Rumor Source Obfuscation on Irregular Trees

Giulia Fanti
University of Illinois at Urbana-Champaign
fanti@berkeley.edu

Peter Kairouz
University of Illinois at Urbana-Champaign
kairouz2@illinois.edu

Sewoong Oh
University of Illinois at Urbana-Champaign
swoh@illinois.edu

Kannan Ramchandran
University of California, Berkeley
kannanr@eecs.berkeley.edu

Pramod Viswanath
University of Illinois at Urbana-Champaign
pramodv@illinois.edu

ABSTRACT

Anonymous messaging applications have recently gained popularity as a means for sharing opinions without fear of judgment or repercussion. These messages propagate anonymously over a network, typically defined by social connections or physical proximity. However, recent advances in rumor source detection show that the source of such an anonymous message can be inferred by certain statistical inference attacks. Adaptive diffusion was recently proposed as a solution that achieves optimal source obfuscation over regular trees. However, in real social networks, the degrees differ from node to node, and adaptive diffusion can be significantly sub-optimal. This gap increases as the degrees become more irregular.

In order to quantify this gap, we model the underlying network as coming from standard branching processes with i.i.d. degree distributions. Building upon the analysis techniques from branching processes, we give an analytical characterization of the dependence of the probability of detection achieved by adaptive diffusion on the degree distribution. Further, this analysis provides a key insight: passing a rumor to a friend who has many friends makes the source more ambiguous. This leads to a new family of protocols that we call Preferential Attachment Adaptive Diffusion (PAAD). When messages are propagated according to PAAD, we give both the MAP estimator for finding the source and also an analysis of the probability of detection achieved by this adversary. The analytical results are not directly comparable, since the adversary's observed information has a different distribution under adaptive diffusion than under PAAD. Instead, we present results from numerical experiments that suggest that PAAD achieves a lower probability of detection, at the cost of increased communication for coordination.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGMETRICS'15, June 15–19, 2015, Portland, OR, USA.

Copyright © 2015 ACM 978-1-4503-3486-0/15/06 ...$15.00.

http://dx.doi.org/10.1145/2745844.2745866.

Categories and Subject Descriptors

G.2.2 [**Graph Theory**]: Network problems, Graph algorithms

Keywords

Anonymous Social Media; Rumor Spreading; Privacy

1. INTRODUCTION

People have a right to share their thoughts without fear of political or economic repercussion. In cyberspace, freedom of expression often depends on a person's ability to remain anonymous—to one's family, peers, or even government. Several anonymous messaging apps have emerged and evolved in recent years that allow users to post contents anonymously – first-generation apps like Whisper [1] and Yik Yak [2], second-generation ones like Secret [3](which is now out of business) and third generation ones like Blind [4]. These apps build upon an underlying connectivity network between their users, representing a social graph or a physical proximity graph, for instance. When a user posts a message, the message spreads to the user's neighbors, or 'friends', on the connectivity graph. If a friend approves the message by 'liking' it, the message propagates to the friend's friends, and so forth, spreading over the network.

The centralized architecture of existing anonymous messaging apps makes them vulnerable to deanonymization, since authorship information is stored on central servers. Third parties could access that information via hacking or government subpoena. A distributed architecture enables users to propagate messages directly to one another, circumventing centralized storage.

Perhaps surprisingly, even under distributed architectures, the source of a message can still be de-anonymized by global adversaries performing statistical inference. Recent work in rumor source detection shows that the spreading pattern itself reveals a great deal about the true message source [32, 30]. Existing platforms transmit messages to all neighbors immediately upon approval (e.g. when a user clicks 'like'). There is local randomness in each message's spread due to the time it takes for a user to see the message, and uncertainty regarding whether the user will like the message. This random process is typically modeled by the standard *random diffusion* process on graphs.

Rumor source detection algorithms, such as those proposed in [32, 30], exploit the inherent symmetry in how random diffusion propagates. In particular, it is shown in [32] that if an adversary, who knows the underlying contact network, observes which nodes received the message at a certain time, the source can be identified with probability bounded away from zero.

Under this vulnerability against statistical inference attacks, a natural question is, "how can a platform designer intervene with the spread of messages, in order to make rumor source inference difficult?" In other words, how can we add artificial delays on top of the natural human delays to obfuscate the source?

This question was asked in a recent work [13], which proposes a protocol called *adaptive diffusion.* This protocol protects the author's anonymity against the kinds of global adversaries typically assumed in rumor source detection literature. Adaptive diffusion breaks the symmetry of random diffusion by spreading messages faster in some directions on the underlying graph than others (detailed description in Section 2). It is shown in [13] that adaptive diffusion achieves perfect obfuscation when: *(a)* the underlying contact network is a regular tree, and *(b)* the adversary, who knows the underlying contact network, also observes the snapshot of who has seen the message at a certain point in time. A protocol is said to achieve perfect obfuscation, if it successfully hides among all nodes that have seen the message. Precisely, this happens if the probability of the true source being detected is $1/n$ when n nodes have received the message at the time of the attack.

