Finding Super-spreaders in Network Cascades

Elchanan Mossel* Anirudh Sridhar†

March 8, 2024

Abstract

Suppose that a cascade (e.g., an epidemic) spreads on an unknown graph, and only the infection times of vertices are observed. What can be learned about the graph from the infection times caused by multiple distinct cascades? Most of the literature on this topic focuses on the task of recovering the *entire* graph, which requires $\Omega(\log n)$ cascades for an n-vertex bounded degree graph. Here we ask a different question: can the important parts of the graph be estimated from just a few (i.e., constant number) of cascades, even as n grows large?

In this work, we focus on identifying super-spreaders (i.e., high-degree vertices) from infection times caused by a Susceptible-Infected process on a graph. Our first main result shows that vertices of degree greater than $n^{3/4}$ can indeed be estimated from a constant number of cascades. Our algorithm for doing so leverages a novel connection between vertex degrees and the second derivative of the cumulative infection curve. Conversely, we show that estimating vertices of degree smaller than $n^{1/2}$ requires at least $\log(n)/\log\log(n)$ cascades. Surprisingly, this matches (up to $\log\log n$ factors) the number of cascades needed to learn the *entire* graph if it is a tree.

1 Introduction

Cascading behaviors in networks are ubiquitous. In human interaction networks, the spread of diseases through local interactions can quickly escalate, causing population-level pandemics [7]. Similarly, the structure of social networks plays a fundamental role in the mass dissemination of ideas, such as product adoption or misinformation [32, 47, 44, 22]. In all of these examples, understanding the key factors which facilitate cascades is crucial for the analysis and control of these processes.

The structure of the underlying network plays a fundamental role in the evolution of a cascading process. Consider, for instance, the impact of a high-degree vertex. Once such a vertex is affected by the cascade, they may, in turn, disseminate the cascade's effects on a massive scale to their neighbors. Due to their potential to accelerate the course of a cascade in this manner, high-degree vertices are of special interest in the analysis of spreading processes. In the context of epidemics, quarantining or immunizing high-degree individuals, also known as *super-spreaders*, can greatly mitigate the damage caused by a pandemic [40, 16]. In social networks, firms may want to leverage the impact of high-degree vertices, also known as *influencers*, to accelerate the speed of product adoption [25].

A fundamental hurdle in carrying out these ideas in practice is that the underlying network may be partially or fully unknown. In epidemiology, disease-spreading contacts between individuals are often not measured. In marketing, even if retailers have access to certain social networks, only some of the observed links may be effective in facilitating a cascade, and many links may also be unobserved (e.g., if two individuals communicate through other means). While it is possible to learn aspects of the network through contact tracing or interviewing, such approaches are quite intensive and time-consuming [46]. To mitigate this issue, we aim to understand the extent to which more efficient *data-driven* approaches can be used to learn about the network. Specifically, we study how network structure can be estimated from the set of "infection times" caused by a cascade, which we call the *cascade trace*. Typically, a single cascade trace may not be enough to learn meaningful information, but much more can be said by combining information from multiple different cascades on the same network.

*Department of Mathematics, MIT; elmos@mit.edu
†Department of Mathematics, MIT; anisri@mit.edu

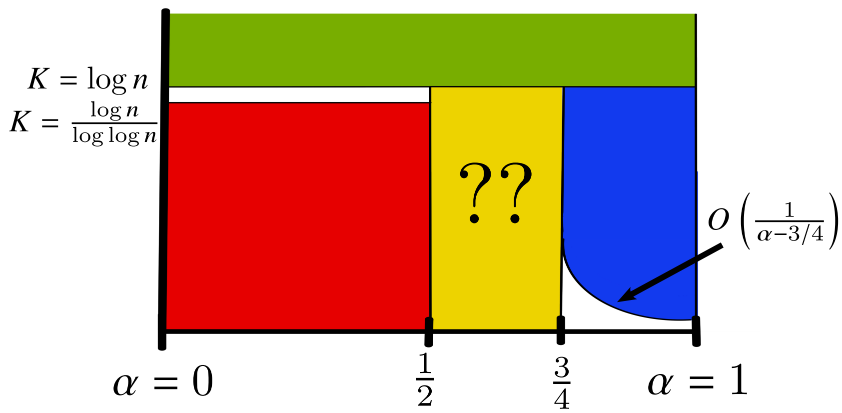


Figure 1: Phase diagram for the possibility and impossibility of estimating high-degree vertices in a graph G. *Blue region:* High-degree vertices can be estimated from $O((\alpha - 3/4)^{-1})$ traces. *Red region:* Estimating high-degree vertices is impossible, even when G is known to be a tree. *Green region:* Full recovery of G is possible if G is a tree, hence estimation of high-degree vertices is also possible. *White region:* Regimes with small gap between the bounds provided by our analysis. *Yellow region:* The main open problem left given our work is to determine the sample complexity in this region.

The task of estimating a network from multiple cascade traces, known as structure learning, has received considerable attention in the past decade [23, 19, 39, 1, 15, 30, 27, 24, 26, 51]. Almost all the literature on this topic focuses on exactly recovering the *entire* network under sparsity constraints (i.e., bounds on the maximum degree). However, the number of cascade traces required to do so grows as a function of the network size [39, 1, 15]. This can be prohibitive in applications such as epidemiology where it may be costly to allow many cascades to spread. Motivated by this limitation, we ask a different question:

Can the <u>*important*</u> *parts of the network, such as the location of high-degree vertices,*
be learned from just a few cascade traces?

We find that the answer is quite subtle. In an n-vertex graph, we show that vertices of degree at least $n^{3/4}$ can indeed be correctly estimated from a constant number of cascade traces. Conversely, estimating vertices of degree smaller than $\sqrt{n}$ is almost as hard as learning the *entire* network, in terms of the number of cascade traces required.

1.1 Summary of contributions

In a bit more detail, we assume that each cascade spreads on an n-vertex graph G with the following structure: a constant number of (high-degree) vertices have degree at least n^α for some $\alpha \in (0, 1)$, and all other vertices have degree at most $n^{o(1)}$. The goal is to understand how many cascade traces are needed to estimate the set of high-degree vertices with probability $1 - o(1)$.

Our first main contribution is an estimator for the set of high-degree vertices which requires only a constant number of cascade traces. Our estimator is quite simple, and relies only on properties of the *infection curve* $I(t)$, which counts the number of infections that occur before time t. The key idea behind our method is a novel connection between vertex degrees and the second derivative of the infection curve. Specifically, we show that if v becomes infected at time $T(v)$, then a discretized version of the second derivative of $I(t)$ evaluated at time $T(v)$ is a (nearly) unbiased estimator for the degree of v. Motivated by this insight, our estimator computes the (appropriately discretized) second derivative of $I(t)$ at $T(v)$ for each v, and compares it to a threshold. The set of estimated high-degree vertices is then given by all vertices which pass the threshold in at least half of the observed cascade traces (see Algorithm 1 for more details). Our analysis shows that this estimator succeeds provided $\alpha > 3/4$, and we provide some evidence that this is the best that can be achieved with our methods. Notably, our algorithm is a significant departure from existing methods in the literature on learning in graphical models. Indeed, prior algorithms typically focus on learning the precise edges and neighborhoods in the graph [39, 1, 15, 23, 19, 24, 51]. In contrast, we demonstrate how degree information can be *directly* estimated from just a few traces without needing to infer neighborhoods.

A natural follow-up question is whether one can identify high-degree vertices of degree much smaller than $n^{3/4}$, while still using only a constant number of traces. We provide a negative answer, showing that when $\alpha \in (0, 1/2)$, there exist "hard" instances for which $\log(n)/\log\log(n)$ traces are needed to estimate high-degree vertices, no matter the estimator. In other words, it is *information-theoretically impossible* to correctly estimate high-degree vertices in this case. In our proof, we show that estimating high-degree vertices in a particular ensemble of graphs is equivalent to a variant of the *sparse mixture detection problem* [17, 12, 28, 18, 14]. Following the approach of Cai and Wu [14], we solve our version of the sparse mixture detection problem, which leads to a lower bound of $\log(n)/\log\log(n)$ traces for the estimation of high-degree vertices. A surprising implication of our impossibility result is that for $\alpha \in (0, 1/2)$, estimating high-degree vertices is nearly as hard as learning the entire network. Indeed, using $\log(n)$ traces, the Tree Reconstruction algorithm of Abrahao, Chierichetti, Kleinberg and Panconesi [1] is able to exactly recover G (provided it is a tree). On the other hand, the graphs in our ensemble of hard instances are all trees, hence $\log(n)/\log\log(n)$ traces are required to estimate high-degree vertices even if G is known to be a tree. In summary, if G is known to be a tree, the sample complexity (with respect to the number of cascade traces) of estimating high-degree vertices matches that of learning G *exactly*, up to at most a $\log\log(n)$ factor.

Together, our two results highlight a curious phase transition in α (see Figure 1). When $\alpha \in (3/4, 1)$, high-degree vertices can be identified using essentially minimal data about the cascade. When $\alpha \in (0, 1/2)$, we have the other extreme: learning high-degree vertices is nearly as challenging as learning the entire network, when G is a tree. Our work leaves open the intriguing regime of $\alpha \in [1/2, 3/4]$.

1.2 Related Work

The problem of inferring a network from cascade traces was first studied in the context of epidemiology [49] and information flow in blogs [2]. Since then, there has been a large body of theoretical and empirical work on the subject. On the empirical side, Gomez-Rodriguez, Leskovec and Krause [23] initiated a study of scalable and mathematically principled methods with their NetInf algorithm. Since then, a number of follow-up works have tackled the exact network inference problem using a combination of likelihood-based approaches, convex optimization, message passing and machine learning [15, 23, 19, 24, 51]. In a related vein, there has been significant interest from the machine learning community in the task of learning Hawkes processes from vertex activation times; see [34] and references therein. We expect that our methods may be useful in this setting as well.

The first results on the sample complexity (i.e., number of traces needed) of exactly recovering the underlying graph were by Netrapalli and Sanghavi [39], which applied to discrete-time cascades (i.e., the independent cascade model) spreading on bounded-degree graphs. This work was soon followed by Abrahao, Chierichetti, Kleinberg and Panconesi [1], who studied the sample complexity of exact network inference using a variant of the SI process to model the cascade. Since their work, a number of extensions have been explored, including inference from noisy infection times [27], correlated cascades [30] and learning mixtures of graphs [26]. A common technique underlying these is an analysis of maximum likelihood estimator for the underlying network. We note that our algorithm is a significant departure from these methods, as it is able to directly estimate degree information *without* knowledge of the precise topology.

Our work falls under the broader umbrella of literature studying the sample complexity of learning structures in graphical models. Our work deals with observations from *transient* models, though learning from *stationary* models such as Gaussian Graphical Models and Ising models have also received considerable attention [42, 35, 36, 50, 13, 29, 31, 52, 48, 45, 8, 11]. While such frameworks are naturally quite different from ours, there are also many qualitative similarities. For the task of recovering the precision matrix of a n-dimensional Gaussian graphical model, it is known that $\Theta(\log n)$ independent samples are necessary and sufficient under some sparsity conditions [36, 50, 13, 29]. Similarly, $\Theta(\log n)$ independent samples are necessary and sufficient for learning bounded-width Ising models [10, 52, 48, 45, 8] as well as general classes of Markov Random Fields [11]. These results parallel the sample complexity of exactly recovering the graph from cascade traces, and many algorithmic approaches are similar in flavor.

While the work above largely focuses on recovering the underlying network, we mention a few recent papers which also aim to learn the "important" parts of the relevant graphical model, though with significantly different interpretations of importance. Boix-Adserà, Bresler and Koehler [6] as well as Bresler and Karzand [9] focus on the task of learning tree-structured Ising models with a good prediction accuracy with respect to the ground-truth model. Eckles, Esfandiari, Mossel and Rahimian [20] leverage cascade traces from a discrete-time cascade model to perform influence maximization. These works, in conjunction with ours, demonstrate the potential to learn key information about distributions and systems *without* fully learning the underlying model.

1.3 Definitions and Main Results

Notation. For a set S, we let $|S|$ denote its cardinality. For a graph $G = (V, E)$, we denote the number of vertices by $n := |V|$ and let $\mathcal{N}(v)$ denote the neighborhood of a vertex v. For a subset of vertices $S \subset V$, we denote $\mathsf{cut}(S) := |\{(u, v) \in E : u \in S, v \notin S\}|$. Throughout the paper, we use standard asymptotic notation and all limits are as $n \to \infty$ unless otherwise specified.

The Susceptible-Infected process. We model the cascade using the well-known Susceptible-Infected (SI) process, which is a continuous-time Markov process in which susceptible vertices in a graph become infected upon interacting with infected neighbors. In more detail, we let G be the graph upon which the cascade spreads. We denote by $\mathcal{I}(t)$ the set of vertices which are infected at a time index $t \geq 0$, and we let $\lambda > 0$ represent the *interaction rate* between individuals. Initially, we assume that only a single vertex v_0 is infected; that is, $\mathcal{I}(0) = \{v_0\}$. Given $\mathcal{I}(t)$, the cascade evolves as follows: for $v \in V \setminus \mathcal{I}(t)$, it holds that as $\epsilon \to 0$,

$$\mathbb{P}\left(v \in \mathcal{I}(t+\epsilon) \middle| \mathcal{F}_t \right) = \epsilon \lambda |\mathcal{N}(v) \cap \mathcal{I}(t)| + o(\epsilon), \tag{1.1}$$

where $\{\mathcal{F}_t\}_{t \geq 0}$ is the natural filtration corresponding to the stochastic evolution of the cascade and $o(\epsilon) \to 0$ faster than $\epsilon \to 0$. In words, the rate at which a susceptible vertex becomes infected is proportional to the number of its infected neighbors and the interaction rate λ. Given the evolution of the cascade $\mathcal{I}(t)$, a few important quantities include the *infection curve* $I(t) := |\mathcal{I}(t)|$ as well as $I[a, b]$ for $0 \leq a \leq b$, which denotes the number of infections occurring in the interval $[a, b]$.

We remark that the SI process (1.1) is closely related to a number of other well-studied models. It is known to be equivalent to First Passage Percolation with $\mathrm{Exp}(\lambda)$ edge weights, a popular model in mathematical physics [5, 43]. It is also equivalent to the Contact Process without recovery [33, 37], and has been used as a rigorous foundation for standard population-based epidemiological models [7, Chapter 9]. In recent years, a large body of work has also investigated generalizations of (1.1) with edge-dependent interaction rates (e.g., due to mask-wearing tendencies, multiple viral strains); see, e.g., [4, 3, 21].

For a given vertex $v \in V$, we formally define its *infection time* in the SI process to be

$$T(v) := \inf\{t \geq 0 : v \in \mathcal{I}(t)\}.$$

We call the collection of infection times $\mathbf{T} := \{T(v)\}_{v \in V}$ the *cascade trace.*

Learning from cascade traces. Suppose that the underlying graph G is unknown, but we observe multiple independent cascade traces on G, given formally by

$$(\mathbf{T}_1, \ldots, \mathbf{T}_K) \sim \mathcal{SI}(G, v_{0,1}) \otimes \ldots \otimes \mathcal{SI}(G, v_{0,K}), \tag{1.2}$$

where $v_{0,i}$ is the source vertex for the ith cascade. For brevity, for a graph G and collection of source vertices $v_0 := (v_{0,1}, \ldots, v_{0,K})$, we denote $\mathbb{P}_{G,v_0}$ to be the probability measure corresponding to (1.2). Broadly, we are interested in the following question of structural inference: *What can be learned about G from K cascade traces?*

While this has been a question of significant interest in the past decade, the existing literature on the topic almost exclusively focuses on *exactly* recovering G. In the context of cascade traces, this task was first studied by Netrapalli and Sanghavi [39] for a discrete-time cascade model. In particular, they proved that for n-vertex graphs G with maximum degree at most Δ, $K = \mathrm{poly}(\Delta) \log n$ is necessary and sufficient to exactly recover G. Abrahao, Chierichetti, Kleinberg and Panconesi [1] studied the exact recovery problem for a slightly more general version of the SI process (1.1), showing that $K = \text{poly}(\Delta)\log(n)$ suffices for general graphs (provided Δ does not scale as a function of n), and that $K = \Theta(\log n)$ traces suffices if G is a tree.

A key takeaway from this literature on exact recovery of G is that the number of cascade traces K must grow as a function of the network size n. However, this requirement may be prohibitive in settings such as epidemiology where one must make decisions based on a relatively small amount of data. In light of this issue, we study whether it is possible to learn the *important* parts of a network – such as the location of high-degree vertices – from just a few cascade traces. To the best of our knowledge, this angle has not been previously investigated.

Estimating high-degree vertices. To formalize the task of estimating high-degree vertices, we assume that most of the vertices in G have a relatively small degree, except for a constant number of vertices with substantially larger degree. The latter type is the set of high-degree vertices we wish to estimate. We mathematically capture this structural property with the class of graphs $\mathcal{G}(n, m, d, D)$, described below.

Definition 1.1. *We say that $G \in \mathcal{G}(n, m, d, D)$ if and only if (1) G is a connected graph on n vertices, (2) there are at most m vertices of degree at least D, (3) all other vertices have degree at most d and (4) no two vertices of degree at least D are neighbors. We denote the set of high-degree vertices in $G \in \mathcal{G}(n, m, d, D)$ to be* $\mathsf{highdeg}(G) := \{v \in V : \deg(v) \geq D\}$.

We remark that the assumption of connectivity ensures that all vertices are eventually infected, which is in general necessary for the estimation of high-degree vertices. Indeed, if a high-degree vertex never becomes infected, it is impossible to estimate it. Additionally, a subtle aspect of $\mathcal{G}(n, m, d, D)$ is the requirement that no two high-degree vertices are neighbors. This is a technical condition that is important for bounding the speed of the cascade (see, e.g., Lemma 3.7 in Section 3). Next, we assume the following about the scaling of the parameters m, d, D with respect to n.

