with offspring distribution $\mathcal{D}$ are defined by the following experiment. We start with one node which spawns $\boldsymbol{d}_0 \sim \mathcal{D}$ children. Recursively, any of the children $w_1, \dots w_{\boldsymbol{d}_0}$ spawns $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_{\boldsymbol{d}_0} \sim \mathcal{D}$ children (and so on). We call such a process super-critical, if with positive probability, the vertices spawned during the process form an infinite tree. It is well known that a (not too dense) instance of an Erdős–Rényi random graph $\mathbb{G}\big(n, \frac{d}{n}\big)$ *locally* looks like a Galton-Watson tree with offspring distribution $\mathrm{Po}(d)$. In contrast, a random d-regular graph looks locally like a d-regular tree, provided d is not too large.

We assume that a teacher fixes a time t at which she observes the network. We assume that the student is not aware of the time when the process started. We define $\boldsymbol{X}^\star$ as the set of nodes active in step t of the ICM (starting from an unknown and randomly chosen source $\boldsymbol{\omega}$). Note that $\boldsymbol{X}^\star$ can be empty. The set of candidate nodes $\mathcal{C}$ are all nodes v that have the same distance in G to each node in $\boldsymbol{X}^\star$, i.e, for all $x, x' \in \boldsymbol{X}^\star$ it holds that $\mathrm{dist}(x, v) = \mathrm{dist}(x', v)$. Note that the set $\mathcal{C}$ is not empty since $\boldsymbol{\omega} \in \mathcal{C}$ and that $\mathcal{C}$ is the whole graph if $\boldsymbol{X}^\star = \emptyset$. The *closest candidate* $\boldsymbol{\omega}_c \in \mathcal{C}$ is defined as the node with minimum distance to all nodes in $\boldsymbol{X}^\star$, see Figure 1 for an example. Our *source detection heuristic* calculates the closest candidate $\boldsymbol{\omega}_c \in \mathcal{C}$ and returns it as the estimated rumors source. If $\boldsymbol{X}^\star = \emptyset$ (the process died out before time t) or contains at most one node, the heuristic returns a failure. The following results describe the probability of success or failure for this source detection heuristic with respect to the models parameter.

The first theorem shows a phase transition between weak

and strong detection for d-regular trees. We show that, for small $p \cdot (d-1)$ it is not possible to infer the source of the rumor. This is due to the huge likelihood that, in this setting, the process dies out and $\boldsymbol{X}^\star = \emptyset$ or the corresponding set $\boldsymbol{X}^\star$ is very small which makes the inference impossible.

For large $p \cdot (d-1)$ we show that the source node is the closest candidate with probability $1 - o_d(1)$. Note that in this scenario, each active node will infect several nodes and it is unlikely that the process dies out. The size of $\boldsymbol{X}^\star$ grows as a function of d and t which makes the prediction more and more reliable. For intermediate $p \cdot (d-1)$ we show that the closest candidate $\boldsymbol{\omega}_c$ is the source of the rumor with a constant probability. Furthermore, we bound the probability that the real rumor source $\boldsymbol{\omega}$ is far away from $\boldsymbol{\omega}_c$. Intuitively this means that inference of $\boldsymbol{\omega}$ gets easier for increasing values of $(d-1) \cdot p$.

Theorem 2 (d-regular trees)**.** *Let $G = (V, E)$ be an infinite d-regular tree and let $\boldsymbol{X}^\star$ be the set of active nodes generated by the ICM with spreading parameter p after $t = \omega(1)$ steps. Then, the following phase-transitions occur.*

- *If $(d-1) \cdot p \leq 1$, any estimator fails at weak detection with probability $1 - o_t(1)$.*[2]
- *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 $\mathrm{dist}(\boldsymbol{\omega}_c, \boldsymbol{\omega}) > k$ 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).*

We also study the patient zero problem on $\mathrm{Po}(\lambda)$-Galton-Watson trees. This model describes the local structure of Erdős-Rényi random graphs very well. We find a similar phase transition as before.

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

Theorem 3 (Galton-Watson processes)**.** *Let $G = (V, E)$ be an infinite tree generated by a $\mathrm{Po}(\lambda)$-Galton-Watson process. Let $\boldsymbol{X}^\star$ be the set of active nodes generated by the ICM with spreading parameter p after $t = \omega(1)$ steps. Then, the following phase-transition occurs.*

- *If $\lambda p \leq 1$, 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 $\mathrm{dist}(\boldsymbol{\omega}_c, \boldsymbol{\omega}) > k$ 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).*

In Theorems 2 and 3 we assume that the underlying tree network is infinitely large. This is conceptually necessary. Indeed, the trivial algorithm that outputs one node uniformly at random succeeds at weak detection in finite networks. In the next section, we provide extensive simulations that verify the asymptotic statements of the theorems on small networks. Furthermore, we present simulation results on non-acyclic networks such as random geometric graphs and show that (the natural extension of) the closest candidate heuristics works well.