Although adaptive diffusion achieves the best possible source obfuscation, the protocol design assumes a network that is regular and cycle-free. The assumption of cycle-free (i.e. tree-structured) network can be defended by observing that a message's spread on *any* connectivity graph will always be a tree embedded in the true connectivity graph; we assume here that that nodes cannot "be infected", or receive the message, more than once. With out loss of generality, a protocol designer can extract a spanning tree from a given social network, and use only those edges in the tree for spreading the messages. All the privacy guarantees we provide will naturally hold. Having more edges on top of this tree only makes it more difficult for the adversary to locate the source.

On the other hand, experimental results (Figures 6 and 7 and also in [13, Figure 8]) suggest that the performance of adaptive diffusion degrades significantly on certain classes of irregular trees.

In this paper, we ask the fundamental question of how the probability of detection depends on the topology of underlying network, when the network is an *irregular tree.* A precise characterization of the dependency will reveal why adaptive diffusion fails on irregular trees, and provide a guideline for designing novel spreading protocols that improve upon adaptive diffusion.

Model. We follow the setting introduced in [13] for modeling anonymous messaging. At time $t = 0$, a single user $v^* \in V$ starts to spread a message on a contact network $G = (V, E)$ where users and contacts are represented by nodes and edges, respectively. We assume a discrete-time system and model the delays due to user approval and intermittent network access via a deterministic delay of one time unit. Upon receiving the message, the messaging platform can choose to send the message to any of its neighbors the next time step, or add additional delay and wait. Therefore, if we do not intervene with how the messages are spread, then a message always propagates with a delay of one time unit per hop. If the contact network is an infinite d-regular tree, this process spreads to $(d-1)^T$ nodes at time T, but the source is trivially detected as the center of the snapshot. The validity of this model is discussed in Section 5.

Adaptive diffusion is introduced in [13] for hiding the source. The key idea is to add appropriate random delays in order to break the symmetry of the spread. This protocol is shown to achieve perfect obfuscation and is described in the next section.

After T time steps, let $V_T \subseteq V$, G_T, and $N_T \triangleq |V_T|$ denote the set of infected nodes, the subgraph of G containing only V_T, and the number of infected nodes, respectively. At a certain time T, an adversary observes the infected subgraph G_T (as well as the underlying connectivity graph G) and produces an estimate $\hat{v}$ of the source v^* of the message (with probability of detection $P_D = \mathbb{P}(\hat{v} = v^*)$).

There is a tradeoff between hiding the source and increased delay. Assuming a d-regular tree, adaptive diffusion spreads to $(d-1)^{T/2}$ nodes at time T, and it is shown to achieve perfect obfuscation in [13]. We say a protocol achieves a *perfect obfuscation* if the probability of source detection for the maximum likelihood estimator conditioned on n nodes being infected is upper bounded by

$$\mathbb{P}\big(\, \hat{v} = v^* | N_T = n \,\big) \quad = \quad \frac{1}{n} + o\Big(\frac{1}{n}\Big) \, . \tag{1}$$

However, the detection probability can be significantly larger than $1/n$ when the degrees are not regular. To quantify this gap, we analyze the average probability of detection when the underlying contact network is generated from standard random irregular trees with i.i.d. degrees.

Contributions. We make the following contributions in this paper:

- We quantify the sub-optimality of adaptive diffusion, when the underlying contact network is a random irregular tree generated from i.i.d. degree distribution. We give an exact characterization of the detection probability, where the randomness is due to both the underlying network and also the protocol. In the process of the analyses, we prove a new concentration result for an extremal value on a Galton-Watson tree, which may be of independent interest.

- We give a general expression for the adversary's maximum a posteriori (MAP) detection rule for a broad class of protocols, that are generalizations of adaptive diffusion.

- We introduce a family of protocols that we call preferential attachment adaptive diffusion (PAAD). We characterize the probability of detection for this class of protocols. We present numerical experiments suggesting that PAAD achieves better obfuscation than adaptive diffusion.

Related Work. Anonymous communication dates back to Chaum's famous dining cryptographers' (DC) problem [7]. However, most research has so far focused on anonymous *point-to-point* communication, leading to the emergence of Tor [12], Freenet [8], Free Haven [11], and Tarzan [17]. In contrast, we study anonymous *broadcast* messaging. Anonymous broadcast communication has been studied extensibly in the context of DC nets [7, 9, 36, 19, 20, 37]. Our work differs from this body of work by considering: (a) a different class of solutions, based on statistical spreading models rather than cryptographic encoding, and (b) an arbitrary network structure, instead of a fully connected network.

Within the realm of statistical message spreading models, the problem of detecting the origin of an epidemic or the source of a rumor has been studied under the *diffusion* model. Recent advances in [34, 33, 38, 31, 16, 25, 39, 28, 29, 27, 15] show that it is possible to identify the source within a few hops with high probability. Drawing an analogy to epidemics, we refer to a person who has received the message as 'infected' and the act of passing the message as 'spreading the infection'. Just as infectious diseases often spread radially about patient zero geographically, under diffusion, the message spreads in a "ball" around the true source over the network. Thus, the author is very close to the center of that ball, making identification easy. Moreover, this is true independent of the size of that ball.

