Supplementary material: Metadata-conscious anonymous messaging

A. Warm-up Example: Line Graph

We begin by considering the special contact network of a line graph. This example highlights how severely metadata can hurt anonymity; nonetheless, Section 3 illustrates that our seemingly-negative result on lines does not extend to higher-degree trees.

Consider a line graph $G(V, E)$ in which $V = \{0, 1, \ldots, n, n+1\}$, nodes $s_1 = 0$ and $s_2 = n+1$ are spies, and $E = \{(i, i+1) \mid i \in \{0, \ldots, n\}\}$. One of the n nodes between the spies is chosen uniformly at random as a source, denoted by $v^* \in \{1, \ldots, n\}$. When the message reaches a spy s_i, the spy collects at least two pieces of metadata: the timestamp T_{s_i} and the parent node p_{s_i} that relayed the message. We let t_0 denote the time the source starts propagating the message according to some global reference clock. Let $T_{s_1} = T_1 + t_0$ and $T_{s_2} = T_2 + t_0$ denote the timestamps when the two spy nodes receive the message, respectively. Knowing the spreading protocol and the metadata, the adversary uses the maximum likelihood estimator to optimally estimate the source.

In this section, we first show that under standard diffusion, the probability of source detection scales as $1/\sqrt{n}$. We also show that if spy nodes observed only timestamps and parent nodes, adaptive diffusion would achieve the optimal detection probability of $1/n$. However, adaptive diffusion passes extra metadata, which we call a *control packet*, to coordinate the message spread (details below). Control packets allow a spy to identify the source with probability 1. To overcome this challenge, we propose a new implementation of adaptive diffusion that provably achieves $1/\sqrt{n}$ (Proposition A.1). It is an open question if a smaller probability of detection can be achieved on a line.

Standard diffusion. Consider a standard discrete-time random diffusion with a parameter $q \in (0, 1)$ where each uninfected neighbor is infected with probability q. The adversary observes T_{s_1} and T_{s_2}. Knowing the value of q, it computes the ML estimate $\hat{v}_{\mathrm{ML}} = \arg\max_{v \in [n]} \mathbb{P}_{T_1 - T_2 | V^*}(T_{s_1} - T_{s_2} | v)$, which is optimal assuming uniform prior on v^*. Since t_0 is not known, the adversary can only use the difference $T_{s_1} - T_{s_2} = T_1 - T_2$ to estimate the source. We can exactly compute the corresponding probability of detection; Figure 5 (bottom) illustrates that the posterior (and the likelihood) is concentrated around the ML estimate, and the source can only hide among $O(\sqrt{n})$ nodes. The detection probability correspondingly scales as $1/\sqrt{n}$ (top).

Adaptive diffusion on a line. Adaptive diffusion introduced in (Fanti et al., 2015) on a line is a random message spreading model governed by the location of a *virtual source* v_t at any (even) time t. At time 0, the source determines either the left or the right neighbor to be the next virtual source with equal probability. The message is propagated to the chosen node at time $t = 1$. At $t = 2$, the new virtual source v_2 propagates the message to its uninfected neighbor. At this point, three nodes are infected, with the virtual source v_2 at the center. At any given even time t, the infected subgraph is a subset of $t + 1$ nodes, centered around the virtual source v_t. At each even time t, the protocol has two options: keep the virtual source where it is, or pass it to the only neighbor who has not yet been a virtual source. The protocol keeps the current virtual source with probability $\frac{2\delta_H(v_t, v^*)}{t+2}$, where $\delta_H(v_t, v^*)$ denotes the hop distance between the source and the virtual source, and passes it otherwise. The control packet therefore contains two pieces of information: $\delta_H(v_t, v^*)$ and t. In the next two time steps, the message spreads in such a way that two more nodes are infected, and the virtual source is again at the center of the infected subgraph. This choice of virtual-