3 SIMULATIONS

In this section we present simulation results that support and complement our main theorems for cyclic graphs, implementation details are given in the supplementary material. Our algorithm is generalized to cyclic graphs as follows. For $v \in V$ and $t' \geq 0$ let $N_{t'}(v)$ denote the set of nodes $w \in V$ that have distance at most t' to v. First we calculate

$$N_{t'} = \bigcap_{u \in \boldsymbol{X}^\star} N_{t'}(u).$$

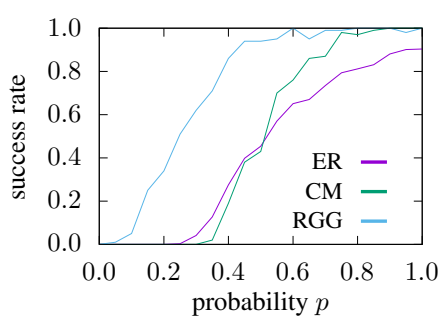

Figure 2: Visualization of success rates for Erdős-Rényi graphs (ER), random regular graphs (CM), and random geometric graphs (RGG).

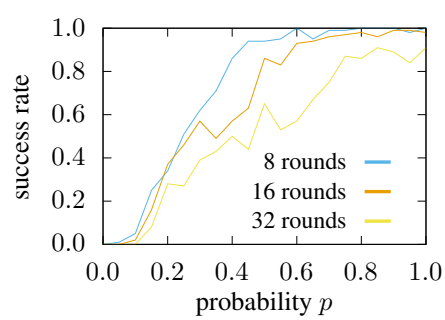

Figure 3: Visualization of success rates in a random geometric graph with expected node degree 16 after 8, 16, 32 rounds of the ICM

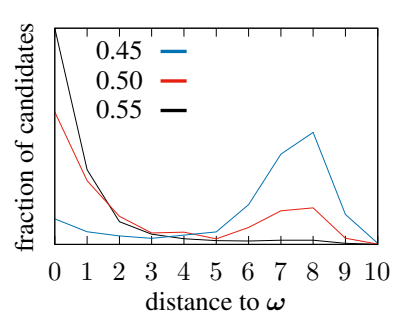

Figure 4: Histogram of the distribution of the distances of the candidates returned by our heuristic to $\boldsymbol{\omega}$ for $p = 0.45, 0.5, 0.55$.

Hence, N'_t is the set of nodes with distance at most t' to every node in $\boldsymbol{X}^\star$. We then pick the minimum t' such that $N_{t'} \neq \emptyset$ and return $N_{t'}$ as the candidate set. Note that such a t' exists since $\boldsymbol{\omega} \in N_t$.

To generate our data we execute 100 independent simulation runs for each network, where we simulate the ICM for 8 rounds. The success rates are visualized in Figure 2. In Figure 3 we present simulation data for random geometric graphs with expected node degree 16, where we increased the numbers of rounds of the ICM. We show the success rates after 8, 16, and 32 rounds. Figure 4 highlights the phase transition behavior of our algorithm.

4 ANALYSIS

In this section we formally prove our main results, Theorems 2 and 3. The proof of Theorem 1 is given in the supplementary material. The main proof strategy of Theorems 2 and 3 is to interpret the transmission process as a special type of percolation on the underlying network. As in Theorem 3 the underlying network itself is random, it turns out to be technically non-trivial to pin down the exact distributions involved in this process due to subtle rare events that might yield either very small or large node degrees. The Poisson thinning technique allows us to carry out the calculations smoothly.

4.1 PROOF OF THEOREM 2

A crucial observation in the proof of Theorem 2 is the following. If node v gets activated during the spreading process by node w, it has $d-1$ additional neighbors $v_1, \ldots, v_{d-1}$ except w which we call *children* of v. Any of those children gets activated with probability p independently from everything else in the next step. Suppose without loss of generality that $v_1, \ldots, v_{\boldsymbol{d}_0}$ are the activated children where $\boldsymbol{d}_0 \sim \mathrm{Bin}(d-1, p)$. In every of those children v_i, a new and independent rumor spreading process starts in the tree rooted at v_i and directed away from v. As this tree is, itself, d-regular, this process is distributed equally as starting $\boldsymbol{d}_0$ independent Galton-Watson processes with offspring distribution $\mathrm{Bin}(d-1, p)$. Depending on $(d-1)p$, few, some or many of those processes will die out eventually.

To prove our result, we need some additional notation, see Figure 5. Given a node v, we can direct the tree away from v and denote the set of subtrees rooted at v's children by $\mathcal{T}^v$. Most interesting to our proof are the subtrees that contain active nodes. We denote them by $\mathcal{T}^v_{\boldsymbol{X}^\star}$ and denote

$$Y_v = Y_v(\boldsymbol{X}^\star) = |\mathcal{T}^v_{\boldsymbol{X}^\star}|.$$

Note that all candidates but one have at most one subtree containing active nodes. Only the closest candidate can have more than one. Finally, let $t_v^{\boldsymbol{X}^\star}$ denote the distance from v to any of the vertices in $\boldsymbol{X}^\star$.

