

Quickest Inference of Network Cascades with Noisy Information*

Anirudh Sridhar, H. Vincent Poor

November 15, 2022

Abstract

We study the problem of estimating the source of a network cascade given a time series of noisy information about the spread. Initially, there is a single vertex affected by the cascade (the source) and the cascade spreads in discrete time steps across the network. Although the cascade evolution is hidden, one observes a noisy measurement of the evolution at each time step. Given this information, we aim to reliably estimate the cascade source as fast as possible.

We investigate Bayesian and minimax formulations of the source estimation problem, and derive near-optimal estimators for simple cascade dynamics and network topologies. In the Bayesian setting, samples are taken until the error of the Bayes-optimal estimator falls below a threshold. For the minimax setting, we design a novel multi-hypothesis sequential probability ratio test. These optimal estimators require $\log \log n / \log(k - 1)$ observations for a k -regular tree network, and $(\log n)^{\frac{1}{\ell+1}}$ observations for a ℓ -dimensional lattice. We then discuss conjectures on source estimation in general topologies. Finally, we provide simulations which validate our theoretical results on trees and lattices, and illustrate the effectiveness of our methods for estimating the sources of cascades on Erdős-Rényi graphs.

1 Introduction

Network-based interactions lie at the core of many dynamic systems, including social behavior, biological processes and wireless communications. Unfortunately, the decentralized nature of networks often make them susceptible to *cascading failures* in which behaviors or information originating from a small subset of nodes diffuse rapidly throughout the rest of the network. Examples include viral spread in contact networks (see e.g., [8]), misinformation in social networks [5, 14, 16, 41] and malware in cyber-physical networks [22, 26, 45, 51]. In all of these scenarios, the rapid spread of the cascade can have devastating effects. It is therefore of the utmost importance to track the cascade and contain it as fast as possible.

A fundamental challenge in accomplishing this task is that information about the cascade is usually noisy or uncertain in real-time settings. To illustrate this point, suppose that a virus is spreading over a contact network. When the number of individuals is large, it may be infeasible to force everyone to quarantine, hence diagnostic tests may be administered to track and contain the spread. If there are not enough diagnostic tests to test the entire population at a given point in time, there is uncertainty in the status of individuals who are not tested. Moreover, diagnostic

*This work was supported by the United States National Science Foundation (NSF) under RAPID Grant IIS-2026982. This article was presented in part at the 54th Asilomar Conference on Signals, Systems and Computers [39] and at the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [40].

A. Sridhar and H. V. Poor are with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 (email: anirudhs@princeton.edu, poor@princeton.edu).

tests are typically not perfectly accurate, so even among the tested individuals there may be false positives and negatives.

Nevertheless, by observing the results of many rounds of testing over time, it is natural to expect that one can accurately estimate the spread of the virus using the right testing and information aggregation strategies. On the other hand, if one waits too long to obtain reasonable estimates, the cascade will spread to a large subset of the population, which is undesirable. The goal of this work is to characterize inference algorithms which achieve the *optimal* tradeoff between the estimation error and the time until estimation. Moreover, we study how the structure of the underlying network influences the design and performance of such algorithms.

1.1 Summary of contributions

For the most part, existing theoretical work on estimating the source of a network cascade takes the perspective of a *reconstruction* problem: given a large, *known* set of infected nodes, the goal is to identify the source among them [23, 13, 35, 46, 47, 36, 37, 38, 50, 49]. In contrast, we study source estimation from the novel perspective of *real-time inference*: by monitoring real-time signals from each node, we aim to find the source *before* the number of affected nodes is large. The two paradigms of source estimation are fundamentally different, and as such, require drastically different models and methods.

We mathematically formalize the task of real-time source estimation as follows. Consider a statistical model of network cascades with noisy observations where, at discrete timesteps, each node produces a signal that is an independent sample from a pre-change distribution Q_0 if the node has not yet been affected by the cascade, else the signal is an independent sample from a post-change distribution Q_1 . Initially, a single unknown vertex (the source) is affected by the cascade, and the cascade propagates to neighbors of affected vertices at each timestep. Our objective is to design algorithms that estimate the unknown source as fast as possible. We provide the first solution to this problem, to the best of our knowledge, and derive optimal source estimators from Bayesian and minimax perspectives.

To develop a concrete characterization of optimal source estimators, we focus on simple cascade dynamics and networks. The cascade dynamics we consider are *deterministic*: at each timestep, the cascade spreads to all neighbors of currently-affected nodes. We assume the network topology is either a k -regular tree or a ℓ -dimensional lattice; we do so because such networks are simple to describe, they represent a diverse family of topologies, and they enjoy convenient symmetry properties which simplify our analysis considerably. We further assume that there is a known set of n *candidate nodes* which contains the unknown source. When n is large, we show that at least $\log \log n / \log(k - 1)$ timesteps of noisy observations are required for reliable source estimation in k -regular trees, while at least $(\log n)^{\frac{1}{\ell+1}}$ timesteps are required for ℓ -dimensional lattices.

We then derive *optimal* estimation algorithms whose performance matches the lower bounds described above. We show that the optimal algorithm in the Bayesian formulation of the source estimation problem is a simple procedure that continues to observe noisy observations of the cascade propagation until the Bayes-optimal estimator is sufficiently accurate. In the minimax formulation, we phrase source estimation as a n -ary hypothesis testing problem among the n candidate nodes and show that a natural test based on likelihood ratios – called the multi-hypothesis sequential probability ratio test (MSPRT) – is optimal. Interestingly, the design of the MSPRT which matches the lower bounds can be viewed as a *multi-scale search procedure*: it simultaneously identifies the general area of the source while also performing a local, fine-grained analysis to obtain more precise estimates.

Admittedly, our setting of deterministic cascade dynamics on regular trees or lattices is simplistic

compared to more realistic cascade and network models [20, 27, 30, 3, 25, 3, 7]. However, we find that our setting leads to a mathematically rich problem and serves as an important starting point for understanding the source estimation problem for more complex propagation dynamics and networks. On a more technical level, we present a mostly unified treatment of optimal source estimation algorithms on regular trees and lattices, with only minor differences between the two. This suggests that our methods could be generalized to describe optimal source estimators for *arbitrary topologies*, though this requires significantly more effort so we leave it to future work. We discuss in detail the potential extensions of our work to arbitrary topologies, providing conjectures on the structure and performance of optimal algorithms. Finally, we assess the performance of the estimators we develop through simulations. In addition to validating our theory for tree and lattice topologies, we show that our estimators perform well on natural models of random networks (Erdős-Rényi graphs). Strikingly, even when there is a moderate amount of noise in vertex signals, our estimators can reliably locate the cascade source in Erdős-Rényi graphs before 40 vertices are infected for networks with up to 2000 vertices. This provides strong evidence that our methods may be applicable quite broadly.

1.2 Related work

Source estimation from a noiseless snapshot. Perhaps the most well-known work on estimating the source of a cascade is by Shah and Zaman [36, 37, 38]. In their formulation of the problem, the cascade spreads randomly via the Susceptible-Infected process, and a single snapshot of the set of infected vertices is observed at a later point in time. They derive an expression for the maximum likelihood estimate of the source in trees and study properties of the estimator. Many authors have expanded on these ideas and methods in subsequent work, studying for instance the effect of multiple observations, multiple sources, confidence sets for the source, different network models, and different cascade models [23, 13, 35, 46, 47, 50, 49]. We emphasize that while this literature on source estimation is similar in spirit to the problem we consider in this paper, it is fundamentally different from modeling and algorithmic perspectives. For instance, the literature cited above is of a *static* nature, where we have a single (or a fixed number of) perfect-information snapshot(s) of a large cascade. On the other hand, we consider *dynamic* settings where we obtain noisy and incomplete measurements of a small but growing cascade. Moreover, the methods developed in the literature cited above (e.g., rumor centrality, Jordan centrality) have no obvious counterpart in our setting, since they are computed based on known infections. However, in the model of noisy, real-time measurements considered in this paper, it is impossible to know exactly which vertices are infected.

Cascade inference from a noisy time series. A growing body of literature uses the data model (2.1) to perform inference of cascades, including detecting the presence of a cascade [54, 52, 53, 34, 33, 15, 48], estimating the source [39, 40] and controlling its spread [17, 18, 29]. The closest work to ours in terms of methods and analysis is by Zou, Veeravalli, Li and Towsley [54], who studied the following quickest *detection* problem: a cascade spreads via unknown dynamics, and the goal is to stop sampling once the cascade affects a given number of vertices. Their test, which can be viewed as an adaptation of the CUSUM procedure, is agnostic to the spreading dynamics of the cascade and is optimal in the regime where samples are taken much frequently than the growth of the cascade. By contrast, we consider the regime of large networks and where samples are taken at a comparative rate to the growth of the cascade. Moreover, our results reveal the effect of the network topology on the performance of inference procedures, which is not the case in [54].

Finally, we remark that compared to our prior conference submissions on the source estimation

problem [39, 40], the current paper provides a unified and substantially more general solution. In particular, [40] only provided a Bayesian solution for lattices and [39] established results for the minimax setting under a somewhat unnatural, but mathematically simpler, constraint on the stopping time and estimator.

1.3 Notation

Let \mathbb{R} and \mathbb{Z} denote the set of reals and integers, respectively. For a graph $G = (V, E)$, let V denote the set of vertices and let E denote the set of edges. For $u, v \in V$, $d(u, v)$ represents the shortest path distance between u and v in G . For $v \in V$ and a non-negative integer s , $\mathcal{N}_v(s)$ is the s -hop neighborhood of v ; that is, $\mathcal{N}_v(s) := \{u \in V : d(u, v) \leq s\}$.

We utilize standard asymptotic notation throughout. In particular, for two functions $g(n)$ and $h(n)$, we say $g(n) \lesssim h(n)$ if there is $c > 0$ such that for n sufficiently large, $g(n) \leq ch(n)$. We say $g(n) \asymp h(n)$ (in words, $g(n)$ and $h(n)$ are orderwise equivalent) if and only if there are $c_1, c_2 > 0$ such that $c_1 h(n) \leq g(n) \leq c_2 h(n)$ for n sufficiently large. We say $g(n) \sim h(n)$ (in words, $g(n)$ is equal to $h(n)$ up to first-order terms) if $\lim_{n \rightarrow \infty} g(n)/h(n) = 1$.

1.4 Organization

The rest of the paper is organized as follows. In Section 2, we formally define our model of cascade evolution with noisy observations, as well as the Bayesian and minimax optimality criteria. In Section 3, we provide a description and overview of our results on optimal estimation in regular trees and lattices, as well as a discussion on how one might extend our techniques to general topologies. In Section 4, we provide numerical results on the performance of optimal estimators from simulations on trees, lattices and Erdős-Rényi graphs. The remaining sections are devoted to the proofs of our main results. The proofs of main results on the Bayesian setting are in Section 5, and proofs for the main results in the minimax setting are in Section 6. Sections 7 and 8 contain supporting results for the proofs in Sections 5 and 6. We conclude in Section 9. Additional combinatorial results concerning the topology of regular trees and lattices can be found in Appendix A and B.

2 Problem formulation

We begin by describing the most general formulation of the source estimation problem. Let G be a graph with vertex set and edge set given by V and E , respectively. Initially, a single vertex $v^* \in V$ is affected by the cascade; we call this vertex the *cascade source*. From v^* , the cascade spreads over time via the edges of the graph according to a known random or deterministic discrete-time process. Examples of cascade dynamics include variants of the susceptible-infected (SI) process, the independent cascade model and the linear threshold model (see [20] and references therein).

For any $v \in V$ and any time index $t \geq 0$, let $x_v(t) \in \{0, 1\}$ denote the *private state* of v , where $x_v(t) = 1$ if v is affected by the cascade at time t , otherwise $x_v(t) = 0$. The private states are not observable, but the system instead monitors the *public signals* $\{y_u(t)\}_{u \in V}$, defined as

$$y_u(t) \sim \begin{cases} Q_0 & x_u(t) = 0; \\ Q_1 & x_u(t) = 1, \end{cases} \quad (2.1)$$

where Q_0 and Q_1 are two mutually absolutely continuous probability measures. We can think of $y_u(t) \sim Q_0$ being typical behavior and $y_u(t) \sim Q_1$ as anomalous behavior caused by the cascade.

As a shorthand, we denote $y(t) := \{y_u(t)\}_{u \in V}$ to be the collection of all public states at time t . See Figure 1 for an illustration of this data model. We remark that this data model has been studied in recent literature in the context of cascade source estimation [39, 40], quickest detection of cascades [54, 52, 53, 34, 33, 15, 48], and control of cascades [18, 29, 17].

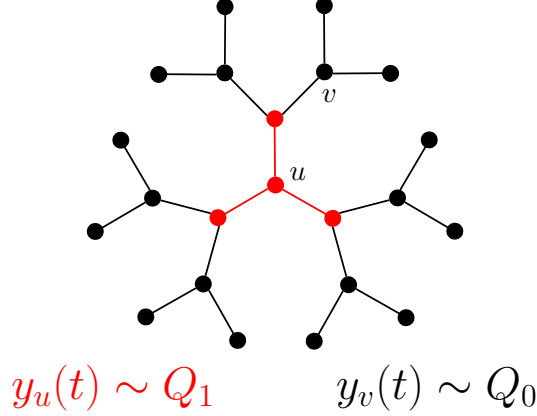


Figure 1: Illustration of the data model at a given time t . The nodes in red have been affected by the cascade, and the black nodes are unaffected, though this information is hidden from the observer. The public signals of the red nodes are sampled from Q_1 , while the public signals of the black nodes are sampled from Q_0 .

Remark 2.1. *The data model in (2.1) can capture a variety of realistic scenarios. In the context of viral spread for instance, a common symptom of sickness is a fever. If the public signals correspond to the body temperature of individuals in the population, one may expect that $y_v(t)$ will be close to the typical body temperature of the individual represented by vertex v if they do not carry the virus, else $y_v(t)$ is expected to be significantly higher if the individual does carry the virus.*

Another practical example of (2.1) is diagnostic testing with errors, which is used for malware detection in computer networks [2] and tracking the spread of infectious diseases [9]. Suppose that at a given point in time, each vertex is given a diagnostic test with probability p , independently over all vertices. If a test is taken, the output is either 0 (the vertex is not affected) or 1 (the vertex is affected). With probability ϵ , the result of the test will be incorrect. To formulate this in terms of (2.1), let the support of Q_0 and Q_1 be $\{0, 1, \times\}$, where 0 indicates a test result of 0, 1 indicates a test result of 1, and \times indicates that a test was not taken. The distributions $Q_0 := (q_0(0), q_0(1), q_0(\times))$ and $Q_1 := (q_1(0), q_1(1), q_1(\times))$ are given by

$$\begin{cases} q_0(0) = p(1 - \epsilon) \\ q_0(1) = p\epsilon \\ q_0(\times) = 1 - p \end{cases} \quad \text{and} \quad \begin{cases} q_1(0) = p\epsilon \\ q_1(1) = p(1 - \epsilon) \\ q_1(\times) = 1 - p. \end{cases}$$

Given the data model (2.1), the problem of estimating the cascade source can be phrased as a sequential multi-hypothesis testing problem: given the collection of hypotheses $\{H_v\}_{v \in V}$ where H_v is the hypothesis that v is the source, our goal is to output a hypothesis with a small probability of error. At the same time, it is also important that we come to a decision as fast as possible in order to minimize the number of vertices affected by the cascade. This reveals a fundamental tradeoff: when more samples are taken, one can obtain more reliable estimates of the source at the cost of

allowing the cascade to spread further. An *optimal* procedure will achieve the best possible tradeoff between the estimation error and the number of samples needed.

We shall proceed by formalizing these ideas. Observe that any source estimator can be represented by the pair $(T, \hat{\mathbf{v}})$, where T is a stopping time indicating when to stop sampling and $\hat{\mathbf{v}} := \{\hat{v}(t)\}_{t \geq 0}$ is a sequence of source estimators so that $\hat{v}(t)$ is the estimate of the source given the data at time t . The final source estimate produced by $(T, \hat{\mathbf{v}})$ is $\hat{v}(T)$. We shall also assume that a *candidate set* $U \subset V$ is known, so that the unknown source is an element of U . We remark that the size of U , denoted by $|U|$, measures in a sense the initial uncertainty around the source location. As a matter of notation, we let \mathbb{P}_v be the probability measure corresponding to the hypothesis H_v (v is the source). Similarly, \mathbb{E}_v denotes the expectation with respect to the measure \mathbb{P}_v . If the location of the source is given by a probability distribution $\pi := \{\pi_v\}_{v \in V}$ (where π_v is the probability that the source is v), we denote \mathbb{P}_π and \mathbb{E}_π to be the probability measure and expectation operator with respect to π , respectively. Formally, we may write

$$\mathbb{P}_\pi(\cdot) := \sum_{v \in V} \pi_v \mathbb{P}_v(\cdot) \quad \text{and} \quad \mathbb{E}_\pi[\cdot] := \sum_{v \in V} \pi_v \mathbb{E}_v[\cdot]. \quad (2.2)$$

We remark that often in this paper, we will consider the operator $\mathbb{E}_{\pi(t)}[\cdot]$ where $\pi(t)$ is the posterior distribution of the source after observing the public signals $y(0), \dots, y(t)$. In such a case, $\mathbb{E}_{\pi(t)}[\cdot]$ would be a random variable, since it is equal to the conditional expectation $\mathbb{E}[\cdot | y(0), \dots, y(t)]$.

We next define the performance metrics used to evaluate the effectiveness of a source estimator. For a source estimator $(T, \hat{\mathbf{v}})$, we shall study the *expected number of samples*, given by $\mathbb{E}_v[T]$ when v is the source. The *estimation error* is the expected distance between $\hat{v}(T)$ and the source, given by $\mathbb{E}_v[d(v, \hat{v}(T))]$ when v is the source. Here, we recall from Section 1.3 that $d(\cdot, \cdot)$ denotes the shortest-path distance between two vertices in G .

We study two natural ways to capture the tradeoff between estimation error and the expected number of samples.

A Bayesian perspective. Denote the source vertex by v^* , and suppose that the prior distribution for the source is uniform over the elements of the candidate set U ; we denote this prior by $\pi(U)$. We say that the optimal procedure solves the following optimization problem:

$$\inf_{T, \hat{\mathbf{v}}} \mathbb{E}_{\pi(U)} [d(v^*, \hat{v}(T)) + T] = \inf_{T, \hat{\mathbf{v}}} \frac{1}{|U|} \sum_{v \in U} \mathbb{E}_v [d(v, \hat{v}(T)) + T], \quad (2.3)$$

where we recall that $\mathbb{E}_{\pi(U)}$ denotes the expectation operator with respect to the measure $\pi(U)$. In words, (2.3) is the sum of the estimation error and the expected number of samples. If only the first term in (2.3) was present, the optimal strategy would be to set $T = \infty$, since more samples can only help in bringing down the estimation error. On the other hand, if only the second term in (2.3) was present, the optimal strategy would be to set $T = 0$. The estimator that solves (2.3) therefore achieves the best tradeoff between the two extremes. We remark that it is standard in Bayesian formulations of sequential testing problems to minimize the sum of the error and expected number of samples [32, 31]. Furthermore, we remark that one may consider other ways to quantify the tradeoff between estimation error and time – for instance, by replacing T with $h(T)$ for some increasing function h . While we focus on the formulation (2.3) for simplicity and ease of exposition, the methods we develop can also handle a large class of functions h .

A minimax perspective. As an alternative to the Bayesian approach, one can formalize the tradeoff between the estimation error and expected number of samples via the following optimization

problem:

$$\inf_{T, \hat{v}} \max_{v \in U} \mathbb{E}_v[T] \quad \text{subject to} \quad \max_{v \in U} \mathbb{E}_v[d(v, \hat{v}(T))] \leq \alpha, \quad (2.4)$$

where $\max_{v \in U} \mathbb{E}_v[d(v, \hat{v}(T))]$ is the *worst-case estimation error*, α is a specified bound on the worst-case estimation error and $\max_{v \in U} \mathbb{E}_v[T]$ is the *worst-case expected runtime* of the procedure. As in the Bayesian case, we may consider two extremes. When $\alpha = \infty$, the optimal choice is $T = 0$, whereas when $\alpha = 0$ the optimal choice is $T = \infty$.¹ For intermediate values of α , the optimal algorithm indeed achieves a tradeoff between the estimation error and the worst-case expected runtime.

3 Results

The goal of our work is to characterize optimal estimators based on the formulations in (2.3) and (2.4). We are particularly interested in how the structure and performance of optimal estimators depend on the network topology. In order to provide a tractable theoretical analysis, we focus on simple networks and cascade dynamics. The cascade dynamics we consider is outlined in the following assumption.

Assumption 3.1 (Cascade dynamics). *Initially, a single vertex v^* (the source) is affected by the cascade. The cascade then spreads deterministically in discrete time steps, so that vertex v is affected by the cascade at time t if and only if $d(v, v^*) \leq t$.*

We consider two classes of networks – regular trees and lattices – which are defined formally below.

Definition 3.2 (Infinite k -regular tree). *Let v_r be a designed root vertex, and let $\mathcal{T}_k(1)$ be the tree with k leaves attached to v_r . Given $\mathcal{T}_k(m)$, we construct $\mathcal{T}_k(m+1)$ by attaching $k-1$ leaves to each leaf in $\mathcal{T}_k(m)$. The infinite k -regular tree \mathcal{T}_k is the limiting graph obtained when $m \rightarrow \infty$; that is, (u, v) is an edge in \mathcal{T}_k if and only if (u, v) is an edge in $\mathcal{T}_k(m)$ for some positive integer m .*

Definition 3.3 (Infinite ℓ -dimensional lattice). *Label elements of the vertex set by \mathbb{Z}^ℓ . There is an edge between vertices u, v in the infinite ℓ -dimensional lattice if and only if $\sum_{i=1}^\ell |u_i - v_i| = 1$.²*

See Figure 2 for an illustration of the cascade dynamics and the evolution of observed signals on regular trees and lattices. We choose to study regular trees and lattices for several reasons. For one, they have strong symmetry properties (e.g., the local structure around all vertices are the same) which makes it easier to explicitly determine the performance of optimal algorithms. Second, we present a unified treatment of source estimation on regular trees and lattices (except for minor differences), even though the two families of graphs have extremely different topologies; perhaps the most obvious difference is that trees are acyclic while lattices contain many cycles of varying lengths. This indicates that it may be possible to generalize our methods to other topologies as well (see Section 3.3 for further discussion on this point). We also remark that it is a common assumption in the theoretical analysis of cascade models and inference tasks that the underlying graph has infinitely many vertices [28, 40, 39, 19, 6, 37, 38, 23, 13, 35]. Moreover, the infinite graph setting allows us to capture scenarios where the size of the cascade is small compared to the total population without unnecessarily complicating our mathematical analysis.

¹More precisely, if there exists an estimator \hat{v} such that $\hat{v}(t) \rightarrow v^*$ as $t \rightarrow \infty$, then the stopping time $T = \infty$ is optimal. If such an estimator does not exist, there is no feasible solution to (2.4) when $\alpha = 0$.

²The 2-regular tree \mathcal{T}_2 is the same as the 1-dimensional lattice. Henceforth, we shall identify \mathcal{T}_2 as the 1-dimensional lattice and always consider k -regular trees with $k \geq 3$. Indeed, from our analysis, it can be seen that the relevant properties of \mathcal{T}_2 make the graph most naturally associated with the class of lattices.