Recently, [26] modeled the rumor spreading and source identification problem as a "hide and seek" game. [26] studied optimal strategies for both the source and adversary under tree networks from a game theoretic perspective, and derived conditions under which a Nash equilibrium exists. In a different context, [5] studies privacy aware network formation games, where actions are adding/removing edges and the reward is measured by the utility of having friends subtracted by the loss in privacy.

Our problem formulation is closely related to the recent work of [13] where adaptive diffusion was presented, and its optimality under regular tree networks was shown. However, [13] does not study the performance of adaptive diffusion under irregular tree networks. In this work, we (a) prove that adaptive diffusion is sub-optimal under irregular trees, (b) characterize the sub-optimality gap, and (c) present a new class of statistical spreading protocols that improves over adaptive diffusion.

Outline. In Section 2, we describe adaptive diffusion protocol introduced in [13]. We then present our theoretical analysis of adaptive diffusion over irregular random trees in Section 3. Section 4 describes our proposed preferential-attachment adaptive diffusion algorithm, and demonstrates through simulation that it outperforms regular adaptive diffusion on irregular random trees. We discuss some implications of this work and open questions in Section 5. We give the proofs of the main results in Section 6 and in Appendix.

2. ADAPTIVE DIFFUSION

For completeness, adaptive diffusion is described in full detail in the appendix (see Protocol 1). Adaptive diffusion as introduced in [13] refers to a family of protocols parametrized by d_0, where d_0 indicates the degree of the regular tree network for which the protocol is customized. This choice of d_0 determines also the probability of keeping or passing the virtual source in the original adaptive diffusion protocol. Since we consider irregular trees in this paper, the parameter d_0 must be chosen; we set $d_0 = \infty$ throughout this paper. This is equivalent to deciding to pass the virtual source all the time. First of all, this choice is universal, independent of the topology of the underlying graph. Further, it has been suggested via numerical experiments that this choice of d_0 achieves a performance close to the optimal choice (e.g. Figure 5 in [13]). Henceforth, we refer to adaptive diffusion with the choice of $d_0 = \infty$ as simply 'adaptive diffusion'. However, it should be noted that in practice one would prefer choosing a finite d_0 in order to ensure that the messages spread in all directions eventually, as we discuss in Section 5.

As illustrated in Figure 1, adaptive diffusion ensures that at every even time step the infected subtree is a balanced tree with the true source v^* at one of the leaves. At even time T, the infected balanced-tree has radius $T/2$ and the center of this sub-tree is called the *virtual source*, denoted by v_T. Notice that two time steps are needed to spread the infection from one balanced tree of radius $T/2$ rooted at v_T to another balanced tree of radius $T/2+1$ rooted at the next virtual source v_{T+2}, which is adjacent to v_T. The odd time steps are intermediate steps, necessary to make such transitions. Since the propagation of the messages is fully described by the dynamics of the virtual sources, adaptive diffusion only needs to provide a (randomized) rule for choosing the virtual source's location at each (even) time.

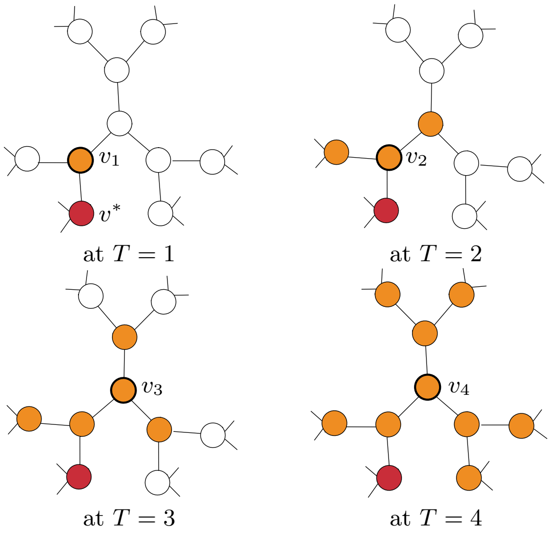

Figure 1: A sample path of adaptive diffusion.

At time $t = 0$, the source v^* chooses a neighbor to be the next virtual source uniformly at random, and infects it. Let v_1 denote this new virtual source as shown in Figure 1. The next time step is spent spreading the infection one more hop to appropriate neighbors, in order to maintain the balanced tree of depth one rooted at the new virtual source, which remains $v_2 = v_1$. In subsequent even time steps, a new virtual source is chosen from one of the current virtual source's neighbor uniformly at random, excluding previous virtual sources. The odd time steps are spent maintaining the balanced infection structure.

3. ANALYSIS OF ADAPTIVE DIFFUSION

At a certain time T, the adversary attacks and observes the set of infected nodes thus far. Given the snapshot at time T, the underlying network, and the knowledge of what protocol is used, the attacker performs statistical inference to detect the source using maximum a posteriori (MAP) estimator. It is proved in [13] that adaptive diffusion achieves (near) perfect obfuscation for *regular* trees. The key idea is that, by construction, all leaves in the infection are equally likely to have been the source, and there are as many leaf nodes in the boundary as in the interior of the infection.

On *irregular* trees, adaptive diffusion is known to be sub-optimal, and the gap depends on the underlying topology of the irregular tree. In order to quantify this gap and characterize the dependence on the underlying tree, we analyze adaptive diffusion on the following model of a random tree and provide rigorous analysis for the average case performance, where the randomness is due to both the underlying contact network as well as the protocol.