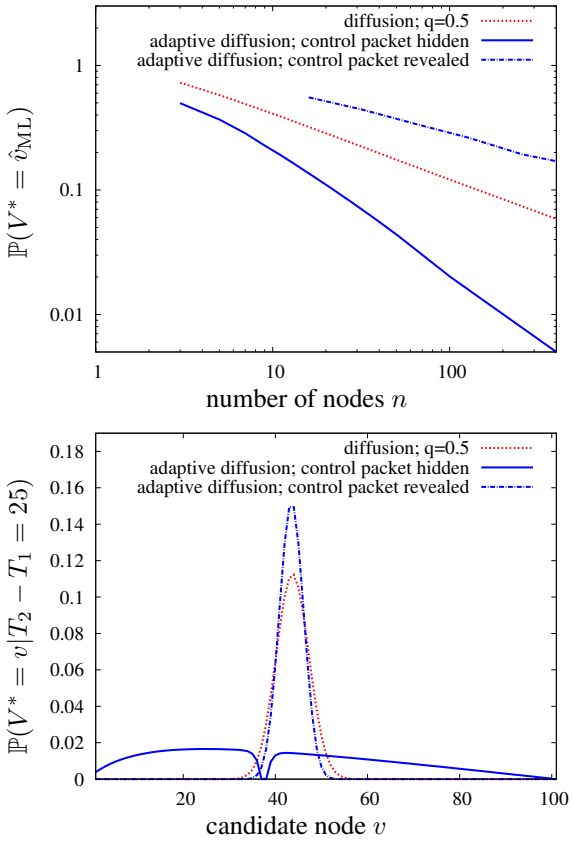

Figure 5. Comparisons of probability of detection as a function of n (top) and the posterior distribution of the source for an example with $n = 101$ and $T_2 - T_1 = 25$ (bottom). The line with 'control packet revealed' uses the Pólya's urn implementation.

source-spreading probability is optimal against a snapshot adversary, guaranteeing perfect obfuscation of the source.

Suppose spy nodes only observed timestamps and parent nodes but *not* control packets. The adversary could then numerically compute the ML estimate $\hat{v}_{\mathrm{ML}} = \arg\max_{v\in[n]} \mathbb{P}_{T_1-T_2|V^*}(T_{s_1} - T_{s_2}|v)$. Figure 5 shows the posterior is close to uniform (bottom) and the probability detection would scale as $1/n$ (top), which is the best one can hope for. Of course, spies *do* observe control packets, including the information to generate the randomness. This reveals the distance to the true source $\delta_H(v_T, v^*)$, and the true source is exactly identified with probability 1. We therefore introduce a new implementation (tailored for the line graph) that is robust to control packet information.

Adaptive diffusion via Pólya's urn. The random process governing the virtual source's propagation is identical to a Pólya's urn process (Johnson & Kotz, 1977). We propose the following alternative implementation of adaptive diffusion. At $t = 0$ the protocol decides whether to pass the virtual source left ($D = \ell$) or right ($D = r$) with probability half. Let D denote this random choice. Then, a latent variable q is drawn from the uniform distribution over $[0, 1]$. Thereafter, at each even time t, the virtual source is passed with probability q or kept with probability $1-q$. The Bayesian interpretation of Pólya's urn processes shows that this process is equivalent to the adaptive diffusion process.

Further, in practice, the source could simulate the whole process in advance. The control packet would simply reveal to each node how long it should wait before further propagating the message. Under this implementation, spy nodes only observe timestamps T_{s_1} and T_{s_2}, parent nodes, and control packets containing the infection delay for the spy and all its descendants in the infection. Given this, the adversary can exactly determine the timing of infection with respect to the start of the infection T_1 and T_2, and also the latent variables D and q. A proof of this statement and the following proposition is provided in Section A.1 of the Supplementary Material. Precisely, we provide an upper bound on the detection probability for such an adversary.

Proposition A.1. *When the source is uniformly chosen from n nodes between two spy nodes, the ML estimator achieves a detection probability upper bounded by*

$$\mathbb{P}\big(V^* = \hat{v}_{\mathrm{ML}}\big) \leq \frac{\pi\sqrt{8}}{\sqrt{n}} + \frac{2}{n}\,.$$

Equipped with the ML estimator, we can also simulate adaptive diffusion on a line. Figure 5 (top) illustrates that even with access to control packets, the adversary achieves probability of detection scaling as $1/\sqrt{n}$ – similar to standard diffusion. For a given value of T_1, the posterior and the likelihood are concentrated around the ML estimate, and the source can only hide among $O(\sqrt{n})$ nodes, as shown in the bottom panel for $T_1 = 58$. In the realistic adversarial setting where control packets are revealed at spy nodes, adaptive diffusion can only hide as well as standard diffusion over a line.