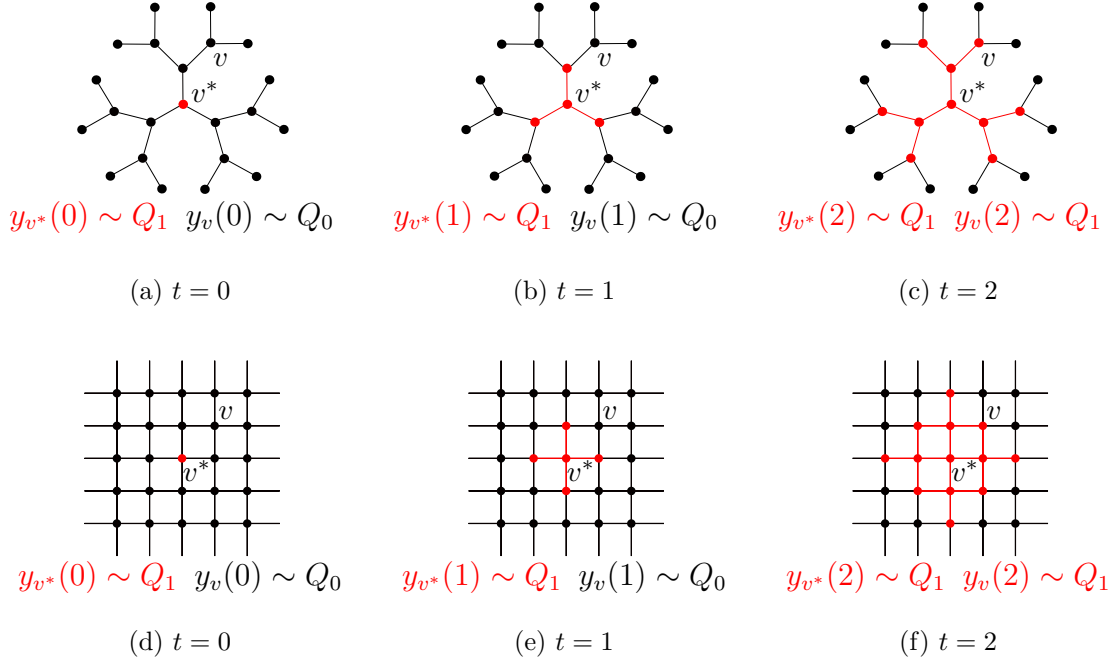


Figure 2: Illustration of cascade propagation and observations in 3-regular trees (a-c) and 2-dimensional lattices (d-f). In each image, the red nodes are those affected by the cascade, and the black are unaffected, though this information is hidden from the observer. Notice that the public signals for the cascade source v^* are always sampled from Q_1 , whereas the public signals for another vertex v are initially sampled from Q_0 and change to samples from Q_1 once the cascade spreads sufficiently far.

There are several network and cascade models that are arguably more realistic than the ones we study in this paper; see for instance [20, 27, 30, 3, 25, 3, 7]. However, even for the simple networks and cascade dynamics we consider, we expect that an exact characterization of optimal source estimators is mathematically intractable. The reason for this is that we may interpret the problem of source estimation as a *sequential multi-hypothesis testing problem*, where different hypotheses correspond to different potential sources. In the two-hypothesis case, the optimal hypothesis test is known to be the *sequential probability ratio test* (SPRT), which is a relatively simple procedure that tracks the cumulative log-likelihood ratio over time and stops when it achieves a particular threshold [44]. When there are more than two hypotheses, the optimal test has a complicated form and is difficult to analyze [4]. To carry out a tractable analysis, we therefore characterize optimal source estimators in *asymptotic regimes*, where the number of possible source vertices – in other words, the size of the candidate set – tends to infinity. Formally, we consider a sequence of candidate sets and study asymptotic properties of optimal estimators when the size of the candidate set grows large. As we shall see, the analysis of optimal estimators depends not only on the *size* of the candidate set, but also its topology. For instance, if two vertices in the candidate set are adjacent, there is a lot of overlap in the set of potential infections caused by each vertex. Hence it takes more effort and information to decide between the two vertices. On the other extreme, if two vertices in the candidate set are very far apart, it takes comparatively less effort to distinguish between them. One can therefore imagine that a *worst-case* candidate set is one where all vertices are as close to each other as possible. This idea is formally captured in the following assumption.

Assumption 3.4. We assume the sequence of candidate sets $\{V_n\}_n$ satisfies the following:

1. For all positive integers n , $V_n \subset V$ and $|V_n| = n$;
2. There is a designated vertex $v_0 \in V$ and a sequence of integers $\{r_n\}_n$ such that for all positive integers n ,

$$\mathcal{N}_{v_0}(r_n) \subseteq V_n \subset \mathcal{N}_{v_0}(r_n + 1).$$

Above, we recall that $\mathcal{N}_{v_0}(r)$ is the r -hop neighborhood of v_0 . The second condition in Assumption 3.4, which states that V_n is approximately a neighborhood of some arbitrary vertex, correctly captures the notion of a worst-case candidate set in the sense that it maximizes the number of vertex pairs that are close to each other. On a more technical note, by assuming a *specific* topological structure for the candidate set, we have enough detail to carry out a precise mathematical analysis of optimal estimators.

The value of r_n used in Assumption 3.4 can be made explicit. For the graphs of interest to us, we can employ straightforward combinatorial arguments to show

$$r_n \sim \begin{cases} \frac{\log n}{\log(k-1)} & G \text{ is a } k\text{-regular tree, } k \geq 3; \\ \left(\frac{\ell!}{2^\ell} n\right)^{1/\ell} & G \text{ is a } \ell\text{-dimensional lattice, } \ell \geq 1. \end{cases} \quad (3.1)$$

For details, see Corollaries A.2 and A.4 in Appendix A.

As a final remark on the candidate set, we emphasize that Assumption 3.4 is made only for the purposes of studying the performance of optimal estimators; it need not be satisfied to apply our estimators to more realistic, finite networks. See the part of Section 4 concerning cascade source estimation on Erdős-Rényi random graphs for more details on this point.

3.1 Results on Bayesian estimation

For a stopping time T , sequence of estimators $\hat{\mathbf{v}}$ and a candidate set V_n , define the quantities

$$\begin{aligned} \text{val}_B(T, \hat{\mathbf{v}}) &:= \mathbb{E}_{\pi(V_n)}[d(v^*, \hat{v}(T)) + T] \\ \text{val}_B^*(V_n) &:= \inf_{T, \hat{\mathbf{v}}} \text{val}_B(T, \hat{\mathbf{v}}). \end{aligned}$$

Note in particular that $\text{val}_B^*(V_n)$ is the optimal value of (2.3) when the candidate set is V_n . Our main result on the Bayesian formulation is the following theorem.

Theorem 3.5. When G is a k -regular tree with $k \geq 3$,

$$\text{val}_B^*(V_n) \sim \frac{\log \log n}{\log(k-1)}. \quad (3.2)$$

On the other hand, when G is a ℓ -dimensional lattice, there exist constants a, b depending only on ℓ, Q_0, Q_1 such that for n sufficiently large,

$$a(\log n)^{\frac{1}{\ell+1}} \leq \text{val}_B^*(V_n) \leq b(\log n)^{\frac{1}{\ell+1}}. \quad (3.3)$$

In words, (3.2) pins down the exact first-order asymptotic behavior of $\text{val}_B^*(V_n)$ when n is large in the case of regular trees. For lattices, (3.3) captures the *orderwise* behavior of $\text{val}_B^*(V_n)$ when n is large. While Theorem 3.5 focuses on how $\text{val}_B^*(V_n)$ scales with n , we remark that the (appropriately defined) distance between Q_0 and Q_1 plays a role in the performance of optimal estimators. In

the case of k -regular trees, it appears in the *second-order* expansion of $\text{val}_B^*(V_n)$. For ℓ -dimensional lattices, the distance between Q_0 and Q_1 influences the constants a and b , with both blowing up to infinity as the distance between Q_0 and Q_1 becomes small.

Proof summary. For any vertex $v \in V$ and integer $s \geq 0$, recall that $\mathcal{N}_v(s)$ is the set of vertices within distance s of v . From the cascade dynamics defined in Assumption 3.4, $\mathcal{N}_{v^*}(s)$ is precisely the set of vertices which have public signals distributed according to Q_1 , rather than Q_0 . The number of *total* public signals distributed according to Q_1 in the first t timesteps is therefore given by

$$f(t) := \sum_{s=0}^t |\mathcal{N}_{v^*}(s)|. \quad (3.4)$$

Due to the symmetry of regular trees and lattices, $|\mathcal{N}_u(s)| = |\mathcal{N}_v(s)|$ for any $u, v \in V$ and $s \geq 0$. Hence $f(t)$ does not depend on v^* , which is why we do not include v^* in the notation. The interpretation of $f(t)$ as the number of public signals distributed according to Q_1 implies that, in an abstract sense, $f(t)$ is a measure of the amount of information an observer has about the spread of the cascade.³ On the other hand, the initial uncertainty around the location of v^* can be measured by the entropy of $\pi(V_n)$, which is $\log n$. One may then expect that when $f(t) \lesssim \log n$, the information about the cascade propagation is not enough to overcome the uncertainty around the source location. It turns out that this intuition does indeed hold: for any $(T, \hat{\mathbf{v}})$, $\mathbb{E}_{\pi(V_n)}[d(v^*, \hat{v}(T))]$ is large (order $\log n$ in regular trees and $n^{1/\ell}$ in ℓ -dimensional lattices) when $f(T) \lesssim \log n$. It follows that accurate source estimation is only possible when $f(T) \gtrsim \log n$ or equivalently, when $T \gtrsim f^{-1}(\log n)$. This leads to the lower bound $\text{val}_B^*(V_n) \gtrsim f^{-1}(\log n)$, which is $\log \log n / \log(k-1)$ in k -regular trees up to first order terms and is of order $(\log n)^{\frac{1}{\ell+1}}$ in ℓ -dimensional lattices.

An upper bound on $\text{val}_B^*(V_n)$ is then derived by characterizing the performance of a given source estimator. Consider $(T_{th}, \hat{\mathbf{v}}_B)$, given formally by

$$\hat{v}_B(t) \in \arg \min_{v \in V_n} \mathbb{E}_{\pi(t)}[d(v^*, v)] \quad (3.5)$$

$$T_{th} := \min \{t \geq 0 : \mathbb{E}_{\pi(t)}[d(v^*, \hat{v}_B(t))] \leq 1\}. \quad (3.6)$$

Above, the measure $\pi(t)$ is the *posterior distribution* of the source v^* after observing the sequence of public signals $\mathbf{y}(0), \dots, \mathbf{y}(t)$, hence $\mathbb{E}_{\pi(t)}[d(v^*, v)]$ can be viewed as a conditional expectation. The interpretations of T_{th} and $\hat{\mathbf{v}}_B$ are quite intuitive. In words, $\hat{v}_B(t)$ is a vertex which achieves the minimum estimation error, conditioned on the observed information until time t . The estimator $\hat{\mathbf{v}}_B$ can therefore be thought of as the *Bayes-optimal source estimator*, as it minimizes the conditional estimation error.⁴ The stopping time T_{th} will keep sampling until the conditional estimation error of the optimal estimator falls below the threshold 1 (the subscript *th* references the fact that we stop once the estimation error is below a *threshold*). In characterizing the performance of the estimator $(T_{th}, \hat{\mathbf{v}}_B)$, we show that $\text{val}_B(T_{th}, \hat{\mathbf{v}}_B) \sim \log \log n / \log(k-1)$ in k -regular trees and $\text{val}_B(T_{th}, \hat{\mathbf{v}}_B) \asymp (\log n)^{\frac{1}{\ell+1}}$ in ℓ -dimensional lattices. Remarkably, these match the lower bounds previously established for $\text{val}_B^*(V_n)$, leading to (3.2) and (3.3). Moreover, our analysis shows that the estimator $(T_{th}, \hat{\mathbf{v}}_B)$ enjoys near-optimal performance when n is large.

As a final remark, we note that our proof methods are quite general and can also handle the case where T is replaced with $h(T)$ in (2.3) for any h that increases slower than an exponential

³We later make this more formal by showing that the Kullback-Liebler (KL) divergence between the measures \mathbb{P}_u and \mathbb{P}_v pertaining to the variables $y(0), \dots, y(t)$ is proportional to $f(t)$ for most pairs $u, v \in V$.

⁴We formalize this idea in Lemma 5.1, where we show that if the stopping time T is fixed, $\hat{\mathbf{v}}_B$ achieves $\inf_{\hat{\mathbf{v}}} \text{val}_B(T, \hat{\mathbf{v}})$.

function. Moreover, the *same* estimator $(\widehat{\mathbf{v}}_B, T_{th})$ also enjoys near-optimal performance in this case. For details, see Remark 5.4.

3.2 Results on minimax estimation

We begin by defining some notation. Let $\alpha > 0$ be fixed, and suppose V_n is the candidate set. Define the class of estimators

$$\Delta(V_n, \alpha) := \left\{ (T, \widehat{\mathbf{v}}) : \max_{v \in V_n} \mathbb{E}_v[d(v, \widehat{v}(T))] \leq \alpha \right\}.$$

In words, $\Delta(V_n, \alpha)$ is the class of source estimators which have a worst-case estimation error of at most α . In particular, $\Delta(V_n, \alpha)$ is the set of feasible estimators in the minimax formulation (2.4). The optimal value of the minimax formulation is denoted by

$$\text{val}_M^*(V_n, \alpha) := \inf_{(T, \widehat{\mathbf{v}}) \in \Delta(V_n, \alpha)} \max_{v \in V_n} \mathbb{E}_v[T].$$

The results we obtain for $\text{val}_M^*(V_n, \alpha)$ are essentially the same as in the Bayesian formulation. Specifically, we prove the following theorem.

Theorem 3.6. *Let α be fixed. When G is a k -regular tree,*

$$\text{val}_M^*(V_n, \alpha) \sim \frac{\log \log n}{\log(k-1)}. \quad (3.7)$$

On the other hand, when G is a ℓ -dimensional lattice, there exist constants a', b' depending only on ℓ, Q_0, Q_1 such that for n sufficiently large,

$$a'(\log n)^{\frac{1}{\ell+1}} \leq \text{val}_M^*(V_n, \alpha) \leq b'(\log n)^{\frac{1}{\ell+1}}. \quad (3.8)$$

As in Theorem 3.5, (3.7) provides an exact first-order characterization of $\text{val}_M^*(V_n, \alpha)$ when n is large, and (3.8) describes the *orderwise* behavior of $\text{val}_M^*(V_n, \alpha)$. We remark that the constants a' and b' used in (3.8) are potentially distinct from the constants a, b used in the Bayesian analogue (3.3). However, we make no attempt to optimize the constants, instead focusing on the orderwise behavior as n grows large. As in Theorem 3.5, the (appropriately defined) distance between Q_0 and Q_1 plays a role in the second order terms of $\text{val}_M^*(V_n, \alpha)$ in regular trees. In lattices, the constants a' and b' blow up to infinity when the distance between Q_0 and Q_1 is small.

Proof summary. As in the Bayesian case, we focus on establishing lower bounds for $\text{val}_M^*(V_n, \alpha)$ and derive matching upper bounds by studying the performance of a carefully designed estimator which lies within the feasible set $\Delta(V_n, \alpha)$.

To derive lower bounds for $\text{val}_M^*(V_n, \alpha)$, we observe that the Bayesian objective value – which measures *average-case* behaviors of source estimators – is less strict than the minimax objective, which measures *worst-case* behaviors of source estimators. Hence the lower bound for $\text{val}_B^*(V_n)$ also holds for $\text{val}_M^*(V_n, \alpha)$, provided α is constant with respect to n .

Next, we establish an upper bound for $\text{val}_M^*(V_n, \alpha)$ by characterizing the performance of a *specific* stopping time and estimator. Unfortunately, we cannot use $(T_{th}, \widehat{\mathbf{v}}_B)$ (used in the Bayesian setting) for this task since it is unclear whether it is an element of the class $\Delta(V_n, \alpha)$. We therefore take a different approach in designing an appropriate estimator within $\Delta(V_n, \alpha)$ which matches the lower bounds for $\text{val}_M^*(V_n, \alpha)$.

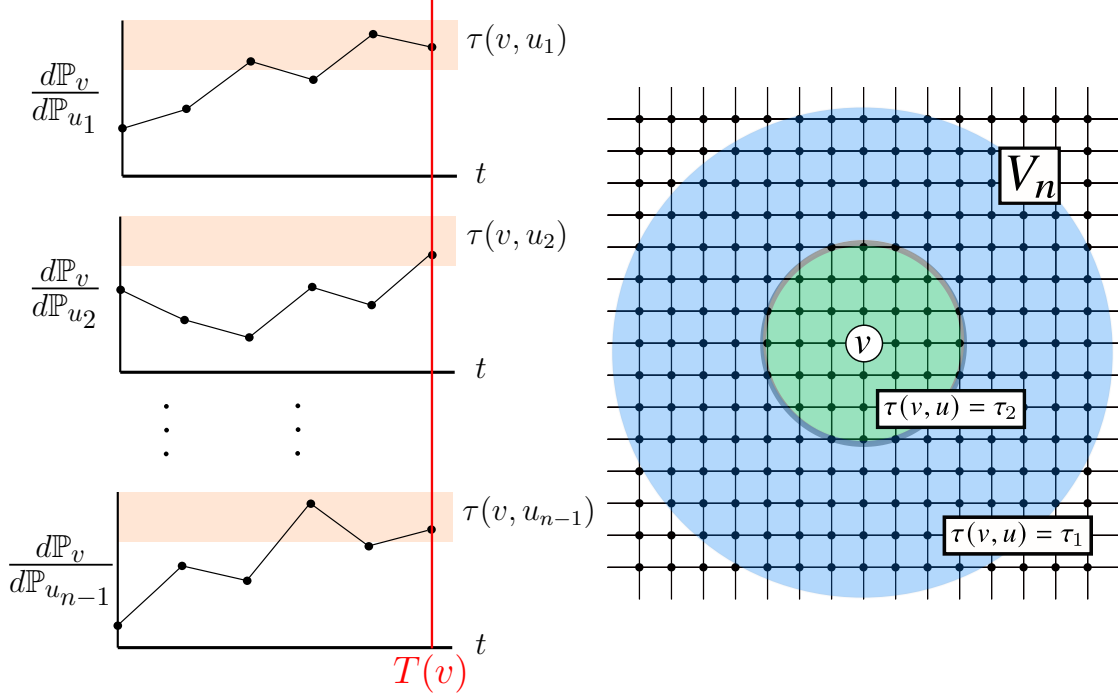


Figure 3: Schematic representation of the MSPRT we design to achieve near-optimal source estimation. *Left*: Plots of the likelihood ratios with respect to vertex v , where we have enumerated $V_n \setminus \{v\}$ as u_1, \dots, u_{n-1} . We halt when $d\mathbb{P}_v/d\mathbb{P}_{u_i}$ crosses the threshold $\tau(v, u_i)$ (shown in orange) for all $1 \leq i \leq n-1$. This stopping time, labeled $T(v)$, is shown in red. The source estimator is the vertex v which achieves the smallest value of $T(v)$. *Right*: Design of the thresholds $\tau(v, u)$, visualized here for the 2-dimensional lattice. For vertices $u, v \in V_n$ that are far (blue region), we set $\tau(v, u) = \tau_1$ and for vertices $u, v \in V_n$ that are close (green region), we set $\tau(v, u) = \tau_2$.

To this end, recall that the problem of source estimation can be viewed as a sequential multi-hypothesis testing problem, where each hypothesis corresponds to the possibility of a particular vertex being the source. Motivated by the optimality of SPRTs for the two-hypothesis setting, we consider a natural extension to the multiple hypothesis setting known as the *multi-hypothesis sequential probability ratio test* (MSPRT), described below:

- For each pair of distinct $u, v \in V_n$, specify a threshold $\tau(v, u)$, which is a positive real number.
- Let $T(v)$ be the stopping time that halts as soon as

$$\frac{d\mathbb{P}_v}{d\mathbb{P}_u}(y(0), \dots, y(t)) \geq \tau(v, u), \quad \forall u \in V_n \setminus \{v\}.$$

Here, we recall that \mathbb{P}_v is the probability measure conditioned on $v^* = v$.

- The final source estimator is $\arg \min_{v \in V_n} T(v)$; that is, the vertex whose stopping time halts first.

For general multi-hypothesis testing problems, it is known that MSPRTs enjoy near-optimal performance when the number of hypotheses is fixed and the bound on the estimation error, α , is

small [4, 42, 11, 24, 12, 10]. Although our setting is different, since α is fixed and the number of hypotheses are large, it is natural to expect that MSPRTs still have good performance. Indeed, we provide a novel way to design MSPRTs with worst-case expected runtime that match the lower bounds for $\text{val}_M^*(V_n, \alpha)$: for $u, v \in V_n$ that are “far”, we set $\tau(v, u) = \tau_1$ and for $u, v \in V_n$ that are “close” we set $\tau(v, u) = \tau_2$ where τ_1, τ_2 are pre-determined parameters depending on the graph structure and n . This design can be interpreted as a *multi-scale search strategy*: an analysis of the likelihood ratios $d\mathbb{P}_v/d\mathbb{P}_u$ for u, v far apart determine the general location of the source, and an analysis of the likelihood ratios $d\mathbb{P}_v/d\mathbb{P}_u$ for u, v close give us a more fine-grained estimate. We show that with the right definition of “closeness” as well as an appropriate choice of τ_1, τ_2 , an MSPRT designed in this way achieves the upper bounds for $\text{val}_M^*(V_n, \alpha)$ described in Theorem 3.6. A diagram illustrating the key ideas of the MSPRT we have described can be found in Figure 3.

3.3 Conjectures on optimal estimators in general graphs

In this work, we primarily focus on regular trees and lattices for a few key reasons. For one, they enjoy strong symmetry properties. In particular, the local structure of all vertex neighborhoods are isomorphic, leading to conceptually simpler proofs and near-exact computations of $\text{val}_B^*(V_n)$ as well as $\text{val}_M^*(V_n, \alpha)$. Second, regular trees and lattices have drastically different topological structure, yet most of our proofs work equally well for both topologies, with just minor differences. This suggests that our methods can also be used to describe optimal source estimators for *general topologies*. Below, we discuss how Theorems 3.5 and 3.6 may change for general topologies.

We start by defining relevant notation. Let G be a graph with (countably) infinite vertices that is locally finite (i.e., all vertex degrees are finite). As discussed earlier, we study infinite graphs since it allows us to consider scenarios where the cascade is small relative to the network size without complicating our analysis. Define the *vertex-dependent* neighborhood growth function

$$f_v(t) := \sum_{s=0}^t |\mathcal{N}_v(s)|.$$

We then have the following conjecture concerning optimal source estimation in general topologies.

Conjecture 3.7. *Suppose that G is a graph with countably infinite vertices that is locally finite. Let the sequence of candidate sets $\{V_n\}_{n \geq 1}$ satisfy Assumption 3.4. Then*

$$\text{val}_B^*(V_n) \asymp \frac{1}{n} \sum_{v \in V_n} f_v^{-1}(\log n).$$

Additionally, for any $\alpha > 0$ that is constant with respect to n ,

$$\text{val}_M^*(V_n, \alpha) \asymp \max_{v \in V_n} f_v^{-1}(\log n).$$

We expect that Conjecture 3.7 can be proved by a straightforward generalization of our techniques. Following analogous arguments as the proof summary for Theorem 3.5, if $v^* = v$ then $f_v(t)$ is the total number of public signals distributed according to Q_1 until timestep t . We therefore expect that the uncertainty in the source location is too large to reasonably estimate the source in the regime $f_v(t) \lesssim \log n$ (equivalently, $t \lesssim f_v^{-1}(\log n)$), since the entropy of the prior $\pi(V_n)$ is $\log n$. Hence any algorithm must observe for at least $f_v^{-1}(\log n)$ timesteps to reliably estimate the source. Averaging over $v \in V_n$ leads to the lower bound $\text{val}_B^*(V_n) \gtrsim \frac{1}{n} \sum_{v \in V_n} f_v^{-1}(\log n)$. On the other hand, the minimax setting captures the *worst-case* expected number of samples as opposed to the average-case number of samples, hence $\text{val}_M^*(V_n, \alpha) \gtrsim \max_{v \in V_n} f_v^{-1}(\log n)$.

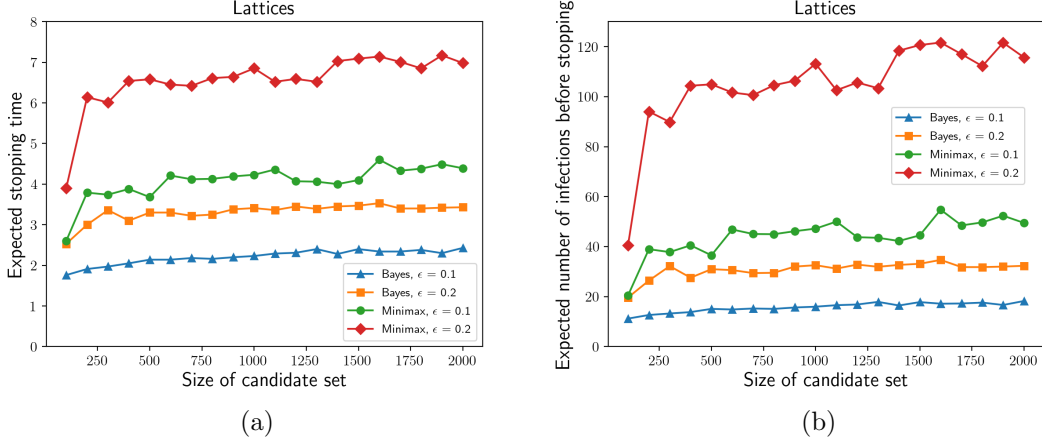


Figure 4: Plots of the expected stopping time (a) and the expected number of infections (b) as a function of n , the size of the candidate set. Pictured here are the performances of the Bayes and minimax-optimal source estimators in 2-dimensional lattices.