We assume adaptive diffusion spreads over a random irregular tree according to a branching process with i.i.d. degrees according to some distribution D. Specifically, at time $t=0$, the source v^* draws a degree d_{v^*} from D, and generates d_{v^*} child nodes. The source picks one of these neighbors uniformly at random to be the new virtual source. The infection spreads as per adaptive diffusion, and each infected node draws its degree from D generating $D-1$ new (uninfected) children nodes.

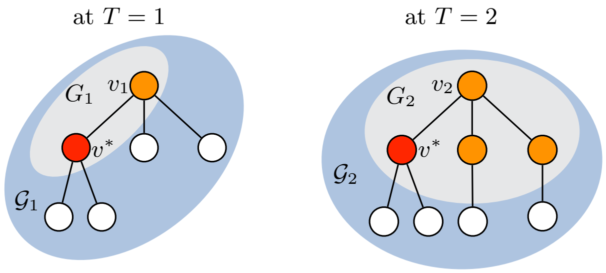

Figure 2: Example of the branching process modeling adaptive diffusion on irregular trees.

We denote the sub-tree of infected nodes as G_T and the branching tree generated at time T as $\mathcal{G}_T$. The structure of the contact network outside of $\mathcal{G}_T$ is independent of G_T conditioned on the uninfected neighbors $\partial\mathcal{G}_T$, where $\partial\mathcal{G}_T$ denotes the leaves of the tree $\mathcal{G}_T$. To ensure that the graph grows indefinitely, we assume that the minimum degree of a node is bounded below by two.

3.1 Probability of Detection Given a Snapshot

The adversary observes this random process at time T, i.e. $\mathcal{G}_T$, knowing that the interior G_T are the infected nodes, and estimates one of the leaf node as an estimate of the true source which started the random process. The following theorem analyzes the probability of detecting the true source for any estimate $\hat{v}$, given a snapshot $\mathcal{G}_T$.

Theorem 3.1. *Under the above described random process of adaptive diffusion, an adversary observes the snapshot $\mathcal{G}_T$ at an even time $T > 0$ and estimates $\hat{v} \in \partial G_T$. For any estimator $\hat{v}$, the conditional probability of detection is*

$$\mathbb{P}(\hat{v} = v^* | \mathcal{G}_T) \;=\; \frac{1}{d_{v_T}} \prod_{\substack{w \in \phi(\hat{v}, v_T) \\ \setminus \{v_T, \hat{v}\}}} \frac{1}{(d_w - 1)} \;, \qquad (2)$$

where v_T is the center of $\mathcal{G}_T$, $\phi(\hat{v}, v_T)$ is the (unique) path from $\hat{v}$ to v_T, G_T is the interior of $\mathcal{G}_T$ which is the infected sub-tree, and ∂G_T is the set of leaves of G_T.

A proof is provided in Appendix A. Intuitively, Equation (2) is the probability that the virtual source starting from $\hat{v}$ ends up at v_T (up to some constant factor for normalization). This gives a simple rule for the adversary to achieve the best detection probability by computing the MAP estimate:

$$\hat{v}^{(T)}_{\text{MAP}} \;\in\; \arg\max_{\hat{v}} \; \mathbb{P}(\hat{v}^{(T)} = v^* | \mathcal{G}_T) \;. \qquad (3)$$

Corollary 3.2. *Under the hypotheses of Theorem 3.1, the MAP estimator in* (3) *can be computed as*

$$\hat{v}^{(T)}_{\text{MAP}} \;=\; \arg\min_{v \in \partial G_T} \prod_{\substack{w \in \phi(v, v_T) \\ \setminus \{v_T, v\}}} (d_w - 1) \;, \qquad (4)$$

achieving a conditional probability of detection

$$\mathbb{P}(\hat{v}^{(T)}_{\text{MAP}} = v^* | \mathcal{G}_T) \;=\; \max_{v \in \partial G_T} \frac{1}{d_{v_T} \prod_{\substack{w \in \phi(v, v_T) \\ \setminus \{v_T, v\}}} (d_w - 1)} . \qquad (5)$$

When applied to regular trees, this recovers known results of [13], which confirms that adaptive diffusion provides strong anonymity guarantees under d-regular trees. But more importantly, Corollary 3.2 characterizes how the anonymity guarantee depends on the general topology of the snapshot. We illustrate this in two extreme examples: a regular tree and an extreme example in Figure 3.

For a d-regular tree, where all nodes have the same degree, the size of infection at even time T is the number of nodes in a d-regular tree of depth $T/2$:

$$N_T \;=\; \frac{d(d-1)^{T/2}}{d-2} + \frac{2}{d-2} \;. \qquad (6)$$

To achieve a perfect obfuscation, we want the probability of detection to decay as $1/N_T$. We can apply Corollary 3.2 to this d-regular tree and show the probability of detection is $((d-1)/d)(d-1)^{-T/2})$, which recovers one of the known results in [13, Proposition 2.2]. This confirms that adaptive diffusion achieves near-perfect obfuscation, up to a small factor of $(d-1)/(d-2)$.