A.1. Proof of Proposition A.1

The control packet at spy node s_1 includes the amount of delay at $s_1 = 0$ and all descendants of s_1, which is the set of nodes $\{-1, -2, \ldots\}$. The control packet at spy node s_2 includes the amount of delay at $s_2 = n + 1$ and all descendants of s_2, which is the set of nodes $\{n + 2, n + 3, \ldots\}$. Given this, it is easy to figure out the whole trajectory of the virtual source for time $t \geq T_1$. Since the virtual source follows i.i.d. Bernoulli trials with probability q, one can exactly figure out q from the infinite Bernoulli trials. Also the direction D is trivially revealed.

To lighten the notation, suppose that $T_1 \leq T_2$ (or equivalently $T_{s_1} \leq T_{s_2}$). Now using the difference of the observed time stamps $T_{s_2} - T_{s_1}$ and the trajectory of the virtual source between T_{s_1} and T_{s_2}, the adversary can also learn the time stamp T_1 with respect to the start of the infection. Further, once the adversary learns T_1 and the location of the virtual source v_{T_1}, the timestamp T_2 does not provide any more information. Hence, the adversary performs ML estimate using T_1, D and q. Let $B(k, n, q) = \binom{n}{k} q^k (1 - q)^{n-k}$ denote the pmf of the binomial distribution. Then, the likelihood can be computed for T_1 as

$$\mathbb{P}^{(\mathrm{adaptive})}_{T_1|V^*,Q,D}\big(\,t_1\big|v^*, q, \ell\big) = \begin{cases} qB(v^* - \frac{t_1}{2} - 2, \frac{t_1}{2} - 2, q)\mathbb{I}_{(v^*\in[2+\frac{t_1}{2}, t_1])} & \text{if } t_1 \text{ even} \\ B(v^* - \frac{t_1+3}{2}, \frac{t_1-3}{2}, q)\,\mathbb{I}_{(v^*\in[\frac{t_1+3}{2}, t_1])} & \text{if } t_1 \text{ odd} \end{cases} \tag{2}$$

$$\mathbb{P}^{(\mathrm{adaptive})}_{T_1|V^*,Q,D}\big(\,t_1\big|\,v^*, q, r\big) = \begin{cases} 0 & \text{if } t_1 \text{ even} \\ (1-q)B(\frac{t_1-1}{2} - v^*, \frac{t_1-3}{2}, q)\mathbb{I}_{(v^*\in[1, \frac{t_1-1}{2}])} & \text{if } t_1 \text{ odd.} \end{cases} \tag{3}$$

This follows from the construction of the adaptive diffusion. The protocol follows a binomial distribution with parameter q until $(T_1 - 1)$. At time T_1, one of the following can happen: the virtual source can only be passed (the first equation in (2)), it can only stay (the second equation in (3)), or both cases are possible (the second equation in (2)).

Given T_1, Q and D, which are revealed under the adversarial model we consider, the above formula implies that the posterior distribution of the source also follows a binomial distribution. Hence, the ML estimate is the mode of a binomial distribution with a shift, for example when t_1 is even, ML estimate is the mode of $2 + (t_1/2) + Z$ where $Z \sim \mathrm{Binom}((t_1/2) - 2, q)$. The adversary can compute the

ML estimate:

$$\hat{v}_{\mathrm{ML}} = \begin{cases} \frac{T_1+2}{2} + \left\lfloor q\left(\frac{T_1-2}{2}\right)\right\rfloor & \text{if } T_1 \text{ even \& } D=\ell\,, \\ \frac{T_1+3}{2} + \left\lfloor q\left(\frac{T_1-1}{2}\right)\right\rfloor & \text{if } T_1 \text{ odd \& } D=\ell\,, \\ 1 + \left\lfloor (1-q)\left(\frac{T_1-1}{2}\right)\right\rfloor & \text{if } T_1 \text{ odd \& } D=r\,. \end{cases} \tag{4}$$