Establishing upper bounds for $\text{val}_B^*(V_n)$ and $\text{val}_M^*(V_n, \alpha)$ that are orderwise tight requires an analysis of specific source estimators. Since our analysis of $(T_{th}, \hat{\mathbf{v}}_B)$ is quite similar for both regular trees and lattices (with only minor differences), we expect that it should achieve optimal performance in general as well. We also believe that a properly designed MSPRT can achieve optimal performance in the minimax setting as well; we provide further details on this point in Section 6.2 (see Remark 6.8).

4 Simulations

In this section, we complement our theoretical results through simulations which reveal the *non-asymptotic* performance of our source estimators. Specifically, we study the performance of two estimators: the Bayes estimator described in (3.5) and (3.6), and the MSPRT used to prove the achievability results in Theorem 3.6. At a high level, our simulations show that even in non-asymptotic regimes, our estimators are able to locate the source while ensuring that only a small number of individuals are infected, thus validating our theoretical results on trees and lattices. We further apply our estimators to cascades spreading on natural models of random graphs (the Erdős-Rényi model), showing that our estimators can be successfully applied to broader scenarios.

Signal distributions. We consider the case of noisy and incomplete testing, described in Remark 2.1; we briefly recap the model here. Interpret the network cascade as an infection, and assume that at every timestep, each individual tests for infection with probability p . The test outputs the correct result (i.e., positive or negative) with probability $1 - \epsilon$. In our simulations, we let $p = 0.5$ and $\epsilon \in \{0.1, 0.2\}$. The distributions Q_0, Q_1 derived from this scenario are formally described in Remark 2.1.

Lattices. To make our simulations as close to our theoretical setup as possible, our base graph G was taken to be a 2-dimensional 100×100 lattice (10,000 vertices). If the size of the candidate set is n , in accordance with Assumption 3.4, we chose the candidate set to be the n closest vertices to the center of the lattice. We emphasize that choosing the candidate set in this way captures

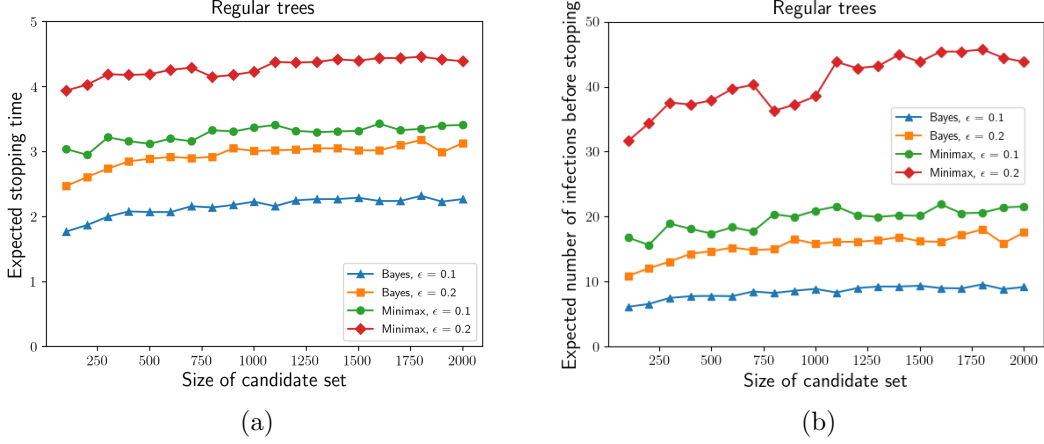


Figure 5: Plots of the expected stopping time (a) and the expected number of infections (b) as a function of n , the size of the candidate set. Pictured here are the performances of the Bayes and minimax-optimal source estimators in 3-regular trees.

the notion of a *worst-case candidate set* (see the discussion surrounding Assumption 3.4), and we choose vertices close to the lattice center only to avoid boundary effects (i.e., to ensure that the cascade will evolve similarly from all potential source vertices, given that G is finite). In our simulations, the cascade begins at the lattice center and spreads via the deterministic dynamics described in Assumption 3.1, producing random observed vertex-level signals according to (2.1). Although we could, in principle, choose any source vertex in the candidate set, we consistently choose the lattice center in order to reduce the variance of the estimators' performance across independent simulations. For each choice of n (from 100 to 2000, collected at regular intervals of 100) and $\epsilon \in \{0.1, 0.2\}$, we carried out 100 independent simulations of the cascade. We averaged over the stopping time and the number of total infections until the stopping time to generate the plots in Figure 4. The design of the MSPRT weights for the minimax estimator follows Figure 3; for the specific threshold values, see Theorem 6.7 in Section 6.2.

Figure 4 highlights important *finite-size* behaviors of the Bayes and minimax estimators. Notice that each curve is quite flat: the expected stopping time as well as the number of infections changes little with respect to n . This weak dependence extends to the asymptotic regime $n \rightarrow \infty$ as well; Theorems 3.5 and 3.6 show that the expected stopping time scales as $(\log n)^{1/3}$. Moreover, the Bayes estimator has strictly better performance than the minimax estimator in all cases – notably, nearly 100 infections are prevented in the case $\epsilon = 0.2$ for large n when comparing the Bayes and minimax estimators.

Trees. Our base graph G was taken to be a 3-regular balanced tree with 16,383 vertices. If the size of the candidate set is n , we choose the candidate set to be the n closest vertices to the root of balanced tree G for similar reasons as in the case of lattices. In our simulations, the cascade begins at the root of G and spreads via the deterministic dynamics described in Assumption 3.1, producing random observed vertex-level signals according to (2.1). For each choice of n and ϵ , we carried out 100 independent simulations of the cascade and average over the stopping time as well as the number of total infections to generate the plots in Figure 5. For the minimax estimator, we use an MSPRT with *constant* weights, which is proved to be asymptotically optimal; see Theorem 6.5 in Section 6.2 for details.

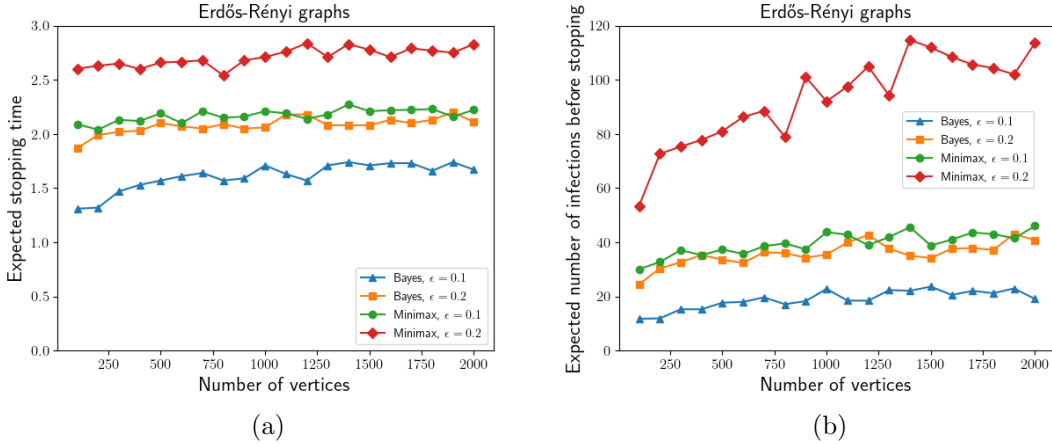


Figure 6: Plots of the expected stopping time (a) and the expected number of infections (b) as a function of n , the size of the candidate set. Pictured here are the performances of the Bayes and minimax-optimal source estimators in the Erdős-Rényi graph $G(n, 5/n)$.

Similar conclusions as in the case of lattices can be drawn for trees based on Figure 5. Interestingly, although the Bayes and minimax estimators take $\log \log n$ samples in light of Theorems 3.5 and 3.6, Figure 5 shows that the Bayes-optimal estimator stops earlier in finite regimes; we believe this is due to the provable optimality of the estimator (see Lemma 5.1). Furthermore, although we have proved that the distance between Q_0 and Q_1 does not affect the first-order asymptotics of the expected stopping time in both Bayes and minimax settings, it appears to play a significant role in finite regimes. Notably, the time it takes the minimax-optimal estimator to stop is more than doubled when $\epsilon = 0.2$ compared to $\epsilon = 0.1$ for many values of n .

Erdős-Rényi random graphs. Recall that for a positive integer n and $q \in [0, 1]$, an Erdős-Rényi random graph $G(n, q)$ is generated as follows. Let the vertex set V be a set of n labeled vertices, and for each pair of distinct vertices an edge is added between them with probability q , independently across all vertex pairs. Since our work is primarily concerned with *sparse* graphs (i.e., vertex degrees are not too large), we chose $q = 5/n$ to ensure that the average degree of the graph is 5. This choice of q ensures that the largest connected component of $G(n, q)$ (also known as the *giant component*) is most of the graph, while also keeping the average degree relatively small. For a given realization of $G(n, q)$, our candidate set was taken to be the vertices in the giant component.⁵ Figure 6 shows the performance of the Bayes and minimax estimators on Erdős-Rényi graphs, both of which exhibit similar performance to that noted for trees and lattices. For the minimax estimator, since sparse Erdős-Rényi graphs are known to be locally tree-like [1], we use the MSPRT with uniform weights (see Section 6.2) which is optimal in regular trees. To summarize, Figure 6 shows that the estimators we develop are *robust* and apply to a broader class of graphs than the ones we analyze theoretically.

To generate the numerical results in Figure 6, as before we ran 100 independent simulations for each n and ϵ considered. For each simulation, an independent Erdős-Rényi graph was generated. The data in Figure 6 were computed by averaging the stopping times and number of infected

⁵It is known that with high probability, components other than the giant component are of order $\log n$ [1, Chapter 11.9]. As a result, the infection will never reach most of the graph even after a long time passes. Our assumption that the candidate set is the vertex set of the giant component avoids such simple edge cases.

vertices, *conditioned* on the event that the cascade did not spread to all vertices by the time the algorithm stopped. The reason for this is that if the cascade affects all vertices, there is effectively no new information to be learned, and the stopping time would be extremely large with high probability. In almost all cases, however, at most one out of the 100 trials would fall into this category. The only exception was the case of $n = 100$ and $\epsilon = 0.2$ for the minimax estimator, which had 12 trials fall into this category. We expect that this is because the cascade spreads too quickly to detect it on a graph of this small size. Finally, we remark that some of the curves in Figure 6 may appear noisier than the ones in Figures 5 and 4; this is likely due to the randomness of the base graph in the Erdős-Rényi case, compared to the deterministic nature of the other topologies considered.

5 Analysis of the Bayesian setting

5.1 Behavior of the Bayes-optimal estimator

We begin with a discussion of the estimator $\hat{v}_B = \{\hat{v}_B(t)\}_{t \geq 0}$, which is defined formally by

$$\hat{v}_B(t) \in \arg \min_{v \in V_n} \mathbb{E}_{\pi(t)}[d(v^*, v)],$$

where V_n is the candidate set under consideration. It is straightforward to show that \hat{v}_B is *optimal*, in the sense that it minimizes the error of the final source estimator for any choice of stopping time. This is explained more formally in the following result.

Lemma 5.1. *Let the candidate set be V_n . Fix any stopping time T and let $\hat{u} = \{u(t)\}_{t \geq 0}$ be any source estimator so that $u(t)$ is measurable with respect to $y(0), \dots, y(t)$. Then*

$$\mathbb{E}_{\pi(V_n)}[d(v^*, \hat{v}_B(T))] \leq \mathbb{E}_{\pi(V_n)}[d(v^*, \hat{u}(T))].$$

In particular,

$$\inf_{(T, \hat{v})} \text{val}_B(T, \hat{v}) = \inf_T \text{val}_B(T, \hat{v}_B).$$

Proof. For any given time index $t \geq 0$, it follows from the definition of $\hat{v}_B(t)$ that

$$\mathbb{E}_{\pi(t)}[d(v^*, \hat{v}_B(t))] \leq \mathbb{E}_{\pi(t)}[d(v^*, \hat{u}(t))]. \quad (5.1)$$

We can then write

$$\begin{aligned} \mathbb{E}_{\pi(V_n)}[d(v^*, \hat{v}_B(T))] &\stackrel{(a)}{=} \mathbb{E}_{\pi(V_n)} [\mathbb{E}_{\pi(T)}[d(v^*, \hat{v}_B(T))]] \\ &= \mathbb{E}_{\pi(V_n)} \left[\sum_{t=0}^{\infty} \mathbb{E}_{\pi(t)}[d(v^*, \hat{v}_B(t))] \mathbf{1}(T = t) \right] \\ &\stackrel{(b)}{\leq} \mathbb{E}_{\pi(V_n)} \left[\sum_{t=0}^{\infty} \mathbb{E}_{\pi(t)}[d(v^*, \hat{u}(t))] \mathbf{1}(T = t) \right] \\ &= \mathbb{E}_{\pi(V_n)}[d(v^*, \hat{u}(T))]. \end{aligned}$$

Above, (a) is due to the tower rule and (b) is a consequence of (5.1). Taking an infimum over all source estimators (T, \hat{u}) on both sides yields the second statement of the lemma. \square

With the optimal estimator explicitly derived, we focus on characterizing the estimation error of \hat{v}_B , which will in turn aid us in characterizing optimal stopping times. Indeed, the proof of Theorem 3.5 depends on the following two lemmas which characterize the estimation error of \hat{v}_B . Before stating them, we review some basic properties of the neighborhood growth function $f(t)$ (defined in (3.4)). First, it can be shown through straightforward combinatorial arguments that

$$f(t) \sim \begin{cases} \frac{k(k-1)}{(k-2)^2} (k-1)^t & G \text{ is a } k\text{-regular tree, } k \geq 3; \\ \frac{2^\ell}{(\ell+1)!} t^{\ell+1} & G \text{ is a } \ell\text{-dimensional lattice, } \ell \geq 1. \end{cases} \quad (5.2)$$

A proof of (5.2) can be found in Appendix A – see in particular Lemmas A.1 and A.3. Importantly, (5.2) can be used to study the asymptotics of the *inverse function* $F = f^{-1}$. Indeed, it follows that

$$F(z) \sim \begin{cases} \frac{\log z}{\log(k-1)} & G \text{ is a } k\text{-regular tree, } k \geq 3; \\ \left(\frac{(\ell+1)!}{2^\ell} z \right)^{\frac{1}{\ell+1}} & G \text{ is a } \ell\text{-dimensional lattice, } \ell \geq 1. \end{cases} \quad (5.3)$$

We are now ready to state our results on the estimation error. The first establishes a lower bound for the estimation error when t is not too large.

Lemma 5.2. *Suppose G is a k -regular tree with $k \geq 3$ and that the sequence of candidate sets $\{V_n\}_{n \geq 1}$ satisfies Assumption 3.4. There are constants $a_1 = a_1(k, Q_0, Q_1)$ and $b_1 = b_1(k)$ such that*

$$\lim_{n \rightarrow \infty} \max_{v \in V_n} \mathbb{P}_v \left(\min_{0 \leq t \leq F(a_1 \log n)} \mathbb{E}_{\pi(t)}[d(v^*, \hat{v}_B(t))] \leq b_1 \log n \right) = 0.$$

Next suppose that G is a ℓ -dimensional lattice and $\{V_n\}_{n \geq 1}$ satisfies Assumption 3.4. There are constants $a'_1 = a'_1(Q_0, Q_1)$ and $b'_1 = b'_1(\ell)$ such that

$$\lim_{n \rightarrow \infty} \max_{v \in V_n} \mathbb{P}_v \left(\min_{0 \leq t \leq F(a'_1 \log n)} \mathbb{E}_{\pi(t)}[d(v^*, \hat{v}_B(t))] \leq b'_1 n^{1/\ell} \right) = 0.$$

It can be shown through straightforward combinatorial arguments (see Lemmas B.3 and B.5 in Appendix B) that the initial estimation error satisfies

$$\mathbb{E}_{\pi(0)}[d(v^*, \hat{v}_B(0))] \asymp \begin{cases} \log n & G \text{ is a } k\text{-regular tree;} \\ n^{1/\ell} & G \text{ is a } \ell\text{-dimensional lattice.} \end{cases} \quad (5.4)$$

In light of (5.4), Lemma 5.2 states that with high probability, the estimation error does not significantly decrease for $t \lesssim F(\log n)$. At a high level, this is because the information from the public signals corresponding to the true spread of the cascade is not enough to offset the uncertainty in the source location. These ideas are formalized by computing the mean and variance of $\mathbb{E}_{\pi(t)}[d(v^*, \hat{v}(t))]$ and applying Chebyshev's inequality. We also remark that as a consequence of our proofs, the constant a_1 in Lemma 5.2 depends on the average degree for regular trees, while the constant a'_1 does not depend on the average degree in lattices. For details, see Section 7.2.

The next result establishes an upper bound on the estimation error once t is sufficiently large.

Lemma 5.3. *Suppose that G is a k -regular tree or a ℓ -dimensional lattice and that the sequence of candidate sets $\{V_n\}_{n \geq 1}$ satisfies Assumption 3.4. There are constants $a_2 = a_2(Q_0, Q_1)$ and $b_2 = b_2(Q_0, Q_1)$ such that if $t \geq F(a_2 \log n)$,*

$$\max_{v \in V_n} \mathbb{P}_v \left(\mathbb{E}_{\pi(t)}[d(v^*, \hat{v}_B(t))] \geq e^{-b_2 t} \right) \leq e^{-b_2 t}. \quad (5.5)$$

The proof relies on large-deviations bounds which show that $\pi_u(t)/\pi_{v^*}(t)$ tends to 0 at an exponential rate for any $u \neq v^*$. For details, see Section 7.3.

5.2 Putting everything together: Proof of Theorem 3.5

Combined, Lemmas 5.2 and 5.3 show that the estimation error exhibits a sharp transition: it is large for $t \lesssim F(\log n)$ and it is exponentially decaying to zero for $t \gtrsim F(\log n)$. The optimality and performance of the estimator $(T_{th}, \hat{\mathbf{v}}_B)$ is essentially obtained from this observation.

Proof of Theorem 3.5. Define the pair of constants a, A so that $(a, A) = (a_1, b_1 \log n)$ if G is a k -regular tree and $(a, A) = (a'_1, b'_1 n^{1/\ell})$ if G is a ℓ -dimensional lattice (see Lemma 5.2 for definitions of these constants). For every $t \geq 0$, we also define the event

$$\mathcal{E} := \left\{ \min_{0 \leq t \leq F(a \log n)} \mathbb{E}_{\pi(t)}[d(v^*, \hat{v}_B(t))] \geq A \right\}.$$

For any stopping time T , we have, for any $v \in V_n$,

$$\begin{aligned} \mathbb{E}_v [\mathbb{E}_{\pi(T)}[d(v^*, \hat{v}_B(T))] + T] &= \mathbb{E}_v \left[\sum_{t=0}^{\infty} \mathbb{E}_{\pi(t)}[d(v^*, \hat{v}_B(t))] \mathbf{1}(T = t) + T \right] \\ &\geq \mathbb{E}_v \left[\sum_{t=0}^{F(a \log n)} \mathbb{E}_{\pi(t)}[d(v^*, \hat{v}_B(t))] \mathbf{1}(\{T = t\} \cap \mathcal{E}) + T \mathbf{1}(T > F(a \log n)) \right] \\ &\stackrel{(a)}{\geq} A \sum_{t=0}^{F(a \log n)} \mathbb{P}_v(\{T = t\} \cap \mathcal{E}) + F(a \log n) \cdot \mathbb{P}_v(T > F(a \log n)) \\ &\geq A \cdot \mathbb{P}_v(\{T \leq F(a \log n)\} \cap \mathcal{E}) + F(a \log n) \cdot \mathbb{P}_v(\{T > F(a \log n)\} \cap \mathcal{E}) \\ &\geq \min\{A, F(a \log n)\} \cdot \mathbb{P}_v(\mathcal{E}) \\ &\stackrel{(b)}{=} F(a \log n) \cdot \mathbb{P}_v(\mathcal{E}). \end{aligned} \tag{5.6}$$

Above, (a) follows from the definition of \mathcal{E} and by lower bounding T by $F(a \log n)$ on the event $\{T > F(a \log n)\}$, and (b) uses $F(a \log n) \leq A$ for n sufficiently large, which follows from the asymptotic behavior of F (see (5.3)). An important consequence of (5.6) is that the value associated with the pair $(T, \hat{\mathbf{v}})$ can be lower bounded as

$$\begin{aligned} \text{val}_B(T, \hat{\mathbf{v}}_B) &= \mathbb{E}_{\pi(V_n)}[d(v^*, \hat{v}_B(T)) + T] \\ &\stackrel{(c)}{=} \mathbb{E}_{\pi(V_n)} [\mathbb{E}_{\pi(T)}[d(v^*, \hat{v}_B(T))] + T] \\ &\stackrel{(d)}{=} \frac{1}{n} \sum_{v \in V_n} \mathbb{E}_v [\mathbb{E}_{\pi(T)}[d(v^*, \hat{v}_B(T))] + T] \\ &\stackrel{(e)}{\geq} F(a \log n) \left(\frac{1}{n} \sum_{v \in V_n} \mathbb{P}_v(\mathcal{E}) \right). \end{aligned} \tag{5.7}$$

Above, (c) is due to the tower rule, (d) follows since $\pi(V_n)$ is a uniform distribution over elements of V_n , and (e) is a consequence of (5.6). Moreover, since (5.7) holds for *any* stopping time T , we have

$$\begin{aligned} \text{val}_B^*(V_n) &= \inf_{T, \hat{\mathbf{v}}} \text{val}_B(T, \hat{\mathbf{v}}) \\ &= \inf_T \text{val}_B(T, \hat{\mathbf{v}}_B) \\ &\geq F(a \log n) \left(\frac{1}{n} \sum_{v \in V_n} \mathbb{P}_v(\mathcal{E}) \right), \end{aligned}$$

where the second equality follows from the optimality of the estimator $\widehat{\mathbf{v}}_B$, proved in Lemma 5.1. Rearranging terms and sending $n \rightarrow \infty$, we arrive at

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{\text{val}_B^*(V_n)}{F(a \log n)} &\geq \liminf_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{v \in V_n} \mathbb{P}_v(\mathcal{E}) \right) \\ &\geq \liminf_{n \rightarrow \infty} \min_{v \in V_n} \mathbb{P}_v(\mathcal{E}) = 1, \end{aligned} \quad (5.8)$$

where the final equality above is a direct consequence of Lemma 5.2.