Assumption 1.2. *We assume that m is constant, $d = n^{o(1)}$ and $D = n^\alpha$ for some $\alpha \in (0, 1)$.*

We remark that this assumption is quite different from the usual assumption made in the literature on structure learning. Indeed, in much of the work on exact recovery of G, it is assumed that its maximum degree is bounded as $n \to \infty$ [39, 1, 15]. A notable exception is the Tree Reconstruction algorithm of Abrahao, Chierichetti, Kleinberg, and Panconesi [1], which has provable guarantees for *any* tree with a sufficiently large number of vertices. However, as we are only interested in estimating high-degree vertices rather than specific connections in the network, we may take significantly more relaxed assumptions compared to the bulk of the structure learning literature.

Our first main result shows that we can correctly identify all vertices in G of degree larger than $n^{3/4}$, using a *constant* number of cascade traces – far less than what is needed to learn the entire graph.

Theorem 1.3. *Suppose that $\alpha \in (3/4, 1)$ and that $D = n^\alpha$. If K satisfies*

$$K \geq \frac{15}{\alpha - 3/4},$$

then there is an estimator $\widehat{\mathrm{HD}}$ such that for any $G \in \mathcal{G}(n, m, d, D)$ and any collection of source vertices $v_0 = (v_{0,1}, \ldots, v_{0,K})$,

$$\mathbb{P}_{G,v_0}\left(\widehat{\mathrm{HD}}\left(\mathbf{T}_1, \ldots, \mathbf{T}_K\right) = \mathsf{highdeg}(G)\right) = 1 - o(1),$$

where $o(1) \to 0$ as $n \to \infty$.

Though Theorem 1.3 shows that a constant number of traces suffices for any $\alpha > 3/4$, K grows as α approaches $3/4$. While this condition is needed in our analysis to handle worst-case examples of G, we remark that in practice K could be smaller and may not explode as α approaches $3/4$. To prove the theorem, we develop a novel algorithm for estimating high-degree vertices from a constant number of cascade traces (Algorithm 1). This algorithm and its underlying principles are described in detail in Section 2.1.

Our second main contribution is a negative result, showing that *any* estimator requires at least $\log(n)/\log\log(n)$ traces to successfully identify high-degree vertices of degree smaller than $\sqrt{n}$.

Theorem 1.4. *Fix $\epsilon > 0$, let $\alpha \in (0, 1/2)$. Suppose that $D = n^\alpha$, $d \geq \log^2 n$ and $m = 1$. Additionally, assume that*

$$K \leq \left(\frac{1 - 2\alpha - \epsilon}{5}\right) \frac{\log n}{\log \log n}. \tag{1.3}$$

Then there exists an ensemble $\mathcal{H} \subset \mathcal{G}(n, m, d, D)$ and a probability distribution μ over graphs in $\mathcal{H}$ and source vertices $v_0 = (v_{0,1}, \ldots, v_{0,K})$ such that, for any estimator HD*,*

$$\mathbb{P}_{G,v_0 \sim \mu} \left(\mathrm{HD}(\mathbf{T}_1, \ldots, \mathbf{T}_K) = \mathsf{highdeg}(G)\right) = o(1).$$

An important implication of the theorem is that if K is smaller than $\log(n)/\log\log(n)$, no estimator can successfully identify high-degree vertices for every graph in the class $\mathcal{G}(n, m, d, D)$. The ensemble $\mathcal{H}$ used in the theorem is a collection of trees with a particular structure; see Sections 2.2 and 6 for more details. This observation, in the context of prior work on the sample complexity of exact recovery of G [1], shows that learning high-degree vertices when $\alpha \in (0, 1/2)$ is *nearly* as hard as learning the full graph in the case of trees. To emphasize this point, we rephrase the result of [1] below.

Theorem 1.5. *Suppose that $K \geq C \log n$, for some universal constant $C > 0$. Then there exists an estimator $\widehat{G}$ such that for any collection of source vertices $v_0 = (v_{0,1}, \ldots, v_{0,K})$ and any tree G, it holds that $\mathbb{P}_{G,v_0}(\widehat{G} = G) = 1 - o(1)$.*

Together, Theorems 1.4 and 1.5 show that there is at most a $\log \log n$ multiplicative gap between the sample complexity of estimating high-degree vertices for $\alpha \in (0, 1/2)$ and that of fully recovering G, in the case where G is a tree.

1.4 Discussion and future work

In this work, we study the sample complexity of estimating super-spreaders (i.e., high-degree vertices) in network cascades, using only the infection times of vertices across multiple cascades. This is the first work to consider this task, to the best of our knowledge. Our results show that estimating vertices with degree at least $n^{3/4}$ can be done using a constant number of cascades, while estimating vertices with degree at most $n^{1/2}$ requires almost as many cascades as learning the entire graph (if the graph is a tree). We discuss several important avenues for future work below.

- **The regime $\alpha \in [1/2, 3/4]$.** In terms of sample complexity, our results show that the regime $\alpha \in (3/4, 1)$ is "easy", i.e., the number of traces needed is minimal. On the other hand, $\alpha \in (0, 1/2)$ captures a "hard" regime, where the sample complexity nearly matches that of learning the entire graph exactly. What is the right sample complexity for $\alpha \in [1/2, 3/4]$? Is it in the easy regime, the hard regime, or something in between?

- **Estimation in general graphs when $\alpha \in (0, 1/2)$.** The Tree Reconstruction algorithm of [1] provides a means of estimating high-degree vertices when the graph is a tree. What about estimating high-degree vertices for general graphs in $\mathcal{G}(n, m, d, D)$?

- **Noisy observations.** In practice, the infection times may be noisy due to delays in reporting or due to quantization (e.g., positive cases being consolidated on a day-by-day basis). To what extent can our methods be adapted to handle these more realistic scenarios?

- **Other motifs.** A high-degree vertex can be viewed as a particular type of motif. What is the sample complexity of detecting and estimating other types of motifs (e.g., cliques)?

- **Other types of graphical models.** Our methods fundamentally leverage the temporal nature of the observed data. How can high-degree vertices be estimated in stationary graphical models, such as Gaussian graphical models or Ising models?

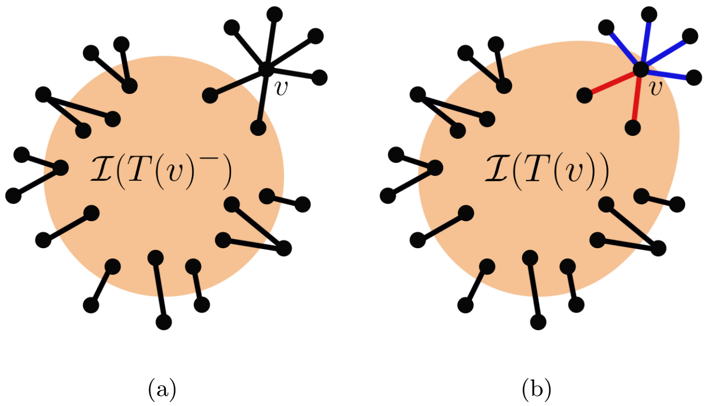


Figure 2: Visualization of the edges contributing to $\mathsf{cut}(\mathcal{I}(t))$ before (a) and after (b) a vertex v becomes infected. In Figure 2b, the number of blue and red edges denote the positive and negative change to $\mathsf{cut}(\mathcal{I}(t))$, respectively, upon v being infected.

2 Technical overview

In this section, we provide a detailed overview of our methods and proof techniques. Section 2.1 motivates our algorithm for estimating high-degree vertices in the regime $\alpha \in (3/4, 1)$ (Algorithm 1) and provides a proof sketch of Theorem 1.3. In Section 2.2, we discuss the key ideas behind our impossibility result (Theorem 1.4). In particular, we elaborate on the connection between the sparse mixture detection problem and the impossibility of estimating high-degree vertices when $\alpha \in (0, 1/2)$.

2.1 Estimating high-degree vertices when $\alpha > 3/4$

The starting point for the design of our estimator is the following simple observation: when a high-degree vertex becomes infected, many of its neighbors will become infected shortly after. In a bit more detail, we could assess whether a vertex v is high-degree by examining the number of infections occurring in a δ-size interval after v is infected, given formally by $I[T(v), T(v) + \delta]$. Since we can write $I[T(v), T(v) + \delta] = I(T(v) + \delta) - I(T(v))$, the quantity $I[T(v), T(v) + \delta]$ can be thought of as the discrete first derivative of the infection curve $I(t)$. This suggests that high-degree vertices could be estimated by examining the points at which the first derivative of $I(t)$ is large. Unfortunately, such an approach fails in general. Indeed, if we observe many infections in the interval $[T(v), T(v) + \delta]$, then it is challenging to discern whether the infections were mainly caused by a *single*, common infection (namely, v), or by many other infected vertices which just happened to be infected around the same time as v. However, it turns out that if one considers *higher-order* information – specifically, the *second* derivative of $I(t)$ – then high-degree vertices can be clearly identified.

To see this concretely, we begin by examining the first derivative of $I(t)$ in more detail. By (1.1),

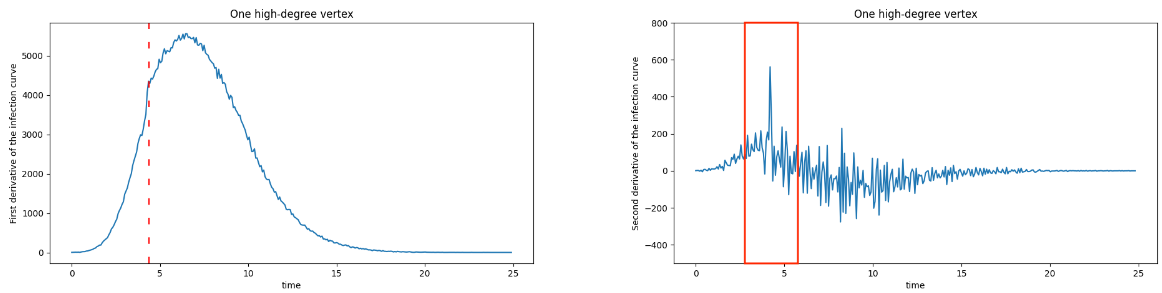
$$J(t) := \mathbb{E}\left[\frac{d}{dt} I(t) \middle| \mathcal{F}_t \right] = \lambda \sum_{v \in V \setminus \mathcal{I}(t)} |\mathcal{N}(v) \cap \mathcal{I}(t)| = \lambda \, \mathsf{cut}(\mathcal{I}(t)).$$

Now suppose that a vertex v joins the cascade at time t (that is, $T(v) = t$). Since no other vertices can be infected at the exact same time due to the continuous-time nature of the cascade, we have

$$J(t) - J(t^-) = \lambda \left(|\mathcal{N}(v) \setminus \mathcal{I}(t)| - |\mathcal{N}(v) \cap \mathcal{I}(t)| \right) = \lambda \deg(v) - 2\lambda |\mathcal{N}(v) \cap \mathcal{I}(t)|. \tag{2.1}$$

The first equality in (2.1) can be explained as follows. When v joins the cascade, all of its susceptible neighbors, i.e., those in $\mathcal{N}(v) \setminus \mathcal{I}(t)$, become exposed. On the other hand, the edges which connect v to $\mathcal{I}(t)$ – namely, edges which connect v to $\mathcal{N}(v) \cap \mathcal{I}(t)$ – can no longer lead to infection events, hence the contribution for these edges is subtracted. See Figure 2 for a visualization of this idea.

Crucially, (2.1) shows that as long as a vertex is infected before most of its neighborhood (i.e., $|\mathcal{N}(v) \cap \mathcal{I}(t)|$ is much smaller than $\deg(v)$), the second-order changes in the infection curve, captured by $J(t) - J(t^-)$, can



(a) (b)

Figure 3: Plots of the discrete first (a) and second derivative (b) of the infection curve with $\delta = 0.075$ generated from a graph G with approximately 500,000 vertices and one high-degree vertex. The infection time of the high-degree vertex can be identified by the red dotted line in (a) and the large peak in the second derivative plot (b), highlighted by the red rectangle. G was chosen to be a balanced, 5-regular tree of height 8, where one vertex in the 6th layer of the tree has degree 7500.

reveal the degree of a vertex. However, since $J(t)$ is not a directly observable quantity, we must approximate it from the cascade traces. A natural way to do so is by choosing $\delta > 0$ and using the discrete derivative $I[t, t+\delta]/\delta$ as a proxy for $J(t)$. Similarly, we can use $I[t-\delta, t]/\delta$ as a proxy for $J(t^-)$. We can therefore approximate the degree of v using

$$\widehat{\deg}_\delta(v) := \frac{I[T(v), T(v)+\delta] - I[T(v)-\delta, T(v)]}{\delta}.$$

In words, we estimate the degree of v by computing the discrete second derivative of the infection curve when v gets infected. As shown through the empirical example in Figure 3, this statistic is quite effective in identifying the infection time of a high-degree vertex from just a single cascade.

In light of (2.1), our estimator for $\mathsf{highdeg}(G)$ returns the set of all vertices v for which $\widehat{\deg}_\delta(v)$ is large in most cascade traces; see Algorithm 1. As the Theorem 2.1 shows, this method succeeds with high probability provided $\alpha > 3/4$.

Algorithm 1 Finding high-degree vertices via second derivative thresholding

Input: Cascade traces $\mathbf{T}_1, \ldots, \mathbf{T}_K$ and parameters $\delta, \tau > 0$
Output: A set $\widehat{\mathrm{HD}} \subset V$

1: For each $i \in [K]$ and $v \in V$ compute

$$\widehat{\deg}_{\delta,i}(v) := \frac{I_i[T_i(v), T_i(v)+\delta] - I_i[T_i(v)-\delta, T_i(v)]}{\delta}.$$

2: $\widehat{\mathrm{HD}} \leftarrow \emptyset$.
3: For each $v \in V$: If $\widehat{\deg}_{\delta,i}(v) \geq \tau$ for at least $K/2$ cascades, then $\widehat{\mathrm{HD}} \leftarrow \widehat{\mathrm{HD}} \cup \{v\}$.
4: Return $\widehat{\mathrm{HD}}$.

Theorem 2.1. *Suppose that $\alpha > 3/4$ and that*

$$\delta = \frac{1}{n^{\alpha - 1/4}}, \qquad \tau = \frac{n^\alpha}{\log n}, \qquad K \geq \frac{15}{\alpha - 3/4}.$$

Additionally assume that $d = n^{o(1)}$ and $D = n^\alpha$. Then for any $G \in \mathcal{G}(n, m, d, D)$, the output of Algorithm 1 is equal to $\mathsf{highdeg}(G)$*, with probability $1 - o(1)$.*

Remark 2.2. *A useful feature of Algorithm 1 is that it is adaptive to the unknown parameters m (the number of high-degree vertices) and λ (the spreading rate of the cascade). In contrast, the prior work on structure learning leverages likelihood-based methods which require precise knowledge of the cascade model [39, 1].*

We note that Theorem 1.3 follows immediately from Theorem 2.1 as a corollary. We provide a sketch of the proof of Theorem 2.1 below, and defer the full details to Sections 4 and 5.

Proof sketch of Theorem 2.1. Let $\gamma < \alpha$ be sufficiently close to α. The strategy of our proof is to show, for all $v \in V$, that

$$\left|\widehat{\deg}_{\delta,i}(v) - \lambda \deg(v)\right| \leq n^{\gamma}, \tag{2.2}$$

for at least half of the cascades. If (2.2) is true for a particular $i \in [K]$ and v is a low-degree vertex, then $\widehat{\deg}_{\delta,i}(v) \leq \lambda \deg(v) + n^{\gamma} \leq 2n^{\gamma}$. Else if v is high-degree, then $\widehat{\deg}_{\delta,i}(v) \geq \lambda \deg(v) - n^{\gamma} \geq \frac{\lambda}{2} n^{\alpha}$. If λ is not known a priori, any choice of threshold satisfying $\tau = n^{\alpha - o(1)}$ works in distinguishing the two cases. In particular, $\tau = n^{\alpha}/\log(n)$ suffices.

Due to the way $\widehat{\deg}_{\delta,i}(v)$ is computed, establishing (2.2) reduces to showing that $I[t, t+\delta]/\delta$ concentrates around $J(t)$ and $I[t, t-\delta]/\delta$ concentrates around $J(t^{-})$ for the values of t corresponding to vertices' infection times. For this task, choosing the right value of δ is crucial. If δ is too small, then $I[t, t+\delta]$ may also be quite small, making meaningful concentration impossible. On the other hand, if δ is too large, then there may be a large number of low-degree vertices that consistently infected around the same time as a high-degree vertex. If u is one such low-degree vertex, then $\widehat{\deg}_{\delta}(u)$ can be quite large across multiple traces, leading to false positives.