Together with the likelihoods in Eqs. (2) and (3), this gives

$$\mathbb{P}^{(\mathrm{adaptive})}_{T_1,D|V^*,Q}\big(t_1,r,\hat{v}_{\mathrm{ML}}=v^*\big|v^*,q\big) = \frac{1}{2}(1-q)\,B\Big(\frac{t_1-1}{2}-v^*,\frac{t_1-3}{2},q\Big)\mathbb{I}_{(\hat{v}_{\mathrm{ML}}=v^*)}\,\mathbb{I}_{(t_1 \text{ is odd})} \tag{5}$$

$$\mathbb{P}^{(\mathrm{adaptive})}_{T_1,D|Q}\big(t_1,r,V^*=\hat{v}_{\mathrm{ML}}\big|q\big) = \frac{1}{2\,n}(1-q)\,B\Big(\frac{t_1-1}{2}-\hat{v}_{\mathrm{ML}},\frac{t_1-3}{2},q\Big)\,\mathbb{I}_{(t_1 \text{ is odd})} \tag{6}$$

$$\leq \frac{(1-q)}{2\,n}\Big(\frac{\sqrt{2}\,\mathbb{I}_{(t_1 \text{ is odd and } t_1>3)}}{\sqrt{\frac{t_1-3}{2}q(1-q)}} + \mathbb{I}_{(t_1=3)}\Big) \tag{7}$$

where $\hat{v}_{\mathrm{ML}} = \hat{v}_{\mathrm{ML}}(t_1,q,r)$ is provided in (4), and the bound on $B(\cdot)$ follows from Gaussian approximation (which gives an upper bound $1/\sqrt{2\pi kq(1-q)}$) and Berry-Esseen theorem (which gives an approximation guarantee of $2\times 0.4748/\sqrt{kq(1-q)}$), for $k=(t_1-3)/2$. Marginalizing out $T_1 \in \{3,5,\ldots,2\lfloor(n-1)/2\rfloor+1\}$ and applying an upper bound $\sum_{i=1}^{k} 1/\sqrt{i} \leq 2\sqrt{k+1}-2 \leq 2\sqrt{k-1}+\sqrt{1/(2(k-1))}-2 \leq \sqrt{4(k-1)}$, we get

$$\mathbb{P}\big(D=r,V^*=\hat{v}_{\mathrm{ML}},T_1 \text{ is odd}\big|Q=q\big) \leq \frac{(1-q)\sqrt{2}}{2\,n\,\sqrt{q(1-q)}}\sqrt{8\left\lfloor\frac{n-1}{2}\right\rfloor}+\frac{1-q}{2\,n}\,. \tag{8}$$

Similarly, we can show that

$$\mathbb{P}\big(D=\ell,V^*=\hat{v}_{\mathrm{ML}},T_1 \text{ is odd}\big|Q=q\big) \leq \frac{\sqrt{2}}{2\,n\,\sqrt{q(1-q)}}\sqrt{8\left\lfloor\frac{n-1}{2}\right\rfloor}+\frac{1}{n}\,, \tag{9}$$

$$\mathbb{P}\big(V^*=\hat{v}_{\mathrm{ML}},T_1 \text{ is even}\big|Q=q\big) \leq \frac{q\sqrt{2}}{2\,n\,\sqrt{q(1-q)}}\sqrt{8\left\lfloor\frac{n}{2}\right\rfloor}+\frac{1+q}{2\,n}\,, \tag{10}$$

Summing up,

$$\mathbb{P}(V^*=\hat{v}_{\mathrm{ML}}|Q=q) \quad\leq\quad \sqrt{\frac{8}{n\,q\,(1-q)}}+\frac{2}{n}\,. \tag{11}$$

Recall Q is uniformly drawn from $[0,1]$. Taking expectation over Q gives

$$\mathbb{P}(V^*=\hat{v}_{\mathrm{ML}}) \quad\leq\quad \pi\sqrt{\frac{8}{n}}+\frac{2}{n}\,, \tag{12}$$

where we used $\int_0^1 1/\sqrt{x(1-x)}dx = \arcsin(1)-\arcsin(-1)=\pi$.