Next, to establish an asymptotic upper bound for the optimal value, we consider the stopping time

$$T_{th} := \min\{t \geq 0 : \mathbb{E}_{\pi(t)}[d(v^*, \widehat{v}_B(t))] \leq 1\},$$

which stops once the estimation error falls below a threshold. We can then bound

$$\begin{aligned} \text{val}_B(T_{th}, \widehat{\mathbf{v}}_B) &\stackrel{(f)}{\leq} \mathbb{E}_{\pi(V_n)}[1 + T_{th}] = 1 + \sum_{t=0}^{\infty} \mathbb{P}_{\pi(V_n)}(T_{th} > t) \\ &\leq 1 + F(a_2 \log n) + \sum_{t=F(a_2 \log n)+1}^{\infty} \mathbb{P}_{\pi(V_n)}(T_{th} > t) \\ &\stackrel{(g)}{\leq} 1 + F(a_2 \log n) + \sum_{t=F(a_2 \log n)+1}^{\infty} \max_{v \in V_n} \mathbb{P}_v(T_{th} > t) \\ &\stackrel{(h)}{\leq} 1 + F(a_2 \log n) + \sum_{t=F(a_2 \log n)+1}^{\infty} e^{-b_2 t} \\ &\leq 1 + F(a_2 \log n) + \frac{1}{b_2} e^{-b_2 F(a_2 \log n)}. \end{aligned}$$

Above, the inequality (f) is due to the definition of T_{th} , (g) follows since $\mathbb{P}_{\pi(V_n)}(\cdot) := \frac{1}{n} \sum_{v \in V_n} \mathbb{P}_v(\cdot) \leq \max_{v \in V_n} \mathbb{P}_v(\cdot)$, and (h) follows from noting that $T_{th} > t$ implies that $\mathbb{E}_{\pi(t)}[d(v^*, \widehat{v}_B(t))] > 1 \geq e^{-b_2 t}$ and applying Lemma 5.3 to bound the latter event. Dividing both sides of the final inequality by $F(a_2 \log n)$ and letting $n \rightarrow \infty$ shows that

$$\limsup_{n \rightarrow \infty} \frac{\text{val}_B(T_{th}, \widehat{\mathbf{v}}_B)}{F(a_2 \log n)} \leq 1. \quad (5.9)$$

The desired result follows from (5.8) and (5.9) by considering the asymptotic behavior of F in trees and lattices (see (5.3)). \square

Remark 5.4 (General temporal cost functions). *It is also interesting to consider the case where the cost of the stopping time in the Bayesian objective (2.3) is given by $h(T)$ instead of T , where h is some increasing function. It turns out that a slight modification of the proof of Theorem 3.5 shows that the estimator $(T_{th}, \widehat{\mathbf{v}}_B)$ is still asymptotically near-optimal as long as h grows slower than any exponential function. Indeed, if we follow the derivation of the bound in (5.6) and (5.7), we obtain that*

$$\mathbb{E}_{\pi(V_n)}[d(v^*, \widehat{v}_B(T) + h(T))] \geq \min\{A, h(F(a \log n))\} \left(\frac{1}{n} \sum_{v \in V_n} \mathbb{P}_v(\mathcal{E}) \right),$$

where A, a, \mathcal{E} are the same as in the proof of Theorem 3.5. Following (5.8), we obtain that

$$\liminf_{n \rightarrow \infty} \frac{\inf_T \mathbb{E}_{\pi(V_n)}[d(v^*, \hat{v}_B(T)) + h(T)]}{\min\{A, h(F(a \log n))\}} \geq 1.$$

In particular, if h increases slower than any exponential function, it can be seen that, for the choices of A and F used for regular trees and lattices, $h(F(a \log n))$ is an asymptotic lower bound on the performance of any estimator.

The proof of the upper bound on the performance of the optimal estimator can be similarly derived. For the same estimator $(\hat{\mathbf{v}}_B, T_{th})$, we may follow the proof of Theorem 3.5 to show that

$$\begin{aligned} \inf_{(\hat{\mathbf{v}}, T_{th})} \mathbb{E}_{\pi(V_n)}[d(\hat{v}(T), v^*) + h(T)] &\leq \mathbb{E}_{\pi(V_n)}[1 + h(T_{th})] \\ &= 1 + \sum_{t=0}^{\infty} \mathbb{P}_{\pi(V_n)}(h(T_{th}) > t) \\ &= 1 + \sum_{t=0}^{\infty} \mathbb{P}_{\pi(V_n)}(T_{th} > h^{-1}(t)) \\ &\leq 1 + h(F(a_2 \log n)) + \sum_{t=h(F(a_2 \log n))+1}^{\infty} e^{-b_2 h^{-1}(t)}. \end{aligned}$$

We can reach the same conclusion as in Theorem 3.5; that is, that

$$\limsup_{n \rightarrow \infty} \frac{\inf_{(\hat{\mathbf{v}}, T)} \mathbb{E}_{\pi(V_n)}[d(\hat{v}(t), v^*) + h(T)]}{h(F(a_2 \log n))} \leq 1,$$

provided $\sum_{t \geq 0} e^{-b_2 h^{-1}(t)} < \infty$. This is the case provided $h^{-1}(t)$ increases faster than $\log t$ or equivalently, if $h(t)$ increases slower than any exponential function. Together, this shows that

$$h(F(a_1 \log n)) \lesssim \inf_{(\hat{\mathbf{v}}, T)} \mathbb{E}_{\pi(V_n)}[d(\hat{v}(t), v^*) + h(T)] \lesssim h(F(a_2 \log n)),$$

provided h increases slower than any exponential function.

6 Analysis of the minimax setting

In this section we prove Theorem 3.6. To do so, we prove lower and upper bounds for $\text{val}_M^*(V_n, \alpha)$ in separate theorems. In Section 6.1, we prove Theorem 6.1, which establishes a lower bound for $\text{val}_M^*(V_n, \alpha)$. The upper bounds for $\text{val}_M^*(V_n, \alpha)$ are achieved by MSPRTs of a particular design: this is proved in Theorem 6.5 (regular trees) and in Theorem 6.7, which can be found in Section 6.2. Remarkably, a simple design in which we set all the thresholds in the MSPRT to be the same value achieves the lower bound for regular trees. However, the same estimator adapted to lattices fails to achieve the lower bound due to key differences in the topology of lattices. To fix this issue, we consider a novel MSPRT design we term *K-level thresholds*, where we assign different thresholds to pairs of vertices u, v satisfying $d(u, v) \leq K$ and those satisfying $d(u, v) > K$. The performance of the resulting estimator does indeed achieve (up to a constant factor) the lower bound for $\text{val}_M^*(V_n, \alpha)$.

6.1 Lower bounding the optimal value

Theorem 6.1 (Lower bound part of Theorem 3.6). *Suppose that the sequence of candidate sets $\{V_n\}_{n \geq 1}$ satisfies Assumption 3.4. If G is a k -regular tree,*

$$\liminf_{n \rightarrow \infty} \frac{\text{val}_M^*(V_n, \alpha)}{\frac{\log \log n}{\log(k-1)}} \geq 1. \quad (6.1)$$

On the other hand, if G is a ℓ -dimensional lattice then there is a constant a_4 depending on ℓ, Q_0, Q_1 such that

$$\liminf_{n \rightarrow \infty} \frac{\text{val}_M^*(V_n, \alpha)}{(\log n)^{\frac{1}{\ell+1}}} \geq a_4. \quad (6.2)$$

Proof. Let $(T, \hat{\mathbf{v}})$ be an estimator in $\Delta(V_n, \alpha)$. Given the prior $\pi(V_n)$, recall from Lemma 5.1 that the estimator

$$\hat{v}_B(t) \in \arg \min_{v \in V_n} \mathbb{E}_{\pi(t)}[d(v^*, v)]$$

minimizes $\mathbb{E}_{\pi(V_n)}[d(v^*, \hat{v}(T))]$ over all estimators $\hat{\mathbf{v}}$, for *any* stopping time T .

For given constants a, A , define the events

$$\begin{aligned} \mathcal{E}_1 &:= \{T \leq F(a \log n)\} \\ \mathcal{E}_2 &:= \left\{ \min_{0 \leq t \leq F(a \log n)} \mathbb{E}_{\pi(t)}[d(v^*, \hat{v}_B(t))] \geq A \right\}. \end{aligned}$$

We can then lower bound the worst-case estimation error as

$$\begin{aligned} \max_{v \in V_n} \mathbb{E}_v[d(v, \hat{v}(T))] &\stackrel{(a)}{\geq} \mathbb{E}_{\pi(V_n)}[d(v^*, \hat{v}(T))] \\ &\stackrel{(b)}{\geq} \mathbb{E}_{\pi(V_n)}[d(v^*, \hat{v}_B(T))] \\ &\stackrel{(c)}{\geq} \mathbb{E}_{\pi(V_n)}[\mathbb{E}_{\pi(T)}[d(v^*, \hat{v}_B(T))] \mathbf{1}(\mathcal{E}_1 \cap \mathcal{E}_2)] \\ &\stackrel{(d)}{\geq} A \cdot \mathbb{P}_{\pi(V_n)}(\mathcal{E}_1 \cap \mathcal{E}_2). \end{aligned} \quad (6.3)$$

Above, (a) follows since $\mathbb{E}_{\pi(V_n)}$ is an average over the collection of operators $\{\mathbb{E}_v\}_{v \in V_n}$ and the maximum is greater than the average; (b) is due to the optimality of $\hat{\mathbf{v}}_B$, proved in Lemma 5.1; inequality (c) is due to the tower rule and inequality (d) follows since $\mathbb{E}_{\pi(T)}[d(v^*, \hat{v}_B(T))] \geq A$ on the event $\mathcal{E}_1 \cap \mathcal{E}_2$.

Noting that $\max_{v \in V_n} \mathbb{E}_{\pi(V_n)}[d(v^*, \hat{v}(T))] \leq \alpha$ for $(T, \hat{\mathbf{v}}) \in \Delta(V_n, \alpha)$, (6.3) implies

$$\mathbb{P}_{\pi(V_n)}(\mathcal{E}_1 \cap \mathcal{E}_2) \leq \frac{\alpha}{A},$$

which in turn implies that

$$\begin{aligned} \mathbb{P}_{\pi(V_n)}(\mathcal{E}_1^c) &\geq \mathbb{P}_{\pi(V_n)}(\mathcal{E}_1^c \cup \mathcal{E}_2^c) - \mathbb{P}_{\pi(V_n)}(\mathcal{E}_2^c) \\ &\geq 1 - \frac{\alpha}{A} - \mathbb{P}_{\pi(V_n)}(\mathcal{E}_2^c). \end{aligned} \quad (6.4)$$

By Lemma 5.2, if we set $(a, A) = (a_1, b_1 \log n)$ if G is a k -regular tree and $(a, A) = (a'_1, b'_1 n^{1/\ell})$ if G is a ℓ -dimensional lattice (see Lemma 5.2 for definitions of these constants) then the right hand side of (6.4) tends to 1 as $n \rightarrow \infty$. We can then lower bound the worst-case expected runtime as

$$\begin{aligned} \max_{v \in V_n} \mathbb{E}_v[T] &\geq \mathbb{E}_{\pi(V_n)}[T] \geq \mathbb{E}_{\pi(V_n)}[T \mathbf{1}(\mathcal{E}_1^c)] \\ &\geq F(a \log n) \mathbb{P}_{\pi(V_n)}(\mathcal{E}_1^c). \end{aligned} \quad (6.5)$$

Since (6.5) holds for *any* element of $\Delta(V_n, \alpha)$, we have

$$\text{val}_M^*(V_n, \alpha) \geq F(a \log n) \mathbb{P}_{\pi(V_n)}(\mathcal{E}_1^c).$$

Dividing by $F(a \log n)$ and sending $n \rightarrow \infty$, we arrive at

$$\liminf_{n \rightarrow \infty} \frac{\text{val}_M^*(V_n, \alpha)}{F(a \log n)} \geq 1.$$

The desired result follows from the asymptotic behavior of F (see (5.3)). \square

6.2 Achieving the lower bound with MSPRTs

An important question is whether the lower bound in Theorem 6.1 is achievable. To answer this question, we focus on a class of sequential estimation procedures called *multi-hypothesis sequential probability ratio tests* (MSPRTs). For convenience, we define them formally below.

Given vertices $u, v \in V$, define the log-likelihood ratio

$$Z_{vu}(t) := \log \frac{d\mathbb{P}_v}{d\mathbb{P}_u}(y(0), \dots, y(t)).$$

In our model, the likelihood of observing the sequence of public signals $\mathbf{y}(0), \dots, \mathbf{y}(t)$ under the measure \mathbb{P}_v has the form

$$d\mathbb{P}_v(y(0), \dots, y(t)) = \prod_{s=0}^t \left(\left[\prod_{w \in \mathcal{N}_v(s)} dQ_1(y_w(s)) \right] \left[\prod_{w \notin \mathcal{N}_v(s)} dQ_0(y_w(s)) \right] \right). \quad (6.6)$$

Above, the product containing terms of the form $dQ_1(y_w(s))$ computes the likelihood of nodes that have been affected by the cascade, and the product containing the terms of the form $dQ_0(y_w(s))$ computes the likelihood of nodes that have not yet been affected by the cascade. In light of (6.6), $Z_{vu}(t)$ can be written as

$$\begin{aligned} Z_{vu}(t) &= \sum_{s=0}^t \log \frac{\left(\prod_{w \in \mathcal{N}_v(s)} dQ_1(y_w(s)) \right) \left(\prod_{w \notin \mathcal{N}_v(s)} dQ_0(y_w(s)) \right)}{\left(\prod_{w \in \mathcal{N}_u(s)} dQ_1(y_w(s)) \right) \left(\prod_{w \notin \mathcal{N}_u(s)} dQ_0(y_w(s)) \right)} \\ &= \sum_{s=0}^t \left(\sum_{w \in \mathcal{N}_v(s) \setminus \mathcal{N}_u(s)} \log \frac{dQ_1}{dQ_0}(y_w(s)) - \sum_{w \in \mathcal{N}_u(s) \setminus \mathcal{N}_v(s)} \log \frac{dQ_1}{dQ_0}(y_w(s)) \right). \end{aligned}$$

In words, the log-likelihood ratio $\log dQ_1/dQ_0$ measures how likely it is that a certain public signal came from the distribution Q_1 as opposed to the distribution Q_0 . It follows that summations

of the form $\sum_{w \in S} \log \frac{dQ_1}{dQ_0}(y_w(s))$ measures the *net likelihood* that the public signals in S are caused by the cascade. Hence if it is more likely that the public signals in $\mathcal{N}_v(s) \setminus \mathcal{N}_u(s)$ are caused by the cascade rather than the public signals in $\mathcal{N}_u(s) \setminus \mathcal{N}_v(s)$, the log-likelihood ratio $Z_{vu}(t)$ is positive; else it is negative.

We are now ready to define the MSPRT.

Definition 6.2 (Multi-hypothesis sequential probability ratio test). *Fix a positive integer n and specify a threshold function $\tau : V_n \times V_n \rightarrow (0, \infty)$. Consider the stopping time*

$$T(v) := \min\{t \geq 0 : Z_{vu}(t) \geq \log \tau(v, u), \forall u \in V_n \setminus \{v\}\}.$$

The corresponding MSPRT (T, \hat{v}) is defined via

$$T := \min_{v \in V_n} T(v) \quad \text{and} \quad \hat{v}(T) := \arg \min_{v \in V_n} T(v).$$

In words, the output of the MSPRT is the first vertex v for which all log-likelihood ratios $Z_{vu}(t)$ pass the thresholds $\log \tau(v, u)$ for all $u \in V_n \setminus \{v\}$. If (T, \hat{v}) is a MSPRT with threshold function τ , we have the following useful relation, which is due to Tartakovsky [42, Theorem 3.1].

Lemma 6.3. *For any distinct $u, v \in V_n$,*

$$\mathbb{P}_v(\hat{v}(T) = u) \leq \frac{1}{\tau(u, v)}$$

Since the proof is short, we provide it here for completeness.

Proof of Lemma 6.3.

$$\begin{aligned} \mathbb{P}_v(\hat{v}(T) = u) &= \mathbb{E}_v[\mathbf{1}(\hat{v}(T) = u)] \\ &= \mathbb{E}_u[\mathbf{1}(\hat{v}(T) = u)e^{-Z_{uv}(T)}] \\ &\leq e^{-\log \tau(u, v)} \mathbb{E}_u[\mathbf{1}(\hat{v}(T) = u)] \\ &\leq \frac{1}{\tau(u, v)}. \end{aligned}$$

Above, the equality in the second line follows from $e^{-Z_{uv}(T)} = \frac{d\mathbb{P}_v}{d\mathbb{P}_u}(y(0), \dots, y(T))$. The inequality in the third line follows since $Z_{uv}(T) \geq \log \tau(u, v)$ on the event $\{\hat{v}(T) = u\}$ by the definition of a MSPRT. \square

Using Lemma 6.3, we can bound the worst-case estimation error of the MSPRT (T, \hat{v}) as

$$\begin{aligned} \max_{v \in V_n} \mathbb{E}_v[d(v, \hat{v}(T))] &= \max_{v \in V_n} \sum_{u \in V_n} d(u, v) \mathbb{P}_v(\hat{v}(T) = u) \\ &\leq \max_{v \in V_n} \sum_{u \in V_n} \frac{d(u, v)}{\tau(u, v)}. \end{aligned}$$

To ensure that $(T, \hat{v}) \in \Delta(V_n, \alpha)$, it suffices to check that

$$\max_{v \in V_n} \sum_{u \in V_n} \frac{d(u, v)}{\tau(u, v)} \leq \alpha. \quad (6.7)$$

Perhaps the simplest weight design which satisfies (6.7) is a *uniform weights design*, in which the $\tau(v, u)$'s all take on the same value.

Definition 6.4 (Uniform weights design). *The MSPRT $(T_{n,\alpha}, \hat{\mathbf{v}}_{n,\alpha})$ is designed with uniform weights if*

$$\tau(v, u) := \frac{n^2}{\alpha}, \quad \forall u, v \in V_n : u \neq v. \quad (6.8)$$

With the uniform weights design, we have

$$\max_{v \in V_n} \sum_{u \in V_n} d(u, v) \frac{\alpha}{n^2} \leq \max_{v \in V_n} \sum_{u \in V_n} \frac{\alpha}{n} = \alpha.$$

Hence $(T_{n,\alpha}, \hat{\mathbf{v}}_{n,\alpha}) \in \Delta(V_n, \alpha)$. Note that in the first equality above, we have used the (loose) bound $\max_{u,v \in V_n} d(v, u) \leq n$.

When G is a regular tree, the following result shows that the performance of the MSPRT with uniform weights matches the lower bound in Theorem 6.1.

Theorem 6.5. *Let G be a k -regular tree with $k \geq 3$ and fix $\alpha > 0$. If $(T_{n,\alpha}, \hat{\mathbf{v}}_{n,\alpha})$ is the MSPRT with uniform weights,*

$$\limsup_{n \rightarrow \infty} \frac{\max_{v \in V_n} \mathbb{E}_v[T_{n,\alpha}]}{\frac{\log \log n}{\log(k-1)}} \leq 1.$$

Combined with Theorem 6.1, Theorem 3.6 for regular trees follows as an immediate consequence.

Proof of Theorem 3.6 for regular trees. We have the series of inequalities

$$\begin{aligned} 1 &\stackrel{(a)}{\leq} \liminf_{n \rightarrow \infty} \frac{\text{val}_M^*(V_n, \alpha)}{\frac{\log \log n}{\log(k-1)}} \\ &\leq \limsup_{n \rightarrow \infty} \frac{\text{val}_M^*(V_n, \alpha)}{\frac{\log \log n}{\log(k-1)}} \\ &\stackrel{(b)}{\leq} \limsup_{n \rightarrow \infty} \frac{\max_{v \in V_n} \mathbb{E}_v[T_{n,\alpha}]}{\frac{\log \log n}{\log(k-1)}} \\ &\stackrel{(c)}{\leq} 1. \end{aligned}$$

Above, (a) is due to Theorem 6.1, (b) follows since $(T_{n,\alpha}, \hat{\mathbf{v}}_{n,\alpha}) \in \Delta(V_n, \alpha)$ and (c) is due to Theorem 6.5. Since both the start and the end of the chain of inequalities is 1, the inequalities are all *equality*. Hence the following limit is well-defined:

$$\lim_{n \rightarrow \infty} \frac{\text{val}_M^*(V_n, \alpha)}{\frac{\log \log n}{\log(k-1)}} = 1,$$

which proves the desired result. \square

We provide a brief sketch of the proof of Theorem 6.5, and defer the details to Section 8.2. Suppose that $v \in V_n$ is the true source. When t is sufficiently large, we show that it holds for all $u \in V_n \setminus \{v\}$ that

$$Z_{vu}(t) \sim \mathbb{E}_v[Z_{vu}(t)] \asymp \sum_{s=0}^t |\mathcal{N}_v(s) \setminus \mathcal{N}_u(s)| \asymp (k-1)^t.$$

This in particular implies that once $t \gtrsim \frac{\log \log(n^2/\alpha)}{\log(k-1)}$, all the log-likelihood ratios $\{Z_{vu}(t)\}_{u \in V_n \setminus \{v\}}$ will cross the threshold $\log n^2/\alpha$. Since $\log \log(n^2/\alpha) \sim \log \log n$ when n is large, Theorem 6.5 follows. The key technical ingredient of this proof is a large-deviations-type inequality for $Z_{vu}(t)$.

Unfortunately, the MSPRT with a uniform weights design is *not* optimal in lattices. Suppose that the dimension of the lattice is ℓ . For $u, v \in V_n$ which are far apart, we have $Z_{vu}(t) \asymp t^{\ell+1}$, but for u, v which are relatively close together, $Z_{vu}(t) \asymp t^\ell$. As the log-likelihood $Z_{vu}(t)$ grows at a slower rate in this latter case, this ends up being the primary contributor to the behavior of the stopping time. As a result, we obtain an upper bound for $T^*(V_n, \alpha)$ of order $(\log n)^{1/\ell}$, whereas the lower bound established in Theorem 6.1 is of order $(\log n)^{1/(\ell+1)}$. To close this gap, we shall consider a design for the MSPRT weights which places different thresholds for pairs of vertices that are close and pairs that are far.

Definition 6.6 (K -level weights). *Let K be a non-negative integer. The MSPRT $(T_{n,\alpha}^K, \hat{\mathbf{v}}_{n,\alpha}^K)$ is designed with K -level weights if*

$$\tau(v, u) := \begin{cases} \frac{2K|\mathcal{N}(K)|}{\alpha} & 0 < d(u, v) \leq K \\ \frac{2n^2}{\alpha} & \text{else.} \end{cases}$$

It is straightforward to show that the K -level weights satisfy (6.7):

$$\begin{aligned} \max_{v \in V_n} \sum_{u \in V_n} \frac{d(u, v)}{\tau(u, v)} &= \max_{v \in V_n} \left(\sum_{u \in V_n \cap \mathcal{N}_v(K)} \frac{d(u, v)}{\tau(u, v)} + \sum_{u \in V_n \setminus \mathcal{N}_v(K)} \frac{d(u, v)}{\tau(u, v)} \right) \\ &\leq \max_{v \in V_n} \left(\sum_{u \in V_n \cap \mathcal{N}_v(K)} \frac{\alpha}{2|\mathcal{N}(K)|} + \sum_{u \in V_n \setminus \mathcal{N}_v(K)} \frac{\alpha}{2n} \right) \\ &\leq \alpha, \end{aligned}$$

where to derive the inequality on the second line, we used $d(u, v) \leq K$ for $u \in \mathcal{N}_v(K)$ and $d(u, v) \leq n$ for $u \in V_n$. Putting everything together, we have shown that $(T_{n,\alpha}^K, \hat{\mathbf{v}}_{n,\alpha}^K) \in \Delta(V_n, \alpha)$.

At a high level, the MSPRT designed with K -level weights can be thought of as a *multi-scale* source estimation algorithm. When the $Z_{vu}(t)$'s are large for many vertices u far from v (specifically, $d(u, v) > K$), this indicates that v must be relatively close to the source. Assuming K is not too large, there are roughly n vertices far from v , which means that the threshold for the log-likelihood ratios *must* be on the order of $\log n$ in order to reliably narrow down the general location of the source. Simultaneously, the $Z_{vu}(t)$'s for u close to v (specifically, $d(u, v) \leq K$) provide *fine-grained* information about the location of the source within the localized region $\mathcal{N}_v(K)$. To compensate for the slower growth of $Z_{vu}(t)$ in this case, we require that they pass the much smaller threshold of $\log(2K|\mathcal{N}(K)|/\alpha)$, hence removing the bottleneck found in the uniform weights design.

In the following result, we show that for the right value of K , the MSPRT designed with K -level weights is *orderwise optimal* in lattices.

Theorem 6.7. *Let G be an ℓ -dimensional lattice and set $K = (\log n)^{1/\ell}$. Then there is a constant b_4 depending only on ℓ, Q_0, Q_1 such that*

$$\limsup_{n \rightarrow \infty} \frac{\max_{v \in V_n} \mathbb{E}_v[T_{n,\alpha}^K]}{(\log n)^{\frac{1}{\ell+1}}} \leq b_4,$$

where a_4 is the constant used in (6.2).

Combined with Theorem 6.1, Theorem 3.6 for lattices immediately follows.