Examining these conditions in more detail exposes why we must assume $\alpha > 3/4$ for Algorithm 1 to succeed. Let us first consider the behavior of $I[t, t+\delta]/\delta$ (the analysis of $I[t-\delta, t]/\delta$ is similar). Since $I[t, t+\delta]$ is derived from a counting process, we can deduce a subgaussian-type concentration of $I[t, t+\delta]$ around $\delta J(t)$ (see Lemma 5.3 in Section 5.1). It follows that

$$I[t, t+\delta] = \delta J(t) \pm O\left(\sqrt{\delta J(t)}\right), \tag{2.3}$$

where the second term represents the random fluctuations around $J(t)$. A direct consequence of (2.3) is that the deviation of $I[t, t+\delta/\delta]/\delta$ from $J(t)$ is at most $O(\sqrt{J(t)/\delta})$. In order to correctly estimate high-degree vertices through (2.1), this deviation needs to be smaller than $D = n^{\alpha}$. Since all but a $o(1)$ fraction of vertices in G has degree at most $d = n^{o(1)}$, it follows from (2.1) that $J(t) \leq n^{1+o(1)}$. Hence we require that $\sqrt{n^{1+o(1)}/\delta} = o(n^{\alpha})$, which is equivalent to

$$\delta = \omega\left(n^{-(2\alpha-1)-o(1)}\right). \tag{2.4}$$

Next, we need to ensure that there is no low-degree vertex u that is infected around the same time as a high-degree vertex in at least half the cascade traces. Otherwise, as we mentioned earlier, $\widehat{\deg}_{\delta,i}(u)$ could be quite large, which would imply that the output of Algorithm 1 contains false positives. With this scenario in mind, suppose that a low-degree u becomes infected first at time t, and that point in time, a high-degree v has M infected neighbors. By (1.1), it holds for small δ that

$$\mathbb{P}(|T(u) - T(v)| \leq \delta | \mathcal{F}_t) = \mathbb{P}(v \in \mathcal{I}(t+\delta) | \mathcal{F}_t) = \delta M + o(\delta).$$

The probability that u and v are close in at least half of the cascade traces is therefore at most $O((\delta M)^{K/2})$. A union bound over pairs of vertices shows that

$$\mathbb{P}(\exists u, v \in V : |T_i(u) - T_i(v)| \leq \delta \text{ in at least half the cascades}) \leq O\left(n^2 (\delta M)^{K/2}\right). \tag{2.5}$$

To show that the right hand side of (2.5) is $o(1)$, we first need to choose δ such that $\delta M = o(1)$. Although M can in principle be as large as $\deg(v)$ (i.e., if all of v's neighbors are infected before v), we show that with high probability $M \leq \sqrt{n}$ (see Lemma 3.7 in Section 3). Hence we may choose δ satisfying

$$\delta = o\left(n^{-1/2}\right). \tag{2.6}$$

Writing $\delta = n^{-\beta}$ for $\beta > 0$, (2.4) and (2.6) imply that $\beta \in (1/2, 2\alpha - 1)$. This interval is nonempty when $\alpha > 3/4$, and we may choose $\beta := \alpha - 1/4$ to satisfy (2.4) and (2.6). Finally, since $\delta = n^{-(\alpha - 1/4)}$ and $M \leq \sqrt{n}$, $K \geq \Omega((\alpha - 3/4)^{-1})$ is needed for the right hand side of (2.5) to be $o(1)$. $\square$

2.2 Impossibility of estimation when $\alpha \in (0, 1/2)$

Connections to sparse mixture detection. The cornerstone of our proof is a correspondence between high-degree estimation and the problem of sparse mixture detection. In a bit more detail, let G' be a graph on $n/2$ vertices of maximum degree $n^{o(1)}$ (i.e., all its vertices are low-degree). The graph G will be formed by adding an additional $n/2$ vertices labelled $1, \ldots, n/2$ to G', along with a single edge per added vertex which connects to a vertex in G'. A natural way of generating a graph G *without* any high-degree vertices is to let each of the new $n/2$ vertices connect to a uniform random parent in G'. Through standard probabilistic arguments, it can be seen that no more than $O(\log n)$ neighbors are added to any particular vertex in G', so G contains no high-degree vertices.

A simple modification of this procedure can also generate graphs with a high-degree vertex. Suppose we fix a vertex v in G'. For each of the additional $n/2$ vertices, with probability $2D/n$, it connects to v. Otherwise, with probability $1 - 2D/n$, it connects to a uniform random vertex in G'. In expectation, v receives D additional connections (so, in particular, $\deg(v) \geq D$ in expectation). As before, all of the other vertices receive at most $O(\log n)$ additional neighbors, with high probability. Hence we obtain a graph G in which v is the only high-degree vertex.

Observe that in the second procedure, the distribution of the infection times of each of the added $n/2$ vertices can be viewed as a mixture of two distributions. To be precise, let P be the distribution of a vertex's infection times in the observed cascade traces if they connect to a uniform random neighbor in G'. Similarly, let Q be the distribution of a vertex's infection times if they connect to v. Then the task of differentiating between the two generative models for G – or equivalently, of testing whether v is high-degree – can be phrased as the following hypothesis testing problem:

$$H_0 : \mathbf{T}(1), \ldots \mathbf{T}(n/2) \sim P \qquad H_1 : \mathbf{T}(1), \ldots, \mathbf{T}(n/2) \sim ((1 - \epsilon_n) P + \epsilon_n Q), \tag{2.7}$$

where we set $\epsilon_n := 2D/n$ above, and we denote $\mathbf{T}(j) := (T_1(j), \ldots, T_K(j))$ where $T_i(j)$ is the infection time of vertex j in the ith cascade.

It turns out that (2.7) is a special case of a well-studied problem known as *sparse mixture detection*, which dates back to the work of Dobrusin [17]. Burnashev [12] studied a variant related to Gaussian processes, and Ingster [28] identified a subtle phase transition between the possibility and impossibility of detection for mixtures of Gaussian distributions. The problem was popularized by the work of Donoho and Jin [18], who formulated an adaptive method in place of the likelihood ratio called *higher criticism*, again for mixtures of Gaussians. Later, Cai and Wu [14] developed an analysis of the sparse mixture problem for generic distributions. We remark that Mossel and Roch [38] used sparse mixture detection in the inference of combinatorial structure, though in a very different context.

Impossibility of detecting high-degree vertices. We solve our instance of the sparse mixture detection problem (2.7) using the techniques developed by Cai and Wu [14]. That is, we bound the total variation distance between the two distributions in (2.7) by the χ^2 divergence since it tensorizes and because it is easier to study for problems involving mixtures. Our analysis shows that if K satisfies (1.3), then the χ^2 divergence tends to zero. As a result, it is impossible to distinguish between scenarios where there are no high-degree vertices (H_0) and where v is a high-degree vertex (H_1).

Putting everything together. Finally, we show that the impossibility of detection implies that no estimator can successfully output the correct high-degree vertex with probability greater than $o(1)$. We remark that the ensemble $\mathcal{H}$ and distribution μ stated in the theorem essentially corresponds to the random constructions for G described in this subsection. For more details, see Section 6.

3 Preliminaries: Cascade dynamics and infection times

In this section we derive several useful results on the behavior of the SI process (1.1) and the distribution of infection times. Throughout, we will assume that $\mathcal{I}(t)$ is a realization of $\mathcal{SI}(G, v_0)$ for a particular graph $G \in \mathcal{G}(n, m, d, D)$ and vertex $v_0 \in V$.

We begin by discussing an important equivalence between (1.1) and First Passage Percolation on G with $\mathrm{Exp}(\lambda)$ edge weights. In a bit more detail, let $\{\mathcal{F}_t\}_{t\geq 0}$ denote the natural filtration corresponding to the SI process. Fix $t \geq 0$, and let $\{F(e)\}_{e\in E}$ be a collection of i.i.d. $\mathrm{Exp}(\lambda)$ random variables that are independent of $\mathcal{F}_t$. For a path[1] P in the graph, we define the *weight* of the path to be $\mathsf{weight}(P) := \sum_{e\in P} F(e)$. Then we have the following result.

Proposition 3.1. *Let $t \geq 0$, and condition on $\mathcal{F}_t$. Suppose that $v \in V \setminus \mathcal{I}(t)$. Let $\mathcal{P}(v,t)$ be the collection of paths from $\mathcal{I}(t)$ to v where only the first edge in the path has an endpoint in $\mathcal{I}(t)$. Then the conditional distribution of $T(v)$ with respect to $\mathcal{F}_t$ is given by*

$$T(v) = t + \min_{P\in\mathcal{P}(v,t)} \mathsf{weight}(P).$$

This result is standard, and essentially follows from the memoryless property of exponential distributions. We defer the reader to [5, Chapter 6], or to Section 7 for full details of the proof.

Using the representation in Proposition 3.1, we can derive a number of useful results on the infection times of vertices. We start by establishing probabilistic upper and lower bounds on conditional infection times of vertices.

Lemma 3.2. *Let $t \geq 0$, and condition on $\mathcal{F}_t$. Suppose that $v \in V \setminus \mathcal{I}(t)$. Then $T(v) - t$ is stochastically dominated by a $\mathrm{Exp}(\lambda|\mathcal{N}(v) \cap \mathcal{I}(t)|)$ random variable.*

Proof. Conditioned on $\mathcal{F}_t$, Proposition 3.1 implies that

$$T(v) - t = \min_{P\in\mathcal{P}(v,t)} \mathsf{weight}(P) \leq \min_{P\in\mathcal{P}(v,t):|P|=1} \mathsf{weight}(P) = \min_{u\in\mathcal{N}(v)\cap\mathcal{I}(t)} F(u,v).$$

Due to properties of exponential random variables, we have that

$$\min_{u\in\mathcal{N}(v)\cap\mathcal{I}(t)} F(u,v) \sim \mathrm{Exp}\left(\lambda|\mathcal{N}(v) \cap \mathcal{I}(t)|\right),$$

which proves the claimed result. □

Lemma 3.3. *Let $t \geq 0$, condition on $\mathcal{F}_t$, and suppose that $v \in V \setminus \mathcal{I}(t)$. Then $T(v) - t$ stochastically dominates a $\mathrm{Exp}(\lambda|\mathcal{N}(v)|)$ random variable.*

Proof. Conditioned on $\mathcal{F}_t$, we have that

$$T(v) - t = \min_{P\in\mathcal{P}(v,t)} \mathsf{weight}(P) \geq \min_{u\in\mathcal{N}(v)} F(u,v),$$

where the inequality follows since any path ending in v must include one of the edges incident to v. The claim follows since $\min_{u\in\mathcal{N}(v)} F(u,v) \sim \mathrm{Exp}(\lambda|\mathcal{N}(v)|)$. □

We next define a useful event concerning the number of infected neighbors of high-degree vertices. We remark that the control of this quantity is especially important for us, as it is equal to the bias of our estimate of vertex degrees (see (2.1)).

Definition 3.4. *The event $\mathcal{A}$ holds if and only if*

$$|\mathcal{N}(v) \cap \mathcal{I}(T(v))| \leq 2d\sqrt{n}\log^2 n,$$

for all $v \in V$.

Remark 3.5. *As we assume throughout this paper that $d = n^{o(1)}$, we will often use that $|\mathcal{N}(v) \cap \mathcal{I}(T(v))| \leq n^{1/2+o(1)}$ when using the event $\mathcal{A}$.*

Remark 3.6. *Assuming that $G \in \mathcal{G}(n,m,d,D)$, the event $\mathcal{A}$ is only relevant for high-degree vertices, since*

$$|\mathcal{N}(v) \cap \mathcal{I}(T(v))| \leq \deg(v) = n^{o(1)} \leq 2d\sqrt{n}\log^2 n,$$

otherwise.

[1] Recall that a path is a finite sequence of edges $e_1, \ldots, e_m$ such that for each $i \geq 1$, e_i and e_{i+1} share exactly one endpoint.

Lemma 3.7. *Suppose that $m = o(\log^2 n)$. Then $\mathbb{P}(\mathcal{A}) = 1 - o(1)$.*

Proof. If $\deg(v) \le 2d\sqrt{n}\log^2 n$, then

$$|\mathcal{N}(v) \cap \mathcal{I}(T(v))| \le |\mathcal{N}(v)| \le 2d\sqrt{n}\log^2 n,$$

which satisfies the condition in $\mathcal{A}$ corresponding to v. It therefore suffices to consider the case $\deg(v) \ge 2d\sqrt{n}\log^2 n$. Define the event

$$\mathcal{E}_v := \left\{|\mathcal{N}(v) \cap \mathcal{I}(T(v))| \ge 2d\sqrt{n}\log^2 n\right\},$$

and let T be the stopping time at which $\sqrt{n}\log n$ neighbors of v are first infected.

Let $t \ge 0$, and condition on $\mathcal{F}_t$. Let us also assume that $T = t$ (an event that is $\mathcal{F}_t$-measurable), so that $|\mathcal{N}(v) \cap \mathcal{I}(t)| = \sqrt{n}\log n$. If $T(v) \le t$, then we have that $\mathbb{P}(\mathcal{E}_v|\mathcal{F}_t) = 0$. Otherwise, Lemma 3.2 implies that $T(v) - t$ is stochastically dominated by a $\mathrm{Exp}(\lambda\sqrt{n}\log n)$ random variable. Hence

$$\mathbb{P}\left(T(v) \ge t + \frac{1}{\lambda\sqrt{n}} \,\middle|\, \mathcal{F}_t\right) \le \frac{1}{n}. \tag{3.1}$$

Next, we upper bound the number of neighbors of v that become infected in the interval $[t, t + 1/(\lambda\sqrt{n})]$. Since no two high-degree vertices are neighbors for $G \in \mathcal{G}(n, m, d, D)$ (see Definition 1.1), all neighbors of v have degree at most d. It follows that for $u \in \mathcal{N}(v) \setminus \mathcal{I}(t)$, we have that

$$\begin{aligned}\mathbb{P}\left(\{T(u) \le T(v)\} \cap \left\{T(v) < t + \frac{1}{\lambda\sqrt{n}}\right\} \,\middle|\, \mathcal{F}_t\right) &\le \mathbb{P}\left(T(u) < t + \frac{1}{\lambda\sqrt{n}} \,\middle|\, \mathcal{F}_t\right)\\ &\le 1 - e^{-d/\sqrt{n}} \le \frac{d}{\sqrt{n}}.\end{aligned}$$

Above, the first inequality on the second line follows since $T(u) - t$ stochastically dominates a $\mathrm{Exp}(\lambda d)$ random variable, in light of Lemma 3.3. Next, summing over all $u \in \mathcal{N}(v) \setminus \mathcal{I}(t)$ shows that

$$\mathbb{E}\left[|(\mathcal{N}(v) \cap \mathcal{I}(T(v))) \setminus \mathcal{I}(t)|\mathbf{1}\left(T(v) < t + \frac{1}{\lambda\sqrt{n}}\right) \,\middle|\, \mathcal{F}_t\right] \le \frac{d|\mathcal{N}(v) \setminus \mathcal{I}(t)|}{\sqrt{n}} \le d\sqrt{n},$$

where the final inequality uses $|\mathcal{N}(v)| \le n$. As a consequence of the previous display, we have that

$$\begin{aligned}&\mathbb{E}\left[|\mathcal{N}(v) \cap \mathcal{I}(T(v))|\,\mathbf{1}\left(T(v) < t + \frac{1}{\lambda\sqrt{n}}\right) \,\middle|\, \mathcal{F}_t\right]\\ &\quad\le |\mathcal{N}(v) \cap \mathcal{I}(t)| + \mathbb{E}\left[|(\mathcal{N}(v) \cap \mathcal{I}(T(v))) \setminus \mathcal{I}(t)|\mathbf{1}\left(T(v) < t + \frac{1}{\lambda\sqrt{n}}\right) \,\middle|\, \mathcal{F}_t\right]\\ &\quad\le \sqrt{n}\log n + d\sqrt{n} \le 2d\sqrt{n}\log n,\end{aligned}$$

where, in the final expression, we have used our assumption $T = t$. Next, Markov's inequality implies

$$\mathbb{P}\left(\mathcal{E}_v \cap \left\{T(v) < t + \frac{1}{\lambda\sqrt{n}}\right\} \,\middle|\, \mathcal{F}_t\right) \le \frac{1}{\log n}.$$

The display above combined with (3.1) shows that $\mathbb{P}(\mathcal{E}_v|\mathcal{F}_t) \le 1/\log(n) + 1/n \le 2/\log(n)$.

Next, since $T < \infty$ almost surely (since all vertices in the graph become infected in finite time), it holds that

$$\mathbb{P}(\mathcal{E}_v|\mathcal{F}_T) = \int_0^\infty \mathbb{P}(\mathcal{E}|\mathcal{F}_t)\mathbf{1}(T = t)dt,$$

which, in light of the bound $\mathbb{P}(\mathcal{E}_v|\mathcal{F}_t) \le 2/\log(n)$, implies that $\mathbb{P}(\mathcal{E}_v|\mathcal{F}_T) \le 2/\log(n)$. It follows that $\mathbb{P}(\mathcal{E}_v) = \mathbb{E}[\mathbb{P}(\mathcal{E}_v|\mathcal{F}_T)] \le 2/\log(n)$. Finally, we note that

$$\mathcal{A}^c = \bigcup_{v \in V : \deg(v) \ge 2d\sqrt{n}\log^2 n} \mathcal{E}_v,$$

hence a union bound over these at most m events shows that

$$\mathbb{P}(\mathcal{A}^c) \le \frac{2m}{\log n} = o(1).$$

$\square$

As the following result shows, the event $\mathcal{A}$ is crucially used to obtain sharper characterizations of infection times than what Lemma 3.3 can provide, especially for high-degree vertices.

Lemma 3.8. *Let $v \in V$ and let $\delta > 0$ satisfy $\lambda^2\delta^2 n \le 1$. Fix $t \ge 0$, and let $v \in V \setminus \mathcal{I}(t)$. Then*

$$\mathbb{P}\left(\{T(v) \in [t, t+\delta]\} \cap \mathcal{A} \middle| \mathcal{F}_t\right) \le 3d\lambda\delta\sqrt{n}\log^2(n).$$

Proof. Define the event

$$\mathcal{A}_t := \left\{\text{For all } v \in V \text{ such that } T(v) > t, \text{ it holds that } |\mathcal{N}(v) \cap \mathcal{I}(t)| \le 2d\sqrt{n}\log^2 n\right\}.$$

Notice that $\mathcal{A}_t$ is $\mathcal{F}_t$-measurable, and $\mathcal{A} \subset \mathcal{A}_t$. Moreover, for v such that $T(v) > t$,

$$\mathbb{P}\left(\{T(v) \le t+\delta\} \cap \mathcal{A} | \mathcal{F}_t\right) \le \mathbb{P}\left(\{T(v) \le t+\delta\} \cap \mathcal{A}_t | \mathcal{F}_t\right) = \mathbb{P}(T(v) \le t+\delta | \mathcal{F}_t)\mathbf{1}(\mathcal{A}_t). \tag{3.2}$$

We will therefore proceed by bounding $\mathbb{P}(T(v) \le t+\delta|\mathcal{F}_t)$ under the assumption that $\mathcal{A}_t$ holds.