B. Regular Tree Analysis

B.1. Proof of Theorem 1

We begin by expanding some points regarding the ML estimator in Algorithm 2 that were omitted in Section 3. First, note that it is possible to derive an ML estimate without requiring the presence of a spine spy; the estimator described here uses a spine spy purely for ease of exposition. The omitted details are: (1) given a spy node s, how does the estimator find that spy node's pivot ℓ_s? (2) Why does timing information enable the estimator to disregard any subtree neighboring ℓ_{min} that contains at least one spy?

To answer the first question, consider the first spine spy s_0 and all spies in the feasible subtree. For each spy s in the feasible subtree (none of which lies on the spine), there exists a unique path between s and s_0. There exists a unique node on this path that is both part on the spine and closer to the true source than any other node in the path—this is precisely the pivot node. The estimator uses the observed metadata to infer the pivot, as well as its level in the infected subtree, for each spy in the feasible subtree. This inference proceeds by solving a system of equations:

$$h_{s,\ell_s}+h_{\ell_s,s_0} = |\mathcal{P}(s,s_o)|$$
$$h_{\ell_s,s_0}-h_{s,\ell_s} = T_{s_0}-T_s$$

where $\mathcal{P}(s,s_o)$ denotes the path between s and s_o, $h_{s,\ell_s}=\delta_H(s,\ell_s)$ denotes the distance from spy s_i to the pivot node ℓ_s, and h_{ℓ_s,s_0} is equal to $\delta_H(\ell_s,s_o)$ by construction. This system of equations always has a unique solution; hence the uniqueness of ℓ_s given s and s_0. The first equation holds by construction. The second equation holds because conditioned on the time at which the pivot receives the message T_{ℓ_s}, s_0 receives the message at time $T_{\ell_s}+h_{\ell_s,s_0}$, and s receives it at $T_{\ell_s}+h_{s,\ell_s}$.

Let L denote the set of pivots corresponding to each spy in the feasible subtree; in the example in Figure 2, $L=\{1,2\}$. Define $\ell_{min} = \mathrm{argmin}_{\ell\in L} m_\ell$. That is, ℓ_{min} denotes the pivot closest to the true source in hop distance, i.e., whose level is lowest. Now consider the subtrees of depth $m_{\ell_{min}}-1$ rooted at the neighbors of ℓ_{min}. The subtree including s_0 cannot contain the true source because we know the message traveled from ℓ_{min} to s_0. The source must therefore lie in one of the remaining $d-1$ neighbor subtrees, which we refer to as *candidate subtrees*.

We now argue that the estimator can rule out any candidate subtree of ℓ_{min} that contains at least one spy node. Suppose otherwise: there is a candidate subtree containing a spy s, and the source v^* is contained in that subtree. Then the path $\mathcal{P}(v^*,s)$ cannot pass through ℓ_{min} because ℓ_{min} does not belong to any of its own neighboring subtrees by construction. Then there must exist some node ℓ' on the spine such that $|\mathcal{P}(\ell',s)| < |\mathcal{P}(\ell_{min},s)|$. But this is a contradiction

because ℓ_{min} is chosen as the minimum-level pivot across all spies, and each spy has a unique pivot on the spine.

Since we can now rule out candidate subtrees with at least one spy, let $X+1$, $X \in \mathbb{N}$ be the number of candidate subtrees containing no spies. We use this notation because there will always be at least one candidate subtree with no spies (the one containing the true source). In Figure 2, $X = 0$. Thus, the ML estimator chooses one of the leaves in the remaining $X+1$ candidate subtrees uniformly at random. All remaining nodes in $V \backslash U$ have likelihood 0.