Proof of Theorem 3.6 for lattices. Let G be the ℓ -dimensional lattice. We have the series of inequalities

$$\begin{aligned}
a_4 &\stackrel{(a)}{\leq} \liminf_{n \rightarrow \infty} \frac{\text{val}_M^*(V_n, \alpha)}{(\log n)^{\frac{1}{\ell+1}}} \\
&\leq \limsup_{n \rightarrow \infty} \frac{\text{val}_M^*(V_n, \alpha)}{(\log n)^{\frac{1}{\ell+1}}} \\
&\stackrel{(b)}{\leq} \limsup_{n \rightarrow \infty} \frac{\max_{v \in V_n} \mathbb{E}_v[T_{n,\alpha}^K]}{(\log n)^{\frac{1}{\ell+1}}} \\
&\stackrel{(c)}{\leq} b_4.
\end{aligned}$$

Above, (a) is due to Theorem 6.1, (b) follows since $(T_{n,\alpha}^K, \hat{\mathbf{v}}_{n,\alpha}^K) \in \Delta(V_n, \alpha)$, and (c) is due to Theorem 6.7. In particular, the chain of inequalities implies that for n sufficiently large,

$$\begin{aligned}
\frac{a_4}{2} (\log n)^{\frac{1}{\ell+1}} &\leq \text{val}_M^*(V_n, \alpha) \\
&\leq \max_{v \in V_n} \mathbb{E}_v[T_{n,\alpha}^K] \\
&\leq 2b_4 (\log n)^{\frac{1}{\ell+1}}.
\end{aligned}$$

The theorem follows from setting $a' := a_4/2$ and $b' := 2b_4$. \square

The proof of Theorem 6.7 follows similar reasoning as the proof of Theorem 6.5. In the case that v is the true source of the cascade, we separately consider the performance of $Z_{vu}(t)$ for $u \in \mathcal{N}_v(K)$ and $u \in V_n \setminus \mathcal{N}_v(K)$ and use large-deviations-type results to characterize when the log-likelihood ratios cross the thresholds specified by the K -level weights design. For full details, see Section 8.3.

Remark 6.8 (MSPRTs for general topologies). *The K -level weights introduced in Definition 6.6 can be generalized to arbitrary topologies if we set*

$$\tau(v, u) := \begin{cases} \frac{2K|\mathcal{N}_v(K)|}{\alpha} & 0 < d(v, u) \leq K; \\ \frac{2n^2}{\alpha} & \text{else.} \end{cases}$$

Above, we write $|\mathcal{N}_v(K)|$ instead of $|\mathcal{N}(K)|$ since the size of the K -hop neighborhood of a vertex v can depend strongly on v in general topologies. Since the proof of Theorem 6.7 is quite generic for the most part, we expect that a MSPRT with K -level weights can achieve the lower bound for $\text{val}_M^*(V_n, \alpha)$ in general. The choice of K , however, will depend on the topology of interest.

7 Analysis of the Bayesian estimation error: Proof of Lemmas 5.2 and 5.3

7.1 Preliminary results: Properties of the posterior distribution

Before proving Lemmas 5.2 and 5.3, we introduce some simple supporting results. Recall that the posterior distribution is given by $\pi(t) = \{\pi_v(t)\}_{v \in V_n}$, where

$$\pi_v(t) := \mathbb{P}_{\pi(V_n)}(v^* = v \mid y(0), \dots, y(t)).$$

From Bayes' formula, it holds for any distinct $u, v \in V_n$ that

$$\frac{\pi_v(t)}{\pi_u(t)} = \frac{\pi_v(t-1)}{\pi_u(t-1)} \cdot \frac{d\mathbb{P}_v(y(t))}{d\mathbb{P}_u(y(t))}. \quad (7.1)$$

Recall that under \mathbb{P}_v , $y_w(t) \sim Q_1$ if $w \in \mathcal{N}_v(t)$, else $y_w(t) \sim Q_0$. Since the public signals $y(0), \dots, y(t)$ are independent conditioned on the source, the likelihood ratio $d\mathbb{P}_v(y(t))/d\mathbb{P}_u(y(t))$ can be written as

$$\begin{aligned} \frac{d\mathbb{P}_v(y(t))}{d\mathbb{P}_u(y(t))} &= \frac{\left(\prod_{w \in \mathcal{N}_v(t)} dQ_1(y_w(t)) \right) \left(\prod_{w \notin \mathcal{N}_v(t)} dQ_0(y_w(t)) \right)}{\left(\prod_{w \in \mathcal{N}_u(t)} dQ_1(y_w(t)) \right) \left(\prod_{w \notin \mathcal{N}_u(t)} dQ_0(y_w(t)) \right)} \\ &= \frac{\prod_{w \in \mathcal{N}_v(t) \setminus \mathcal{N}_u(t)} \frac{dQ_1}{dQ_0}(y_w(t))}{\prod_{w \in \mathcal{N}_u(t) \setminus \mathcal{N}_v(t)} \frac{dQ_1}{dQ_0}(y_w(t))} \end{aligned} \quad (7.2)$$

Combining (7.1) and (7.2), we have

$$\frac{\pi_v(t)}{\pi_u(t)} = \frac{\prod_{s=0}^t \prod_{w \in \mathcal{N}_v(s) \setminus \mathcal{N}_u(s)} \frac{dQ_1}{dQ_0}(y_w(s))}{\prod_{s=0}^t \prod_{w \in \mathcal{N}_u(s) \setminus \mathcal{N}_v(s)} \frac{dQ_1}{dQ_0}(y_w(s))} \quad (7.3)$$

$$= \frac{\prod_{s=0}^t \prod_{w \in \mathcal{N}_v(s)} \frac{dQ_1}{dQ_0}(y_w(s))}{\prod_{s=0}^t \prod_{w \in \mathcal{N}_u(s)} \frac{dQ_1}{dQ_0}(y_w(s))}. \quad (7.4)$$

Equation (7.3) follows directly from (7.1) and (7.2). The difference between (7.3) and (7.4) is that we take a product over $w \in \mathcal{N}_u(s) \setminus \mathcal{N}_v(s)$ and $w \in \mathcal{N}_v(s) \setminus \mathcal{N}_u(s)$ in the former, and $w \in \mathcal{N}_u(s)$ and $w \in \mathcal{N}_v(s)$ in the latter. The expressions are equivalent since the vertices in $\mathcal{N}_u(s) \cap \mathcal{N}_v(s)$ are cancelled out. We display both equations, as each will be useful in different contexts.

Equation (7.4) implies that the posterior probabilities can be written as

$$\pi_v(t) = \frac{1}{Y(t)} \prod_{s=0}^t \prod_{w \in \mathcal{N}_v(s)} \frac{dQ_1}{dQ_0}(y_w(s)),$$

where the normalizing constant, $Y(t)$, is given by

$$Y(t) := \sum_{v \in V_n} \prod_{s=0}^t \prod_{w \in \mathcal{N}_v(s)} \frac{dQ_1}{dQ_0}(y_w(s)).$$

It will be convenient to use the notation $\pi_v(t) = X_v(t)/Y(t)$, where $X_v(t)$ is explicitly given by

$$X_v(t) := \prod_{s=0}^t \prod_{w \in \mathcal{N}_v(s)} \frac{dQ_1}{dQ_0}(y_w(s)).$$

With this notation, $Y(t) = \sum_{u \in V_n} X_u(t)$. The following lemma establishes some basic properties of the collection $\{X_u(t)\}_{u \in V_n}$.

Lemma 7.1. *Denote*

$$\beta := \mathbb{E}_{\mathcal{A} \sim Q_1} \left[\frac{dQ_1}{dQ_0}(\mathcal{A}) \right].$$

Then $\beta > 1$ and for any $u, v \in V_n$ and $t \geq 0$,

$$\mathbb{E}_v[X_u(t)] = \beta^{\sum_{s=0}^t |\mathcal{N}_v(s) \cap \mathcal{N}_u(s)|}.$$

Proof. We start by proving $\beta > 1$. We can write

$$\begin{aligned} \mathbb{E}_{\mathcal{A} \sim Q_1} \left[\frac{dQ_1}{dQ_0}(\mathcal{A}) \right] &= \mathbb{E}_{\mathcal{B} \sim Q_0} \left[\left(\frac{dQ_1}{dQ_0}(\mathcal{B}) \right)^2 \right] \\ &\geq \mathbb{E}_{\mathcal{B} \sim Q_0} \left[\frac{dQ_1}{dQ_0}(\mathcal{B}) \right]^2 = 1. \end{aligned}$$

Above, the first equality is due to a change of measure (a valid operation since Q_0, Q_1 are mutually absolutely continuous), and the inequality is due to Jensen's inequality. Above, the inequality is strict since the equality case only occurs if $\frac{dQ_1}{dQ_0}(\mathcal{B})$ is a constant (equivalently, $Q_0 = Q_1$).

Since $\{y_w(s)\}_{w \in V, 0 \leq s \leq t}$ is a collection of independent random variables conditioned on $v^* = v$, we have

$$\mathbb{E}_v[X_u(t)] = \mathbb{E}_v \left[\prod_{s=0}^t \prod_{w \in \mathcal{N}_u(s)} \frac{dQ_1}{dQ_0}(y_w(s)) \right] \quad (7.5)$$

$$= \prod_{s=0}^t \prod_{w \in \mathcal{N}_u(s)} \mathbb{E}_v \left[\frac{dQ_1}{dQ_0}(y_w(s)) \right]. \quad (7.6)$$

For each term in the product, we have

$$\mathbb{E}_v \left[\frac{dQ_1}{dQ_0}(y_w(s)) \right] = \begin{cases} \beta & d(v, w) \leq s \\ 1 & \text{else,} \end{cases} \quad (7.7)$$

where we have used the fact that $y_w(s) \sim Q_1$ in the first case, and $y_w(s) \sim Q_0$ in the second. The desired statement follows from substituting (7.7) into (7.5). \square

The following lemma bounds the covariance between $X_u(t)$ and $X_v(t)$. As a matter of notation, we recall that the *neighborhood growth function* (defined in (3.4)) is given by $f(t) := \sum_{s=0}^t |\mathcal{N}(s)|$.

Lemma 7.2. *For any $u, v, w \in V_n$ and $t \geq 0$, there is a constant $\lambda = \lambda(Q_0, Q_1)$ such that $\text{Cov}_v(X_u(t), X_w(t)) = 0$ if $d(u, w) > 2t$ and $\text{Cov}_v(X_u(t), X_w(t)) \leq \lambda^{f(t)}$ if $d(u, w) \leq 2t$.*

Proof. Since $X_u(t)$ depends only on the signals in $\mathcal{N}_u(t)$, it is clear that $X_u(t)$ and $X_w(t)$ are independent under \mathbb{P}_v if $d(u, w) > 2t$. Hence $\text{Cov}_v(X_u(t), X_w(t)) = 0$ in this case. To handle the case where $d(u, w) \leq 2t$, we first define

$$\lambda_1 := \mathbb{E}_{\mathcal{A} \sim Q_1} \left[\left(\frac{dQ_1}{dQ_0}(\mathcal{A}) \right)^2 \right], \lambda_0 := \mathbb{E}_{\mathcal{B} \sim Q_0} \left[\left(\frac{dQ_1}{dQ_0}(\mathcal{B}) \right)^2 \right].$$

We have the following bound on the covariance due to the Cauchy-Schwartz inequality.

$$\begin{aligned} \text{Cov}_v(X_u(t), X_w(t)) &\leq \mathbb{E}_v[X_u(t)X_w(t)] \\ &\leq \mathbb{E}_v[X_u(t)^2]^{1/2} \mathbb{E}_v[X_w(t)^2]^{1/2}. \end{aligned}$$

To bound $\mathbb{E}_v[X_u(t)^2]$, we can write

$$\begin{aligned}\mathbb{E}_v[X_u(t)^2] &= \prod_{s=0}^t \prod_{a \in \mathcal{N}_u(s)} \mathbb{E}_v \left[\left(\frac{dQ_1}{dQ_0}(y_a(s)) \right)^2 \right] \\ &= \lambda_1^{\sum_{s=0}^t |\mathcal{N}_u(s) \cap \mathcal{N}_v(s)|} \lambda_0^{\sum_{s=0}^t |\mathcal{N}_u(s) \setminus \mathcal{N}_v(s)|} \\ &\leq (\max\{\lambda_0, \lambda_1\})^{\sum_{s=0}^t |\mathcal{N}_u(s)|} \\ &= (\max\{\lambda_0, \lambda_1\})^{f(t)}.\end{aligned}$$

Since the bound we have derived holds for any $u \in V_n$, it follows that $\text{Cov}_v(X_u(t), X_w(t)) \leq (\max\{\lambda_0, \lambda_1\})^{f(t)}$, which proves the desired claim with $\lambda := \max\{\lambda_0, \lambda_1\}$. \square

The results we have established allow us to prove the following concentration result for $Y(t)$ when t is not too large.

Lemma 7.3. *Recall the constants $\beta = \beta(Q_0, Q_1)$ (defined in Lemma 7.1) and $\lambda = \lambda(Q_0, Q_1)$ (defined in 7.2). Furthermore let $F := f^{-1}$ be the inverse of the neighborhood growth function. If*

$$t \leq F \left(\frac{\log n}{4 \log(\max\{\beta, \lambda\})} \right),$$

then for any $\epsilon > 0$,

$$\max_{v \in V_n} \mathbb{P}_v(|Y(t) - n| \geq \epsilon n) \leq \frac{4}{\epsilon^2 \sqrt{n}}. \quad (7.8)$$

Proof. Let $\beta = \beta(Q_0, Q_1)$ be the constant defined in Lemma 7.1. We begin by computing the expectation of $Y(t)$ with respect to \mathbb{P}_v .

$$\begin{aligned}\mathbb{E}_v[Y(t)] &= \sum_{u \in V_n} \mathbb{E}_v[X_u(t)] \\ &= \sum_{u \in V_n} \beta^{\sum_{s=0}^t |\mathcal{N}_v(s) \cap \mathcal{N}_u(s)|} \\ &= \sum_{u \in \mathcal{N}_v(2t)} \beta^{\sum_{s=0}^t |\mathcal{N}_v(s) \cap \mathcal{N}_u(s)|} + |V_n \setminus \mathcal{N}_v(2t)|.\end{aligned}$$

Since $\beta > 1$ and $\sum_{s=0}^t |\mathcal{N}_v(s) \cap \mathcal{N}_u(s)| \leq f(t)$, we have the bounds

$$n \leq \mathbb{E}_v[Y(t)] \leq n + |\mathcal{N}(2t)|\beta^{f(t)}.$$

From Lemmas A.1 and A.3, we have the asymptotics

$$\begin{aligned}|\mathcal{N}(2t)| &\asymp \begin{cases} (k-1)^{2t} & G \text{ is a } k\text{-regular tree;} \\ t^\ell & G \text{ is a } \ell\text{-dimensional lattice} \end{cases} \\ f(t) &\asymp \begin{cases} (k-1)^t & G \text{ is a } k\text{-regular tree;} \\ t^{\ell+1} & G \text{ is a } \ell\text{-dimensional lattice.} \end{cases}\end{aligned}$$

Hence we have, for t sufficiently large, the simpler upper bound of $n + \beta^{2f(t)}$ for $\mathbb{E}_v[Y(t)]$. Next, suppose that t satisfies $f(t) \leq \frac{\log n}{4 \log \beta}$ so that $\beta^{2f(t)} \leq \sqrt{n}$. Then

$$\begin{aligned}\mathbb{P}_v(|Y(t) - n| \geq \epsilon n) &\leq \mathbb{P}_v\left(|Y(t) - \mathbb{E}_v[Y(t)]| \geq \frac{\epsilon}{2}n\right) \\ &\leq \frac{4 \cdot \text{Var}_v(Y(t))}{\epsilon^2 n^2},\end{aligned} \quad (7.9)$$

where the first inequality holds if $\epsilon \geq \frac{2}{\sqrt{n}}$ and the second inequality is due to Chebyshev's inequality. Using Lemma 7.2, we can bound the variance of $Y(t)$ as

$$\begin{aligned}\text{Var}_v(Y(t)) &= \sum_{u,w \in V_n} \text{Cov}_v(X_u(t), X_w(t)) \\ &\leq \sum_{u \in V_n} |\mathcal{N}(2t)| \lambda^{f(t)} \\ &\leq 2nf(t)^2 \lambda^{f(t)} \\ &\leq n\lambda^{2f(t)},\end{aligned}$$

where we have used $|\mathcal{N}(2t)| \leq f(2t) \leq 2f(t)^2$ and $\lambda = \lambda(Q_0, Q_1)$ is defined in Lemma 7.2. Moreover, if $f(t) \leq \frac{\log n}{4 \log \lambda}$, then $\lambda^{2f(t)} \leq \sqrt{n}$ and $\text{Var}_v(Y(t)) \leq n^{3/2}$. The desired result follows. \square

Our final result establishes exponential lower tail bounds for the ratio $\pi_v(t)/\pi_u(t)$ using Chernoff bounds. Before stating the result, we define some notation. Let

$$D(Q_0, Q_1) := \mathbb{E}_{\mathcal{A} \sim Q_1} \left[\log \frac{dQ_1}{dQ_0}(\mathcal{A}) \right] + \mathbb{E}_{\mathcal{B} \sim Q_0} \left[\log \frac{dQ_0}{dQ_1}(\mathcal{B}) \right] \quad (7.10)$$

denote the *symmetrized Kullback-Liebler divergence*, and define the rate function

$$I(x) := \sup_{\lambda \geq 0} \left\{ -\lambda(D(Q_0, Q_1) - x) - \log \mathbb{E} \left[\left(\frac{dQ_0}{dQ_1}(\mathcal{A}) \right)^\lambda \left(\frac{dQ_1}{dQ_0}(\mathcal{B}) \right)^\lambda \right] \right\}, \quad (7.11)$$

where $\mathcal{A} \sim Q_1$ and $\mathcal{B} \sim Q_0$ are independent. Finally, for $u, v \in V$, define the *neighborhood difference function*

$$f_{vu}(t) := \sum_{s=0}^t |\mathcal{N}_v(s) \setminus \mathcal{N}_u(s)|.$$

Lemma 7.4. *Let $u, v \in V$ be any two vertices. If $x > 0$, then $I(x) > 0$ and*

$$\mathbb{P}_v \left(\frac{\pi_v(t)}{\pi_u(t)} \leq e^{(D(Q_0, Q_1) - x)f_{vu}(t)} \right) \leq e^{-I(x)f_{vu}(t)}, \quad x > 0. \quad (7.12)$$

Proof. Conditioned on \mathcal{A}, \mathcal{B} , define the function

$$g(\lambda) := \left(\frac{dQ_0}{dQ_1}(\mathcal{A}) \frac{dQ_1}{dQ_0}(\mathcal{B}) \right)^\lambda.$$

Note that $g(\lambda)$ is differentiable, and by the mean-value theorem, we have for any $\lambda_1 \in [0, 1]$ that there exists $\eta \in [0, \lambda_1]$ so that

$$\begin{aligned}\frac{|g(\lambda_1) - 1|}{\lambda_1} &= |g'(\eta)| = \eta \left| \frac{dQ_0}{dQ_1}(\mathcal{A}) \frac{dQ_1}{dQ_0}(\mathcal{B}) \right|^{\eta-1} \\ &\leq 1 + \left| \frac{dQ_0}{dQ_1}(\mathcal{A}) \frac{dQ_1}{dQ_0}(\mathcal{B}) \right|.\end{aligned} \quad (7.13)$$

In the final inequality, we have used $\eta \leq 1$ as well as the fact that $a^{\eta-1} \leq a$ if $a \geq 1$ else $a^{\eta-1} \leq 1$ if $0 \leq a \leq 1$. Furthermore, note that due to the independence of \mathcal{A} and \mathcal{B} ,

$$\mathbb{E}_{\mathcal{A} \sim Q_1, \mathcal{B} \sim Q_0} \left[\frac{dQ_0}{dQ_1}(\mathcal{A}) \frac{dQ_1}{dQ_0}(\mathcal{B}) \right] = \mathbb{E}_{\mathcal{A} \sim Q_1} \left[\frac{dQ_0}{dQ_1}(\mathcal{A}) \right] \mathbb{E}_{\mathcal{B} \sim Q_0} \left[\frac{dQ_1}{dQ_0}(\mathcal{B}) \right] = 1.$$

It then follows from the definition of the Lebesgue integral that

$$\mathbb{E} \left[\left| \frac{dQ_0}{dQ_1}(\mathcal{A}) \frac{dQ_1}{dQ_0}(\mathcal{B}) \right| \right] < \infty. \quad (7.14)$$

In particular, the dominating function for $|g(\lambda_1) - 1|/\lambda_1$ in (7.13) is integrable. Together, (7.13) and (7.14) along with the Lebesgue Dominated Convergence Theorem imply that $\mathbb{E}[g(\lambda)]$ is differentiable at $\lambda = 0$. Moreover, the derivative of $\log \mathbb{E}[g(\lambda)]$ at $\lambda = 0$ is equal to $-D(Q_0, Q_1)$. Next, if we define the function

$$h(\lambda) := -\lambda(D(Q_0, Q_1) - x) - \log \mathbb{E}[g(\lambda)],$$

then the results we have established thus far imply $h'(0) = x > 0$. Since $h(0) = 0$, it follows that $h(\lambda)/\lambda > 0$ for sufficiently small λ , and in particular $h(\lambda) > 0$. Since $I(x) = \sup_{\lambda \geq 0} h(\lambda)$, the claim $I(x) > 0$ follows.

We now show how one can use the rate function $I(x)$ to obtain the inequality (7.12). Recall that under the measure \mathbb{P}_v , the variables $\{y_w(s)\}_{w \in V, 0 \leq s \leq t}$ are independent, with

$$y_w(s) \sim \begin{cases} Q_1 & w \in \mathcal{N}_v(s) \\ Q_0 & \text{else.} \end{cases}$$

Using the representation (7.3), the following distributional identity holds under the measure \mathbb{P}_v :

$$\frac{\pi_v(t)}{\pi_u(t)} \stackrel{d}{=} \prod_{i=1}^{f_{vu}(t)} W_i, \quad (7.15)$$

where the W_i 's are i.i.d. with distribution given by

$$W_i \stackrel{d}{=} \frac{dQ_1}{dQ_0}(\mathcal{A}) \left(\frac{dQ_1}{dQ_0}(\mathcal{B}) \right)^{-1}, \quad (7.16)$$

for independent $\mathcal{A} \sim Q_1$ and $\mathcal{B} \sim Q_0$. A Chernoff-type bound implies that

$$\begin{aligned} \mathbb{P}_v \left(\frac{\pi_v(t)}{\pi_u(t)} \leq e^{(D(Q_0, Q_1) - x)f_{vu}(t)} \right) &= \inf_{\lambda \geq 0} \mathbb{P}_v \left(\left(\frac{\pi_v(t)}{\pi_u(t)} \right)^{-\lambda} \geq e^{-\lambda(D(Q_0, Q_1) - x)f_{vu}(t)} \right) \\ &\leq \inf_{\lambda \geq 0} \exp \left(\lambda(\tilde{D}(Q_0, Q_1) - x)f_{vu}(t) + \log \mathbb{E} \left[\left(\frac{\pi_v(t)}{\pi_u(t)} \right)^{-\lambda} \right] \right) \\ &= \exp(-I(x)f_{vu}(t)), \end{aligned}$$

where the final expression follows from the distributional representation for $\pi_v(t)/\pi_u(t)$ in (7.15) and (7.16). \square

7.2 Lower bounding the estimation error: Proof of Lemma 5.2

At a high level, the proof strategy is to first establish a probabilistic lower bound for $\mathbb{E}_{\pi(t)}[d(v^*, u)]$ where $u \in V_n$ is *fixed*. Through union bounds, this will lead to a probabilistic lower bound for $\min_{u \in V_n} \mathbb{E}_{\pi(t)}[d(v^*, u)]$. We remark that the proof of Lemma 5.2 makes use of some combinatorial properties of trees and lattices, the proofs of which may be found in Appendix B.