Suppose that $\deg(v) \le d$. Then by Lemma 3.3, $T(v) - t$ stochastically dominates a $\mathrm{Exp}(\lambda d)$ random variable, so that

$$\mathbb{P}(T(v) \le t+\delta|\mathcal{F}_t) \le 1 - e^{-\delta\lambda d} \le \delta\lambda d. \tag{3.3}$$

Next, suppose that $\deg(v) \ge D$, and let $\mathcal{P}_2(v)$ be the set of paths of length 2 starting from v. Then by Proposition 3.1,

$$\begin{aligned} T(v) - t &= \min_{P \in \mathcal{P}(v,t)} \mathsf{weight}(P) && (3.4)\\ &\ge \min\left\{\min_{u \in \mathcal{N}(v)\cap\mathcal{I}(t)} F(u,v), \min_{P \in \mathcal{P}_2(v)} \mathsf{weight}(P)\right\}, && (3.5) \end{aligned}$$

which can be explained as follows. If the minimizing path $P \in \mathcal{P}(v,t)$ in (3.4) has length 1, then the minimum is achieved by $F(u,v)$ for some $u \in \mathcal{N}(v) \cap \mathcal{I}(t)$. Otherwise, the minimizing path in (3.4) has length at least 2, which must include a path of length 2 starting at v. In light of (3.5),

$$\begin{aligned} \mathbb{P}(T(v) \le t+\delta|\mathcal{F}_t) &\le \mathbb{P}\left(\min\left\{\min_{u \in \mathcal{N}(v)\cap\mathcal{I}(t)} F(u,v), \min_{P \in \mathcal{P}_2(v)} \mathsf{weight}(P)\right\} \le \delta \middle| \mathcal{F}_t\right)\\ &\le \sum_{u \in \mathcal{N}(v)\cap\mathcal{I}(t)} \mathbb{P}(F(u,v) \le \delta|\mathcal{F}_t) + \sum_{P \in \mathcal{P}_2(v)} \mathbb{P}(\mathsf{weight}(P) \le \delta|\mathcal{F}_t)\\ &\overset{(a)}{\le} |\mathcal{N}(v) \cap \mathcal{I}(t)|\lambda\delta + |\mathcal{P}_2(v)|\lambda^2\delta^2\\ &\overset{(b)}{\le} 2d\lambda\delta\sqrt{n}\log^2(n) + d\lambda^2\delta^2 n\\ &\overset{(c)}{\le} 3d\lambda\delta\sqrt{n}\log^2(n), \end{aligned} \tag{3.6}$$

where the inequality (a) follows since $F(u,v) \sim \mathrm{Exp}(\lambda)$ and $\mathsf{weight}(P) \sim \mathrm{Gamma}(2,\lambda)$ for $P \in \mathcal{P}_2(v)$. The inequality (b) uses the bound $|\mathcal{N}(v) \cap \mathcal{I}(t)| \le 2d\sqrt{n}\log^2 n$ which holds on the event $\mathcal{A}_t$, and also uses that the number of paths of length 2 starting from v is at most dn, since we assume that high-degree vertices are not neighbors in G (see Definition 1.1) and the degree of any vertex is at most n. Finally, (c) holds by our assumption that $\lambda^2\delta^2 n \le 1$. $\square$

4 Analysis of Algorithm 1: Proof of Theorem 2.1

The main result behind the performance guarantees of Algorithm 1 is the following lemma, which shows that $\widehat{\deg}_{\delta,i}(v)$ is a good approximation for $\lambda \deg(v)$ in at least half of the cascades, for all $v \in V$.

Lemma 4.1. *Suppose that β, γ satisfies $\beta > 1/2$ and $\gamma > (1+\beta)/2$. Additionally assume*

$$K \geq \frac{15}{\beta - 1/2}.$$

Then with probability $1 - o(1)$, for each $v \in V$ there exists a set $S_v \subset [K]$ such that $|S_v| \geq \frac{K}{2}$ and

$$\left|\widehat{\deg}_{\delta,i}(v) - \lambda \deg(v)\right| \leq n^\gamma, \qquad \forall i \in S_v.$$

Before proving the lemma, we show how it can be used to prove Theorem 2.1.

Proof of Theorem 2.1. Let $\gamma \in (\alpha/2 + 3/8, \alpha)$. It is readily verified that the choice of α, β, γ, K satisfies the assumptions of Lemma 4.1. In the case where $v \in \mathsf{highdeg}(G)$, it holds that for all $i \in S_v$,

$$\widehat{\deg}_{\delta,i}(v) \geq \lambda \deg(v) - n^\gamma \geq \lambda n^\alpha - n^\gamma \geq \frac{n^\alpha}{\log n}.$$

In particular, $\widehat{\deg}_{\delta,i}(v) \geq n^\alpha / \log(n)$ for at least half of the cascades. On the other hand, if $\deg(v) \leq d$, it holds for all $i \in S_v$ that

$$\widehat{\deg}_{\delta,i}(v) \leq \lambda \deg(v) + n^\gamma < \frac{n^\alpha}{\log n}.$$

In particular, $\widehat{\deg}_{\delta,i}(v) < n^\alpha / \log(n)$ for at least half of the cascades. The correctness of Algorithm 1 follows. $\square$

In the rest of this section, we focus on proving Lemma 4.1 through a discretization argument. Formally, we define the (finite) set of time indices

$$\mathbb{T}_n := \left\{ \frac{k}{n^5} : k \in \mathbb{Z}_{\geq 0} \right\} \cap [0, n^2].$$

For $t \in \mathbb{T}_n$, let us also define the ground-truth discretized second derivative

$$dI_\beta(t) := \lambda \left(\mathsf{cut}(\mathcal{I}(t)) - \mathsf{cut}\left(\mathcal{I}\left(t - \frac{1}{n^5} \right) \right) \right),$$

as well as the empirical second derivative

$$\widehat{dI_\beta}(t) := n^\beta \left(I\left[t, t + \frac{1}{n^\beta}\right] - I\left[t - \frac{1}{n^\beta}, t\right] \right).$$

Note in particular that $\widehat{dI_\beta}(T(v)) = \widehat{\deg}_\delta(v)$ with $\delta = n^{-\beta}$. Our strategy is to show that for any *fixed* $t \in \mathbb{T}_n$, $\widehat{dI_{\beta,i}}(t)$ is close to $dI_{\beta,i}(t)$ in at least half of the cascade traces. We will then take a union bound over all $t \in \mathbb{T}_n$. Finally, Lemma 4.1 follows from showing that the discretization in $\mathbb{T}_n$ is fine enough so that $\widehat{\deg}_{\delta,i}(v)$ is close to $\widehat{dI_{\beta,i}}(t)$ and $dI_{\beta,i}(t)$ is close to $\lambda \deg(v)$ for some $t \in \mathbb{T}_n$ and all $i \in [K]$.

To execute this strategy, we start by showing that $\widehat{dI_\beta}(t)$ and $dI_\beta(t)$ are close for a fixed $t \geq 0$, with high probability. We also recall the definition of the event $\mathcal{A}$ from Definition 3.4, which we use in the lemma.

Lemma 4.2. *Let $G \in \mathcal{G}$, and let β, γ satisfy $\beta > 1/2$ and $\gamma > (1+\beta)/2$. For any $t \in \mathbb{T}_n$, it holds that*

$$\mathbb{P}\left(\left\{ \left| \widehat{dI_\beta}(t) - dI_\beta(t) \right| \geq n^\gamma \right\} \cap \mathcal{A} \right) \leq n^{-(\beta - 1/2) + o(1)}.$$

Since the proof is somewhat involved, we defer the details to Section 5.

Notice that the probability bound in Lemma 4.2 is not small enough to handle a union bound over the $O(n^7)$ elements of $\mathbb{T}_n$. In the following result, we therefore consider multiple cascade traces to sufficiently reduce the probability of "bad" events.

Lemma 4.3. *Suppose that β, γ satisfies $\beta > 1/2$ and $\gamma > (1+\beta)/2$. Additionally assume*

$$K \geq \frac{15}{\beta - 1/2}.$$

Then with probability $1 - o(1)$, there exists a set $S_t \subset [K]$ for each $t \in \mathbb{T}_n$ such that $|S_t| \geq \frac{K}{2}$ and

$$\left|\widehat{dI_{\beta,i}}(t) - dI_{\beta,i}(t)\right| \leq n^{\gamma}, \qquad i \in S_t.$$

In words, it holds for all $t \in \mathbb{T}_n$ that in at least half of the cascade traces, $\widehat{dI_\beta}(t)$ is a good approximation for $dI_\beta(t)$. However, it need not be the case that the same cascade provides a good approximation for all t simultaneously.

Proof of Lemma 4.3. For $i \in [K]$, let $\mathcal{A}_i$ be the event $\mathcal{A}$ corresponding to the cascade $\mathcal{I}_i$. Observe that on the complement of the event described in Lemma 4.3, there must exist $t \in \mathbb{T}_n$ and a set $S \subset [K]$ with $|S| \geq K/2$ such that $|\widehat{dI_{\beta,i}}(t) - dI_{\beta,i}(t)| \geq n^\gamma$, for all $i \in S$. For a fixed $t \in \mathbb{T}_n$, we can bound this event as follows.

$$\begin{aligned}
&\mathbb{P}\left(\left\{\exists S \subset [K] : |S| \geq \frac{K}{2} \text{ and } \min_{i \in S} |\widehat{dI_{\beta,i}}(t) - dI_{\beta,i}(t)| \geq n^\gamma\right\} \cap \bigcap_{i \in [K]} \mathcal{A}_i\right) \\
&\quad \leq \sum_{S \subset [K] : |S| \geq K/2} \mathbb{P}\left(\left\{\min_{i \in S} |\widehat{dI_{\beta,i}}(t) - dI_{\beta,i}(t)| \geq n^\gamma\right\} \cap \bigcap_{i \in S} \mathcal{A}_i\right) \\
&\quad \stackrel{(a)}{=} \sum_{S \subset [K] : |S| \geq K/2} \prod_{i \in S} \mathbb{P}\left(\left\{|\widehat{dI_{\beta,i}}(t) - dI_{\beta,i}(t)| \geq n^\gamma\right\} \cap \mathcal{A}_i\right) \\
&\quad \stackrel{(b)}{\leq} \sum_{S \subset [K] : |S| \geq K/2} n^{-(\beta - 1/2)K/2 + o(1)} \stackrel{(c)}{=} o\left(n^{-7}\right).
\end{aligned}$$

Above, (a) follows from the independence of the K cascades, (b) uses Lemma 4.2, and (c) follows from our choice of K and since the number of subsets of $[K]$ is $2^K = n^{o(1)}$. Taking a union bound over the n^7 elements of $\mathbb{T}_n$ shows that

$$\mathbb{P}\left(\left\{\exists t \in \mathbb{T}, S \subset [K] : |S| \geq \frac{K}{2} \text{ and } \min_{i \in S} |\widehat{dI_{\beta,i}}(t) - dI_{\beta,i}(t)| \geq n^\gamma\right\} \cap \bigcap_{i \in [K]} \mathcal{A}_i\right) = o(1).$$

Finally, the desired result follows since

$$\mathbb{P}\left(\left(\bigcap_{i \in [K]} \mathcal{A}_i\right)^c\right) \leq \sum_{i=1}^K \mathbb{P}(\mathcal{A}_i^c) = o(1),$$

where the final expression is due to Lemma 3.7. $\square$

Next, we argue that $t \in \mathbb{T}_n$ can be replaced with $\{T(v)\}_{v \in V}$ in the statement of Lemma 4.3, which will prove Lemma 4.1. Towards this goal, we show that the discretization $\mathbb{T}_n$ is rich enough to capture the continuous-time dynamics of the cascade.

Lemma 4.4. *With probability $1 - o(1)$, for each $v \in V$ there exists a unique $t \in \mathbb{T}_n$ such that $t - 1/n^5 \leq T(v) \leq t$.*

Proof. We first show that $\max_{v\in V} T(v) \leq n^2$. From the representation in Proposition 3.1, we can upper bound $\max_{v\in V} T(v)$ by $\sum_{e\in E(G)} F(e)$, so we focus on characterizing the latter quantity. In particular, we have that

$$\mathbb{E}\left[\sum_{e\in E(G)} F(e)\right] = \frac{E(G)}{\lambda}.$$

We can upper bound the expectation by noting that $E(G) \leq dn + mn = n^{1+o(1)}$, which holds since there are at most n vertices of degree at most d and at most m vertices of degree at most n for $G \in \mathcal{G}(n,m,d,D)$. Markov's inequality now implies

$$\mathbb{P}\left(\max_{v\in V} T(v) \geq n^2\right) \leq n^{-1+o(1)}.$$

Hence there must exist some interval of the form $[t, t+1/n^5]$ which contains $T(v)$, for every $v \in V$.

We next show that in each interval of the form $[t, t+1/n^5]$, there exists at most one infection event. Condition on $\mathcal{F}_t$. If there are at least 2 infection events in $[t, t+1/n^5]$, then there must exist distinct $e_1, e_2 \in E(G)$ such that $F(e_1), F(e_2) \leq n^{-5}$ (if not, at most one vertex is infected in the interval). The probability of this occurring is at most

$$\binom{|E(G)|}{2}\left(1 - e^{-\lambda/n^5}\right)^2 \leq n^{2+o(1)}\left(\frac{\lambda}{n^5}\right)^2 = n^{-8+o(1)}.$$

We conclude by taking a union bound over the $O(n^7)$ elements of $\mathbb{T}_n$. $\square$

We are now ready to prove the main result.

Proof of Lemma 4.1. Assume that the events described in Lemmas 4.3 and 4.4 as well as $\mathcal{A}_i$ hold for all $i \in [K]$, which is the case with probability $1 - o(1)$. Additionally, for a given $v \in V$, let $t \in \mathbb{T}_n$ satisfy $t - n^{-5} \leq T_i(v) \leq t$. Then we have that

$$n^{-\beta}\left(\widehat{\deg}_{\delta,i}(v) - \widehat{dI_{\beta,i}}(t)\right) = 2I[T_i(v), t] - I\left[T_i(v) + \frac{1}{n^\beta}, t + \frac{1}{n^\beta}\right] - I\left[T_i(v) - \frac{1}{n^\beta}, t - \frac{1}{n^\beta}\right].$$

On the event described in Lemma 4.4, no other vertex is infected in the interval $[t - n^{-5}, t]$, hence $I[T_i(v), t] = 1$. Furthermore, since the interval $[T(v) + n^{-\beta}, t - n^{-\beta}]$ intersects at most two intervals of the form $[k/n^5, (k+1)/n^5]$, the second term on the right hand side is at most 2 in absolute value. The same argument holds for the interval $[T_i(v) - n^{-\beta}, t + n^{-\beta}]$. As a result, we have that

$$\left|\widehat{\deg}_{\delta,i}(v) - \widehat{dI_{\beta,i}}(t)\right| \leq 6n^\beta.$$

On the other hand, since v is the only vertex which becomes infected in the interval $[t - n^{-5}, t]$, we have that

$$\mathsf{cut}(\mathcal{I}_i(t)) = \mathsf{cut}(\mathcal{I}_i(T_i(v))) \text{ and } \mathsf{cut}\left(\mathcal{I}_i\left(t - \frac{1}{n^5}\right)\right) = \mathsf{cut}(\mathcal{I}(T_i(v)^-)).$$

As a result,

$$dI_{\beta,i}(t) = \lambda\left(\mathsf{cut}(\mathcal{I}_i(T_i(v))) - \mathsf{cut}(\mathcal{I}_i(T_i(v)^-))\right) = \lambda\left(\deg(v) - 2|\mathcal{N}(v) \cap \mathcal{I}_i(T_i(v))|\right),$$

where the second equality is due to (2.1).

On the event $\bigcap_{i\in[K]} \mathcal{A}_i$, it holds that $|\mathcal{N}(v) \cap \mathcal{I}_i(T_i(v))| \leq n^{1/2+o(1)}$. Putting everything together, it holds with probability $1 - o(1)$ that

$$\left|\widehat{\deg}_{\delta,i}(v) - \lambda\deg(v)\right| \leq \left|\widehat{dI_{\beta,i}}(t) - dI_{\beta,i}(t)\right| + 6n^\beta + n^{1/2+o(1)}.$$

In particular, for $i \in S_t$ we have that $|\widehat{\deg}_i(v) - \lambda\deg(v)| \leq 3n^\gamma$. As this holds for all $v \in V$, the desired result follows. $\square$

5 Concentration of the second derivative: Proof of Lemma 4.2

The proof of Lemma 4.2 relies on the following two results. Both results use the event $\mathcal{A}$ (see Definition 3.4).

Lemma 5.1. *Let $G \in \mathcal{G}$, let $\beta > 1/2$ and assume that $\gamma > (1-\beta)/2$. Then for any $t \geq 0$, it holds that*

$$\mathbb{P}\left(\left\{\left|I\left[t, t+\frac{1}{n^\beta}\right] - \frac{\lambda}{n^\beta}\mathsf{cut}(\mathcal{I}(t))\right| \geq n^\gamma\right\} \cap \mathcal{A}\right) \leq n^{-(\beta-1/2)+o(1)}.$$

Lemma 5.2. *Let $\beta > 1/2$ and $\gamma > \beta$. Then*

$$\mathbb{P}\left(\left\{\left|\mathsf{cut}\left(\mathcal{I}\left(t-\frac{1}{n^5}\right)\right) - \mathsf{cut}\left(\mathcal{I}\left(t-\frac{1}{n^\beta}\right)\right)\right| \geq n^\gamma\right\} \cap \mathcal{A}\right) \leq n^{-(\beta-1/2)+o(1)}.$$

Before proving the two lemmas, we show how they can be combined to prove the main result.