On the other hand, when there exists a path to a leaf node consisting of low-degree nodes, adaptive diffusion can be sub-optimal, and the gap to optimality can be made arbitrarily large. Figure 3 illustrates such an example. This is a tree where all nodes have the same degree $d = 5$, except for those nodes along the path from the center v_T to a leaf node v, including v_T and excluding v. The center v_T has degree two and the nodes in the path have degree three. Hence, the shaded triangles indicate d-regular subtrees of appropriate heights. The size of this infection is $N_T = ((d-1)^{T/2+1}/(d-2)^2)(1+o(1))$. Ideally, one might hope to achieve a probability of detection that scales as $1/(d-1)^{T/2}$. However, Corollary 3.2 shows that the adaptive diffusion achieves probability of detection $1/2^{T/2}$, with the leaf node v achieving this maximum in Equation (5). Hence, there is a multiplicative gap of $((d-1)/2)^{T/2}$. By increasing d, the gap can be made arbitrarily large. On the other hand, such an extreme topology is rare under the i.i.d tree model.

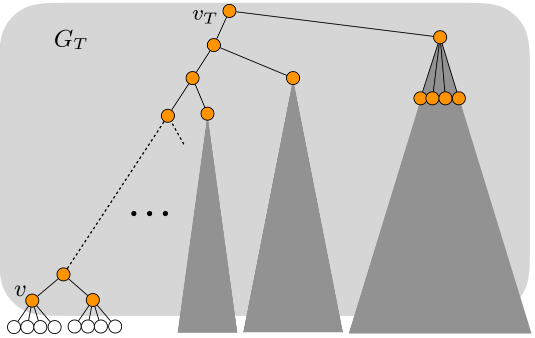

Figure 3: An example of a snapshot emphasizing the sub-optimality of the adaptive diffusion.

We want to emphasize that the results in [13] only use elementary proof techniques and only work for regular trees. In comparison, we develop new proof techniques for the combinatorial problem of counting the number of instances that can start from a leaf v and generate a snapshot $\mathcal{G}_T$.

3.2 Concentration of Probability of Detection

Depending on the topology, adaptive diffusion can be significantly sub-optimal. A natural question is "what is the typical topology for a graph resulting from the random tree model?" Under the model introduced in the beginning of Section 3, we give a concrete answer. Perhaps surprisingly, we can characterize the typical topology as a solution of a simple convex optimization.

We are interested in the following extremal value

$$\Lambda_{G_T} \;\equiv\; d_{v_T} \min_{v\in\partial G_T} \prod_{\substack{w\in\phi(v,v_T)\\ \setminus\{v_T,v\}}} (d_w-1)\;, \tag{7}$$

which captures the topology of the snapshot. We want to characterize the typical value of this function over random tree G_T resulting from the adaptive diffusion process.

Observe that the distribution of the balanced tree G_T follows a simple branching process known as Galton-Watson process. This is because G_T resulting from adaptive diffusion has the same distribution, independent of the location of the source v^*. We consider a given degree distribution D. We use D to denote both a random variable and its distribution—the distinction should be clear from context. The random variable D has support $\boldsymbol{f}=(f_1,\ldots,f_\eta)$ associated with probability $\boldsymbol{p}=(p_1,\ldots,p_\eta)$ such that the degree of node v is i.i.d. with

$$d_v \;=\; \begin{cases} f_1 & \text{with probability } p_1\;,\\ \;\vdots & \;\vdots \\ f_\eta & \text{with probability } p_\eta\;,\end{cases} \tag{8}$$

where $2<f_1<f_2<\cdots<f_\eta$ are integers and the positive p_i's sum to one. We also assume D's support set has at least two elements, i.e., $\eta\ge 2$.

Note that the adaptive diffusion always passes the virtual source token and also chooses where to move it uniformly at random. Hence, for a fixed T, the resulting graph G_T is generated by generating each node degree i.i.d, taking into account the randomness in the social network as well. It is not hard to show that the following branching process has the same distribution over graphs as adaptive diffusion starting from a leaf node v^*: at time $T=0$ a root node, which we denote as the virtual source v_T, creates D offspring. At each subsequent even time step, each leaf node in G_T creates new offspring independently according to $D-1$ (where we subtract one because each leaf is already connected to its parent). This process is repeated until time step T, which generates a random tree G_T.

The following theorem provides a concentration inequality on the extremal quantity Λ_{G_T}, which in turn determines the probability of detection as provided by Corollary 3.2:

$$\mathbb{P}(\hat v^{(T)}_{\rm MAP}=v^*|G_T) \;\;=\;\; \frac{1}{\Lambda_{G_T}}\;. \tag{9}$$

Theorem 3.3. *For an even $T>0$, suppose a random tree G_T is generated from the root v_T according to the Galton-Watson process with i.i.d. degree distribution D, where $\boldsymbol{f}$ and $\boldsymbol{p}$ are defined as in (8), then the following results hold:*

(*a*) *If $p_1(f_1-1)>1$, for any positive $\delta>0$, there exists positive constants $C_{D,\delta}$ and $C'_{D,\delta}$ that depend only on the degree distribution and the choice of δ such that*

$$\mathbb{P}\left(\left|\frac{\log(\Lambda_{G_T})}{T/2}-\log(f_1-1)\right|>\delta\right)\le e^{-C_{D,\delta}T}\;, \tag{10}$$

for an even time $T\ge C'_{D,\delta}$.