Proposition 4. *If ρ_t is the probability that at time t there are no more active vertices under the spreading model on an infinite d-regular tree with infection probability p, we get*

$$\rho_t = (1 - p + p\rho_{t-1})^{d-1}$$

and the ultimate extinction probability of the spreading process is the smallest fixed-point of $x \mapsto (1-p+px)^{d-1}$. Furthermore for $v \in \mathcal{C}$

$$\mathbb{P}(Y_v(\boldsymbol{X}^\star) = k \,|\, \boldsymbol{\omega} = v) = \begin{cases} \sum_{d_0=k}^{d} \mathbb{P}(\mathrm{Bin}(d,p) = \boldsymbol{d}_0) \binom{\boldsymbol{d}_0}{k} \rho_{t_v^{\boldsymbol{X}^\star}}^{\boldsymbol{d}_0 - k} (1 - \rho_{t_v^{\boldsymbol{X}^\star}})^k & \text{if } v = \boldsymbol{\omega}_c \\ \sum_{d_0=1}^{d} \mathbb{P}(\mathrm{Bin}(d,p) = \boldsymbol{d}_0) \binom{\boldsymbol{d}_0}{1} \rho_{t_v^{\boldsymbol{X}^\star}}^{\boldsymbol{d}_0 - 1} (1 - t_v^{\boldsymbol{X}^\star}) & \text{if } v \neq \boldsymbol{\omega}_c. \end{cases}$$

Proof. The first part, namely that $\rho_t = (1-p+p\rho_{t-1})^{d-1}$, is an application of probability generating functions and their connection to branching processes. The probability generating function $f_{\mathrm{Bin}(n,p)}$ of a Binomial distribution with parameters n and p reads

$$f_{\mathrm{Bin}(n,p)}(s) = \mathbb{E}\Big[s^{\mathrm{Bin}(n,p)}\Big] = (1-p+ps)^n.$$

Now, let $p_k = \mathbb{P}(\mathrm{Bin}(d-1,p) = k)$. It is immediate that

$$\rho_{t+1} = p_0 + p_1\rho_{t-1} + \ldots + p_{d-1}\rho_{t-1}^{d-1} = \sum_{k=0}^{d-1} p_k \rho_{t-1}^k$$

which is exactly the probability generating function of the Binomial distribution. A detailed explanation and a formal proof of this statement can be, for instance, found in [Grimmett and Stirzaker, 2020].

Let us now suppose that $v = \boldsymbol{\omega}_c$. Let $\mathcal{V}_0$ be the event that v has exactly $k \leq \boldsymbol{d}_0 \leq d$ children that get activated by v. Clearly, $\mathbb{P}(\mathcal{V}_0) = \mathbb{P}(\mathrm{Bin}(d,p) = \boldsymbol{d}_0)$ and of course, $\boldsymbol{d}_0$ needs to be at least k as differently, the probability of having k active sub-trees was zero. Given $\mathcal{V}_0$, we find that the probability of observing exactly k active sub-trees is the probability that exactly k out of $\boldsymbol{d}_0$ independent Galton-Watson processes with offspring distribution $\mathrm{Bin}(d-1,p)$ survived the first $t_v^{\boldsymbol{X}^\star}$ steps. Therefore, the number of active sub-trees at time t is distributed as $\mathrm{Bin}\Big(\boldsymbol{d}_0, \rho_{t_v^{\boldsymbol{X}^\star}}\Big)$ given $\mathcal{V}_0$ and the first part of the formula follows.

If, on contrary, v is not the closest candidate but a node that has a different distance from $\boldsymbol{X}^\star$, we observe that from the originally $1 \leq \boldsymbol{d}_0 \leq d$ Galton-Watson processes originated in the children of v, exactly one process needed to survive and $\boldsymbol{d}_0 - 1$ needed to become extinct at time $t_v^{\boldsymbol{X}^\star}$. The proposition follows. □

Proof of Theorem 2 (i). To prove the first part of Theorem 2, it suffices to apply the first part of Proposition 4. Indeed,

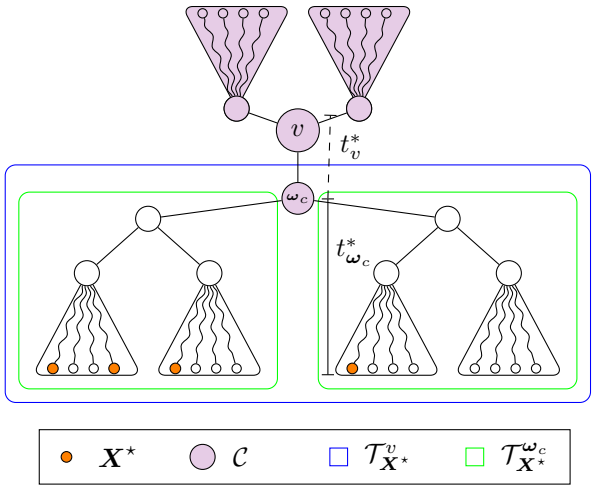

Figure 5: Here, $\boldsymbol{\omega}_c$ spawned three sub-trees out of which two contain active elements of $\boldsymbol{X}^\star$ (orange) and one does not contain active elements (purple). Thus, $\mathcal{C}$ consists of all vertices in the purple sub-tree rooted at $\boldsymbol{\omega}_c$.

if $(d-1)p \leq 1$, we find that the smallest fixed-point of $x \mapsto (1-p+px)^{d-1}$ is $x = 1$. Therefore, $\rho_t = 1 - o_t(1)$. Furthermore, as p is a constant, we have that in this case $d = \Theta(1)$ as well. Therefore, a union bound over the at most d possible independent Galton-Watson processes with offspring distribution $\mathrm{Bin}(d-1, p)$ originated in the children of $\boldsymbol{\omega}$, yields that with probability $1 - o_t(1)$, we find $\boldsymbol{X}^\star = \emptyset$. In this case, detection clearly fails with high probability. □