Proof of Lemma 4.2. Let $\gamma > \beta$, and define the events

$$\begin{aligned}
\mathcal{E}_1 &:= \left\{\left|I\left[t, t+\frac{1}{n^\beta}\right] - \frac{\lambda}{n^\beta}\mathsf{cut}(\mathcal{I}(t))\right| \leq n^{\gamma-\beta}\right\} \\
\mathcal{E}_2 &:= \left\{\left|I\left[t-\frac{1}{n^\beta}, t\right] - \frac{\lambda}{n^\beta}\mathsf{cut}\left(\mathcal{I}\left(t-\frac{1}{n^\beta}\right)\right)\right| \leq n^{\gamma-\beta}\right\} \\
\mathcal{E}_3 &:= \left\{\left|dI_\beta(t) - \lambda\left(\mathsf{cut}(\mathcal{I}(t)) - \mathsf{cut}\left(\mathcal{I}\left(t-\frac{1}{n^\beta}\right)\right)\right)\right| \leq n^\gamma\right\}.
\end{aligned}$$

On $\mathcal{E}_1 \cap \mathcal{E}_2$ it holds that

$$\left|\widehat{dI_\beta}(t) - \lambda\left(\mathsf{cut}(\mathcal{I}(t)) - \mathsf{cut}\left(\mathcal{I}\left(t-\frac{1}{n^\beta}\right)\right)\right)\right| \leq 2n^\gamma.$$

Additionally if $\mathcal{E}_3$ holds, we have by the triangle inequality that $|\widehat{dI_\beta}(t) - dI_\beta(t)| \leq 3n^\gamma = n^{\gamma+o(1)}$. It therefore suffices to study the probability of $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$. By Lemma 5.1, it holds for $\gamma > (1+\beta)/2$ that

$$\mathbb{P}\left((\mathcal{E}_1 \cap \mathcal{E}_2)^c \cap \mathcal{A}\right) \leq \mathbb{P}(\mathcal{E}_1^c \cap \mathcal{A}) + \mathbb{P}(\mathcal{E}_2^c \cap \mathcal{A}) \leq n^{-(\beta-1/2)+o(1)}. \tag{5.1}$$

It remains to study $\mathcal{E}_3$. Lemma 5.2 as well as the definition of $dI_\beta(t)$ implies that

$$\mathbb{P}(\mathcal{E}_3^c \cap \mathcal{A}) = \mathbb{P}\left(\left\{\left|\mathsf{cut}\left(\mathcal{I}\left(t-\frac{1}{n^5}\right)\right) - \mathsf{cut}\left(\mathcal{I}\left(t-\frac{1}{n^\beta}\right)\right)\right| \geq n^\gamma\right\} \cap \mathcal{A}\right) \leq n^{-(\beta-1/2)+o(1)}. \tag{5.2}$$

Together, the bounds (5.1) and (5.2) prove the lemma. □

The rest of this section is devoted to the proofs of the two lemmas. The proof of Lemma 5.1 is in Section 5.1, and the proof of Lemma 5.2 is in Section 5.2.

5.1 Approximating the cut: Proof of Lemma 5.1

We start by recalling and defining some notation. For $0 \leq a \leq b$, recall that $\mathcal{I}[a,b]$ denotes the set of vertices infected in the interval $[a,b]$. In particular, $I[a,b] = |\mathcal{I}[a,b]|$. For $t \geq 0$, we also define the *infection boundary*

$$\partial\mathcal{I}(t) := \{v \in V \setminus \mathcal{I}(t) : \mathcal{N}(v) \cap \mathcal{I}(t) \neq \emptyset\}.$$

In words, $\partial\mathcal{I}(t)$ is the set of vertices which are not yet infected at time t, but have at least one infected neighbor. Conditioned on $\mathcal{F}_t$, we partition $\mathcal{I}[t, t+n^{-\beta}] = \mathcal{J}_1 \cup \mathcal{J}_2$, where

$$\begin{aligned}
\mathcal{J}_1 &:= \left\{v \in \partial\mathcal{I}(t) : \min_{u \in \mathcal{N}(v)\cap\mathcal{I}[0,t]} F(u,v) \leq \frac{1}{n^\beta}\right\} \\
\mathcal{J}_2 &:= \left\{v \in \mathcal{I}\left[t, t+\frac{1}{n^\beta}\right] : P \in \mathcal{P}(v,t) \text{ and } \mathsf{weight}(P) \leq \frac{1}{n^\beta} \Rightarrow |P| \geq 2\right\} = \mathcal{I}\left[t, t+\frac{1}{n^\beta}\right] \setminus \mathcal{J}_1.
\end{aligned}$$

We proceed by studying the sizes of $\mathcal{J}_1$ and $\mathcal{J}_2$ separately. Our first result characterizes $|\mathcal{J}_1|$, essentially through an application of Bernstein's inequality.

Lemma 5.3. *Let G be a graph on n vertices, let $\beta > 1/2$ and assume that $\gamma > (1-\beta)/2$. Recall the event $\mathcal{A}$ from Definition 3.4. Then for any $t \geq 0$, it holds that*

$$\mathbb{P}\left(\left\{\left||\mathcal{J}_1| - \frac{\lambda}{n^\beta}\mathsf{cut}(\mathcal{I}(t))\right| \geq n^\gamma\right\} \cap \mathcal{A}\right) = o\left(\frac{1}{n}\right).$$

Proof. Define the event

$$\mathcal{A}_t := \left\{\text{For all } v \in V \setminus \mathcal{I}(t), \text{ it holds that } |\mathcal{N}(v) \cap \mathcal{I}(t)| \leq n^{1/2+o(1)}\right\}.$$

Note that $\mathcal{A}_t$ is $\mathcal{F}_t$-measurable, and $\mathcal{A} \subset \mathcal{A}_t$. In particular,

$$\begin{aligned}\mathbb{P}\left(\left\{\left||\mathcal{J}_1| - \frac{\lambda}{n^\beta}\mathsf{cut}(\mathcal{I}(t))\right| \geq n^\gamma\right\} \cap \mathcal{A}\middle| \mathcal{F}_t\right) &\leq \mathbb{P}\left(\left\{\left||\mathcal{J}_1| - \frac{\lambda}{n^\beta}\mathsf{cut}(\mathcal{I}(t))\right| \geq n^\gamma\right\} \cap \mathcal{A}_t\middle| \mathcal{F}_t\right)\\ &= \mathbb{P}\left(\left||\mathcal{J}_1| - \frac{\lambda}{n^\beta}\mathsf{cut}(\mathcal{I}(t))\right| \geq n^\gamma\middle| \mathcal{F}_t\right)\mathbf{1}(\mathcal{A}_t). \end{aligned} \tag{5.3}$$

We proceed by bounding the probability on the right hand side. For each $v \in \partial\mathcal{I}(t)$, notice that

$$\min_{u \in \mathcal{N}(v)\cap\mathcal{I}(t)} F(u,v) \sim \mathrm{Exp}(\lambda|\mathcal{N}(v) \cap \mathcal{I}(t)|),$$

so that

$$\mathbb{P}\left(v \in \mathcal{J}_1 | \mathcal{F}_t\right) = 1 - \exp\left(-\lambda n^{-\beta}|\mathcal{N}(v) \cap \mathcal{I}(t)|\right).$$

Moreover, by the independence of edge weights, we have that $|\mathcal{J}_1| \overset{d}{=} \sum_{v \in \partial\mathcal{I}(t)} X(v)$, where the $X(v)$'s are independent with $X(v) \sim \mathrm{Bern}\left(1 - e^{-\lambda n^{-\beta}|\mathcal{N}(v)\cap\mathcal{I}(t)|}\right)$. We can then bound

$$\begin{aligned}\mathbb{P}\left(||\mathcal{J}_1| - \mathbb{E}[|\mathcal{J}_1||\mathcal{F}_t]| \geq \frac{1}{2}n^\gamma\middle| \mathcal{F}_t\right) &\overset{(a)}{\leq} 2\exp\left(-\frac{n^{2\gamma}/8}{\sum_{v\in\partial\mathcal{I}(t)}(1 - \exp(-\lambda n^{-\beta}|\mathcal{N}(v)\cap\mathcal{I}(t)|)) + n^\gamma/6}\right)\\ &\overset{(b)}{\leq} 2\exp\left(-\frac{n^{2\gamma}/8}{n^{1-\beta+o(1)} + n^\gamma/6}\right)\\ &\overset{(c)}{=} o\left(\frac{1}{n}\right). \end{aligned} \tag{5.4}$$

Above, (a) is due to Bernstein's inequality. Step (b) uses the inequality

$$\sum_{v\in\partial\mathcal{I}(t)}\left(1 - e^{-\lambda n^{-\beta}|\mathcal{N}(v)\cap\mathcal{I}(t)|}\right) \leq \frac{\lambda}{n^\beta}\sum_{v\in\partial\mathcal{I}(t)}|\mathcal{N}(v)\cap\mathcal{I}(t)| \leq \frac{\lambda}{n^\beta}(dn + nm) = n^{1-\beta+o(1)}, \tag{5.5}$$

where the second inequality above holds since there are at most n vertices v with $|\mathcal{N}(v)| \leq d$ and m vertices v with $|\mathcal{N}(v)| \leq n$; the final expression uses that $d = n^{o(1)}$ and $m = n^{o(1)}$. Finally, (c) is due to our assumption that $\gamma > (1-\beta)/2$.

It remains to study $\mathbb{E}[|\mathcal{J}_1|\mathcal{F}_t]$. We have that

$$\begin{aligned}\left|\mathbb{E}[|\mathcal{J}_1||\mathcal{F}_t] - \frac{\lambda}{n^\beta}\mathsf{cut}(\mathcal{I}(t))\right| &= \left|\sum_{v\in\partial\mathcal{I}(t)}\left(1 - \exp(\lambda n^{-\beta}|\mathcal{N}(v)\cap\mathcal{I}(t)|) - \frac{\lambda}{n^\beta}|\mathcal{N}(v)\cap\mathcal{I}(t)|\right)\right|\\ &\leq \sum_{v\in\partial\mathcal{I}(t)}\left|1 - \exp(\lambda n^{-\beta}|\mathcal{N}(v)\cap\mathcal{I}(t)| - \frac{\lambda}{n^\beta}|\mathcal{N}(v)\cap\mathcal{I}(t)|\right|\\ &\overset{(c)}{\leq} \frac{\lambda^2}{n^{2\beta}}\sum_{v\in\partial\mathcal{I}(t)}|\mathcal{N}(v)\cap\mathcal{I}(t)|\\ &\overset{(d)}{\leq} \lambda^2 n^{1-2\beta+o(1)}\\ &\overset{(e)}{\leq} \frac{1}{2}n^\gamma. \end{aligned} \tag{5.6}$$

Above, (c) is due to Taylor's theorem (in particular, the inequality holds as long as $\lambda n^{-\beta}|\mathcal{N}(v) \cap \mathcal{I}(t)| = o(1)$, which is the case since $\mathcal{A}_t$ holds and $\beta > 1/2$). Step (d) follows from the bound $\sum_{v \in \partial \mathcal{I}(t)} |\mathcal{N}(v) \cap \mathcal{I}(t)| \leq n^{1+o(1)}$, which was proved in (5.5). Finally, (e) follows since $\gamma > (1-\beta)/2 > 1-2\beta$ (where the second strict inequality holds when $\beta > 1/2$).

Combining (5.3), (5.4) and (5.6), and then taking an expectation over $\mathcal{F}_t$ proves the lemma. $\square$

Our next result shows that with high probability, $\mathcal{J}_2 = \emptyset$.

Lemma 5.4. *Suppose that $\beta > 1/2$. Then for any $t \geq 0$, it holds that*

$$\mathbb{P}\left(\{|\mathcal{J}_2| \geq 1\} \cap \mathcal{A}\right) \leq n^{-\left(\beta - \frac{1}{2}\right) + o(1)}.$$

Proof. As a shorthand, let $\delta := n^{-\beta}$. Let $t \geq 0$, and condition on $\mathcal{F}_t$. Observe that if $\mathcal{J}_2 \neq \emptyset$, then there must exist a path P of length 2 and weight at most δ for which only the first vertex in P is an element of $\mathcal{I}(t)$. Additionally, if we define the event

$$\mathcal{B}_t := \{\forall v \in \mathsf{highdeg}(G), T(v) \notin (t, t+\delta]\},$$

then all vertices in P must have degree at most d. Otherwise, there would exist a path of weight at most δ connecting a vertex in $\mathcal{I}(t)$ to some $u \in \mathsf{highdeg}(G)$, which would imply that $T(u) \in (t, t+\delta]$. However, this cannot happen on the event $\mathcal{B}_t$. Importantly, $\mathcal{B}_t$ holds with high probability provided $\mathcal{A}$ holds; indeed, by Lemma 3.8 we have that

$$\mathbb{P}(\mathcal{A} \cap \mathcal{B}_t^c | \mathcal{F}_t) \leq \sum_{v \in \mathsf{highdeg}(G)} \mathbb{P}(\mathcal{A} \cap \{T(v) \in (t, t+\delta]\} | \mathcal{F}_t) \leq m n^{-\left(\beta - \frac{1}{2}\right) + o(1)} = n^{-\left(\beta - \frac{1}{2}\right) + o(1)}.$$

To see the implications of the discussion above, let $\mathcal{Q}$ be the set of paths P in G of length 2 for which the first vertex in P is the only one in $\mathcal{I}(t)$, and all vertices in P have degree at most d.

Notice that we can bound

$$|\mathcal{Q}| \leq nd^2 + mnd = n^{1+o(1)}. \tag{5.7}$$

The first inequality can be explained as follows. If $v \in \mathcal{I}(t)$ and v has degree at most d, the number of paths $P \in \mathcal{Q}$ which begin at v is at most d^2. As there are at most n vertices in $\mathcal{I}(t)$ of degree at most d, the total number of $P \in \mathcal{Q}$ starting at such a vertex is nd^2. On the other hand, if $v \in \mathsf{highdeg}(G) \cap \mathcal{I}(t)$, then the total number of paths of length 2 beginning at v is at most nd, since $\deg(v) \leq n$. As there are m vertices in $\mathsf{highdeg}(G)$, the number of paths in $\mathcal{Q}$ for this case is at most mnd. Summing both together leads to the claimed bound.

By our arguments above,

$$\begin{aligned}\mathbb{P}\left(\{|\mathcal{J}_2| \geq 1\} \cap \mathcal{A} \cap \mathcal{B}_t | \mathcal{F}_t\right) &\leq \mathbb{P}\left(\exists P \in \mathcal{Q} : \mathsf{weight}(P) \leq \delta | \mathcal{F}_t\right) \leq \sum_{P \in \mathcal{Q}} \mathbb{P}(\mathsf{weight}(P) \leq \delta | \mathcal{F}_t) \\ &\overset{(a)}{\leq} |\mathcal{Q}| \lambda^2 \delta^2 \overset{(b)}{\leq} n^{1+o(1)} \delta^2 = n^{-(2\beta - 1) + o(1)}.\end{aligned}$$

Above, (a) follows since $\mathsf{weight}(P) \sim \mathrm{Gamma}(2, \lambda)$ and (b) uses (5.7).

Putting everything together, we have that

$$\begin{aligned}\mathbb{P}\left(\{|\mathcal{J}_2| \geq 1\} \cap \mathcal{A} | \mathcal{F}_t\right) &\leq \mathbb{P}\left(\{|\mathcal{J}_2| \geq 1\} \cap \mathcal{A} \cap \mathcal{B}_t | \mathcal{F}_t\right) + \mathbb{P}(\mathcal{A} \cap \mathcal{B}_t^c | \mathcal{F}_t) \\ &\leq n^{-(2\beta - 1) + o(1)} + n^{-(\beta - 1/2) + o(1)} = n^{-\left(\beta - \frac{1}{2}\right) + o(1)}.\end{aligned}$$

$\square$

Lemma 5.1 now follows readily.

Proof of Lemma 5.1. By the definition of $\mathcal{J}_1$ and $\mathcal{J}_2$, we have that $I[t, t+n^{-\beta}] = |\mathcal{J}_1|$ if $\mathcal{J}_2 = \emptyset$. It follows that

$$\begin{aligned}&\mathbb{P}\left(\left\{\left| I\left[t, t+\frac{1}{n^\beta}\right] - \frac{\lambda}{n^\beta} \mathsf{cut}(\mathcal{I}(t))\right| \geq n^\gamma\right\} \cap \mathcal{A}\right) \\ &\qquad \leq \mathbb{P}\left(\left\{\left| |\mathcal{J}_1| - \frac{\lambda}{n^\beta} \mathsf{cut}(\mathcal{I}(t))\right| \geq n^\gamma\right\} \cap \mathcal{A}\right) + \mathbb{P}(\{|\mathcal{J}_2| \geq 1\} \cap \mathcal{A}) \leq n^{-(\beta - 1/2) + o(1)}.\end{aligned}$$

$\square$

5.2 Changes in the cut: Proof of Lemma 5.2

As a shorthand, denote $t_1 := t - n^{-\beta}$ and $t_2 := t - n^{-5}$. Consider the event

$$\mathcal{B} := \{\forall v \in \mathsf{highdeg}(G), T(v) \notin (t_1, t_2]\}.$$

On this event, all vertices in $\mathcal{I}(t_2) \setminus \mathcal{I}(t_1)$ have degree at most d, hence

$$|\,\mathsf{cut}(\mathcal{I}(t_2)) - \mathsf{cut}(\mathcal{I}(t_1))| \leq \sum_{v \in \mathcal{I}(t_2) \setminus \mathcal{I}(t_1)} \deg(v) \leq d|\mathcal{I}(t_2) \setminus \mathcal{I}(t_1)|.$$

Above, the first inequality follows since the size of the cut changes by at most $\deg(v)$ vertices when v is added to the cascade.