For a fixed vertex $u \in V_n$, we can write

$$\begin{aligned}\mathbb{E}_{\pi(t)}[d(v^*, u)] &= \sum_{w \in V_n} d(w, u) \pi_w(t) \\ &= \frac{1}{Y(t)} \sum_{w \in V_n} d(w, u) X_w(t).\end{aligned}$$

Lemma 7.3 has already established a concentration inequality for $Y(t)$, so we will proceed by establishing a probabilistic lower bound for $\sum_{w \in V_n} d(u, w) X_w(t)$. For any $v \in V_n$, we have

$$\begin{aligned}\mathbb{E}_v \left[\sum_{w \in V_n} d(u, w) X_w(t) \right] &= \sum_{w \in V_n} d(u, w) \mathbb{E}_v[X_w(t)] \\ &\geq \sum_{w \in V_n} d(u, w),\end{aligned}\tag{7.17}$$

where the inequality is due to $\mathbb{E}_v[X_w(t)] \geq 1$, which was proved in Lemma 7.1. We can also upper bound the variance as

$$\begin{aligned}\text{Var}_v \left(\sum_{w \in V_n} d(u, w) X_w(t) \right) &= \sum_{w_1 \in V_n} \sum_{w_2 \in V_n} d(u, w_1) d(u, w_2) \text{Cov}_v(X_{w_1}(t), X_{w_2}(t)) \\ &\stackrel{(a)}{\leq} \lambda^{f(t)} \sum_{w_1 \in V_n} \sum_{w_2 \in V_n: d(w_1, w_2) \leq 2t} d(u, w_1) d(u, w_2) \\ &\stackrel{(b)}{\leq} \lambda^{f(t)} |\mathcal{N}(2t)| \sum_{w_1 \in V_n} d(u, w_1) (d(u, w_1) + 2t) \\ &= \lambda^{f(t)} |\mathcal{N}(2t)| \left(\sum_{w \in V_n} d(u, w)^2 + 2t \sum_{w \in V_n} d(u, w) \right) \\ &\stackrel{(c)}{\leq} \lambda^{f(t)} |\mathcal{N}(2t)| (1 + 2t) \sum_{w \in V_n} d(u, w)^2,\end{aligned}\tag{7.18}$$

where (a) follows from Lemma 7.2, (b) is due to the inequality $d(u, w_2) \leq d(u, w_1) + d(w_1, w_2) \leq d(u, w_1) + 2t$, and (c) follows from bounding $d(u, w) \leq d(u, w)^2$.

Next, an application of inequality (7.17) and Chebyshev's inequality yields

$$\mathbb{P}_v \left(\sum_{w \in V_n} d(u, w) X_w(t) \leq \frac{1}{2} \sum_{w \in V_n} d(u, w) \right) \leq 4\lambda^{f(t)} |\mathcal{N}(2t)| (1 + 2t) \cdot \frac{\sum_{w \in V_n} d(u, w)^2}{\left(\sum_{w \in V_n} d(u, w) \right)^2},\tag{7.19}$$

where the right hand side uses the variance upper bound (7.18). To proceed, we bound the right hand side of (7.19) when G is a regular tree or a lattice. Although we treat these cases separately for convenience, the methodology is the same.

Case 1: G is a k -regular tree.

In this case, Lemma B.3 shows that

$$\begin{aligned}\sum_{w \in V_n} d(u, w) &\geq \frac{n \log n}{k \log(k-1)} \\ \sum_{w \in V_n} d(u, w)^2 &\leq \frac{4n \log^2 n}{\log^2(k-1)}.\end{aligned}$$

Substituting the above bounds into (7.19) shows that

$$\begin{aligned} \mathbb{P}_v \left(\sum_{w \in V_n} d(u, w) X_w(t) \leq \frac{n \log n}{2k \log(k-1)} \right) &\leq \lambda^{f(t)} |\mathcal{N}(2t)| (1+2t) \frac{16k^2}{n} \\ &\leq \frac{50t \lambda^{f(t)} |\mathcal{N}(2t)| k^2}{n}. \end{aligned} \quad (7.20)$$

A naïve method for deriving a probabilistic bound for $\min_{u \in V_n} \sum_{w \in V_n} d(u, w) X_w(t)$ is to take a union bound over the events pertaining to $\sum_{w \in V_n} d(u, w) X_w(t)$ for all $u \in V_n$. However, the probability bound in (7.20) is not small enough for a union bound to work, since $|V_n| = n$. Fortunately, as we shall see, it suffices to take a union bound over a much smaller set of vertices. To this end, define

$$m := \left\lceil r_n + 1 - \frac{\log n}{6k \log(k-1)} \right\rceil.$$

Since $r_n \sim \log(n)/\log(k-1)$ (see (3.1)), $m \sim (1 - \frac{1}{6k}) r_n$. In addition, it holds for n sufficiently large that

$$\begin{aligned} |\mathcal{N}_{v_0}(m)| &\stackrel{(d)}{=} 1 + \frac{k}{k-2} ((k-1)^m - 1) \\ &\stackrel{(e)}{\leq} 1 + \frac{k}{k-2} (k-1)^{(1-\frac{1}{12k}) \frac{\log n}{\log(k-1)}} \\ &\stackrel{(f)}{\leq} 2kn^{1-\frac{1}{12k}}, \end{aligned}$$

where (d) follows from Lemma A.1, (e) holds since $m \leq (1 - \frac{1}{12k}) \frac{\log n}{\log(k-1)}$ for n sufficiently large due to the asymptotics of r_n , and (f) follows from upper bounding the coefficient on the first-order term in the second line. Combining (7.20) with a union bound over elements of $\mathcal{N}_{v_0}(m)$ implies

$$\mathbb{P}_v \left(\min_{u \in \mathcal{N}_{v_0}(m)} \sum_{w \in V_n} d(u, w) X_w(t) \leq \frac{n \log n}{2k \log(k-1)} \right) \leq 100t \lambda^{f(t)} |\mathcal{N}(2t)| k^3 n^{-\frac{1}{12k}}. \quad (7.21)$$

Next, define the event

$$\mathcal{E} := \left\{ \frac{n}{2} \leq Y(t) \leq \frac{3n}{2} \right\}.$$

If the event \mathcal{E} holds, we have the series of implications

$$\begin{aligned} \min_{u \in \mathcal{N}_{v_0}(m)} \sum_{w \in V_n} d(u, w) \pi_w(t) &\leq \frac{\log n}{3k \log(k-1)} \Rightarrow \min_{u \in \mathcal{N}_{v_0}(m)} \frac{1}{3n/2} \sum_{w \in V_n} d(u, w) X_w(t) \leq \frac{\log n}{3k \log(k-1)} \\ &\Rightarrow \min_{u \in \mathcal{N}_{v_0}(m)} \sum_{w \in V_n} d(u, w) X_w(t) \leq \frac{n \log n}{2k \log(k-1)}. \end{aligned}$$

Above, the first implication uses $\pi_w(t) = X_w(t)/Y(t)$ and the fact that $Y(t) \leq 3n/2$ on \mathcal{E} . We then have, for $t \leq F \left(\frac{\log n}{4 \log \max\{\beta, \lambda\}} \right)$,

$$\begin{aligned} \mathbb{P}_v \left(\min_{u \in \mathcal{N}_{v_0}(m)} \sum_{w \in V_n} d(u, w) \pi_w(t) \leq \frac{\log n}{3k \log(k-1)} \right) &\leq \mathbb{P}_v \left(\min_{u \in \mathcal{N}_{v_0}(m)} \sum_{w \in V_n} d(u, w) X_w(t) \leq \frac{n \log n}{2k \log(k-1)} \right) \\ &\quad + \mathbb{P}_v(\mathcal{E}^c) \\ &\leq 100t \lambda^{f(t)} |\mathcal{N}(2t)| k^3 n^{-\frac{1}{12k}} + \frac{16}{\sqrt{n}}. \end{aligned} \quad (7.22)$$

The final inequality above follows from Lemma 7.3 as well as the bound in (7.21). If we additionally have $t \leq F\left(\frac{\log n}{24k \log \lambda}\right)$, we have the bounds

$$\begin{aligned} t &\leq (1 + o_n(1)) \frac{\log \log n}{\log(k-1)} \\ \lambda^{f(t)} &\leq n^{\frac{1}{24k}}, \\ |\mathcal{N}(2t)| &\leq \log^{2+o_n(1)} n, \end{aligned}$$

where $o_n(1) \rightarrow 0$ as $n \rightarrow \infty$. This shows in particular that the final expression in (7.22) can be bounded for n sufficiently large by

$$\frac{16}{\sqrt{n}} + \frac{200k^3(\log \log n)(\log^3 n)}{\log(k-1)} n^{-\frac{1}{24k}} \leq n^{-\frac{1}{30k}}.$$

To put everything together, the way we have defined m implies that for every $u \in V_n$ there is $u' \in \mathcal{N}_{v_0}(m)$ such that $d(u, u') \leq \frac{\log n}{6k \log(k-1)}$. Moreover,

$$\begin{aligned} \sum_{w \in V_n} d(u, w) \pi_w(t) &\geq \sum_{w \in V_n} (d(u', w) - d(u, u')) \pi_w(t) \\ &\geq \sum_{w \in V_n} d(u', w) \pi_w(t) - \frac{\log n}{6k \log(k-1)}. \end{aligned}$$

We therefore have

$$\min_{u' \in \mathcal{N}_{v_0}(m)} \sum_{w \in V_n} d(u', w) \pi_w(t) > \frac{\log n}{3k \log(k-1)} \Rightarrow \min_{u \in V_n} \sum_{w \in V_n} d(u, w) \pi_w(t) > \frac{\log n}{6k \log(k-1)}.$$

To summarize, we have shown that if we set

$$a_1 := (\max\{4 \log \beta, 24k \log \lambda\})^{-1},$$

then for $t \leq F(a_1 \log n)$,

$$\mathbb{P}_v \left(\min_{u \in V_n} \sum_{w \in V_n} d(u, w) \pi_w(t) \leq \frac{\log n}{6k \log(k-1)} \right) \leq n^{-\frac{1}{30k}}.$$

Taking a union bound over $0 \leq t \leq F(a_1 \log n)$, and recalling that $\mathbb{E}_{\pi(t)}[d(v^*, u)] = \sum_{w \in V_n} d(u, w) \pi_w(t)$, we arrive at

$$\mathbb{P}_v \left(\max_{0 \leq t \leq F(a_1 \log n)} \mathbb{E}_{\pi(t)}[d(v^*, \hat{v}_B(t))] \leq \frac{\log n}{6k \log(k-1)} \right) \leq F(a_1 \log n) n^{-\frac{1}{30k}}.$$

Since $F(a_1 \log n) \sim \frac{\log \log n}{\log(k-1)}$, the right hand side tends to 0 uniformly over $v \in V_n$ as $n \rightarrow \infty$, and the result of Lemma 5.2 for trees follows.

Case 2: G is a ℓ -dimensional lattice.

By Lemma B.5, there exist constants $c_3, c_4 > 0$ depending only on ℓ such that

$$\begin{aligned} \sum_{w \in V_n} d(u, w) &\geq c_3 n^{1+\frac{1}{\ell}} \\ \sum_{w \in V_n} d(u, w)^2 &\leq c_4 n^{1+\frac{2}{\ell}}. \end{aligned}$$

Substituting the above bounds into (7.19) shows that

$$\begin{aligned} \mathbb{P}_v \left(\sum_{w \in V_n} d(u, w) X_w(t) \leq \frac{c_3}{2} n^{1+\frac{1}{\ell}} \right) &\leq 4\lambda^{f(t)} |\mathcal{N}(2t)| (1+2t) \frac{c_4 n^{1+2/\ell}}{c_3^2 n^{2+2/\ell}} \\ &\leq \left(\frac{12c_4}{c_3} \right) \frac{t\lambda^{f(t)} |\mathcal{N}(2t)|}{n}. \end{aligned}$$

As in the case of regular trees, we need to take a union bound over a small set of vertices. The following combinatorial lemma guarantees the existence of such a set.

Lemma 7.5. *For every n , there exists a set S_n whose size can be bounded as a function of ℓ only, such that for every $u \in V_n$, there exists $u' \in S_n$ such that $d(u, u') \leq \frac{c_3}{6} n^{1/\ell}$.*

Before proving the lemma, we shall show how we can use it to prove Lemma 5.2 for lattices. Mirroring the steps of (7.22) in the case of lattices, if $t \leq F \left(\frac{\log n}{4 \log \max\{\beta, \lambda\}} \right)$ we arrive at the probability bound

$$\mathbb{P}_v \left(\min_{u \in S_n} \sum_{w \in V_n} d(u, w) \pi_w(t) \leq \frac{c_3}{3} n^{1/\ell} \right) \leq \frac{16}{\sqrt{n}} + \frac{12c_4}{c_3^2} |S_n| \cdot \frac{t\lambda^{f(t)} |\mathcal{N}(2t)|}{n}. \quad (7.23)$$

If we additionally have $t \leq F \left(\frac{\log n}{3 \log \lambda} \right)$, we have the bounds

$$\begin{aligned} t &\leq O \left((\log n)^{\frac{1}{\ell+1}} \right) \\ \lambda^{f(t)} &\leq n^{1/3} \\ |\mathcal{N}(2t)| &\leq O \left((\log n)^{\frac{\ell}{\ell+1}} \right). \end{aligned}$$

The big- O bounds follow from the asymptotic behavior of F (see (5.3)) as well as the asymptotic behavior of $|\mathcal{N}(t)|$ (see (A.8)). For n sufficiently large, we can therefore bound the right hand side in (7.23) by

$$\frac{16}{\sqrt{n}} + O((\log n) n^{-2/3}) \leq \frac{20}{\sqrt{n}}, \quad (7.24)$$

where the final inequality holds for n sufficiently large. Putting everything together, since for every $u \in V_n$ we can find $u' \in S_n$ such that $d(u, u') \leq \frac{c_3}{6} n^{1/\ell}$, we have

$$\min_{u' \in S_n} \sum_{w \in V_n} d(u', w) \pi_w(t) > \frac{c_3}{3} n^{1/\ell} \Rightarrow \min_{u \in V_n} \sum_{w \in V_n} d(u, w) \pi_w(t) > \frac{c_3}{6} n^{1/\ell}.$$

Hence, if we set

$$a'_1 := (\max\{4 \log \beta, 4 \log \lambda\})^{-1},$$

then (7.23) and (7.24) imply that for $t \leq F(a'_1 \log n)$,

$$\mathbb{P}_v \left(\min_{u \in V_n} \sum_{w \in V_n} d(u, w) \pi_w(t) \leq \frac{c_3}{6} n^{1/\ell} \right) \leq \frac{20}{\sqrt{n}}.$$

Taking a union bound over $0 \leq t \leq F(a'_1 \log n)$ and recalling the definition of $\mathbb{E}_{\pi(t)}[d(v^*, u)]$, we arrive at

$$\mathbb{P}_v \left(\max_{0 \leq t \leq F(a'_1 \log n)} \mathbb{E}_{\pi(t)}[d(v^*, \widehat{v}_B(t))] \leq \frac{c_3}{6} n^{1/\ell} \right) \leq \frac{20F(a'_1 \log n)}{\sqrt{n}}.$$

Due to the asymptotic behavior of F (see (5.3)), the right hand side tends to 0 uniformly over $v \in V_n$ as $n \rightarrow \infty$, and the result of Lemma 5.2 for lattices follows.

We now turn to the proof of Lemma 7.5.

Proof of Lemma 7.5. Let m be a positive integer. We define an m -covering of V_n to be $S \subseteq V_n$ such that for any $u \in V_n$, there exists $v \in S$ such that $d(u, v) \leq m$. We also define an m -packing of V_n to be $S \subseteq V_n$ such that for any $u, v \in S$, $d(u, v) > m$. We say a S is a *maximal* m -packing of V_n if it has the maximum possible cardinality. A fundamental result on coverings and packings is that a maximal m -packing is also a valid covering [43, Lemma 5.12], so the proof focuses on bounding the size of a maximal packing. Our proof is based on [43, Lemma 5.13].

Set $m := \lfloor \frac{c_3}{8} n^{1/\ell} \rfloor$ and let $S \subset V_n$ be a $2m$ -packing. This in particular implies that $\{\mathcal{N}_u(m)\}_{u \in S}$ is a collection of disjoint sets satisfying

$$\bigcup_{u \in S} \mathcal{N}_u(m) \subseteq \bigcup_{u \in V_n} \mathcal{N}_u(m) \subseteq \mathcal{N}_{v_0}(r_n + 1 + m).$$

This in turn implies

$$\sum_{u \in S} |\mathcal{N}_u(m)| = |S| \cdot |\mathcal{N}(m)| \leq |\mathcal{N}(r_n + 1 + m)|. \quad (7.25)$$

Since $r_n \asymp n^{1/\ell}$ (see (3.1)) and $m \asymp n^{1/\ell}$, we can find constants $C_1, C_2 > 0$ depending only on ℓ such that

$$C_1 n^{1/\ell} \leq m \leq r_n + 1 + m \leq C_2 n^{1/\ell}.$$

Next, recall that $|\mathcal{N}(t)| \sim c_\ell t^\ell$, where c_ℓ is a constant depending only on ℓ (see (A.8)). Hence

$$|S| \leq \frac{|\mathcal{N}(r_n + 1 + m)|}{|\mathcal{N}(m)|} \sim \frac{(r_n + 1 + m)^\ell}{m^\ell} \leq \left(\frac{C_2}{C_1} \right)^\ell.$$

Note that the right hand side is of constant order even as $n \rightarrow \infty$. Moreover, the bound holds for *all* $2m$ -packings, including *maximal* packings that are also coverings. This guarantees the existence of a $2m$ -covering of size bounded by a constant depending on ℓ even as $n \rightarrow \infty$. We conclude by noting that $2m \leq \frac{c_3}{4} n^{1/\ell}$. \square

7.3 Upper bounding the estimation error: Proof of Lemma 5.3

For any vertex $v \in V_n$, we can write

$$\begin{aligned} \mathbb{E}_{\pi(t)}[d(v^*, \widehat{v}_B(t))] &\stackrel{(a)}{\leq} \mathbb{E}_{\pi(t)}[d(v^*, v)] = \sum_{w \in V_n} d(w, v) \pi_w(t) \\ &\stackrel{(b)}{\leq} \sum_{w \in V_n} d(w, v) \frac{X_w(t)}{X_v(t)}, \end{aligned}$$

where (a) follows since $\mathbb{E}_{\pi(t)}[d(v^*, \widehat{v}_B(t))] = \min_{u \in V_n} \mathbb{E}_{\pi(t)}[d(v^*, u)]$ and (b) follows since $\pi_w(t) = X_w(t)/Y(t)$ and $Y(t) \geq X_v(t)$. For distinct vertices $w, v \in V$, recall the notation $f_{vw}(t) := \sum_{s=0}^t |\mathcal{N}_v(s) \setminus \mathcal{N}_w(s)|$ and further recall that $D(Q_0, Q_1)$ is the symmetrized Kullback-Liebler divergence between Q_0 and Q_1 (see (7.10)). As a shorthand, denote $\theta := D(Q_0, Q_1)/2$. Next, define the event

$$\mathcal{E}_{vw} := \left\{ \frac{X_v(t)}{X_w(t)} \geq e^{\theta f_{vw}(t)} \right\}.$$

By Lemma 7.4, $\mathbb{P}_v(\mathcal{E}_{vw}^c) \leq e^{-I(\theta)f_{vw}(t)}$, where $I(\cdot)$ is the large-deviations rate function defined in Lemma 7.4. On the event $\mathcal{E}_v := \bigcup_{w \in V_n \setminus \{v\}} \mathcal{E}_{vw}$, we have the bound

$$\begin{aligned} \mathbb{E}_{\pi(t)}[d(v^*, \hat{v}_B(t))] &\leq \sum_{w \in V_n} d(w, v) \frac{X_w(t)}{X_v(t)} \\ &\leq \sum_{w \in V_n} d(w, v) e^{-\theta f_{vw}(t)}. \end{aligned} \quad (7.26)$$

To bound the final summation in (7.26), we split the summation into two parts: w such that $d(v, w) \leq 2t$ and w such that $d(v, w) > 2t$. To handle the first part, it is useful to define the function

$$f_1(t) := \sum_{s=0}^t |\mathcal{N}_a(s) \setminus \mathcal{N}_b(s)|$$

where a, b are any two neighboring vertices (since the graph is vertex-transitive, we obtain the same formula for any two neighboring a, b). We may now bound the summation over w such that $d(v, w) \leq 2t$ as

$$\sum_{w \in V_n: d(w, v) \leq 2t} d(w, v) e^{-\theta f_{vw}(t)} \leq 2t |\mathcal{N}(2t)| e^{-\theta f_1(t)}. \quad (7.27)$$

To handle the summation over w such that $d(v, w) > 2t$, first note that

$$f_{vw}(t) = \sum_{s=0}^t |\mathcal{N}_v(s) \setminus \mathcal{N}_w(s)| \stackrel{(c)}{=} \sum_{s=0}^t |\mathcal{N}_v(s)| = f(t), \quad (7.28)$$

where the equality (c) follows since $\mathcal{N}_v(s) \cap \mathcal{N}_w(s) = \emptyset$ for $0 \leq s \leq t$ because $d(v, w) > 2t$. Hence we can bound the second part of the summation by

$$\begin{aligned} \sum_{w \in V_n: d(w, v) > 2t} d(w, v) e^{-\theta f(t)} &\leq \sum_{w \in V_n: d(w, v) > 2t} n e^{-\theta f(t)} \\ &\leq n^2 e^{-\theta f(t)}, \end{aligned} \quad (7.29)$$

where we have used the coarse bound $d(w, v) \leq n$ above. Putting everything together, the total bound on the estimation error on the event \mathcal{E}_v is

$$2t |\mathcal{N}(2t)| e^{-\theta f_1(t)} + n^2 e^{-\theta f(t)}.$$

The remaining element of the proof is to bound $\mathbb{P}_v(\mathcal{E}_v^c)$. We can write

$$\begin{aligned} \mathbb{P}_v(\mathcal{E}_v^c) &\stackrel{(d)}{\leq} \sum_{w \in V_n \setminus \{v\}} \mathbb{P}_v(\mathcal{E}_{vw}^c) \\ &\stackrel{(e)}{\leq} \sum_{w \in V_n \setminus \{v\}} e^{-I(\theta)f_{vw}(t)} \\ &\stackrel{(f)}{\leq} \sum_{w \in V_n: d(w, v) \leq 2t} e^{-I(\theta)f_1(t)} + \sum_{w \in V_n: d(w, v) > 2t} e^{-I(\theta)f(t)} \\ &\leq |\mathcal{N}(2t)| e^{-I(\theta)f_1(t)} + n e^{-I(\theta)f(t)}. \end{aligned} \quad (7.30)$$

Above, (d) is due to a union bound, (e) follows from Lemma 7.4, and (f) uses $f_1(t) \leq f_{vw}(t)$ as well as $f_{vw}(t) = f(t)$ if $d(v, w) > 2t$ (see (7.28)). Putting everything together, we have shown that for any $v \in V_n$,

$$\begin{aligned} \mathbb{P}_v \left(\mathbb{E}_{\pi(t)}[d(v^*, \widehat{v}_B(t))] > 2t|\mathcal{N}(2t)|e^{-\theta f_1(t)} + n^2 e^{-\theta f(t)} \right) \\ \leq |\mathcal{N}(2t)|e^{-I(\theta)f_1(t)} + ne^{-I(\theta)f(t)}. \end{aligned}$$

Note in particular that the probability bound above holds uniformly over all $v \in V_n$. Focusing on the special cases of regular trees and lattices, we will simplify the bound on the estimation error as well as the probability bound.

Case 1: G is a k -regular tree. Lemma A.1 provides the asymptotics of various combinatorial quantities related to neighborhood sizes, summarized below:

$$\begin{aligned} |\mathcal{N}(2t)| &\sim \frac{k}{k-2}(k-1)^{2t} \\ f_1(t) &\sim \frac{(k-1)^{t+1}}{k-2} \\ f(t) &\sim \frac{k}{(k-2)^2}(k-1)^{t+1} \\ F(z) &\sim \frac{\log z}{\log(k-1)}. \end{aligned}$$

The terms $2t|\mathcal{N}(2t)|e^{-\theta f_1(t)}$ and $|\mathcal{N}(2t)|e^{-I(\theta)f_1(t)}$ can therefore be bounded by $e^{-O((k-1)^t)}$ for t sufficiently large. If $t \geq F\left(\frac{4\log n}{\theta}\right)$, then $n^2 e^{-\theta f(t)} \leq e^{-\frac{1}{2}\theta f(t)} = e^{-O((k-1)^t)}$, where the hidden factors in the big O do not depend on n . Similarly, if $t \geq F\left(\frac{2\log n}{I(\theta)}\right)$, then $ne^{-I(\theta)f(t)} \leq e^{-O((k-1)^t)}$, where again, the hidden factors do not depend on n .

Putting everything together, we have shown that for $t \geq F\left(\frac{4\log n}{\min\{\theta, I(\theta)\}}\right)$,

$$\mathbb{P}_v \left(\mathbb{E}_{\pi(t)}[d(v^*, \widehat{v}_B(t))] > e^{-O((k-1)^t)} \right) \leq e^{-O((k-1)^t)},$$

which implies the desired result for k -regular trees.