Proof of Theorem 2 (ii). We start with the part of the theorem that claims that weak detection succeeds by the source detection heuristic, namely that $\mathbb{P}(\boldsymbol{\omega}_c = \boldsymbol{\omega}) = \Omega(1)$. We find that $\boldsymbol{\omega}_c = \boldsymbol{\omega}$ with probability one if the set of possible candidate nodes $\mathcal{C}$ has size 1. Therefore, it suffices to prove that $\mathbb{P}(|\mathcal{C}| = 1) = \Omega(1)$. This is the case if (and only if), the rumor's source $\boldsymbol{\omega}$ propagated the rumor to all of its d children and all d independent Galton-Watson processes with offspring distribution $\mathrm{Bin}(d-1, p)$ originated in the children of $\boldsymbol{\omega}$ did not become extinct. Let $\boldsymbol{d}_0$ denote the number of children of $\boldsymbol{\omega}$ that get activated. Clearly,

$$\mathbb{P}(\boldsymbol{d}_0 = d) = \mathbb{P}(\mathrm{Bin}(d, p) = d) = p^d = \Omega(1) \tag{1}$$

as, by assumption, p and d are constants. Furthermore, since $1 < (d-1)p = \Theta(1)$ holds, the smallest fixed point of $x \mapsto (1-p+px)^{d-1}$ is a real number between zero and one. Therefore, by Proposition 4, there are constants $0 < \gamma_1 \leq \gamma_2 < 1$ such that

$$\gamma_1 - o_t(1) < \rho_t < \gamma_2 + o_t(1). \tag{2}$$

Thus, it follows that the source detection heuristic succeeds at weak detection if $1 < (d-1)p = \Theta(1)$ with probability $1 - o_t(1)$ by (1) – (2).

It is left to prove that under the same assumptions we have

$$\mathbb{P}(\mathrm{dist}(\boldsymbol{\omega}_c, \boldsymbol{\omega}) \geq k) \leq \exp(-\Omega(k)). \tag{3}$$

Suppose that $\left|t^{\boldsymbol{X}^\star}_{\boldsymbol{\omega}} - t^{\boldsymbol{X}^\star}_{\boldsymbol{\omega}_c}\right| = \mathrm{dist}(\boldsymbol{\omega}_c, \boldsymbol{\omega}) = k > 3$. Therefore, there is a unique path $P_{\boldsymbol{\omega}_c, \boldsymbol{\omega}}$ in G that connects $\boldsymbol{\omega}_c$ and $\boldsymbol{\omega}$ with $k-2$ internal vertices. All of those internal vertices will get activated exactly once during the spreading process. Therefore, for any of those $k-2$ steps, the process needs to either activate only exactly one child or, it activates more than one child, but the remaining processes have died out at the observation time. The probability that a node spawns exactly one active child is given by

$$\mathbb{P}(\mathrm{Bin}(d, p) = 1) = p(1-p)^{d-1} \tag{4}$$

which is bounded away from zero and one if $d, p = \Theta(1)$. By (1) – (2) as well as (4), we find that there is a sequence of constants $\{\gamma_i\}_{i=1\ldots k-2}$ dependent only on p, d, and k all of which are uniformly bounded away from zero and one. Therefore,

$$\mathbb{P}(\mathrm{dist}(\boldsymbol{\omega}_c, \boldsymbol{\omega}) \geq k) \leq \min_{i=1\ldots k-2} \gamma_i^k = \exp(-\Omega(k))$$

which implies (3). □

The last part of Theorem 2 states that the source detection heuristic succeeds at strong detection with high probability if $(d-1)p \in \omega(1)$. We start with the following simple observation which is an immediate consequence of Chernoff bounds applied to the $\mathrm{Bin}((d-1), p)$ distribution.

Observation 5. *If $(d-1)p \in \omega(1)$, we find that if node v gets activated, the number of activated children ω_v satisfies $\omega_v \geq (d-1)p/2 = \omega(1)$ with high probability.*

Next, we show that out of $\boldsymbol{d}_0 = \omega(1)$ independent Galton-Watson processes with offspring distribution $\mathrm{Bin}((d-1), p)$ at least a $(1-\varepsilon)$-fraction will not become eventually extinct with high probability for any $\varepsilon > 0$.

Lemma 6. *Suppose that ℓ Galton-Watson processes with offspring distribution $\mathrm{Bin}((d-1), p)$ start independently under the condition that $(d-1)p \in \omega(1)$. Let $\boldsymbol{Y}$ denote the number of processes that do not ultimately become extinct. Then, $\mathbb{P}(\boldsymbol{Y} \geq (1 - o(1))\ell) \geq 1 - o(1)$.*

Proof. By Proposition 4, we have that the probability that one of the processes becomes extinct is $p_e = o(1)$. Thus, the number of not-extinct processes is Binomially distributed with parameters ℓ and $1 - p_e$. Therefore, the lemma follows from Chernoff bounds. □

Proof of Theorem 2 (iii). By Observation 5 and Lemma 6 we directly get that, with high probability, all but $o(dp)$ of the processes started in the children of $\boldsymbol{\omega}$ are still active at the observation time.