We proceed by studying $|\mathcal{I}(t_2) \setminus \mathcal{I}(t_1)|$. We have that

$$\begin{aligned}
\mathbb{E}\left[|\mathcal{I}(t_2) \setminus \mathcal{I}(t_1)|\mathbf{1}(\mathcal{B})|\mathcal{F}_{t_1}\right] &\overset{(a)}{=} \sum_{v \in V \setminus \mathcal{I}(t_1) : \deg(v) \leq d} \mathbb{P}\left(\{T(v) \in (t_1, t)\} \cap \mathcal{B} | \mathcal{F}_{t_1}\right) \\
&\leq \sum_{v \in V \setminus \mathcal{I}(t_1) : \deg(v) \leq d} \mathbb{P}\left(T(v) \in (t_1, t_1 + n^{-\beta}) | \mathcal{F}_{t_1}\right) \\
&\overset{(b)}{\leq} \sum_{v \in V \setminus \mathcal{I}(t_1) : \deg(v) \leq d} \frac{d}{n^\beta} \overset{(c)}{=} n^{1-\beta+o(1)}. \qquad (5.8)
\end{aligned}$$

Above, (a) follows since all vertices in $\mathcal{I}(t_2) \setminus \mathcal{I}(t_1)$ have degree at most d on $\mathcal{B}$, (b) is due to Lemma 3.3 and (c) follows since $d = n^{o(1)}$. Importantly, it follows immediately from (5.8) that the unconditional expectation can be bounded as $\mathbb{E}[|\mathcal{I}(t_2) \setminus \mathcal{I}(t_1)|\mathbf{1}(\mathcal{B})] \leq n^{1-\beta+o(1)}$. Putting everything together, we have that

$$\begin{aligned}
\mathbb{P}\left(\{|\,\mathsf{cut}(\mathcal{I}(t_2)) - \mathsf{cut}(\mathcal{I}(t_1))| \geq n^\gamma\} \cap \mathcal{A} \cap \mathcal{B}\right) &\leq \mathbb{P}\left(\left\{|\mathcal{I}(t_2) \setminus \mathcal{I}(t_1)| \geq \frac{n^\gamma}{d}\right\} \cap \mathcal{B}\right) \\
&\leq \frac{d}{n^\gamma}\, \mathbb{E}\left[|\mathcal{I}(t_2) \setminus \mathcal{I}(t_1)|\mathbf{1}(\mathcal{B})\right] \\
&= n^{1-\beta-\gamma+o(1)} \qquad (5.9) \\
&\leq n^{-(2\beta-1)+o(1)}, \qquad (5.10)
\end{aligned}$$

where the final inequality uses our assumption that $\gamma > \beta$. It remains to bound $\mathbb{P}(\mathcal{A} \cap \mathcal{B}^c)$, which can be done using Lemma 3.8.

$$\mathbb{P}(\mathcal{A} \cap \mathcal{B}^c) \leq \sum_{v \in \mathsf{highdeg}(G)} \mathbb{P}(\mathcal{A} \cap \{T(v) \in (t_1, t_2)\}) \leq mn^{-\left(\beta - \frac{1}{2}\right)+o(1)} = n^{-\left(\beta - \frac{1}{2}\right)+o(1)}, \qquad (5.11)$$

where the final expression uses $m = n^{o(1)}$. The lemma now follows from (5.9) and (5.11). □

6 Lower bounds for $\alpha \in (0, 1/2)$: Proof of Theorem 1.4

To prove the theorem, we will construct a tailored family of graphs $\mathcal{H} \subset \mathcal{G}$ for which we can tractably analyze the distribution of infection times, while also constituting a "hard" instance for estimating high-degree vertices. To construct $\mathcal{H}$, we first describe a particular deterministic graph H.

Definition 6.1. *Fix a positive integer N. Let H' be a connected tree with N vertices, maximum degree at most 3, and diameter at most $2 \log N$ (e.g., a balanced binary tree). For each vertex in $u \in H'$, we add a path of length $\log^5 n$ starting at u; all added paths are disjoint. We call the resulting graph H. The set of endpoints of the paths added to H', which are also the set of leaves of H, is denoted by L.*

The addition of paths which are much longer than the diameter of H' is a crucial aspect of our construction. At a high level, it ensures that the infection times of vertices in L are primarily determined by the infection times of vertices in the corresponding path, rather than the infection times of vertices in H'. Since all the added paths are disjoint, the infection times of vertices in L become *almost* independent, which simplifies much of our analysis. A bit more formally, we have the following useful observation.

Fact 6.2. *The infection times of vertices in L, given by the collection $\{\mathbf{T}(v)\}_{v\in L}$, are conditionally independent with respect to $\{\mathbf{T}(z)\}_{z\in H'}$, the collection of infection times of vertices in H'.*

To construct the graphs of interest, we will introduce additional vertices to H which connect to L in a particular manner.

Definition 6.3. *Suppose that N vertices, labelled from 1 to N, are added to H. Each vertex in $[N]$ forms an edge with a single element of L. Define $\mathcal{H}_\emptyset$ to be the set of resulting graphs for which all vertices in L have degree at most $\log^2 n$. For each $v \in L$, define $\mathcal{H}_v$ to be the set of graphs for which each $u \in L \setminus \{v\}$ has degree at most $\log^2 n$, but v has degree at least $D/2$. The union of all such graphs is given by $\mathcal{H} := \bigcup_{x\in L\cup\{\emptyset\}} \mathcal{H}_x$.*

We make a few simple observations about the family $\mathcal{H}$.

- If $G \in \mathcal{H}_\emptyset$ then $\mathsf{highdeg}(G) = \emptyset$, and if $G \in \mathcal{H}_v$ for $v \in L$ then $\mathsf{highdeg}(G) = \{v\}$.
- The total number of vertices in $G \in \mathcal{H}$ is $n := 2N + N\log^5 N$. In particular, $N = n^{1-o(1)}$.

The first step in our analysis is to show that it is challenging to even detect whether the set of observed infection times are due to a graph in $\mathcal{H}_\emptyset$ or in $\mathcal{H}_v$ for some $v \in L$. To do so, we construct convenient distributions $\mu_\emptyset$ over elements of $\mathcal{H}_\emptyset$ and μ_v over elements of $\mathcal{H}_v, v \in L$. We let $P_\emptyset$ and $P_v, v \in L$ denote the induced distributions over the infection times. That is, for $x \in L \cup \{\emptyset\}$, P_x is the distribution corresponding to $\mathcal{SI}(G, v_0)^{\otimes K}$ where $G \sim \mu_x$ and v_0ßH' is chosen arbitrarily (but is fixed across all cascades). We defer the precise description of the distributions $\{\mu_x\}_{x\in L\cup\{\emptyset\}}$ to Section 6.1, and move forward with the proof for now.

Lemma 6.4. *Let $\epsilon > 0$ and $\alpha \in (0, 1/2)$. Suppose that $D \le n^\alpha$ for some $\alpha \in (0, 1/2)$ and that*

$$K \le \left(\frac{1 - 2\alpha - \epsilon}{5}\right) \frac{\log n}{\log \log n}.$$

Then it holds for all $v \in L$ that $\mathrm{TV}(P_\emptyset, P_v) = o(1)$.

At a high level, we prove Lemma 6.4 by studying the χ^2-divergence between $P_\emptyset$ and P_v for $v \in L$, which provides an upper bound for the total variation distance and has a useful tensorization property which we leverage in our analysis. Before proving the lemma, we show how it can be used to prove the theorem.

Proof of Theorem 1.4. Define the measure μ over the ensemble $\mathcal{H}$ via $\mu := \frac{1}{N+1}\sum_{x\in L\cup\{\emptyset\}} \mu_x$, and similarly define the induced measure over vertex infection times $P := \frac{1}{N+1}\sum_{x\in L\cup\{\emptyset\}} P_x$. Let us also define $\delta := \max_{x,y\in L\cup\{\emptyset\}} \mathrm{TV}(P_x, P_y)$ and note that $\delta = o(1)$ by Lemma 6.4 and the triangle inequality. For any estimator HD, it holds that

$$\begin{aligned}
P(\mathrm{HD} = \mathsf{highdeg}(G)) &= \frac{1}{N+1} \sum_{x\in L\cup\{\emptyset\}} P_x(\mathrm{HD} = \{x\}) \\
&\stackrel{(a)}{=} 1 - \frac{1}{N+1} \sum_{x,y\in L\cup\{\emptyset\}: x\neq y} P_x(\mathrm{HD} = \{y\}) \\
&\stackrel{(b)}{\le} 1 - \frac{1}{N+1} \sum_{x,y\in L\cup\{\emptyset\}: x\neq y} (P_y(\mathrm{HD} = \{y\}) - \delta) \\
&\stackrel{(c)}{=} 1 + N\delta - NP(\mathrm{HD} = \mathsf{highdeg}(G)),
\end{aligned}$$

where (a) is due to the law of total probability, (b) uses the definition of the total variation distance, and (c) follows as each term $P_y(\mathrm{HD} = \{y\})$ is counted N times in the summation. Rearranging terms shows that

$$P(\mathrm{HD} = \mathsf{highdeg}(G)) \le \frac{1 + N\delta}{1 + N} = o(1).$$

□

The remainder of this section is devoted to the proof of Lemma 6.4.

6.1 Distributions of graphs in $\mathcal{H}$

Before proving the lemma, we must describe the distributions $\{\mu_x\}_{x\in L\cup\{\emptyset\}}$ over graphs in $\mathcal{H}$. To this end, it is useful to first define corresponding "auxiliary" distributions $\{\mu'_x\}_{x\in L\cup\{\emptyset\}}$. We say that $G \sim \mu'_\emptyset$ if each vertex in $[N]$ chooses a uniform random element of L with which they form an edge. On the other hand, we say that $G \sim \mu'_v$ if, for each $u \in [N]$, u connects to a uniform random element of L with probability $1 - D/N$, otherwise u connects to v with probability D/N. Finally, for $x \in L \cup \{\emptyset\}$, the measure μ_x is equal to μ'_x conditioned on the sampled graph being an element of $\mathcal{H}_x$.

For $x \in L \cup \{\emptyset\}$, we recall that P_x is the distribution corresponding to $\mathcal{SI}(G, v_0)^{\otimes K}$, where $G \sim \mu_x$. Similarly, we define P'_x to be the distribution corresponding to $\mathcal{SI}(G, v_0)^{\otimes K}$, where $G \sim \mu'_x$. The following result shows that the total variation distance between P_x and P'_x is small.

Lemma 6.5. *It holds for any $x \in L \cup \{\emptyset\}$ that $\mathrm{TV}(P_x, P'_x) = o(1)$.*

The lemma implies that we can essentially replace P_x with P'_x in our analysis moving forward. Indeed, by the triangle inequality for the total variation distance, we have for any $v \in L$ that

$$\mathrm{TV}(P_\emptyset, P_v) \le \mathrm{TV}(P'_\emptyset, P'_v) + \mathrm{TV}(P'_\emptyset, P_\emptyset) + \mathrm{TV}(P'_v, P_v) = \mathrm{TV}(P'_\emptyset, P'_v) + o(1), \tag{6.1}$$

where the final equality is due to Lemma 6.5. Hence it suffices to bound the total variation between the auxiliary distributions.

Proof of Lemma 6.5. Let $x \in L \cup \{\emptyset\}$. By the data processing inequality for the total variation distance, it holds that $\mathrm{TV}(P_x, P'_x) \le \mathrm{TV}(\mu_x, \mu'_x)$. By the definition of μ_x and μ'_x, the two distributions can be coupled precisely when $G \sim \mu'_x$ satisfies $G \in \mathcal{H}_x$. Hence

$$\mathrm{TV}(P_x, P'_x) \le \mathrm{TV}(\mu_x, \mu'_x) = \mu'_x(\mathcal{H}_x^c). \tag{6.2}$$

In light of (6.2), we proceed by proving that $\mu'_x(\mathcal{H}_x^c) = o(1)$. Notice that for $G \sim \mu'_x$, all vertices except for those in L have a deterministic degree which is at most 3, hence it suffices to study the degrees of vertices in L alone. In the case $x = \emptyset$, it holds for every $v \in L$ that $\deg_G(v) - 1 \sim \mathrm{Bin}(N, 1/N)$. Bernstein's inequality implies

$$\mu'_\emptyset\left(\deg_G(v) \ge \log^2 N\right) \le \mu'_\emptyset\left(\deg_G(v) - 1 \ge 1 + \frac{1}{2}\log^2 N\right) \le \exp\left(-\frac{2}{3}\log^2 N\right). \tag{6.3}$$

It follows that

$$\begin{aligned}\mu'_\emptyset\left(\mathcal{H}_\emptyset^c\right) &= \mu'_\emptyset\left(\exists v \in L : \deg_G(v) > \log^2 n\right)\\ &\overset{(a)}{\le} \mu'_\emptyset\left(\exists v \in L : \deg_G(v) > \log^2 N\right)\\ &\overset{(b)}{\le} N\exp\left(-\frac{2}{3}\log^2 N\right) = o(1).\end{aligned} \tag{6.4}$$

Above, (a) follows since $N \le n$, and (b) is due to a union bound over the N elements of L.

We now turn to the case where $x \in L$. If $v \in L \setminus \{x\}$, then $\deg_G(v) - 1 \sim \mathrm{Bin}\left(N, \left(1 - \frac{D}{N}\right)\frac{1}{N}\right)$, which is stochastically bounded by $\mathrm{Bin}(N, 1/N)$. Following the same steps as (6.3), we have that

$$\mu'_x\left(\deg_G(v) \ge \log^2 n\right) \le \mu'_x\left(\deg_G(v) \ge \log^2 N\right) \le \exp\left(-\frac{2}{3}\log^2 N\right).$$

On the other hand, when $v = x$, $\deg_G(v)$ stochastically dominates $\mathrm{Bin}(N, D/N)$. By Bernstein's inequality,

$$\mu'_x\left(\deg_G(v) \le \frac{D}{2}\right) \le \exp\left(-\frac{3}{56}D\right).$$

It follows that

$$\mu_x'\left(\mathcal{H}_x^c\right) = \mu_x'\left(\exists v \in L\setminus\{x\} : \deg_G(v) > \log^2 n \text{ or } \deg_G(x) < \frac{D}{2}\right)$$
$$\leq \sum_{v\in L\setminus\{x\}} \mu_x'\left(\deg_G(v) > \log^2 n\right) + \mu_x'\left(\deg_G(x) < \frac{D}{2}\right)$$
$$\leq N\exp\left(-\frac{2}{3}\log^2 N\right) + \exp\left(-\frac{3}{56}D\right) = o(1). \tag{6.5}$$

Together, (6.4), (6.5) and (6.2) prove the claim. □

6.2 Impossibility of detection: Proof of Lemma 6.4

The advantage of using the auxiliary distributions is that they are product distributions conditioned on the infection times in H, which makes it easier to analyze the infection times in $[N]$. To see this concretely, let us assume without loss of generality that $\lambda = 1$, and define

$$\mathbf{E}(j) = (E_1(j), \ldots, E_K(j)) \overset{i.i.d.}{\sim} \mathrm{Exp}(1)^{\otimes K}, \qquad j \in [N].$$

Let us also define the collection of independent variables $\{U(j)\}_{j\in[N]}$ where $U(j) \sim \mathrm{Unif}(L)$ denotes the random neighbor of the vertex j in the construction of $G \sim P_\emptyset'$. We then have the representation

$$\mathbf{T}(j) = \mathbf{T}(U(j)) + \mathbf{E}(j), \qquad j \in [N]. \tag{6.6}$$

On the other hand, if $G \sim P_v'$ for some $v \in L$, then we have the representation

$$\mathbf{T}(j) = \begin{cases} \mathbf{T}(U(j)) + \mathbf{E}(j) & \text{with probability } 1 - D/N \\ \mathbf{T}(v) + \mathbf{E}(j) & \text{with probability } D/N. \end{cases} \tag{6.7}$$

We can also succinctly characterize the distributions of infection times in $[N]$ for the cases described in (6.6) and (6.7). Define the following (conditional) distributions over $\mathbf{t} := (t_1, \ldots, t_K) \in \mathbb{R}_+^K$,

$$f_v(\mathbf{t}) := \prod_{i=1}^K e^{-(t_i - T_i(v))}\mathbf{1}(t_i \geq T_i(v)) = e^{-\sum_{i=1}^K t_i}\prod_{i=1}^K e^{T_i(v)}\mathbf{1}(t_i \geq T_i(v))$$
$$f_\emptyset(\mathbf{t}) := \frac{1}{N}\sum_{w\in L} f_w(\mathbf{t}) = \frac{e^{-\sum_{i=1}^K t_i}}{N}\sum_{w\in L}\prod_{i=1}^K e^{T_i(w)}\mathbf{1}(t_i \geq T_i(w)).$$

Then it is readily seen that

$$\mathrm{Law}\left(\{\mathbf{T}(j)\}_{j\in[N]} \,\middle|\, \{\mathbf{T}(u)\}_{u\in L}\right) = \begin{cases} f_\emptyset^{\otimes N} & G \sim P_\emptyset' \\ \left[\left(1 - \frac{D}{N}\right) f_\emptyset + \frac{D}{N} f_v\right]^{\otimes N} & G \sim P_v', v \in L. \end{cases} \tag{6.8}$$

Our strategy moving forward is to determine when the distributions in (6.8) are indistinguishable, which amounts to bounding the total variation distance between them. To do so, we will instead bound a larger distance measure – the χ^2 divergence. We recall that the χ^2 divergence between probability measures P and Q is defined as

$$\chi^2(P\|Q) := \int \frac{(dP - dQ)^2}{dQ}.$$

The total variation distance and the χ^2 divergence can be related through the inequality [41, Proposition 7.15]

$$\mathrm{TV}(P, Q) \leq 2\sqrt{\chi^2(P\|Q)}.$$

For our setting, the χ^2 divergence has two advantageous properties. First, it tensorizes; that is, for any positive integer k, it holds that [41, Chapter 7.12]

$$\chi^2\left(P^{\otimes k} \middle\| Q^{\otimes k}\right) = \left(1 + \chi^2(P\|Q)\right)^k - 1.$$

Additionally, the formula for the χ^2 divergence involves the difference between the densities of P and Q, which makes it convenient to analyze mixture distributions, as is the case for our setting.