Case 2: G is a ℓ -dimensional lattice. Lemma A.3 proves the following asymptotic behavior of neighborhood sizes in ℓ -dimensional lattices:

$$|\mathcal{N}(2t)| \asymp t^\ell, \quad f(t) \asymp t^{\ell+1}, \quad F(z) \asymp z^{\frac{1}{\ell+1}}.$$

Additionally, we can lower bound $f_1(t)$ as follows: if u, v are adjacent, then

$$f_1(t) = \sum_{s=0}^t |\mathcal{N}_v(s) \setminus \mathcal{N}_u(s)| \geq t+1, \tag{7.31}$$

where the inequality above uses the fact that $\mathcal{N}_v(s) \setminus \mathcal{N}_u(s) \neq \emptyset$ for all $s \geq 0$. We can therefore bound the terms

$$\begin{aligned} 2t|\mathcal{N}(2t)|e^{-\theta f_1(t)} &\leq O\left(t^{\ell+1}e^{-\theta t}\right) \\ |\mathcal{N}(2t)|e^{-I(\theta)f_1(t)} &\leq O\left(t^\ell e^{-I(\theta)t}\right). \end{aligned}$$

Additionally, as in the previous case, if $t \geq F\left(\frac{4\log n}{\min\{\theta, I(\theta)\}}\right)$ then $n^2e^{-\theta f(t)}$ and $ne^{-I(\theta)f(t)}$ are bounded by $e^{-O(t^{\ell+1})}$, where the hidden factors in the big O do not depend on n . Putting everything together, we have shown that for $t \geq F\left(\frac{4\log n}{\min\{\theta, I(\theta)\}}\right)$,

$$\mathbb{P}_{\pi(V_n)}\left(\mathbb{E}_{\pi(t)}[d(v^*, \hat{v}_B(t))] > O\left(t^{\ell+1}e^{-\theta t}\right)\right) \leq O\left(t^{\ell}e^{-I(\theta)t}\right),$$

which implies the desired result for n sufficiently large if we set $b_2 = 0.5 \cdot \min\{\theta, I(\theta)\}$.

8 Proof of the MSPRT upper bounds

8.1 Useful preliminary results

We start by stating and recalling some useful combinatorial results concerning the sizes of neighborhoods in regular trees and lattices. To begin, for vertices u, v recall that

$$f_{vu}(t) := \sum_{s=0}^t |\mathcal{N}_v(s) \setminus \mathcal{N}_u(s)|.$$

Moreover, recall that the *neighborhood growth function* (originally defined in (3.4)) is

$$f(t) := \sum_{s=0}^t |\mathcal{N}(s)|.$$

We also define, for any pair of adjacent vertices u, v , the function

$$f_1(t) := \sum_{s=0}^t |\mathcal{N}_v(s) \setminus \mathcal{N}_u(s)|.$$

We also define the inverse functions $F_{vu} = f_{vu}^{-1}, F = f^{-1}, F_1 = f_1^{-1}$. These inverse functions are well-defined since f_{vu}, f, f_1 are strictly increasing functions. In k -regular trees, we have the asymptotics

$$F_1(z) \sim \frac{\log z}{\log(k-1)}. \quad (8.1)$$

For a proof, see Lemma A.1. In ℓ -dimensional lattices, we have the *orderwise* asymptotics

$$|\mathcal{N}(t)| \asymp t^{\ell}, \quad f(t) \asymp t^{\ell+1}, \quad F(z) \asymp z^{\frac{1}{\ell+1}}. \quad (8.2)$$

For a more precise statement, see Lemma A.3. Next, we prove a few simple, generic results regarding these functions. The following result provides a simple but useful bounds for f_1 that hold in regular trees and lattices.

Lemma 8.1. *Suppose that G is an infinite regular tree or lattice and $u, v \in V$ are two distinct vertices. If $t_2 \geq t_1 \geq 0$, $f_{vu}(t_2) - f_{vu}(t_1) \geq t_2 - t_1$.*

Proof. For any two distinct vertices u, v and any non-negative integer s , $|\mathcal{N}_v(s) \setminus \mathcal{N}_u(s)| \geq 1$. Hence

$$f_{vu}(t_2) - f_{vu}(t_1) = \sum_{s=t_1+1}^{t_2} |\mathcal{N}_v(s) \setminus \mathcal{N}_u(s)| \geq t_2 - t_1.$$

□

Next, we prove a simple linear upper bound for F_1 , since its exact expression is challenging to compute in lattices.

Lemma 8.2. *Let G be an infinite regular tree or lattice. Then $F_1(z) \leq z$.*

Proof. Lemma 8.1 implies the lower bound

$$f_1(t) = f_1(0) + f_1(t) - f_1(0) \geq t + 1 \geq t.$$

Setting $z = f_1(t)$, we have $F_1(z) = t$ which implies the desired result. \square

The following lemma derives conditions under which $F_{vu} = F$.

Lemma 8.3. *If $z < f(d(u, v)/2)$ then $F(z) = F_{vu}(z)$.*

Proof. For $s < d(u, v)/2$, $\mathcal{N}_v(s)$ and $\mathcal{N}_u(s)$ are disjoint so $\mathcal{N}_v(s) \setminus \mathcal{N}_u(s) = \mathcal{N}_v(s)$. Hence for $t < d(u, v)/2$,

$$f_{vu}(t) = \sum_{s=0}^t |\mathcal{N}_v(s) \setminus \mathcal{N}_u(s)| = \sum_{s=0}^t |\mathcal{N}_v(s)| = f(t).$$

It follows that $F_{vu}(z) = F(z)$ if $F(z) < d(u, v)/2$. Equivalently, $z < f(d(u, v)/2)$ which proves the lemma. \square

Finally, at the core of the analysis is the following large-deviations-type result. It essentially follows as a corollary from the large-deviations result Lemma 7.4 which was used in the analysis of the Bayesian setting.

Lemma 8.4. *Let $u, v \in V$ be any two vertices. For any $x > 0$,*

$$\mathbb{P}_v(Z_{vu}(t) \leq (D(Q_0, Q_1) - x)f_{vu}(t)) \leq e^{-I(x)f_{vu}(t)}.$$

Above, $D(Q_0, Q_1)$ is the symmetrized Kulback-Liebler divergence between Q_0 and Q_1 , and $I(\cdot)$ is the rate function defined in (7.11). Moreover, $I(x) > 0$ for $x > 0$.

Proof. Recall from the analysis of the Bayesian setting that for any vertex $v \in V_n$,

$$\pi_v(t) := \mathbb{P}(v^* = v \mid y(0), \dots, y(t)).$$

Hence, by Bayes' rule,

$$\frac{\pi_v(t)}{\pi_u(t)} = \frac{d\mathbb{P}_v}{d\mathbb{P}_u}(y(0), \dots, y(t)) = e^{Z_{vu}(t)}.$$

The desired result now follows from a direct application of Lemma 7.4. \square

8.2 Performance of the MSPRT in trees: Proof of Theorem 6.5

Our first goal is to establish a probabilistic bound for $T_n(v)$ under the measure \mathbb{P}_v . We can write

$$\begin{aligned} \mathbb{P}_v(T_n(v) > t) &\leq \mathbb{P}_v(\exists u \in V_n \setminus \{v\} \text{ s.t. } Z_{vu}(t) < \log n^2/\alpha) \\ &\leq \sum_{u \in V_n \setminus \{v\}} \mathbb{P}_v(Z_{vu}(t) < \log n^2/\alpha). \end{aligned} \tag{8.3}$$

Next, define $\theta := D(Q_0, Q_1)/2$ and suppose that t is sufficiently large so that

$$\log \frac{n^2}{\alpha} \leq \theta f_1(t). \quad (8.4)$$

Letting $I(\cdot)$ be the large-deviations rate function in Lemma 8.4, we can upper bound the final summation in (8.3) by

$$\begin{aligned} \sum_{u \in V_n \setminus \{v\}} \mathbb{P}_v(Z_{vu}(t) < \theta f_1(t)) &\stackrel{(a)}{\leq} \sum_{u \in V_n \setminus \{v\}} \mathbb{P}_v(Z_{vu}(t) < \theta f_{vu}(t)) \\ &\stackrel{(b)}{\leq} \sum_{u \in V_n \setminus \{v\}} \exp(-I(\theta) f_{vu}(t)) \\ &\stackrel{(c)}{\leq} \exp(\log n - I(\theta) f_1(t)). \end{aligned} \quad (8.5)$$

Above, (a) is due to $f_1(t) \leq f_{vu}(t)$, (b) follows from Lemma 8.4 and (c) is again due to $f_1(t) \leq f_{vu}(t)$ and the observation that the summation is over at most n terms. We now define the quantities

$$C(Q_0, Q_1) := \min\{\theta, I(\theta)\} \text{ and } t_n := F_1\left(\frac{\log n^2/\alpha}{C(Q_0, Q_1)}\right).$$

In particular, if $t \geq t_n$ then (8.4) holds and $\log n \leq I(\theta) f_1(t)$. The expectation of $T_n(v)$ can then be bounded as

$$\begin{aligned} \mathbb{E}_v[T_n(v)] &= \sum_{t=0}^{\infty} \mathbb{P}_v(T_n(v) > t) \\ &\stackrel{(d)}{\leq} t_n + \sum_{t=t_n}^{\infty} \exp(\log n - I(\theta) f_1(t)) \\ &\stackrel{(e)}{\leq} t_n + \exp(\log n - I(\theta) f_1(t_n)) \sum_{s=0}^{\infty} e^{-I(\theta)s} \\ &\stackrel{(f)}{\leq} t_n + \frac{1}{1 - e^{-I(\theta)}}. \end{aligned} \quad (8.6)$$

Above, (d) is due to the upper bound (8.5) on the probabilities in the summation which holds for $t \geq t_n$, (e) uses $f_1(t) - f_1(t_n) \geq t - t_n$ which was proved in Lemma 8.1 and (f) follows from noting $I(\tilde{d}) f_1(t_n) \geq \log n$ and using the geometric sum formula on the summation. Noting that $T_n \leq T_n(v)$, (8.6) implies

$$\limsup_{n \rightarrow \infty} \frac{\max_{v \in V_n} \mathbb{E}_v[T_n]}{t_n} \leq \limsup_{n \rightarrow \infty} \frac{\max_{v \in V_n} \mathbb{E}_v[T_n(v)]}{t_n} \leq 1.$$

From the asymptotic behavior of F_1 (see (8.1)), it follows that

$$t_n \sim \frac{\log \log n}{\log(k-1)},$$

which proves the desired result.

8.3 Performance of the MSPRT in lattices: Proof of Theorem 6.7

As in the proof of Theorem 6.5, we begin by bounding $\mathbb{P}_v(T_n(v) > t)$. A union bound yields

$$\begin{aligned} \mathbb{P}_v(T_n(v) > t) &\leq \sum_{\substack{u \in V_n: \\ 0 < d(v,u) \leq K}} \mathbb{P}_v \left(Z_{vu}(t) < \log \frac{2K|\mathcal{N}(K)|}{\alpha} \right) \\ &\quad + \sum_{\substack{u \in V_n: \\ d(v,u) > K}} \mathbb{P}_v \left(Z_{vu}(t) < \log \frac{2n^2}{\alpha} \right). \end{aligned} \quad (8.7)$$

To bound the summations above, we first recall a few quantities. Define $\theta := \tilde{D}(Q_0, Q_1)/2$ and $C(Q_0, Q_1) := \min\{\theta, I(\theta)\}$, where $I(\cdot)$ is the large-deviations rate function used in Lemma 8.4. Also define

$$\begin{aligned} t_{n,1} &:= F_1 \left(\frac{\log 2K|\mathcal{N}(K)|/\alpha}{C(Q_0, Q_1)} \right) \\ t_{n,2} &:= \max_{u \in V_n: d(u,v) > K} F_{vu} \left(\frac{\log 2n^2/\alpha}{C(Q_0, Q_1)} \right). \end{aligned}$$

For $t \geq t_{n,1}$, we will make use of the following inequalities to bound the first summation in (8.7):

$$\log \frac{2K|\mathcal{N}(K)|}{\alpha} \leq \theta f_1(t) \quad \text{and} \quad \log |\mathcal{N}(K)| \leq I(\theta) f_1(t). \quad (8.8)$$

Using the first inequality in (8.8), we can bound the first summation in (8.7) by

$$\begin{aligned} \sum_{\substack{u \in V_n: \\ 0 < d(v,u) \leq K}} \mathbb{P}_v(Z_{vu}(t) \leq \theta f_1(t)) &\stackrel{(a)}{\leq} \sum_{\substack{u \in V_n: \\ 0 < d(v,u) \leq K}} \mathbb{P}_v(Z_{vu}(t) \leq \theta f_{vu}(t)) \\ &\stackrel{(b)}{\leq} \sum_{\substack{u \in V_n: \\ 0 < d(v,u) \leq K}} e^{-I(\theta) f_{vu}(t)} \\ &\stackrel{(c)}{\leq} \exp(\log |\mathcal{N}(K)| - I(\theta) f_1(t)). \end{aligned} \quad (8.9)$$

Above, (a) and (c) are due to $f_1(t) \leq f_{vu}(t)$, and (b) follows from Lemma 8.4. Similarly, for $t \geq t_{n,2}$ we have, for all $u \in V_n$ satisfying $d(u, v) > K$,

$$\log \frac{2n^2}{\alpha} \leq \theta f_{vu}(t) \quad \text{and} \quad \log n \leq I(\theta) f_{vu}(t). \quad (8.10)$$

Using the same reasoning as in (8.9), we have the following bound on the second summation in (8.7) for $t \geq t_{n,2}$:

$$\exp \left(\log n - I(\theta) \min_{u \in V_n: d(u,v) > K} f_{vu}(t) \right). \quad (8.11)$$

Plugging in the bounds (8.9) and (8.11) into (8.7) shows that, for $t \geq \max\{t_{n,1}, t_{n,2}\} =: t_n$,

$$\mathbb{P}_v(T_n(v) > t) \leq \exp(\log |\mathcal{N}(K)| - I(\theta) f_1(t)) + \exp \left(\log n - I(\theta) \min_{u \in V_n: d(v,u) > K} f_{vu}(t) \right). \quad (8.12)$$

Next, define the quantities

$$\begin{aligned} A_n &:= \log |\mathcal{N}(K)| - I(\theta) f_1(t_n) \\ B_n &:= \log n - I(\theta) \min_{u \in V_n: d(u,v) > K} f_{vu}(t_n), \end{aligned}$$

and notice that $A_n, B_n \leq 0$ in light of the second inequalities in (8.8) and (8.10). Using the relation $\mathbb{E}_v[T_n(v)] = \sum_{t=0}^{\infty} \mathbb{P}_v(T_n(v) > t) \leq t_n + \sum_{t=t_n}^{\infty} \mathbb{P}_v(T_n(v) > t)$, we have

$$\begin{aligned} \mathbb{E}_v[T_n(v)] &\stackrel{(d)}{\leq} t_n + \sum_{t=t_n}^{\infty} e^{\log |\mathcal{N}(K)| - I(\theta) f_1(t)} + \sum_{t=t_n}^{\infty} e^{\log n - I(\theta) \min_{u \in V_n: d(u,v) > K} f_{vu}(t)} \\ &\stackrel{(e)}{\leq} t_n + (e^{A_n} + e^{B_n}) \sum_{s=0}^{\infty} e^{-I(\theta)s} \\ &\stackrel{(f)}{\leq} t_n + \frac{1}{1 - e^{-I(\theta)}}. \end{aligned} \tag{8.13}$$

Above, (d) is a consequence of (8.12), (e) is due to the inequality $f_{vu}(t') - f_{vu}(t) \geq t' - t$ which was proved in Lemma 8.1, and (f) holds since $A_n, B_n \leq 0$ and by applying the geometric sum formula. Next, using the inequality $T_n \leq T_n(v)$, (8.13) implies

$$\limsup_{n \rightarrow \infty} \frac{\max_{v \in V_n} \mathbb{E}_v[T_n]}{t_n} \leq \limsup_{n \rightarrow \infty} \frac{\max_{v \in V_n} \mathbb{E}_v[T_n(v)]}{t_n} \leq 1.$$

It remains to study the asymptotics of t_n as n grows large. From the asymptotic behavior of $|\mathcal{N}(t)|$ (8.2), we have $|\mathcal{N}(K)| \asymp K^\ell = \log n$. Hence

$$t_{n,1} \asymp F_1 \left(\frac{\log \log n}{C(Q_0, Q_1)} \right) = O(\log \log n), \tag{8.14}$$

where the final big- O bound is due to the inequality $F_1(z) \leq z$, proved in Lemma 8.2. Next, we establish the asymptotic behavior of $t_{n,2}$. We have, for n sufficiently large,

$$f \left(\frac{K}{2} \right) \asymp (\log n)^{1+\frac{1}{\ell}} \geq \frac{\log 2n^2/\alpha}{C(Q_0, Q_1)}, \tag{8.15}$$

where the asymptotic behavior of $f(K/2)$ follows from (8.2) and the second inequality holds for n sufficiently large. Equation 8.15 satisfies the condition of Lemma 8.3, so we have

$$t_{n,2} = F \left(\frac{\log 2n^2/\alpha}{C(Q_0, Q_1)} \right) \asymp (\log n)^{\frac{1}{\ell+1}}. \tag{8.16}$$

Above, the asymptotic behavior of F follows from (8.2). Hence $t_n \asymp (\log n)^{\frac{1}{\ell+1}}$, which proves the theorem.

9 Conclusion and future directions

In this paper, we considered the problem of quickest estimation of a cascade source from noisy information. We studied a Bayesian and minimax formulation of this problem and derived optimal estimators in the regime of large networks under simple cascade dynamics and network topologies.

Furthermore, our results exposed the interplay between the network topology and the performance of optimal estimators.

There remain several avenues for future work. Although we examined simple networks and cascade dynamics for mathematical tractability, in important next step is to study source estimation for more realistic networks and cascade dynamics [20, 27, 30, 3, 25, 3, 7]. We remark that in many cascade models, the cascade evolution is *non-deterministic*, hence it will be difficult to compute the estimators proposed in this paper. We expect that tractable relaxations of the estimators we consider may be more amenable to the analysis of more complex scenarios.

Another exciting future direction is sampling with *incomplete* information. In this work we assumed that all public signals at a given point in time are observable, but when the network is large this may be infeasible. A natural question of interest is to characterize optimal source estimators given that only a budget of B public signals can be observed at any timestep. There are many possibilities for choosing the B signals to observe: one may target potential super-spreaders (i.e., high-degree vertices) or choose vertices adaptively.

A Bounds on the size of neighborhoods

We begin by defining and recalling some notation. Given a graph G , a vertex v , and a non-negative integer t , we define

$$\begin{aligned}\partial\mathcal{N}_v(t) &:= \{u \in V : d(u, v) = t\} \\ \mathcal{N}_v(t) &:= \{u \in V : d(u, v) \leq t\}.\end{aligned}$$

Since G is vertex-transitive, $|\partial\mathcal{N}_v(t)|$ does not depend on $v \in V$; for brevity of notation, we will therefore write $|\partial\mathcal{N}(t)|$. The same holds for $|\mathcal{N}_v(t)|$, which we will often write as $|\mathcal{N}(t)|$. Additionally recall the *neighborhood growth function*

$$f(t) := \sum_{s=0}^t |\mathcal{N}(s)|$$

as well as

$$f_1(t) := \sum_{s=0}^t |\mathcal{N}_u(s) \setminus \mathcal{N}_v(s)|,$$

where u, v are adjacent vertices. As explained earlier, due to the vertex-transitivity of the underlying graph, the formula for $f_1(t)$ is the same for *any* pair of adjacent vertices. We also define $F_1 = f_1^{-1}$.

The following result provides exact formulas for $|\partial\mathcal{N}(t)|$, $|\mathcal{N}(t)|$ as well as asymptotic behavior for $f(t)$, $f_1(t)$, $F(z)$ and $F_1(z)$ in regular trees.

Lemma A.1. *Let G be a k -regular tree with $k \geq 3$. Then*

$$|\partial\mathcal{N}(t)| = \begin{cases} 1 & t = 0 \\ k(k-1)^{t-1} & t \geq 1; \end{cases} \quad (\text{A.1})$$

$$|\mathcal{N}(t)| = 1 + \frac{k}{k-2} ((k-1)^t - 1); \quad (\text{A.2})$$

$$f_1(t) \sim \frac{(k-1)^{t+1}}{k-2}; \quad (\text{A.3})$$

$$f(t) \sim \frac{k}{(k-2)^2} (k-1)^{t+1}; \quad (\text{A.4})$$

$$F_1(z) \sim \frac{\log n}{\log(k-1)} \quad (\text{A.5})$$

$$F(z) \sim \frac{\log z}{\log(k-1)}. \quad (\text{A.6})$$

Proof. Fix an arbitrary vertex v , and suppose we root G at v so that $|\partial\mathcal{N}(t)|$ is the number of children at height t from the root. Since $\partial\mathcal{N}_v(0) = \{v\}$ and the root node v is the only vertex with k children while all others have $k-1$ children, we have the formula $|\partial\mathcal{N}(t)| = k(k-1)^{t-1}$ for $t \geq 1$.

Next, to compute $|\mathcal{N}(t)|$, we use the formula for $|\partial\mathcal{N}(t)|$ and the geometric sum formula:

$$\begin{aligned} |\mathcal{N}(t)| &= \sum_{s=0}^t |\partial\mathcal{N}(s)| = 1 + k \sum_{s=0}^{t-1} (k-1)^s \\ &= 1 + \frac{k}{k-2} ((k-1)^t - 1). \end{aligned}$$

The same techniques can be used to derive $f(t)$:

$$\begin{aligned} f(t) &= \sum_{s=0}^t |\mathcal{N}(s)| \\ &= \sum_{s=0}^t \left(1 + \frac{k}{k-2} ((k-1)^s - 1) \right) \\ &= -\frac{2(t+1)}{k-2} + \frac{k}{k-2} \sum_{s=0}^t (k-1)^s \\ &= -\frac{2(t+1)}{k-2} + \frac{k}{(k-2)^2} ((k-1)^{t+1} - 1). \end{aligned}$$

To compute $f_1(t)$, we start by computing $|\mathcal{N}_v(s) \setminus \mathcal{N}_u(s)|$. Let u_1, \dots, u_k be the neighbors of u in G and let S_1, \dots, S_k be a partition of the vertices exactly distance s from u , such that the path connecting u and a vertex in S_i must cross u_i . Simple counting arguments show that $|S_i| = (k-1)^{s-1}$ for each i , and that if we assume without loss of generality that $u_1 = v$,

$$|\mathcal{N}_u(s) \setminus \mathcal{N}_v(s)| = |S_2 \cup S_3 \cup \dots \cup S_k| = (k-1)^s.$$

Hence we have

$$f_1(t) = \sum_{s=0}^t (k-1)^s = \frac{(k-1)^{t+1} - 1}{k-2}.$$

The first-order behavior of $F_1 = f_1^{-1}$ is a direct consequence.

We now study $F(z)$, the inverse function of f . Substituting $t = \frac{\log z}{\log(k-1)}$ in the formula for $f(t)$, we have

$$\begin{aligned} f\left(\frac{\log z}{\log(k-1)}\right) &= \frac{k}{(k-2)^2} \left((k-1)^{\frac{\log z}{\log(k-1)}+1} - 1 \right) - \frac{2}{k-2} \left(\frac{\log z}{\log(k-1)} + 1 \right) \\ &= \frac{k(k-1)}{(k-2)^2} z - \frac{k}{(k-2)^2} - \frac{2}{k-2} \left(\frac{\log z}{\log(k-1)} + 1 \right) \end{aligned}$$

Since $\frac{k(k-1)}{(k-2)^2} > 1$, we have for z sufficiently large (in particular, z is larger than some function of k alone) that $f\left(\frac{\log z}{\log(k-1)}\right) \geq z$, which is equivalent to $F(z) \leq \frac{\log z}{\log(k-1)}$. On the other hand,

$$\begin{aligned} f\left(\frac{\log z}{\log(k-1)} - 3\right) &= \frac{k}{(k-1)^2(k-2)^2} z - \frac{k}{(k-2)^2} - \frac{2}{k-2} \left(\frac{\log z}{\log(k-1)} - 3 \right) \\ &\leq \frac{k}{(k-1)^2(k-2)^2} z \\ &\leq \frac{k}{(k-1)^2} z. \end{aligned}$$

Since $k < (k-1)^2$ for $k \geq 3$, we have $f\left(\frac{\log z}{\log(k-1)} - 3\right) \leq z$, which in turn implies that $F(z) \geq \frac{\log z}{\log(k-1)} - 3$. \square

A useful corollary of (A.2) is a characterization of r_n in k -regular trees.