Suppose that $\boldsymbol{\omega} \neq \boldsymbol{\omega}_c$ and $\mathrm{dist}(\boldsymbol{\omega}, \boldsymbol{\omega}_c) = k \geq 1$. This implies that either k times only exactly one child is activated or, given that multiple children are activated, only exactly one of those spreading processes survived eventually. For a specific step $1 \leq i \leq k-1$, the probability that this occurs is by Observation 5 and Lemma 6 at most $\gamma_i = o(1)$. Therefore, $\mathbb{P}(\boldsymbol{\omega}_c \neq \boldsymbol{\omega} \mid \boldsymbol{X}^\star) = o(1)$ which implies the third part of Theorem 2. □

4.2 PROOF OF THEOREM 3

The main proof strategy is similar to the proof of Theorem 2. However, one fundamental difference makes the analysis more involved: In the d-regular case an activated node spawns a random number of independent Galton-Watson processes with offspring distribution $\mathrm{Bin}(d-1, p)$. This is not the case in the setting of Theorem 3. Here, the underlying network itself is a Galton-Watson process with offspring distribution $\mathrm{Po}(\lambda)$. Fortunately, we can apply the *Poisson thinning principle* [Kingman, 1993].

Observation 7. *Let $\boldsymbol{X} \sim \mathrm{Po}(d)$ and furthermore, given $\boldsymbol{X}$, define $\boldsymbol{Y} = \mathrm{Bin}(\boldsymbol{X}, p)$. Then $\boldsymbol{Y} \sim \mathrm{Po}(\lambda p)$.*

This Poisson thinning principle implies that in distribution we can analyze the following spreading process: Once v gets activated, it spawns $\boldsymbol{d}_0 \sim \mathrm{Po}(\lambda p)$ active children and thus $\boldsymbol{d}_0$ independent Galton-Watson processes with offspring distribution $\mathrm{Po}(\lambda p)$. The following proposition characterizes the extinction probability of such processes, the formal proof can be found in the supplementary materials.

Proposition 8. *If $\bar{x}_t$ is the probability that at time t there are no more active vertices under the spreading model on a super-critical Galton-Watson tree with offspring distribution $\mathrm{Po}(\lambda p)$, we find $\bar{x}_t = \exp(-\lambda p(1-\bar{x}_{t-1}))$ and the ultimate extinction probability of the spreading process is the smallest fixed-point of $\bar{x} \mapsto \exp(-\lambda p(1-\bar{x}))$. Furthermore for $v \in \mathcal{C}$*

$$\mathbb{P}(Y_v(\boldsymbol{X}^\star) = k \,|\, \boldsymbol{\omega} = v) = \begin{cases} \sum_{d_0=k}^{d} \mathbb{P}(\mathrm{Po}(\lambda p) = \boldsymbol{d}_0) \binom{\boldsymbol{d}_0}{k} \rho_{t_v^{\boldsymbol{X}^\star}}^{\boldsymbol{d}_0 - k} (1 - \rho_{t_v^{\boldsymbol{X}^\star}})^k & \text{if } v = \boldsymbol{\omega}_c \\ \sum_{d_0=1}^{d} \mathbb{P}(\mathrm{Po}(\lambda p) \binom{\boldsymbol{d}_0}{1} \rho_{t_v^{\boldsymbol{X}^\star}}^{\boldsymbol{d}_0 - 1} (1 - t_v^{\boldsymbol{X}^\star}) & \text{if } v \neq \boldsymbol{\omega}_c. \end{cases}$$

Proof of Theorem 3 (i). As in the d-regular case, the first part of Theorem 3 follows by the first part of Proposition 8 and a large deviation bound. Details can be found in the supplementary material. □

Proof of Theorem 3 (ii). Again, as in the d-regular case, we start proving the weak detection property of the source detection heuristic. Thus, we aim to prove $\mathbb{P}(|\mathcal{C}| = 1) = \Omega(1)$. This is the case if (and only if) $\boldsymbol{\omega}$ propagated the rumor to all of its $\boldsymbol{d}_{\boldsymbol{\omega}} \mathrm{Po}(\lambda)$ children and all $\boldsymbol{d}_0$ independent Galton-Watson processes with offspring distribution $\mathrm{Po}(\lambda p)$ rooted at the children of $\boldsymbol{\omega}$ did die out eventually. Let $\boldsymbol{d}_0$ denote the number of children of $\boldsymbol{\omega}$ that get activated. We first need to calculate the probability that $\boldsymbol{d}_0 = \boldsymbol{d}_{\boldsymbol{\omega}}$. To this end, let

$$I_0(x) = \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(k+1)} \left(\frac{x}{2}\right)^{2k} = \frac{\exp(x)}{\sqrt{2\pi x}} \left(1 + O\left(\frac{1}{x}\right)\right)$$

denote the modified Bessel function of order zero [Martin et al., 2021].