Probability of Detection: We condition on the lowest-level pivot node, ℓ_{min}, giving $\mathbb{P}(\hat{v}_{\text{ML}} = v^*) = \sum_{\ell_{min}} \mathbb{P}(\hat{v}_{\text{ML}} = v^* | \ell_{min}) \mathbb{P}(\ell_{min})$. Since ℓ_{min} lies on the spine, this is equivalent to conditioning on the distance of ℓ_{min} from the true source.

$$\mathbb{P}(\hat{v}_{\text{ML}} = v^*) = \sum_{k=1}^{\infty} \underbrace{\frac{(1-p)^{(|T_{d,k}|-1)}\, p}{|\partial T_{d,k}|}}_{\ell_{min}\ (k^{th}\ \text{spine node) is a spy}} + \underbrace{(1-p)^{|T_{d,k}|} \mathbb{E}_X\Big[\frac{\mathbb{I}(X \neq d-2)}{(X+1)\,|\partial T_{d,k}|}\Big]}_{\ell_{min}\ (k^{th}\ \text{spine node) not a spy}} \tag{13}$$

where $X \sim \text{Binom}(d-2, (1-p)^{|T_{d,k}|})$, $|T_{d,k}| = \frac{(d-1)^k - 1}{d-2}$ is the number of nodes in each candidate subtree for a pivot at level k, and $|\partial T_{d,k}| = (d-1)^{k-1}$ is the number of leaf nodes in each candidate subtree. Let $w = (1-p)$. We have that

$$\begin{aligned}\mathbb{E}_X\Big[\frac{\mathbb{I}(X \neq d-2)}{(X+1)\,|\partial T_{d,k}|}\Big] &= \frac{1}{|\partial T_{d,k}|}\Big(\mathbb{E}_X\Big[\frac{1}{X+1}\Big] - \frac{1}{d-1} w^{|T_{d,k}|\cdot(d-2)}\Big) \\ &= \frac{1}{|\partial T_{d,k}|}\Big(\frac{1}{(d-1)w^{|T_{d,k}|}}(1-(1-w^{|T_{d,k}|})^{d-1}) - \frac{1}{d-1} w^{|T_{d,k}|\cdot(d-2)}\Big)\end{aligned}$$

where the last line results from the expression for the expectation of $1/(1+X)$ when X is binomially-distributed. Namely if $X \sim \text{Binom}(\tilde{n}, \tilde{p})$, then $\mathbb{E}[1/(X+1)] = \frac{1}{(\tilde{n}+1)\tilde{p}}(1-(1-\tilde{p})^{\tilde{n}+1})$. Simplifying gives

$$\begin{aligned} P_D &= \sum_{k=1}^{\infty} \frac{1}{(d-1)^k}\Big[(d-1)pw^{|T_{d,k}|-1} + 1 - w^{|T_{d,k}|\cdot(d-1)} \\ &\quad -(1-w^{|T_{d,k}|})^{d-1}\Big] \\ &= p + \frac{1}{d-2} + \sum_{k=1}^{\infty} \frac{1}{(d-1)^k}\Big[pw^{|T_{d,k+1}|-1} - \\ &\quad w^{|T_{d,k}|\cdot(d-1)} - (1-w^{|T_{d,k}|})^{d-1}\Big] \\ &= p + \frac{1}{d-2} - \sum_{k=1}^{\infty} \frac{1}{(d-1)^k}\Big[w^{|T_{d,k+1}|} + \\ &\quad (1-w^{|T_{d,k}|})^{d-1}\Big]. \end{aligned}$$

where the last line holds because $|T_{d,k+1}| - 1 = |T_{d,k}| \cdot (d-1)$.

Expected hop distance: In the main paper, we lower bounded the expected hop distance by assuming that the estimator guesses the source exactly whenever (a) the pivot ℓ_{min} is a spy node or (b) the estimator chooses the correct candidate subtree. Therefore, if the pivot ℓ_{min} is at level k, we only consider estimates that are exactly $2k$ hops away. The estimator chooses an incorrect candidate subtree with probability $X/(X+1)$.

$$\mathbb{E}[\delta_H(\hat{v}_{\text{ML}}, v^*)] \geq \sum_{k=1}^{\infty} 2k(1-p)^{|T_{d,k}|} \mathbb{E}_{X_k}\Big[\frac{X_k \cdot \mathbb{I}(X_k \neq d-2)}{(X_k+1)}\Big]. \tag{14}$$

If $X_k \sim \text{Binom}(\tilde{n}, \tilde{p})$, where $\tilde{n}$ and $\tilde{p}$ depend on d and k, we have

$$\mathbb{E}_{X_k}\Big[\frac{X_k \cdot \mathbb{I}(X_k \neq \tilde{n})}{(X_k+1)}\Big] = \frac{1}{(\tilde{n}+1)\tilde{p}}\left[(1-\tilde{p})^{\tilde{n}} + \tilde{p}(1-(1-\tilde{p})^{\tilde{n}} + \tilde{n}) - 1 - \tilde{n}\tilde{p}^{\tilde{n}+1}\right]$$

Simplifying and substituting $\tilde{p} = (1-p)^{|T_{d,k}|}$ and $\tilde{n} = d-2$ gives the expression in the theorem.