(*b*) *If $p_1(f_1-1)<1$, define the mean number of children:*

$$\mu_D \;\equiv\; \sum_{i=1}^{\eta} p_i(f_i-1)\;,$$

and the set

$$\mathcal{R}_D \;=\; \big\{\,\boldsymbol{r}\in S_\eta \;|\; \log(\mu_D)\ge D_{\rm KL}(\boldsymbol{r}\|\boldsymbol{\beta})\,\big\}\;, \tag{11}$$

where S_η denotes the η-dimensional probability simplex, $D_{\rm KL}(\cdot\|\cdot)$ denotes Kullback-Leibler divergence, and $\boldsymbol{\beta}$ is a length-η probability vector in which $\beta_i=p_i(f_i-1)/\mu_D$. Further, define $\boldsymbol{r}^*$ as follows:

$$\boldsymbol{r}^* = \;\; \arg\min_{\boldsymbol{r}\in\mathcal{R}_D} \;\; \big\langle\,\boldsymbol{r}\,,\,\log(\boldsymbol{f}-1)\,\big\rangle\;, \tag{12}$$

where $\langle\boldsymbol{r},\log(\boldsymbol{f}-1)\rangle=\sum_{i=1}^{\eta} r_i\log(f_i-1)$. Then for any $\delta>0$, there exist positive constants $C_{D,\delta}$ and $C'_{D,\delta}$ that only depend on the degree distribution D and the choice of $\delta>0$ such that

$$\mathbb{P}\left(\left|\frac{\log(\Lambda_{G_T})}{T/2}-\langle\boldsymbol{r}^*,\log(\boldsymbol{f}-1)\rangle\right|>\delta\right)\;\le\; e^{-C_{D',\delta}T} \tag{13}$$

for an even time $T\ge C'_{D,\delta}$.

A proof of this theorem is provided in Section 6. Putting it together with (9), it follows that the probability of detection concentrates around

$$-\frac{2}{T}\log\big(\mathbb{P}(\hat v^{(T)}_{\rm MAP}=v^*)\big) \;\;\simeq\;\; \langle\boldsymbol{r}^*\,,\,\log(\boldsymbol{f}-1)\rangle\;,$$

in case (b) and around $\log(f_1-1)$ in case (a). Here $\simeq$ indicates concentration for large enough T. We want to emphasize that $\boldsymbol{r}^*$ can be computed using off-the-shelf optimization tools, since the program in (12) is a convex program of dimension η. This follows from the fact that the objective is linear in $\boldsymbol{r}$ and the feasible region is convex since KL divergence is convex in $\boldsymbol{r}$.

For example, if D is 3 w.p. 0.7 or 4 w.p. 0.3, then this falls under case (a). The theorem predicts the probability of detection to decay as $(3-1)^{-T/2}$. On the other hand, if

$$D = \left\{ \begin{array}{ll} 2 & \text{with probability } 0.3 \\ 3 & \text{with probability } 0.7 \end{array} \right. ,$$

then this falls under case (b) with $\mu_D = 1.7$, $\beta_1 = 0.3/1.7$, and $\beta_2 = 1.4/1.7$. In this case, the exponent is a solution of the following optimization for $\boldsymbol{r} = [r, 1-r]$:

$$\begin{array}{ll} \underset{r\in\mathbb{R}}{\text{minimize}} & r\log 1 + (1-r)\log 2 \\ \text{subject to} & r\log\dfrac{1.7r}{0.3} + (1-r)\log\dfrac{1.7(1-r)}{1.4} \le \log(1.7) \\ & r \in [0,1] \end{array}$$

It follows that the optimal solution is $\boldsymbol{r}^* \simeq [0.64,\ 0.36]$ and the probability of detection decays as $2^{-0.36(T/2)}$. Figure 4 confirms this theoretical prediction with numerical simulations for these two examples.

Theorem 3.3 provides a simple convex program that computes the probability of detection for any degree distribution. For random trees, this quantifies the gap between what adaptive diffusion can guarantee and the perfect obfuscation one desires. We define the rescaled log-multiplicative gap as

$$\Delta_D \;\equiv\; \frac{2}{T}\log\frac{\mathbb{P}(v_{\text{MAP}}^{(T)} = v^*)}{1/\mathbb{E}[|\partial G_T|]} ,$$

where $|\partial G_T|$ is the total number of candidates in a snapshot. It is not difficult to show that $\mathbb{E}[|\partial G_T|] = \mu_D^{T/2}$, and it follows that $\Delta_D \simeq \log\mu_D - \langle \boldsymbol{r}^*, \log(\boldsymbol{f}-1)\rangle$. For example, $\Delta_D = 0$ for regular trees, and $\Delta_D = \log_2 2.3 - \log_2 2 = 0.20$ for the first example under case (a) and $\Delta_D = \log_2 1.7 - 0.36 = 0.41$ for the second example under case (b).