We have

$$\mathbb{P}(\boldsymbol{d}_0 = \boldsymbol{d}_{\boldsymbol{\omega}} \mid \boldsymbol{d}_{\boldsymbol{\omega}}) = \frac{(\lambda p)^{\boldsymbol{d}_{\boldsymbol{\omega}}} \exp(-\lambda p)}{\boldsymbol{d}_{\boldsymbol{\omega}}!}.$$

Therefore, by the law of total probability,

$$\begin{aligned} &\mathbb{P}(\text{all children of } \boldsymbol{\omega} \text{ get activated}) \qquad (5) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(\boldsymbol{d}_{\boldsymbol{\omega}} = k) \frac{(\lambda p)^k \exp(-\lambda p)}{k!} \\ &= \sum_{k=1}^{\infty} \frac{\lambda^k \exp(-\lambda)}{k!} \frac{(\lambda p)^k \exp(-\lambda p)}{k!} \\ &= \exp(-\lambda(1+p)) \sum_{k=1}^{\infty} \frac{(\lambda^2 p)^k}{(k!)^2} \\ &= \exp(-\lambda(1+p))(I_0(2\lambda\sqrt{p}) - 1) \\ &= \exp(-\lambda(1+p)) \left(\frac{\exp(2\lambda\sqrt{p})}{\sqrt{4\pi\lambda\sqrt{p}}} - 1 + O\left((\lambda\sqrt{p})^{-1}\right) \right) \\ &= \frac{\exp(-\lambda(1-\sqrt{p})^2)}{\sqrt{4\pi\lambda\sqrt{p}}} - \exp(-\lambda(1+p)) \\ &\quad + O\left(\exp(-\lambda(1+p))(\lambda\sqrt{p})^{-3/2}\right). \end{aligned}$$

If λ and p are constants, it is immediate from (5) that there is a constant $\gamma > 0$ such that

$$\mathbb{P}(\text{all children of } \boldsymbol{\omega} \text{ get activated}) > \gamma.$$

Finally, since $1 < \lambda p = \Theta(1)$ by assumption, the smallest fixed point of $\bar{x} \mapsto \exp(-\lambda(1-\bar{x}))$ is a real number between zero and one. Therefore, by Proposition 8, there are constants $0 < \gamma_1 \leq \gamma_2 < 1$ such that

$$\gamma_1 - o_t(1) < \bar{x}_t < \gamma_2 + o_t(1). \qquad (6)$$

Therefore, the source detection heuristic succeeds at weak detection if $1 < \lambda p = \Theta(1)$ with probability $1 - o_t(1)$ by (5) – (6).

Again, we are left to prove the decay property, namely

$$\mathbb{P}(\mathrm{dist}(\boldsymbol{\omega}_c, \boldsymbol{\omega}) \geq k) \leq \exp(-\Omega(k)). \qquad (7)$$

Suppose that $\left|t_{\boldsymbol{\omega}}^{\boldsymbol{X}^\star} - t_{\boldsymbol{\omega}_c}^{\boldsymbol{X}^\star}\right| = \mathrm{dist}(\boldsymbol{\omega}_c, \boldsymbol{\omega}) = k > 3$. As before we find a unique path $P_{\boldsymbol{\omega}_c,\boldsymbol{\omega}}$ in G that connects $\boldsymbol{\omega}_c$ and $\boldsymbol{\omega}$ with $k-2$ internal vertices. Exactly as in the d-regular case, these internal vertices will get activated exactly once during the spreading process and so, in every step, the process needs to either activate exactly one child or all but one of the remaining processes ultimately become extinct. The probability that a node spawns exactly one child is given by

$$\mathbb{P}(\mathrm{Po}(\lambda p) = 1) = \lambda \exp(-\lambda) \tag{8}$$

which is a real number bounded away from zero and one if $\lambda = \Theta(1)$. By (5) – (6) and (8), there is a sequence of constants $\{\gamma_i\}_{i=1\ldots k-2}$ dependent only on $\gamma(p, \lambda, k)$ and uniformly bounded away from zero and one, such that

$$\mathbb{P}(\mathrm{dist}(\boldsymbol{\omega}_c, \boldsymbol{\omega}) \geq k) \leq \min_{i=1\ldots k-2} \gamma_i^k = \exp(-\Omega(k))$$

and (7) follows. □

We start the proof of Theorem 3 (iii) with the following lemma.

Lemma 9. *If $\lambda p \to \infty$, then, with high probability, an activated node v satisfies the following.*

- $\deg(v) \geq \frac{\lambda}{2} = \omega(1)$ *with high probability,*
- *The number of activated children $\boldsymbol{\omega}_v$ satisfies $\omega_v \geq \frac{\lambda}{2} = \omega(1)$ with high probability.*

Proof. This is an immediate consequence of Chernoff bounds applied to the $\mathrm{Po}(\lambda)$ and, respectively, $\mathrm{Po}(\lambda p)$ distribution given that $\lambda p \to \infty$. □

As a next observation, we claim that if $\boldsymbol{\omega}_v$ (as given by Lemma 9) Galton-Watson processes with offspring distribution $\mathrm{Po}(\lambda p)$ are initialized independently, at least $(1-\varepsilon)$-fraction will survive with high probability for any $\varepsilon > 0$.