Corollary A.2. *Let G be a k -regular tree and let $\{V_n\}_{n \geq 1}$ be a sequence of candidate sets satisfying Assumption 3.4. Then*

$$r_n = \left\lfloor \frac{\log\left(\frac{k-2}{k}(n-1) + 1\right)}{\log(k-1)} \right\rfloor \sim \frac{\log n}{\log(k-1)}.$$

The following lemma computes the asymptotic behavior for $|\partial\mathcal{N}(t)|$, $|\mathcal{N}(t)|$, $f(t)$ and $F(z)$ in lattices.

Lemma A.3. *Let G be a ℓ -dimensional lattice. Then*

$$|\partial\mathcal{N}(t)| \sim \frac{2^\ell}{(\ell-1)!} t^{\ell-1}; \tag{A.7}$$

$$|\mathcal{N}(t)| \sim \frac{2^\ell}{\ell!} t^\ell; \tag{A.8}$$

$$f(t) \sim \frac{2^\ell}{(\ell+1)!} t^{\ell+1}; \tag{A.9}$$

$$F(z) \sim \left(\frac{(\ell+1)!}{2^\ell} z \right)^{\frac{1}{\ell+1}}. \tag{A.10}$$

Proof. Recall that \mathbb{Z} is the set of integers. For a vector $x \in \mathbb{Z}$, let $\|x\|_0$ denote the number of nonzero entries of x and let $\|x\|_1$ denote the ℓ_1 norm of x . For every integer $1 \leq k \leq \ell$ and an integer $t \geq 0$, define the set

$$\mathcal{S}_k(t) := \left\{ x \in \mathbb{Z}^d : \|x\|_0 = k \text{ and } \|x\|_1 = t \right\}.$$

Since the $\mathcal{S}_k(t)$'s partition $\partial\mathcal{N}_0(t)$, we have $|\partial\mathcal{N}(t)| = \sum_{k=1}^{\ell} |\mathcal{S}_k(t)|$. We proceed by computing the size of $|\mathcal{S}_k(t)|$ via combinatorial arguments. First, we choose the k nonzero coordinates of a vector in $\mathcal{S}_k(t)$; this can be done in $\binom{\ell}{k}$ ways. Next, note that the number of positive integer solutions to $y_1 + \dots + y_k = t$ is exactly $\binom{t-1}{k-1}$ if $t \geq k$ else it is 0; this can be seen through standard counting arguments. Now, since the number of vectors in $\mathcal{S}_k(t)$ for which the absolute value of the entries are given by y_1, \dots, y_k (in that order) is 2^k (since each nonzero entry of x can be positive or negative), we may put everything together to obtain

$$|\mathcal{S}_k(t)| = 2^k \binom{\ell}{k} \binom{t-1}{k-1} \text{ if } t \geq k, \text{ else } 0.$$

When t is large, the first-order term of $|\mathcal{S}_k(t)|$ is $\frac{2^k}{(k-1)!} \binom{\ell}{k} t^{k-1}$. It follows that

$$|\partial\mathcal{N}(t)| \sim |\mathcal{S}_{\ell}(t)| \sim \frac{2^{\ell}}{(\ell-1)!} t^{\ell-1}. \quad (\text{A.11})$$

Next, we use (A.11) to obtain the first-order behavior of $|\mathcal{N}(t)|$. To this end, we first note that for any $p \geq 0$, approximating a summation by an integral gives

$$\begin{aligned} \frac{1}{p+1} \left(k_1^{p+1} - (k_0-1)^{p+1} \right) &= \int_{k_0-1}^{k_1} s^p ds \leq \sum_{k=k_0}^{k_1} k^p \\ &\leq \int_{k_0}^{k_1+1} s^p ds = \frac{1}{p+1} \left((k_1+1)^{p+1} - k_0^{p+1} \right). \end{aligned}$$

In particular, when k_1 is much larger than k_0 ,

$$\sum_{k=k_0}^{k_1} k^p \sim \frac{1}{p+1} k_1^{p+1}.$$

The first-order term of $|\mathcal{N}(t)|$ is therefore

$$|\mathcal{N}(t)| = \sum_{s=0}^t |\partial\mathcal{N}(s)| \sim \sum_{s=\ell}^t \frac{2^{\ell}}{(\ell-1)!} s^{\ell-1} \sim \frac{2^{\ell}}{\ell!} t^{\ell}.$$

Through analogous arguments, $f(t) \sim \frac{2^{\ell}}{(\ell+1)!} t^{\ell+1}$. The first order behavior of F is an immediate consequence. \square

A useful corollary of (A.8) is a characterization of r_n in ℓ -dimensional lattices.

Corollary A.4. *Let G be a ℓ -dimensional lattice $\{V_n\}_{n \geq 1}$ be a sequence of candidate sets satisfying Assumption 3.4. Then*

$$r_n \sim \left(\frac{\ell!}{2^{\ell}} n \right)^{1/\ell}.$$

B Summations of geodesics

The goal of this section is to bound summations of the form $\sum_{w \in V_n} d(w, v)$ and $\sum_{w \in V_n} d(w, v)^2$, which are useful in studying the Bayesian formulation of the quickest source estimation problem.

B.1 Regular trees

We begin by proving a few intermediate results. The following lemma will provide a useful lower bound for $\sum_{w \in V_n} d(w, v)$.

Lemma B.1. *Let G be a regular tree, let $v_0 \in V$ and let r be a positive integer. Then for all $v \in \mathcal{N}_{v_0}(r)$,*

$$\sum_{w \in \mathcal{N}_{v_0}(r)} d(w, v) \geq \sum_{w \in \mathcal{N}_{v_0}(r)} d(w, v_0).$$

To prove the lemma, we will make use of a notion of centrality in trees. Let G_r be the finite k -regular tree restricted to the vertex set $\mathcal{N}_{v_0}(r)$. For a given vertex v of G_r , label the neighbors of v by v^1, \dots, v^k . Let $\mathcal{S}_i(v)$ be the set of vertices in $\mathcal{N}_{v_0}(r)$ such that the path connecting w and v includes v^i . Notice that if we root G_r at v , the sets $\{\mathcal{S}_i(v)\}_{i=1}^k$ correspond to subtrees of the rooted tree and we have the partition

$$\mathcal{N}_{v_0}(r) \setminus \{v\} = \bigcup_{i=1}^k \mathcal{S}_i(v). \quad (\text{B.1})$$

We say that v is a *centroid* of G_r if

$$\max_{1 \leq i \leq k} |\mathcal{S}_i(v)| \leq \frac{|\mathcal{N}_{v_0}(r)|}{2}. \quad (\text{B.2})$$

A consequence of (B.2) is that if v is *not* a centroid and $|\mathcal{S}_1(v)| = \max_{1 \leq i \leq k} |\mathcal{S}_i(v)|$, we must have $|\mathcal{S}_1(v)| \geq |\mathcal{N}_{v_0}(r)|/2 + 1$. Since $\sum_{i=1}^k |\mathcal{S}_i(v)| = |\mathcal{N}_{v_0}(r)| - 1$, it follows that $\sum_{i=2}^k |\mathcal{S}_i(v)| \leq |\mathcal{N}_{v_0}(r)|/2 - 2$ which in turn implies

$$|\mathcal{S}_1(v)| \geq \sum_{i=2}^k |\mathcal{S}_i(v)| + 3. \quad (\text{B.3})$$

In general, a tree may have at most two centroids, in which case the centroids are neighbors [21, Lemma 2.1]. This leads us to the following result.

Proposition B.2. *The unique centroid of G_r is v_0 .*

Proof. We first show that v_0 is indeed a centroid of G_r . Notice that if we root G_r at v_0 , the rooted tree is balanced and in particular, $|\mathcal{S}_i(v_0)| = (|\mathcal{N}_{v_0}(r)| - 1)/k$ for all $i \in \{1, \dots, k\}$. Since $k \geq 2$, (B.2) is satisfied.

Next, suppose by contradiction that v_0 is *not* the unique centroid. Without loss of generality, assume that v_0^1 is also a centroid. However, since G_r rooted at v_0 is balanced, all neighbors of v_0 are isomorphic⁶ so all vertices in the collection $\{v_0^i\}_{i=1}^k$ must also be centroids. But since $k \geq 2$, this implies that there are at least 3 centroids, which is a contradiction. \square

We are now ready to prove Lemma B.1.

⁶More precisely, for each pair of neighbors of v_0 , we can find a graph homomorphism mapping one neighbor to the other.

Proof of Lemma B.1. Without loss of generality, we shall assume that $|\mathcal{S}_1(v)| = \max_{1 \leq i \leq k} |\mathcal{S}_i(v)|$. We can then write

$$\begin{aligned}
\sum_{w \in \mathcal{N}(v_0, r)} d(w, v) &\stackrel{(a)}{=} \sum_{w \in \mathcal{S}_1(v)} d(w, v) + \sum_{i=2}^k \sum_{w \in \mathcal{S}_i(v)} d(w, v) \\
&\stackrel{(b)}{=} \sum_{w \in \mathcal{S}_1(v)} (d(w, v^1) + 1) + \sum_{i=2}^k \sum_{w \in \mathcal{S}_i(v)} (d(w, v^1) - 1) \\
&= \sum_{i=1}^k \sum_{w \in \mathcal{S}_i(v)} d(w, v^1) + |\mathcal{S}_1(v)| - \sum_{i=2}^k |\mathcal{S}_i(v)| \\
&\stackrel{(c)}{=} \sum_{w \in \mathcal{N}(v_0, r)} d(w, v^1) - 1 + |\mathcal{S}_1(v)| - \sum_{i=2}^k |\mathcal{S}_i(v)|, \tag{B.4}
\end{aligned}$$

where (a) and (c) are due to (B.1), and (b) follows since v^1, v are neighbors and v^1 is closer to $\mathcal{S}_1(v)$ than v , and v is closer to $\mathcal{S}_i(v)$ than v^1 for $1 \leq i \leq k$. If v is not a centroid, we can apply (B.3) to (B.4) to obtain

$$\sum_{w \in \mathcal{N}(v_0, r)} d(w, v) \geq \sum_{w \in \mathcal{N}(v_0, r)} d(w, v^1) + 2.$$

In light of Proposition B.2, this shows that if $v \neq v_0$,

$$\sum_{w \in \mathcal{N}(v_0, r)} d(w, v) > \min_{u \in \mathcal{N}_{v_0}(r)} \sum_{w \in \mathcal{N}_{v_0}(r)} d(w, u).$$

The only remaining vertex, v_0 , must therefore be the minimizer. \square

The main result for regular trees follows readily from the intermediate results we have established.

Lemma B.3. *Let G be a k -regular tree and let $v \in V_n$. Then for n sufficiently large,*

$$\sum_{w \in V_n} d(w, v) \geq \frac{n \log n}{k \log(k-1)} \tag{B.5}$$

and

$$\sum_{w \in V_n} d(w, v)^2 \leq \frac{4n \log^2 n}{\log^2(k-1)}. \tag{B.6}$$

Proof. Noting that $\mathcal{N}_{v_0}(r_n) \subseteq V_n$, we have

$$\begin{aligned}
\sum_{w \in V_n} d(w, v_0) &\geq r_n |\partial \mathcal{N}(r_n)| \\
&\stackrel{(a)}{=} r_n k(k-1)^{r_n-1} \\
&\stackrel{(b)}{\geq} \frac{r_n k}{(k-2)^2} \left(\frac{k-2}{k} (n-1) + 1 \right) \\
&\stackrel{(c)}{\geq} \frac{n \log n}{k \log(k-1)}.
\end{aligned}$$

Above, (a) is due to (A.1), (b) uses the formula for r_n in Corollary A.2 and (c) holds for n sufficiently large, as it lower bounds the first-order term in the previous expression. Equation (B.5) now follows from an application of Lemma B.1.

Next, we upper bound the squared sum of the distances. Since $V_n \subset \mathcal{N}_{v_0}(r_n + 1)$, the diameter of V_n is at most $2r_n + 2$. It follows that, for n sufficiently large,

$$\sum_{w \in V_n} d(w, v)^2 \leq n(2r_n + 2)^2 \sim \frac{4n \log^2 n}{\log^2(k-1)},$$

where we have used the asymptotic behavior of r_n derived in Corollary A.2. \square

B.2 Lattices

We first prove an intermediate results. The following lemma is an analogue of Lemma B.1 for the case of lattices.

Lemma B.4. *Let G be a ℓ -dimensional lattice, let $v_0 \in V$ and let r be a positive integer. Then for all $v \in V$,*

$$\sum_{w \in \mathcal{N}_{v_0}(r)} d(w, v) \geq \sum_{w \in \mathcal{N}_{v_0}(r)} d(w, v_0). \quad (\text{B.7})$$

Proof. Assume that vertices are labelled by their coordinates in \mathbb{R}^d . It follows that for $u, v \in V$, $d(u, v) = \|u - v\|_1$. We can then write, for any $v \in V$,

$$\sum_{w \in \mathcal{N}_{v_0}(r)} d(w, v) = \sum_{w \in \mathcal{N}_{v_0}(r)} \|w - v\|_1 = \sum_{i=1}^d \sum_{w \in \mathcal{N}_{v_0}(r)} |v_i - w_i|.$$

The value of v_i that minimizes $\sum_{w \in \mathcal{N}_{v_0}(r)} |v_i - w_i|$ is the median of the collection $\{w_i\}_{w \in \mathcal{N}_{v_0}(r)}$, which is $(v_0)_i$ (the i th component of the vector v_0) due to the symmetry of the set $\mathcal{N}_{v_0}(r)$. As this argument holds for each i , (B.7) follows. \square

The following result contains the desired bounds for $\sum_{w \in V_n} d(w, v)$ and $\sum_{w \in V_n} d(w, v)^2$.

Lemma B.5. *Let G be a ℓ -dimensional lattice. There exist constants $c_3, c_4 > 0$ depending only on d such that for all $v \in V_n$,*

$$\sum_{w \in V_n} d(w, v) \geq c_3 n^{1+\frac{1}{\ell}}; \quad (\text{B.8})$$

$$\sum_{w \in V_n} d(w, v)^2 \leq c_4 n^{1+\frac{2}{\ell}}. \quad (\text{B.9})$$

Proof.

$$\begin{aligned}
\sum_{w \in V_n} d(w, v_0) &\geq \sum_{w \in \mathcal{N}_{v_0}(r_n)} d(w, v_0) \\
&= \sum_{k=1}^{r_n} k |\partial \mathcal{N}(k)| \\
&\stackrel{(a)}{\sim} \frac{2^\ell}{(\ell-1)!} \sum_{k=1}^{r_n} k^\ell \\
&\sim \frac{\ell 2^\ell}{(\ell+1)!} r_n^{\ell+1} \\
&\stackrel{(b)}{\sim} \frac{\ell 2^\ell}{(\ell+1)!} \left(\frac{\ell!}{2^\ell} n \right)^{1+\frac{1}{\ell}}.
\end{aligned}$$

Above, (a) follows from the formula for $|\partial \mathcal{N}(k)|$ in (A.7) and (b) is due to the asymptotics of r_n given in Corollary A.4. Equation (B.8) follows.

Next, we upper bound the sum of the squared distances. Noting that the diameter of V_n is at most $2(r_n + 1)$, we have

$$\sum_{w \in V_n} d(w, v)^2 \leq 4(r_n + 1)^2 n \sim 4 \left(\frac{\ell!}{2^\ell} \right)^{\frac{2}{\ell}} n^{1+\frac{2}{\ell}},$$

where the asymptotics of the final expression are obtained from the asymptotics of r_n found in Corollary A.4. Equation (B.9) follows. \square

References

- [1] N. Alon and J. Spencer. *The Probabilistic Method*. Wiley-Interscience, 2008.
- [2] Ö. Aslan and R. Samet. A comprehensive review on malware detection approaches. *IEEE Access*, 8:6249–6271, 2020.
- [3] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [4] C. Baum and V. Veeravalli. A sequential procedure for multihypothesis testing. *IEEE Transactions on Information Theory*, 40(6):1994–2007, 1994.
- [5] A. Bessi, M. Coletto, G. A. Davidescu, A. Scala, G. Caldarelli, and W. Quattrociocchi. Science vs conspiracy: collective narratives in the age of misinformation. *PLOS ONE*, 10(2):1–17, 02 2015.
- [6] S. Bhamidi, D. Nam, O. Nguyen, and A. Sly. Survival and extinction of epidemics on random graphs with general degree. *The Annals of Probability*, 49(1):244–286, 2021.
- [7] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.
- [8] F. Brauer and C. Castillo-Chavez. *Mathematical Models in Population Biology and Epidemiology*. Springer, 2012.

- [9] S. Das, D. R. Shibib, and M. O. Vernon. The new frontier of diagnostics: molecular assays and their role in infection prevention and control. *American Journal of Infection Control*, 45(2):158–169, 2017.
- [10] V. Dragalin, A. Tartakovsky, and V. Veeravalli. Multihypothesis sequential probability ratio tests. II. accurate asymptotic expansions for the expected sample size. *IEEE Transactions on Information Theory*, 46(4):1366–1383, 2000.
- [11] V. P. Dragalin. Asymptotic solution of a problem of detecting a signal from k channels. *Russian Mathematical Surveys*, 42(3):213–214, 1987.
- [12] V. Draglia, A. Tartakovsky, and V. Veeravalli. Multihypothesis sequential probability ratio tests. I. asymptotic optimality. *IEEE Transactions on Information Theory*, 45(7):2448–2461, 1999.
- [13] G. Fanti, P. Kairouz, S. Oh, and P. Viswanath. Spy vs. spy: rumor source obfuscation. In *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pages 271 – 284, 2015.
- [14] A. Fournery, M. Z. Rácz, G. Ranade, M. Mobius, and E. Horvitz. Geographic and Temporal Trends in Fake News Consumption During the 2016 US Presidential Election. In *Proceedings of the 2017 ACM Conference on Information and Knowledge Management (CIKM)*, pages 2071–2074, 2017.
- [15] T. Halme, E. Nitzan, and V. Koivunen. Bayesian multiple change-point detection of propagating events. In *Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4515–4519, 2021.
- [16] P. Heymann, G. Koutrika, and H. Garcia-Molina. Fighting spam on social web sites: A survey of approaches and future challenges. *IEEE Internet Computing*, 11(6):36–45, 2007.
- [17] J. Hoffman. *Epidemics on graphs under uncertainty*. University of Texas at Austin, 2020.
- [18] J. Hoffmann and C. Caramanis. The cost of uncertainty in curing epidemics. In *Proceedings of the 2018 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pages 1–33, 2018.
- [19] X. Huang and R. Durrett. The contact process on random graphs and Galton Watson trees. *ALEA*, 17:159–182, 2020.
- [20] M. Jalili and M. Perc. Information cascades in complex networks. *Journal of Complex Networks*, 5(5):665–693, 2017.
- [21] V. Jog and P.-L. Loh. Persistence of centrality in random growing trees. *Random Structures & Algorithms*, 52(1):136–157, 2018.
- [22] J. O. Kephart and S. R. White. Directed-graph epidemiological models of computer viruses. In *Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy*, pages 343–359, May 1991.
- [23] J. Khim and P.-L. Loh. Confidence sets for the source of a diffusion in regular trees. *IEEE Transactions on Network Science and Engineering*, 4(1):27–40, 2017.

- [24] G. Lorden. Nearly-optimal sequential tests for finitely many parameter values. *The Annals of Statistics*, 5(1):1 – 21, 1977.
- [25] H. M. Mahmoud. Distances in random plane-oriented recursive trees. *Journal of Computational and Applied Mathematics*, 41(1-2):237–245, 1992.
- [26] G. A. N. Mohamed and N. Ithnin. Survey on representation techniques for malware detection system. *American Journal of Applied Sciences*, 14:1049–1069, 2017.
- [27] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–180, 1995.
- [28] M. E. J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66:016128, 2002.
- [29] H.-C. Ou, A. Sinha, S.-C. Suen, A. Perrault, A. Raval, and M. Tambe. Who and when to screen: multi-round active screening for network recurrent infectious diseases under uncertainty. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems*, pages 992 – 1000, 2020.
- [30] M. Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.
- [31] H. V. Poor. *An Introduction to Signal Detection and Estimation (2nd Ed.)*. Springer-Verlag, Berlin, Heidelberg, 1994.
- [32] H. V. Poor and O. Hadjiladis. *Quickest Detection*. Cambridge University Press, 2008.
- [33] G. Rovatsos, V. V. Veeravalli, and G. V. Moustakides. Quickest detection of a dynamic anomaly in a heterogeneous sensor network. In *Proceedings of the 2020 IEEE International Symposium on Information Theory (ISIT)*, pages 1171–1176, 2020.
- [34] G. Rovatsos, V. V. Veeravalli, D. Towsley, and A. Swami. Quickest detection of growing dynamic anomalies in networks. In *Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8926–8930, 2020.
- [35] M. Z. Rcz and J. Richey. Rumor source detection with multiple observations under adaptive diffusions. *IEEE Transactions on Network Science and Engineering*, 8(1):2–12, 2021.
- [36] D. Shah and T. Zaman. Detecting sources of computer viruses in networks: Theory and experiment. In *Proceedings of the 2010 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pages 203–214, 2010.
- [37] D. Shah and T. Zaman. Rumors in a network: who’s the culprit? *IEEE Transactions on Information Theory*, 57(8):5163–5181, 2011.
- [38] D. Shah and T. Zaman. Finding rumor sources on random trees. *Operations Research*, 64(3):736–755, 2016.
- [39] A. Sridhar and H. V. Poor. Sequential estimation of network cascades. In *Proceedings of the 2020 54th Asilomar Conference on Signals, Systems, and Computers*, pages 1507–1511, 2020.
- [40] A. Sridhar and H. V. Poor. Bayes-optimal methods for finding the source of a cascade. In *Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5190–5194, 2021.

- [41] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro. Some like it hoax: automated fake news detection in social networks. In *Proceedings of the 2nd Workshop on Data Science for Social Good*, pages 1–15, 2017.
- [42] A. G. Tartakovsky. Asymptotic optimality of certain multihypothesis sequential tests: non i.i.d. case. *Statistical Inference for Stochastic Processes*, 1(3):265–295, 1998.
- [43] R. van Handel. Probability in high dimension, 2016. Available online at <https://web.math.princeton.edu/~rvan/APC550.pdf>.
- [44] A. Wald and J. Wolfowitz. Optimum character of the sequential probability ratio test. *Ann. Math. Statist.*, 19(3):326–339, Sept 1948.
- [45] C. Wang, J. Knight, and M. Elder. On computer viral infection and the effect of immunization. In *Proceedings of the 16th Annual Computer Security Applications Conference (ACSAC’00)*, pages 246–256, 2000.
- [46] Z. Wang, W. Dong, W. Zhang, and C. W. Tan. Rumor source detection with multiple observations: fundamental limits and algorithms. In *Proceedings of the 2014 ACM SIGMETRICS International Conference on the Measurement and Modeling of Computer Systems*, pages 1–13, 2014.
- [47] L. Ying and K. Zhu. Diffusion source localization in large networks. *Synthesis Lectures on Communication Networks*, 11(1):1–95, 2018.
- [48] R. Zhang, R. Yao, Y. Xie, and F. Qiu. Online detection of cascading change-points. 2021. Available online at <https://arxiv.org/abs/1911.05610>.
- [49] K. Zhu, Z. Chen, and L. Ying. Catch’Em all: Locating multiple diffusion sources in networks with partial observations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1), 2017.
- [50] K. Zhu and L. Ying. Information source detection in networks: Possibility and impossibility results. In *Proceedings of the 35th Annual IEEE International Conference on Computer Communications*, pages 1–9, 2016.
- [51] C. C. Zou, W. Gong, and D. Towsley. Code red worm propagation modeling and analysis. In *Proceedings of the 9th ACM Conference on Computer and Communications Security*, pages 138–147, 2002.
- [52] S. Zou and V. V. Veeravalli. Quickest detection of dynamic events in sensor networks. In *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6907–6911, 2018.
- [53] S. Zou, V. V. Veeravalli, J. Li, and D. Towsley. Quickest detection of significant events in structured networks. In *Proceedings of the 2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pages 1307–1311, 2018.
- [54] S. Zou, V. V. Veeravalli, J. Li, and D. Towsley. Quickest detection of dynamic events in networks. *IEEE Transactions on Information Theory*, 66(4):2280–2295, 2020.