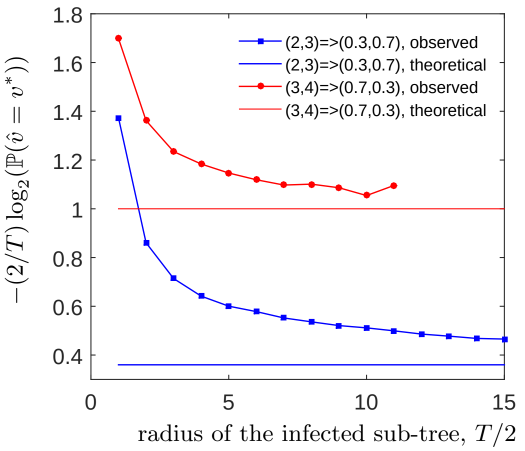

Figure 4: Simulated detection probability converges to the theoretical prediction.

3.3 Simulation Results

The notation $(3,4) => (0.5, 0.5)$ in the legend of Figure 4 indicates that each node in the tree has degree 3 or 4, each with probability 0.5. In this case, we have $p_1(f_1-1) > 1$. This is case (a) in Theorem 3.3, and our analysis predicts that the exponent is $\log_2(f_1-1) = 1$. This is indicated by a solid red line in Figure 4. For the other distribution with support $\boldsymbol{f} = (2,3)$ with probabilities $\boldsymbol{p} = (0.3, 0.7)$, we have $p_1(f_1-1) < 1$. This is case (b) in the theorem, and the analysis predicts that the exponent is 0.36 as we computed in the previous section. This is indicated by a solid blue line.

In this plot, data points represent successive even timesteps of adaptive diffusion—or equivalently, the radius of the infected sub-tree. For both examples, we simulated adaptive diffusion and used the MAP estimator in Equation (4) to detect the source. The resulting average probability of detection, averaged over 10,000 trials, is plotted in Figure 4, where the randomness is both in the protocol as well as the underlying random tree. We observe that the empirical exponent $-\log(\mathbb{P}(\hat{v} = v^*))/(T/2)$ converges to the theoretical prediction, albeit slowly. The size of this experiment was limited by computational considerations, since the graph size grows exponentially in time.

3.4 Sketch of the Proof of Theorem 3.3

We provide a sketch of a proof of Theorem 3.3, first for case (a). We suppose $\eta = 2$ to simplify the notation and highlight the key insights. The random snapshot G_T is distributed according to T steps of Galton-Watson process starting from the root node v_T, with degree distribution D. Note that v_T is the center of G_T and not the source of the message. We want to prove that, with high probability, there exists a path from the root v_T to a leaf v which (mostly) consists of nodes with degree f_1, such that

$$\min_{v} \prod_{w\in\phi(v_T,v)\setminus\{v_T,v\}} (d_w - 1) \;\simeq\; (f_1-1)^{T/2-1} ,$$

where $\simeq$ indicates that we allow for vanishing fraction of nodes to deviate from the minimum degree f_1.

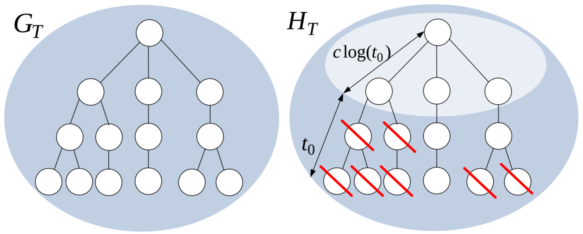

Figure 5: Pruning of a snapshot. We prune all descendants of nodes with degree 3 that are more than $c\log(t_0)$ hops from the root, where $t_0 + c\log t_0 = T/2$.

The main idea is to consider the sub-tree of G_T where we remove all nodes with degree f_2 and also all its descendants. First, if this pruned tree reaches the boundary at $T/2$, which is referred to as surviving until $T/2$ time steps, then we know that in the original G_T there exists a path from the root to a leaf consisting only of nodes with degree f_1. Furthermore, The pruned tree is also a Galton-Watson process, but with a different degree distribution:

$$\tilde{D} \;=\; \left\{ \begin{array}{ll} 1 & \text{with probability } p_2 , \\ f_1 & \text{with probability } p_1 . \end{array} \right. .$$

Under the assumption of case (a) that $(f_1-1)p_1 > 1$, we know from standard analysis of branching processes [22] that this pruned process survives for $T/2$ time steps with a strictly positive probability for any T. By delaying the pruning until an appropriately chosen $O(\log T)$ time steps, we can make this survival probability as large as we want.

In case (b), the same argument proves that the pruned process will not survive, and there will not be a path of minimum degree nodes reaching the boundary. We need to relax the pruning, and allow r_2 proportion of the nodes in a path to have degree f_2 for some $r_2 \in [0,1]$. We search over all possible choices of $\boldsymbol{r} = [1 - r_2,\ r_2]$ such that the process barely survives while minimizing Λ_{G_T}:

$$\begin{aligned} \underset{r_2 \in [0,1]}{\text{minimize}} \quad & (f_1 - 1)^{(1-r_2)T/2} (f_2 - 1)^{(r_2)T/2} \\ \text{subject to} \quad & \mathbb{P}(\text{survival}) > 0 \end{aligned}$$

Note that we have the same objective as in (12), up to a logarithm and scaling. The challenge in analyzing the constraint is that the pruned process is now dynamic, in the sense that we do not fix the number of f_2 nodes allowed in any path (as we did in case (a), allowing zero f_2 nodes), but rather let this allowed number grow proportionally to the tree depth. To analyze this dynamically-pruned process, we define a multi-type branching process, where the type of a node encodes the quantity we are interested in, namely the product of degrees in the path from the root. We analyze all possible pruning in this multi-type branching process, and show that it survives if and only if the conditions in (11) is satisfied. The complete proof is provided in Section 6.