Lemma 10. *Suppose that X Galton-Watson processes with offspring distribution $\mathrm{Po}(\gamma)$ start independently under the condition that $\gamma \to \infty$. Let $\boldsymbol{Y}$ denote the number of such processes that did not ultimately become extinct. Then, $\mathbb{P}(\boldsymbol{Y} \geq (1 - o(1))X) \geq 1 - o(1)$.*

Proof. By Proposition 8, the probability that one specific process out of the X processes gets extinct is $p_e = o(1)$. Thus, as in the d-regular case, we have that the number of not-extinct processes is $\mathrm{Bin}(X, 1 - p_e)$-distributed. The lemma follows from Chernoff bounds. □

Proof of Theorem 3 (iii). As in the d-regular case, the previous lemmas imply that with high probability $\omega(1)$ of the processes started in the children of $\boldsymbol{\omega}$ are still active. Suppose that $\boldsymbol{\omega} \neq \boldsymbol{\omega}_c$ and $\mathrm{dist}(\boldsymbol{\omega}, \boldsymbol{\omega}_c) = k \geq 1$. This implies that either k times only one active child is spawned or, if multiple children are spawned, only exactly one rumor spreading process rooted in those children survives eventually. But by Lemmas 9 and 10, we find that this probability is $o(1)$ for all $k \geq 1$. □

5 DISCUSSION

Comparison with the results of Zhu et al. [2017]. Let us briefly discuss how our results of Theorem 2 extend and strengthen the results obtained by Zhu et al. [2017]. They use a Jordan centrality based estimator and show that if $(d-1)p > 1$, their algorithm outputs a vertex v which is with high probability close to $\boldsymbol{\omega}$. Moreover, the probability that their algorithm outputs a vertex at distance k from $\boldsymbol{\omega}$ decreases polynomially in k. Our Theorem 2 is based on a different estimator which enables us to refine and extend these results. We split the case $(d-1)p > 1$ into two cases. More precisely, if $(d-1)p \gg 1$ our algorithm returns w.h.p. the correct source. If $1 < (d-1)p = O(1)$ the returned node is the correct source with constant probability. Furthermore, we show that the probability for the algorithm to output a node at distance k of $\boldsymbol{\omega}$ decreases exponentially fast in k (opposed to polynomial decrease). Finally, in Theorem 2 (iii), we also prove an impossibility result. Inference is impossible with high probability if $(d-1)p < 1$.

Conclusion. We pin down exact information-theoretic phase-transitions in the source detection task on important tree-network models by proving that as soon as weak detection is possible information-theoretically, the efficient closest candidate heuristics succeeds at this task. Those findings imply, of course, the same result for $\mathbb{G}(n, p)$ and random d-regular graphs as long as they are sufficiently sparse and the spreading process ran for only $o(\ln(n))$ steps as those random networks are then, locally, given by the described tree-networks with high probability.

Furthermore, we show empirically that the source detection heuristic performs well on non-acyclic networks and seems to be a very decent and efficient estimator of the rumor's source. A natural question is whether it is possible to prove similar information-theoretic bounds for non-acyclic networks. While this seems to be a very challenging task in general, it might become accessible if we restrict ourselves to specific random networks or networks with a specific tree-width.

Finally, on the empirical side, it is an interesting question whether the rumor's source of the ICM can be learned by graph neural networks. This seems challenging as only few vertices are active and the network would need to learn possible propagation paths in a graph given only a snapshot of the network.

References

Ruben Becker, Federico Corò, Gianlorenzo D'Angelo, and Hugo Gilbert. Balancing spreads of influence in a social network. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020*, pages 3–10. AAAI Press, 2020. URL https://ojs.aaai.org/index.php/AAAI/article/view/5327.

Indaco Biazzo, Alfredo Braunstein, Luca Dall'Asta, and Fabio Mazza. Epidemic inference through generative neural networks. *CoRR*, abs/2111.03383, 2021. URL https://arxiv.org/abs/2111.03383.

Jacopo Bindi, Alfredo Braunstein, and Luca Dall'Asta. Predicting epidemic evolution on contact networks from partial observations. *PLOS ONE*, 12:1–28, 2017. doi: 10.1371/journal.pone.0176376.

Fred Brauer. Mathematical epidemiology: Past, present, and future. *Infectious Disease Modelling*, 2(2):113–127, 2017. doi: 10.1016/j.idm.2017.02.001.

Tom Britton, Svante Janson, and Anders Martin-Löf. Graphs with specified degree distributions, simple epidemics, and local vaccination strategies. *Advances in Applied Probability*, 39(4):922–948, 2007.

Amin Coja-Oghlan, Oliver Gebhard, Max Hahn-Klimroth, and Philipp Loick. Optimal group testing. In *Conference on Learning Theory, COLT*, volume 125 of *Proceedings of Machine Learning Research*, pages 1374–1388. PMLR, 2020. URL http://proceedings.mlr.press/v125/coja-oghlan20a.html.

Quinlan Dawkins, Tianxi Li, and Haifeng Xu. Diffusion source identification on networks with statistical confidence. In *Proceedings of the 38th International Conference on Machine Learning, ICML*, volume 139 of *Proceedings of Machine Learning Research*, pages 2500–2509. PMLR, 2021. URL http://proceedings.mlr.press/v139/dawkins21a.html.

Elizabeth Gardner and Bernard Derrida. Three unfinished works on the optimal storage capacity of networks. *Journal of Physics A: Mathematical and General*, 22(12): 1983, 1989. doi: 10.1088/0305-4470/22/12/004.

William Goffman and V. Newill. Generalization of epidemic theory: An application to the transmission of ideas. *Nature*, 204(4955):225–228, 1964. doi: 10.1038/204225a0.

Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes*. Oxford University Press, New York, 4 edition, 2020. ISBN 978-0-198-84760-1.

Alankar Jain, Vivek S. Borkar, and Dinesh Garg. Fast rumor source identification via random walks. *Soc. Netw. Anal. Min.*, 6(1):62:1–62:13, 2016. doi: 10.1007/s13278-016-0373-6.