We apply these ideas to the conditional distributions in (6.8). Conditioned on $\{\mathbf{T}(u)\}_{u\in H}$, we have that

$$\begin{aligned}
\mathrm{TV}\left(f_\emptyset^{\otimes N}, \left[\left(1-\frac{D}{N}\right)f_\emptyset + \frac{D}{N}f_v\right]^{\otimes N}\right) &\le 2\sqrt{\chi^2\left(\left[\left(1-\frac{D}{N}\right)f_\emptyset + \frac{D}{N}f_v\right]^{\otimes N} \middle\| f_\emptyset^{\otimes N}\right)} \\
&= 2\sqrt{\left[1+\chi^2\left(\left(1-\frac{D}{N}\right)f_\emptyset + \frac{D}{N}f_v \middle\| f_\emptyset\right)\right]^N - 1} \\
&\le 2\sqrt{\exp\left\{N\chi^2\left(\left(1-\frac{D}{N}\right)f_\emptyset + \frac{D}{N}f_v \middle\| f_\emptyset\right)\right\} - 1}.
\end{aligned}$$

In particular, we see that the conditional total variation tends to 0 if

$$\chi^2\left(\left(1-\frac{D}{N}\right)f_\emptyset + \frac{D}{N}f_v \middle\| f_\emptyset\right) = o\left(\frac{1}{N}\right). \tag{6.9}$$

We proceed by bounding the χ^2 divergence between the two distributions. Letting $\mathbf{t} = (t_1, \dots, t_K) \in \mathbb{R}^K$, it holds that

$$\chi^2\left(\left(1-\frac{D}{N}\right)f_\emptyset + \frac{D}{N}f_v \middle\| f_\emptyset\right) = \left(\frac{D}{N}\right)^2 \int_{\mathbb{R}_+^K} \frac{(f_v(\mathbf{t}) - f_\emptyset(\mathbf{t}))^2}{f_\emptyset(\mathbf{t})} d\mathbf{t} \le \left(\frac{D}{N}\right)^2 \int_{\mathbb{R}_+^K} \left(\frac{f_v(\mathbf{t})}{f_\emptyset(\mathbf{t})}\right) f_v(\mathbf{t}) d\mathbf{t}. \tag{6.10}$$

To bound the χ^2 divergence, we therefore need to study the likelihood ratio $f_v/f_\emptyset$. This is accomplished in the following result.

Lemma 6.6. *For every $v \in L$, it holds with probability $1 - o(1)$ that*

$$\sup_{\mathbf{t}\in\mathbb{R}_+^K} \frac{f_v(\mathbf{t})}{f_\emptyset(\mathbf{t})} \le \log^{5K} n.$$

Before proving Lemma 6.6, we show how it can be used to prove the main result.

Proof of Lemma 6.4. Let $\mathcal{A}_v$ denote the event described in Lemma 6.6; note that $\mathcal{A}_v$ is measurable with respect to $\{\mathbf{T}(u)\}_{u\in L}$. On this event, it follows from (6.10) that

$$\chi^2\left(\left(1-\frac{D}{N}\right)f_\emptyset + \frac{D}{N}f_v \middle\| f_\emptyset\right) \le \left(\frac{D}{N}\right)^2 \int_{\mathbb{R}_+^K} \left(\log^{5K} n\right) f_v(\mathbf{t}) d\mathbf{t} \le \left(\frac{D}{N}\right)^2 \log^{5K} n.$$

Since $D/N = n^{\alpha - 1 + o(1)}$, for any $\epsilon > 0$ the condition (6.9) is satisfied if

$$K \le \frac{1 - 2\alpha - \epsilon}{5}\left(\frac{\log n}{\log\log n}\right). \tag{6.11}$$

We now translate our results on the conditional distributions $f_\emptyset^{\otimes N}$ and $[(1-D/N)f_\emptyset + (D/N)f_v]^{\otimes N}$ to their *unconditional* counterparts $P'_\emptyset$ and P'_v. If (6.11) holds, then

$$\begin{aligned}
\mathrm{TV}(P'_\emptyset, P'_v) &\overset{(a)}{\le} \mathbb{E}\left[\mathrm{TV}\left(f_\emptyset^{\otimes N}, \left[\left(1-\frac{D}{N}\right)f_\emptyset + \frac{D}{N}f_v\right]^{\otimes N}\right)\right] \\
&\overset{(b)}{\le} \mathbb{E}\left[\mathrm{TV}\left(f_\emptyset^{\otimes N}, \left[\left(1-\frac{D}{N}\right)f_\emptyset + \frac{D}{N}f_v\right]^{\otimes N}\right)\mathbf{1}(\mathcal{A}_v)\right] + \mathbb{P}(\mathcal{A}_v^c) = o(1).
\end{aligned}$$

Above, the inequality (a) follows from the property that conditioning increases the total variation [41, Theorem 7.5], and inequality (b) holds since the total variation is always upper bounded by 1. Finally, $\mathrm{TV}(P'_\emptyset, P'_v) = o(1)$ implies that $\mathrm{TV}(P_\emptyset, P_v) = o(1)$ as well in light of (6.1). $\square$

In the remainder of this subsection, we study the likelihood ratio

$$\frac{f_v(\mathbf{t})}{f_\emptyset(\mathbf{t})} = \frac{\prod_{i=1}^K e^{T_i(v)} \mathbf{1}(t_i \geq T_i(v))}{\frac{1}{N}\sum_{w \in L} \prod_{i=1}^K e^{T_i(w)} \mathbf{1}(t_i \geq T_i(w))}.$$

When there exists $i \in [K]$ such that $t_i < T_i(v)$, it is clear from the formula that the likelihood ratio is zero. On the other hand, when $t_i \geq T_i(v)$ for all $i \in [K]$,

$$\begin{aligned}
\frac{f_v(\mathbf{t})}{f_\emptyset(\mathbf{t})} &= \frac{\prod_{i=1}^K e^{T_i(v)}}{\frac{1}{N}\sum_{w \in L} \prod_{i=1}^K e^{T_i(w)} \mathbf{1}(t_i \geq T_i(w))} \\
&\leq \frac{\prod_{i=1}^K e^{T_i(v)}}{\frac{1}{N}\sum_{w \in L} \prod_{i=1}^K e^{T_i(w)} \mathbf{1}(T_i(v) \geq T_i(w))} \\
&= \left(\frac{1}{N} \sum_{w \in L} \prod_{i=1}^K e^{T_i(w) - T_i(v)} \mathbf{1}(T_i(v) \geq T_i(w)) \right)^{-1}.
\end{aligned}$$

Above, the inequality on the second line holds since $\mathbf{1}(T_i(v) \geq T_i(w)) \leq \mathbf{1}(t_i \geq T_i(w))$ if $t_i \geq T_i(v)$. Importantly, the upper bound derived for the likelihood ratio does not depend on $\mathbf{t}$. Hence, to control $\sup_{\mathbf{t} \in \mathbb{R}_+^K} f_v(\mathbf{t})/f_\emptyset(\mathbf{t})$, we establish a lower bound for the quantity

$$\frac{1}{N} \sum_{w \in L} \prod_{i=1}^K e^{T_i(w) - T_i(v)} \mathbf{1}(T_i(v) \geq T_i(w)). \tag{6.12}$$

We remark that the analysis of (6.12) is greatly facilitated by the way in which the graph H is constructed (see Definition 6.1). To see this concretely, for each $v \in L$ let v' be the first vertex in H' on the unique path in H connecting v to H'. Then we can write

$$T_i(v) = T_i(v') + X_i(v), \qquad i \in [K], \tag{6.13}$$

where $X_i(v)$ is the time it takes for the cascade to spread from v' to v. Importantly, as the paths added to H' to form H are all disjoint, $\{X_i(v)\}_{v \in L}$ is a collection of i.i.d. $\mathrm{Gamma}(\log^5 N)$ random variables. Since the diameter of H' is assumed to be much smaller than $\log^5 N$, we can show that $T_i(v') = o(\log^5 N)$, and the behavior of $T_i(v)$ is largely dictated by $X_i(v)$, a random variable that is independent across $v \in L$. As we shall see, this reduction to i.i.d. random variables leads to a tractable analysis of (6.12).

To formalize these ideas, we begin by defining the following "nice" event.

Definition 6.7. *Let $v \in L$. The event $\mathcal{E}_v$ holds if and only if the following conditions are satisfied:*

1. $T_i(z) \leq \log^2 N$ *for all* $z \in H'$ *and* $i \in [K]$.
2. $T_i(v) \geq \frac{2}{3} \log^5 N$ *for all* $i \in [K]$.
3. $\sum_{i=1}^K \left(X_i(v) - \log^5 N\right)^2 \leq 2K \log^5 N$.

Importantly, as the following result shows, $\mathcal{E}_v$ holds with high probability.

Lemma 6.8. *For all $v \in L$, it holds that $\mathbb{P}(\mathcal{E}_v) = 1 - o(1)$.*

Proof. We let $\mathcal{E}_{v,1}, \mathcal{E}_{v,2}$ and $\mathcal{E}_{v,3}$ denote the events corresponding to the first, second and third conditions of $\mathcal{E}_v$ in Definition 6.7. We start by showing that $\mathcal{E}_{v,1}$ holds with probability $1 - o(1)$. Since H' is a tree and the diameter of H' is at most $2\log N$ (see Definition 6.1), we have that $T_i(z)$ is stochastically upper bounded by a Gamma$(2\log N, 1)$ random variable. Hence

$$\mathbb{P}\left(T_i(z) \geq \log^2 N\right) \leq e^{-\frac{1}{2}\log^2 N} 2^{\log N} \leq e^{-\frac{1}{4}\log^2 N},$$

where the first inequality is due to a Chernoff bound, and the second inequality holds for N sufficiently large. Taking a union bound over all $z \in H'$ and $i \in [K]$, we obtain

$$\mathbb{P}(\mathcal{E}_{v,1}^c) = \mathbb{P}\left(\max_{z\in H', i\in[K]} T_i(z) \geq \log^2 N\right) \leq \sum_{z\in H', i\in[K]} \mathbb{P}\left(T_i(z) \geq \log^2 N\right) \leq NKe^{-\frac{1}{4}\log^2 N} = o(1). \tag{6.14}$$

We now turn to the events $\mathcal{E}_{v,2}$ and $\mathcal{E}_{v,3}$. Notice that if $\mathcal{E}_{v,1}$ holds and there exists $i \in [K]$ such that $T_i(v) < \frac{2}{3}\log^5 N$, then

$$\sum_{i=1}^K \left(X_i(v) - \log^5 N\right)^2 \geq \left(\frac{\log^5 N}{4}\right)^2 > 2K\log^5 N, \tag{6.15}$$

where the final inequality holds for N sufficiently large. As a consequence of (6.15), we have that

$$\mathcal{E}_{v,1} \cap \mathcal{E}_{v,2}^c \subseteq \mathcal{E}_{v,3}^c, \tag{6.16}$$

so we focus on bounding the probability of the latter event. We do so through a second moment argument. As a shorthand, let $R := \log^5 N$. The first moment of each of the summands is $\mathbb{E}[(X_i(v)-R)^2] = \mathrm{Var}(X_i(v)) = R$, where we have used the fact that $X_i(v) \sim \mathrm{Gamma}(R, 1)$. The second moment is

$$\begin{aligned}
\mathbb{E}\left[(X_i(v) - R)^4\right] &= \mathbb{E}[X_i(v)^4] - 4R\,\mathbb{E}[X_i(v)^3] + 6R^2\,\mathbb{E}[X_i(v)^2] - 4R^3\,\mathbb{E}[X_i(v)] + R^4 \\
&= R(R+1)(R+2)(R+3) - 4R^2(R+1)(R+2) + 6R^3(R+1) - 4R^4 + R^4 \\
&= 3R^2 + 6R.
\end{aligned}$$

In particular, $\mathrm{Var}((X_i(v) - R)^2) = 2R^2 + 6R \leq 8R^2$, and furthermore,

$$\mathrm{Var}\left(\sum_{i=1}^K (X_i(v) - R)^2\right) = \sum_{i=1}^K \mathrm{Var}((X_i(v) - R)^2) \leq 8KR^2.$$

Finally, by Chebyshev's inequality,

$$\mathbb{P}(\mathcal{E}_{v,3}^c) = \mathbb{P}\left(\sum_{i=1}^K (X_i(v) - R)^2 > 2KR\right) \leq \frac{1}{K^2R^2}\,\mathrm{Var}\left(\sum_{i=1}^K (X_i(v) - R)^2\right) \leq \frac{8}{K} = o(1). \tag{6.17}$$

Together, (6.14), (6.16) and (6.17) prove that $\mathbb{P}(\mathcal{E}_v^c) = o(1)$ as claimed. $\square$

Next, we state a simple consequence of Taylor's theorem that will be useful for the proof of Lemma 6.6.

Lemma 6.9. *Suppose that a, b are real numbers such that $b \neq 0$ and $a/b \geq -1/2$. Then*

$$e^{-a}\left(1 + \frac{a}{b}\right)^b \geq e^{-2a^2/b}.$$

Proof. Taylor's theorem implies that for any $z \geq -1/2$, it holds that $\log(1+z) \geq z - 2z^2$. Using this inequality, we have that

$$\log\left(e^{-a}\left(1 + \frac{a}{b}\right)^b\right) = -a + b\log\left(1 + \frac{a}{b}\right) \geq -a + b\left(\frac{a}{b} - 2\left(\frac{a}{b}\right)^2\right) = -\frac{2a^2}{b}.$$

$\square$

We now prove the main result of this section.

Proof of Lemma 6.6. Throughout this proof, we shall condition on the infection times in H', given by the collection $\{\mathbf{T}(z)\}_{z\in H'}$, as well as v, given by $\mathbf{T}(v)$. We will also assume that the "nice" event $\mathcal{E}_v$ holds, which is measurable with respect to $\{\mathbf{T}(z)\}_{z\in H'}, \mathbf{T}(v)$. Finally, as a shorthand, we will denote $R := \log^5 N$.

Recall from Fact 6.2 that $\{\mathbf{T}(u)\}_{u\in L}$ is a collection of conditionally independent variables with respect to $\{\mathbf{T}(z)\}_{z\in H'}$. As a result, if we condition on both $\{\mathbf{T}(z)\}_{z\in H'}$ and $\mathbf{T}(v)$, then the terms

$$\prod_{i=1}^{K} e^{T_i(w)-T_i(v)}\mathbf{1}(T_i(v)\ge T_i(w)), \qquad w\in L, \tag{6.18}$$

are mutually (conditionally) independent. Moreover, each of the terms in (6.18) are bounded between 0 and 1. Hoeffding's inequality therefore implies that

$$\begin{aligned}&\frac{1}{N}\sum_{w\in L}\prod_{i=1}^{K} e^{T_i(w)-T_i(v)}\mathbf{1}(T_i(v)\ge T_i(w))\\&\qquad\ge \mathbb{E}\left[\frac{1}{N}\sum_{w\in L}\prod_{i=1}^{K} e^{T_i(w)-T_i(v)}\mathbf{1}(T_i(v)\ge T_i(w))\,\middle|\,\{\mathbf{T}(z)\}_{z\in H'},\mathbf{T}(v)\right]-N^{-1/3},\end{aligned} \tag{6.19}$$

with probability at least $1-2\exp(-2N^{1/3})$. We proceed by studying the conditional expectation in (6.19). To this end, for any $w\in L$, we denote w' to be the last vertex in H' on the unique path starting from v_0 and ending at w; such a vertex exists since H is connected. We have that

$$\begin{aligned}&\mathbb{E}\left[\frac{1}{N}\sum_{w\in L}\prod_{i=1}^{K} e^{T_i(w)-T_i(v)}\mathbf{1}(T_i(v)\ge T_i(w))\,\middle|\,\{\mathbf{T}(z)\}_{z\in H'},\mathbf{T}(v)\right]\\&\qquad=\frac{1}{N}\sum_{w\in L}\mathbb{E}\left[\prod_{i=1}^{K} e^{T_i(w)-T_i(v)}\mathbf{1}(T_i(v)\ge T_i(w))\,\middle|\,\{\mathbf{T}(z)\}_{z\in H'},\mathbf{T}(v)\right]\\&\qquad=\frac{1}{N}\sum_{w\in L}\mathbb{E}\left[\prod_{i=1}^{K} e^{T_i(w)-T_i(v)}\mathbf{1}(T_i(v)\ge T_i(w))\,\middle|\,\mathbf{T}(w'),\mathbf{T}(v)\right]\\&\qquad=\frac{1}{N}\sum_{w\in L}\prod_{i=1}^{K}\mathbb{E}\left[e^{T_i(w)-T_i(v)}\mathbf{1}(T_i(v)\ge T_i(w))\,\middle|\,T_i(w'),T_i(v)\right].\end{aligned} \tag{6.20}$$

Above, the second equality uses that the conditional expectation depends only on the random variables $T_i(w), T_i(v)$, and that the distribution of $T_i(w)$ conditioned on $\{\mathbf{T}(z)\}_{z\in H'}$ is the same as the distribution conditioned on $T(w')$ in light of (6.13). The final equality follows from the independence of the $T_i(w)$'s over $i\in[K]$.