Note that this bound is trivially 0 for $d = 3$, since we ignore all nodes in the correct candidate subtree; when $d = 3$, the source's candidate subtree is the only valid option if ℓ_{min} is not a spy. However, for a fixed p with $d \to \infty$, this lower bound approaches the upper bound of $2(1-p)$.

Obtaining a tighter bound is straightforward, but increases the complexity of the expression. These tighter bounds were used for the plots in the main paper. A tighter bound results from considering the cases when (a) the pivot ℓ_{min} is a spy node or (b) the estimator chooses the correct candidate subtree. In both cases, we ignore all but the most

distant estimates. For instance, if ℓ_{min} is on the spine at level k, then the estimate will be at most $2(k-1)$ hops away. Using this rule for both cases (a), we compute the probability of selecting one of the most distant options:

$$a_k \equiv \frac{d-2}{d-1}(1-p)^{|T_{d,k}|(d-1)}$$

and for case (b):

$$b_k \equiv p\frac{d-2}{d-1}(1-p)^{|T_{d,k}|-1}$$

Overall, we get a lower bound of

$$\mathbb{E}[\delta_H(\hat{v}_{\mathrm{ML}}, v^*)] \geq \sum_{k=1}^{\infty} 2(kr_k + (k-1)(a_k+b_k))$$

C. Irregular Tree Analysis

C.1. Proof of Proposition 4.1

All nodes in $V \backslash U$ have likelihood zero, as discussed in the proof of Theorem 1 (recall that V denotes the set of all nodes, and U denotes the set of candidate nodes). The only randomness in adaptive diffusion spreading occurs when a spine node with uninfected neighbors decides which of its neighbors will be added to the spine next. Thus, the (log) likelihood of a candidate source is the sum of the (log) likelihoods of all candidate spine nodes starting at the candidate source. Regardless of which node $u \in U$ is the true source, the spine must pass through ℓ_{min}; since there is a unique path between u and ℓ_{min} over trees, the only feasible sequence of candidate spine nodes starting at candidate u must traverse $\mathcal{P}(u, \ell_{min})$. By the Markov property of the adaptive diffusion spreading mechanism, we only need to consider the likelihood of a candidate spine prior to reaching ℓ_{min}. The propagation of the spine thereafter is conditionally independent of the true source, and therefore equally-likely for all candidates. The maximum likelihood estimator must therefore compute the likelihood of each such candidate sub-spine $\mathcal{P}(u, \ell_{min})$. Since each spine node v chooses one of its uninfected neighbors uniformly at random to be the next spine node, the choice of next spine node is simply $1/(\deg(v)-1)$. Similarly, the likelihood of candidate source u sending the spine in a particular direction is $1/\deg(u)$. The overall likelihood of a candidate is therefore proportional to the product of these degree terms.

C.2. Analysis of spy+snapshot adversarial model

We follow closely the proof of Theorem 1 in Appendix B.1. Given a snapshot at a certain even time T, if there are at least two spy nodes infected at time T, then the adversary can perform the exact same estimation as he did with only spy nodes with $T \to \infty$. We only need to carefully analyze what happens when there are only one spy infected or no spies are infected.