Feng Ji and Wee Peng Tay. An algorithmic framework for estimating rumor sources with different start times. *IEEE Trans. Signal Process.*, 65(10):2517–2530, 2017. doi: 10.1109/TSP.2017.2659643.

S. Jalil Kazemitabar and Arash Ali Amini. Approximate identification of the optimal epidemic source in complex networks. *CoRR*, abs/1906.03052, 2019. URL http://arxiv.org/abs/1906.03052.

J Matt Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008. doi: 10.1515/9781400841035.

David Kempe, Jon M. Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. *Theory Comput.*, 11:105–147, 2015. doi: 10.4086/toc.2015.v011a004.

William Ogilvy Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, 115(772): 700–721, 1927. doi: 10.1098/rspa.1927.0118.

J. F. C. Kingman. *Poisson Processes*. Oxford University Press, 1993.

Kristina Lerman and Rumi Ghosh. Information contagion: An empirical study of the spread of news on digg and twitter social networks. In *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM*. The AAAI Press, 2010. URL http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1509.

Jure Leskovec, Mary McGlohon, Christos Faloutsos, Natalie S. Glance, and Matthew Hurst. Patterns of cascading behavior in large blog graphs. In *Proceedings of the Seventh SIAM International Conference on Data Mining*, pages 551–556. SIAM, 2007. doi: 10.1137/1.9781611972771.60.

Andrey Y. Lokhov. Reconstructing parameters of spreading models from partial observations. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems*, pages 3459–3467, 2016.

Pablo Martin, Eduardo Rojas, Jorge Olivares, and Adrián Sotomayor. Quasi-rational analytic approximation for the modified bessel function i1(x) with high accuracy. *Symmetry*, 13(5), 2021. doi: 10.3390/sym13050741. URL https://www.mdpi.com/2073-8994/13/5/741.

Atalanti-Anastasia Mastakouri and Bernhard Schölkopf. Causal analysis of covid-19 spread in germany. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS*, 2020.

Peter Neal. SIR epidemics on a bernoulli random graph. *Journal of Applied Probability*, 40(3):779–782, 2003.

M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili. Theory of rumour spreading in complex social networks. *Physica A: Statistical Mechanics and its Applications*, 374(1):457—470, 2007. doi: 10.1016/j.physa.2006.07.017.

B. Aditya Prakash, Jilles Vreeken, and Christos Faloutsos. Efficiently spotting the starting points of an epidemic in a large graph . *Knowl. Inf. Syst.*, 38(1):35–59, 2014. doi: 10.1007/s10115-013-0671-5.

Adam Sadilek, Henry A. Kautz, and Vincent Silenzio. Modeling spread of disease from social interactions. In *Proceedings of the Sixth International Conference on Weblogs and Social Media*. The AAAI Press, 2012. URL `http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4493`.

Chintan Shah, Nima Dehmamy, Nicola Perra, Matteo Chinazzi, Albert-László Barabási, Alessandro Vespignani, and Rose Yu. Finding patient zero: Learning contagion source with graph neural networks. *CoRR*, abs/2006.11913, 2020. URL `https://arxiv.org/abs/2006.11913`.

Devavrat Shah and Tauhid Zaman. Detecting sources of computer viruses in networks: theory and experiment. In *SIGMETRICS 2010, Proceedings of the 2010 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pages 203–214. ACM, 2010. doi: 10.1145/1811039.1811063.

Devavrat Shah and Tauhid Zaman. Rumor centrality: a universal source detector. In *ACM SIGMETRICS/PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems, SIGMETRICS '12*, pages 199–210. ACM, 2012. doi: 10.1145/2254756.2254782.

Xincheng Shu, Bin Yu, Zhongyuan Ruan, Qingpeng Zhang, and Qi Xuan. Information source estimation with multichannel graph neural network. In Qi Xuan, Zhongyuan Ruan, and Yong Min, editors, *Graph Data Mining: Algorithm, Security and Application*, pages 1–27. Springer, 2021. doi: 10.1007/978-981-16-2609-8_1.

Zheng Wang, Chaokun Wang, Jisheng Pei, and Xiaojun Ye. Multiple source detection without knowing the underlying propagation model. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 217–223. AAAI Press, 2017. URL `http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14159`.

JA Yorke and HW Hethcote. *Gonorrhea: Transmission dynamics and control. Lecture notes in Biomathematics, 56, 1–105*. Springer-Verlag, Berlin, 1984. ISBN 978-3-662-07544-9. doi: 10.1007/978-3-662-07544-9.

Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016. doi: 10.1080/00018732.2016.1211393.

Jichang Zhao, Junjie Wu, Xu Feng, Hui Xiong, and Ke Xu. Information propagation in online social networks: a tie-strength perspective. *Knowl. Inf. Syst.*, 32(3):589–608, 2012. doi: 10.1007/s10115-011-0445-x.

Kai Zhu and Lei Ying. Information source detection in networks: Possibility and impossibility results. In *35th Annual IEEE International Conference on Computer Communications, INFOCOM*, pages 1–9. IEEE, 2016. doi: 10.1109/INFOCOM.2016.7524363.

Kai Zhu, Zhen Chen, and Lei Ying. Catch'em all: Locating multiple diffusion sources in networks with partial observations. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 1676–1683. AAAI Press, 2017. URL `http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14658`.