We continue by lower bounding the conditional expectations in (6.20). Recall that for $w\in L$, we have from (6.13) that $T_i(w)=T_i(w')+X_i(w)$ where $X_i(w)\sim\text{Gamma}(R,1)$. Then

$$\begin{aligned}&\mathbb{E}\left[e^{T_i(w)-T_i(v)}\mathbf{1}(T_i(v)\ge T_i(w))\,\middle|\,T_i(w'),T_i(v)\right]\\&\qquad=e^{T_i(w')-T_i(v)}\,\mathbb{E}\left[e^{X_i(w)}\mathbf{1}(T_i(v)-T_i(w')\ge X_i(w))\,\middle|\,T_i(w'),T_i(v)\right]\\&\qquad\overset{(a)}{=}\frac{e^{T_i(w')-T_i(v)}}{\Gamma(R)}\int_0^{T_i(v)-T_i(w')}x^{R-1}dx\\&\qquad=\frac{e^{T_i(w')-T_i(v)}(T_i(v)-T_i(w'))^R}{\Gamma(R+1)}\\&\qquad\overset{(b)}{\ge}\frac{e^{T_i(w')-T_i(v)}(T_i(v)-T_i(w'))^R}{3\sqrt{R}(R/e)^R}\\&\qquad=\frac{1}{3\sqrt{R}}e^{-(T_i(v)-T_i(w')-R)}\left(1+\frac{T_i(v)-T_i(w')-R}{R}\right)^R\end{aligned}$$

$$
\begin{aligned}
&\overset{(c)}{\geq} \frac{1}{3\sqrt{R}} \exp\left\{-\frac{2(T_i(v)-T_i(w')-R)^2}{R}\right\} \\
&\overset{(d)}{\geq} \frac{1}{3\sqrt{R}} \exp\left\{-\frac{4}{R}(X_i(v)-R)^2 - \frac{4}{R}\left(T_i(v')+T_i(w')\right)^2\right\} \\
&\overset{(e)}{\geq} \frac{e^{-16}}{3\sqrt{R}} \exp\left\{-\frac{4}{R}(X_i(v)-R)^2\right\}. \qquad (6.21)
\end{aligned}
$$

In the display above, the equality (a) uses that $T_i(v) - T_i(w') \geq 0$, which holds since $T_i(v) \geq 2R/3 \geq \sqrt{R} \geq \max_{z \in G'} T_i(z)$ on the event $\mathcal{E}_v$. The inequality (b) is due to Stirling's approximation. The inequality (c) follows from Lemma 6.9, which can be applied here since $T_i(v) - T_i(w') \geq R/2$ on $\mathcal{E}_v$. The inequality (d) uses $(a+b)^2 \leq 2a^2 + 2b^2$, and the inequality (e) follows since $T_i(v'), T_i(w') \leq \sqrt{R}$ on the event $\mathcal{E}_v$.

Substituting the lower bound (6.21) into (6.20), we obtain

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{N}\sum_{w\in L}\prod_{i=1}^{K} e^{T_i(w)-T_i(v)}\mathbf{1}(T_i(v)\geq T_i(w))\,\middle|\,\{\mathbf{T}(z)\}_{z\in H'}, \mathbf{T}(v)\right] \\
&\qquad\qquad \geq \left(\frac{e^{-16}}{3\sqrt{R}}\right)^K \exp\left\{-\frac{4}{R}\sum_{i=1}^{K}(X_i(v)-R)^2\right\} \overset{(f)}{\geq} \left(\frac{e^{-30}}{\sqrt{R}}\right)^K, \qquad (6.22)
\end{aligned}
$$

where (f) holds on the event $\mathcal{E}_v$.

Together, (6.19) and (6.22) imply that conditioned on $\{\mathbf{T}(z)\}_{z\in H'}, \mathbf{T}(v)$, and provided $\mathcal{E}_v$ holds, we have with probability at least $1 - 2\exp(2N^{1/3})$ that

$$
\frac{1}{N}\sum_{w\in L}\prod_{i=1}^{K} e^{T_i(w)-T_i(v)}\mathbf{1}(T_i(v)\geq T_i(w)) \geq \left(\frac{e^{-30}}{\sqrt{R}}\right)^K - N^{-1/3} \geq R^{-K}.
$$

Finally, the desired result holds since $\mathbb{P}(\mathcal{E}_v) = 1 - o(1)$. □

7 Proof of Proposition 3.1

It suffices to show that the model described in Proposition 3.1 satisfies the equation (1.1). As (1.1) describes a *unique* Markov process, the claim will then follow. We start by showing that it is unlikely for two vertices to become infected in a very small time interval.

Lemma 7.1. *For any $t \geq 0$, it holds that*

$$
\mathbb{P}\left(|\mathcal{I}[t, t+\epsilon]| \geq 2 | \mathcal{F}_t\right) = o(\epsilon).
$$

Proof. Define τ_1 and τ_2 to be the first and second times after t when a new vertex is added to the cascade, respectively. Conditionally on $\mathcal{F}_t$, the memoryless property of Exponential distributions implies that $\tau_1 - t$ is equal in distribution to $\mathrm{Exp}(\lambda|\,\mathsf{cut}(\mathcal{I}[0,t])|)$, which stochastically dominates a $\mathrm{Exp}(\lambda n^2)$ random variable (here, we have used that the number of edges in G is at most n^2). By the memoryless property of Exponential distributions, $\tau_2 - \tau_1$ is independent of $\tau_1 - t$. Through similar argument as above, it holds that $\tau_2 - \tau_1$ stochastically dominates a $\mathrm{Exp}(\lambda n^2)$ random variable as well. Consequently, the probability that at least two vertices join the cascade in the time interval $[t, t+\epsilon]$ is

$$
\begin{aligned}
\mathbb{P}(\tau_2 - t \leq \epsilon | \mathcal{F}_t) &\leq \mathbb{P}\left(\tau_1 - t \leq \epsilon \text{ and } \tau_2 - \tau_1 \leq \epsilon | \mathcal{F}_t\right) \\
&= \mathbb{P}\left(\tau_1 - t \leq \epsilon | \mathcal{F}_t\right) \mathbb{P}\left(\tau_2 - \tau_1 \leq \epsilon | \mathcal{F}_t\right) \\
&\leq \left(1 - e^{-\epsilon\lambda n^2}\right)^2 \\
&\leq \epsilon^2 \lambda^2 n^4 = o(\epsilon).
\end{aligned}
$$

□

We now turn to the proof of the main result of this section.

Proof of Proposition 3.1. We start by defining some relevant notation. Condition on $\mathcal{F}_t$, and let $v \in V \setminus \mathcal{I}[0,t]$. Recalling that $\{F(e)\}_{e \in E}$ is a collection of i.i.d. $\mathrm{Exp}(\lambda)$ random variables, we define

$$W := \min\{F(w,v) : w \in \mathcal{I}[0,t], (w,v) \in E\}.$$

By basic properties of Exponential distributions, we have that $W \sim \mathrm{Exp}(\lambda|\mathcal{N}(v) \cap \mathcal{I}[0,t]|)$. Additionally define the event $\mathcal{E}$ to hold if and only if at most one vertex joins the cascade in the interval $(t, t+\epsilon]$.

We claim that if $\mathcal{E}$ holds, then $v \in \mathcal{I}[0,t+\epsilon] \iff W \leq \epsilon$. Indeed, if $W \leq \epsilon$, then from the description of $T(v)$ in the proposition statement, we have that $T(v) \leq t+\epsilon$, which implies that $v \in \mathcal{I}[t,t+\epsilon]$. On the other hand, if $W > \epsilon$, then we consider two cases: (1) no vertex becomes infected in $[t,t+\epsilon]$ or (2) one vertex becomes infected in $[t,t+\epsilon]$. Note that no other case is possible if $\mathcal{E}$ holds. In the first case, it holds for all $e \in \mathsf{cut}(\mathcal{I}[0,t])$ that $F(e) > \epsilon$, hence the weight of any path starting in $\mathcal{I}[0,t]$ and ending at v is at least ϵ, implying that $T(v) > t+\epsilon$. In the second case, if $v \in \mathcal{I}[t,t+\epsilon]$ despite $W > \epsilon$, there must exist a path $P \in \mathcal{P}(v,t)$ of length at least 2 such that $\mathsf{weight}(P,t) \leq \epsilon$. However, this implies the existence of another vertex u on the path which is *not* an element of $I[0,t]$. Since the portion of the path halting at u has a smaller weight than P, it follows that $T(u) \in \mathcal{I}[t,t+\epsilon]$. However, it is not possible for both u and v to be elements of $\mathcal{I}[t,t+\epsilon]$ if $\mathcal{E}$ holds, hence we must have that $T(v) > t+\epsilon$.

As a result of the argument above, we have that

$$\begin{aligned}
\mathbb{P}\left(\{v \in \mathcal{I}[0,t+\epsilon]\} \cap \mathcal{E} | \mathcal{F}_t\right) &= \mathbb{P}\left(\{W \leq \epsilon\} \cap \mathcal{E} | \mathcal{F}_t\right) \\
&= \mathbb{P}(W \leq \epsilon | \mathcal{F}_t) - \mathbb{P}\left(\{W \leq \epsilon\} \cap \mathcal{E}^c | \mathcal{F}_t\right) \\
&= \epsilon\lambda|\mathcal{N}(v) \cap \mathcal{I}[0,t]| + o(\epsilon).
\end{aligned}$$

Finally, since $\mathbb{P}(\mathcal{E}|\mathcal{F}_t) = 1 - o(\epsilon)$ by Lemma 7.1, it holds that

$$\mathbb{P}\left(v \in \mathcal{I}[0,t+\epsilon] | \mathcal{F}_t\right) = \epsilon\lambda|\mathcal{N}(v) \cap \mathcal{I}[0,t]| + o(\epsilon),$$

hence (1.1) is satisfied. $\square$

References

[1] B. Abrahao, F. Chierichetti, R. Kleinberg, and A. Panconesi. Trace complexity of network inference. In *the Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 491–499, New York, NY, USA, 2013. Association for Computing Machinery.

[2] E. Adar and L. A. Adamic. Tracking information epidemics in blogspace. In *the Proceedings of the IEEE/WIC/ACM Conference on Web Intelligence*, page 207–214, USA, 2005. IEEE Computer Society.

[3] H. Alexander and T. Day. Risk factors for the evolutionary emergence of pathogens. *Journal of The Royal Society Interface*, 7(51):1455–1474, 2010.

[4] A. Allard, P.-A. Noël, L. J. Dubé, and B. Pourbohloul. Heterogeneous bond percolation on multitype networks with an application to epidemic dynamics. *Physical Review E*, 79:036113, 2009.

[5] A. Auffinger, M. Damron, and J. Hanson. *50 Years of First Passage Percolation.* American Mathematical Society, 2017.

[6] E. Boix-Adserà, G. Bresler, and F. Koehler. Chow-liu++: Optimal prediction-centric learning of tree ising models. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 417–426, 2022.

[7] F. Brauer and C. Castillo-Chavez. *Mathematical Models in Population Biology and Epidemiology.* Springer, 2012.

[8] G. Bresler. Efficiently learning ising models on arbitrary graphs. In *the Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 771–782. ACM, 2015.

[9] G. Bresler and M. Karzand. Learning a tree-structured ising model in order to make predictions. *Annals of Statistics*, 48(2), 2020.

[10] G. Bresler, E. Mossel, and A. Sly. Reconstruction of markov random fields from samples: Some easy observations and algorithms. In *the Proceedings of the 11th International Workshop, APPROX 2008, and 12th International Workshop, RANDOM 2008, LNCS 5171*, pages 343–356. Springer, 2008.

[11] G. Bresler, E. Mossel, and A. Sly. Reconstruction of markov random fields from samples: Some observations and algorithms. *SIAM Journal on Computing*, 42(2):563–578, 2013.

[12] M. V. Burnashev and I. A. Begmatov. On a problem of signal detection leading to stable distributions. *Theory of Probability & Its Applications*, 35(3):556–560, 1991.

[13] T. T. Cai, W. Liu, and H. H. Zhou. Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *Annals of Statistics*, 44(2):455–488, 2016.

[14] T. T. Cai and Y. Wu. Optimal detection of sparse mixtures against a given null distribution. *IEEE Transactions on Information Theory*, 60(4):2217–2232, 2014.

[15] H. Daneshmand, M. Gomez-Rodriguez, L. Song, and B. Schoelkopf. Estimating diffusion network structures: Recovery conditions, sample complexity and soft-thresholding algorithm. In E. P. Xing and T. Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32, pages 793–801, Bejing, China, 22–24 Jun 2014. Proceedings of Machine Learning Research.

[16] Z. Dezső and A.-L. Barabási. Halting viruses in scale-free networks. *Physical Review E*, 65:055103, May 2002.

[17] R. L. Dobrusin. A statistical problem arising in the theory of detection of signals in the presence of noise in a multi-channel system and leading to stable distribution laws. *Theory of Probability & Its Applications*, 3(2):161–173, 1958.

[18] D. Donoho and J. Jin. Higher criticism and for detecting sparse heterogeneous mixtures. *Annals of Statistics*, 32(3), 2004.

[19] N. Du, L. Song, M. Yuan, and A. Smola. Learning networks of heterogeneous influence. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[20] D. Eckles, H. Esfandiari, E. Mossel, and M. A. Rahimian. Seeding with costly network information. *Operations Research*, 70(4), 2022.

[21] R. Eletreby, Y. Zhuang, K. M. Carley, O. Yağan, and H. V. Poor. The effects of evolutionary adaptations on spreading processes in complex networks. *Proceedings of the National Academy of Sciences of the U.S.A.*, 117(11):5664–5670, 2020.

[22] A. Friggeri, L. Adamic, D. Eckles, and J. Cheng. Rumor cascades. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1):101–110, May 2014.

[23] M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. *ACM Trans. Knowl. Discov. Data*, 5(4), feb 2012.

[24] S. He, H. Zha, and X. Ye. Network diffusions via neural mean-field dynamics. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 2171–2183. Curran Associates, Inc., 2020.

[25] O. Hinz, B. Skiera, C. Barrot, and J. U. Becker. Seeding strategies for viral marketing: An empirical comparison. *Journal of Marketing*, 75(6):55 – 71, 2011.

[26] J. Hoffmann, S. Basu, S. Goel, and C. Caramanis. Learning mixtures of graphs from epidemic cascades. In H. D. III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 4342–4352. PMLR, 13–18 Jul 2020.

[27] J. Hoffmann and C. Caramanis. Learning graphs from noisy epidemic cascades. *Proceedings of the ACM Measurement and Analysis of Computer Systems*, 3(2), jun 2019.

[28] J. Ingster. *On Some Problems of Hypothesis Testing Leading to Infinitely Divisible Distributions*. Weierstraß-Inst. für Angewandte Analysis und Stochastik, 1996.

[29] J. Kelner, F. Koehler, R. Meka, and A. Moitra. Learning some popular gaussian graphical models without condition number bounds. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 10986–10998. Curran Associates, Inc., 2020.

[30] J. Khim and P.-L. Loh. A theory of maximum likelihood for weighted infection graphs. 2018. Preprint available at `https://arxiv.org/abs/1806.05273`.

[31] A. R. Klivans and R. Meka. Learning graphical models using multiplicative weights. In C. Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 343–354. IEEE Computer Society, 2017.

[32] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Trans. Web*, 1(1):5–es, may 2007.

[33] T. Liggett. *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*, volume 324. Springer, Berlin, Heidelberg, 1999.

[34] R. Lima. Hawkes processes modeling, inference, and control: An overview. *SIAM Review*, 65(2):331–374, 2023.

[35] N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *Annals of Statistics*, 34(3):1436 – 1462, 2006.

[36] S. Misra, M. Vuffray, and A. Y. Lokhov. Information theoretic optimal learning of gaussian graphical models. In J. Abernethy and S. Agarwal, editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 2888–2909. PMLR, 09–12 Jul 2020.

[37] D. Mollison. Markovian Contact Processes. *Advances in Applied Probability*, 10(1):88–108, 1978.

[38] E. Mossel and S. Roch. Distance-based species tree estimation under the coalescent: Information-theoretic trade-off between number of loci and sequence length. *Annals of Applied Probability*, 27(5), October 2017.

[39] P. Netrapalli and S. Sanghavi. Learning the graph of epidemic cascades. *SIGMETRICS Perform. Eval. Rev.*, 40(1):211–222, jun 2012.

[40] R. Pastor-Satorras and A. Vespignani. Immunization of complex networks. *Phys. Rev. E*, 65:036104, Feb 2002.

[41] Y. Polyanskiy and Y. Wu. *Information Theory: From Coding to Learning*. Cambridge University Press, 2024+. Preprint available at `http://www.stat.yale.edu/~yw562/teaching/itbook-export.pdf`.

[42] P. Ravikumar, M. J. Wainwright, G. Raskutti, and B. Yu. High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935 – 980, 2011.

[43] D. Richardson. Random growth in a tessellation. *Proceedings of the Cambridge Philosophical Society*, 74:515–528, 1973.

[44] E. M. Rogers. *Diffusion of Innovations*. Simon and Schuster, 5 edition, 2003.

[45] N. P. Santhanam and M. J. Wainwright. Information-theoretic limits of selecting binary graphical models in high dimensions. *IEEE Transactions on Information Theory*, 58(7):4117–4134, 2012.

[46] K. Spencer, C. Chung, and e. a. Stargel A. COVID-19 case investigation and contact tracing efforts from health departments – united states, june 25 – july 24. *Morbidity and Mortality Weekly Report*, 2020.

[47] D. Strang and S. A. Soule. Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual Review of Sociology*, 24:265 – 290, 1998.

[48] M. Vuffray, S. Misra, and A. Lokhov. Efficient learning of discrete graphical models. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 13575–13585. Curran Associates, Inc., 2020.

[49] J. Wallinga and P. Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*, 6(160):509–516, 2004.

[50] W. Wang, M. J. Wainwright, and K. Ramchandran. Information-theoretic bounds on model selection for gaussian markov random fields. In *2010 IEEE International Symposium on Information Theory*, pages 1373–1377, 2010.

[51] M. Wilinski and A. Lokhov. Prediction-centric learning of independent cascade dynamics from partial observations. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 11182–11192. PMLR, 18–24 Jul 2021.

[52] S. Wu, S. Sanghavi, and A. G. Dimakis. Sparse logistic regression learns all discrete pairwise graphical models. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.