We condition on the lowest-level pivot node, ℓ_{min}, giving $\mathbb{P}(\hat{v}_{\mathrm{ML}} = v^*) = \sum_{\ell_{min}} \mathbb{P}(\hat{v}_{\mathrm{ML}} = v^* | \ell_{min}) \mathbb{P}(\ell_{min})$. Since ℓ_{min} lies on the spine, this is equivalent to conditioning on the distance of ℓ_{min} from the true source. We first define $|S_{d,T}| = 1 + d((d-1)^{T/2}-1)/(d-2)$ as the number of nodes infected at time T, and $|\partial S_{d,T}| = d(d-1)^{(T/2)-1}$ as the number of leaves in the infected subtree. Then,

$$\mathbb{P}(\hat{v}_{\mathrm{ML}} = v^*) = \underbrace{\frac{(1-p)^{|S_{d,T}|-1}}{|\partial S_{d,T}|}}_{\text{no spy}} +$$

$$\sum_{k=1}^{T/2} \Bigg\{ \underbrace{\frac{(1-p)^{(|T_{d,k}|-1)}\, p}{|\partial T_{d,k}|}}_{\ell_{min}\ (k^{th}\text{ spine node) is a spy}} +$$

$$\underbrace{(1-p)^{|T_{d,k}|}(1-(1-p)^{|S_{d,T}|-|T_{d,k+1}|})\mathbb{E}_X\Big[\frac{\mathbb{I}(X \neq d-2)}{(X+1)\,|\partial T_{d,k}|}\Big]}_{\ell_{min}\ (k^{th}\text{ spine node) not a spy}} +$$

$$\underbrace{(1-p)^{|S_{d,T}|-(|T_{d,k+1}|-|T_{d,k}|)}\mathbb{E}_X\Big[\frac{\mathbb{I}(X \neq d-2)}{|\partial S_{d,T}| - (d-2-X)|\partial T_{d,k}|}\Big]}_{\text{all spy descendants of } k\text{-th spine node}} \Bigg\}, \tag{15}$$

where $X \sim \mathrm{Binom}(d-2, (1-p)^{|T_{d,k}|})$, $|T_{d,k}| = \frac{(d-1)^k - 1}{d-2}$ is the number of nodes in each candidate subtree for a pivot at level k, and $|\partial T_{d,k}| = (d-1)^{k-1}$ is the number of leaf nodes in each candidate subtree.

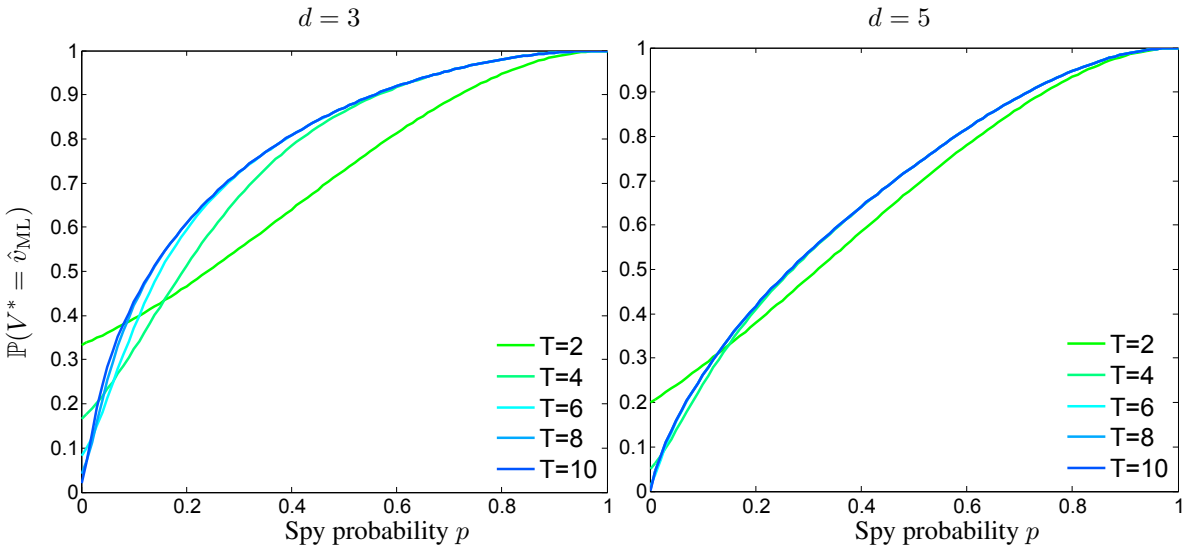

Figure 6. Probability of detection under the spies+snapshot adversarial model. As estimation time and tree degree increase, the effect of the snapshot on detection probability vanishes.