4. PREFERENTIAL ATTACHMENT

Our analysis reveals that adaptive diffusion can be significantly sub-optimal, when the underlying graph degrees are highly irregular. To bridge this gap, we introduce a family of protocols we call *Preferential Attachment Adaptive Diffusion (PAAD)*. We analyze the performance of PAAD and provide numerical simulations showing that PAAD improves over adaptive diffusion when degrees are irregular.

The reason for this gap is that in typical random trees, there are nodes that are significantly more likely to be the source, compared to other typical candidate nodes. To achieve near-perfect obfuscation, we want all candidate nodes to have similar posterior probabilities of being the source. To balance the posterior probabilities of leaf nodes, we suggest passing the virtual source in favor of large-degree nodes. We propose a new family of protocols base on this intuition, and make this intuition precise in Theorem 4.1.

PAAD is based on adaptive diffusion, but we modify how virtual sources are chosen. We parametrize this family of protocols by a non-negative integer g. When a new virtual source is to be chosen, instead of choosing uniformly among its neighbors (except for the previous virtual source), the new virtual source is selected with probability weighted by the size of its g-hop neighborhood. Let $\mathcal{N}_g(v)$ denote the set of g-hop neighbors of node v, and let $\mathcal{N}_g(v, w)$ denote the same set, removing any nodes z for which $w \in \phi(z, v)$, where $\phi(z, v)$ denotes the path between z and v. Then for instance, if $g = 1$, then each time the virtual source is passed from v_T to v_{T+2}, it is passed to a neighbor $w \in \mathcal{N}_1(v_T, v_{T-2})$ with probability proportional to $d_w - 1$:

$$\mathbb{P}(v_{T+2} = w) = \frac{d_w - 1}{\sum_{w' \in \mathcal{N}_1(v_T, v_{T-2})} (d_{w'} - 1)} \ .$$

For general g, the probability is proportional to the size of the candidate w's g-hop local neighborhood, excluding those in the direction of the current virtual source v_T. Each virtual source v_T chooses the next virtual source as follows: for any node $w \in \mathcal{N}_1(v_T, v_{T-2})$,

$$P(v_{T+2} = w) = \frac{|\mathcal{N}_g(w, v_T)|}{\sum_{w' \in \mathcal{N}_1(v_T, v_{T-2})} |\mathcal{N}_g(w', v_T)|} \ .$$

PAAD encourages the virtual source to traverse high-degree nodes. This balances the posterior probabilities, by strengthening the probability of leaf nodes whose path contain high-degree nodes, while weakening those with low-degree nodes.

This intuition is made precise in the following theorem, which analyzes the probability of detection for a given snapshot. Define the probability that the sequence of decisions on choosing the virtual sources results in the path from a source v to the current virtual source v_T as

$$Q(\mathcal{G}_T, v) \quad \equiv \quad \prod_{t=1}^{T/2} \mathbb{P}(v_{2t} = w_t) \ ,$$

where $\phi(v, v_T) = (w_0 = v, w_1, w_2, \ldots, w_{T/2-1}, w_{T/2} = v_T)$. The specific probability depends on the choice of g and the topology of the underlying tree. Note that the progression of the virtual source now depends on g-hop neighborhood, and we therefore define $\mathcal{G}_T$ to include the current infected subgraph G_T and its $(g+1)$-hop neighborhood.

THEOREM 4.1. *Suppose a node v^* starts to spread a message at time $t = 0$ according to PAAD, where the underlying irregular tree is generated according to the random branching process described in the beginning of Section 3. At a certain even time $T \geq 0$, an adversary observes the snapshot of the infected subtree $\mathcal{G}_T$ and computes a MAP estimate of the source v^*. Then, the following results hold:*

(a) *The MAP estimator is*

$$\hat{v}_{\text{MAP}} \ = \ \arg\max_{v \in \partial G_T} \ d_v \, Q(\mathcal{G}_T, v) \qquad (14)$$

where ∂G_T denotes the leaves of G_T.

(b) *The conditional probability of detection achieved by the MAP estimator is*

$$\mathbb{P}(\hat{v}_{\text{MAP}} = v^* | \mathcal{G}_T) \ = \ \frac{\max_{v \in \partial G_T} d_v \, Q(\mathcal{G}_T, v)}{\sum_{w \in \partial G_T} d_w \, Q(\mathcal{G}_T, w)} \qquad (15)$$

The proof relies on the techniques developed to prove Theorem 3.1, and is omitted due to space limitation. The example from Figure 3 illustrates the power of PAAD. For this class of snapshots, it is straightforward to show that under adaptive diffusion, $P_D^{AD} = 2^{-T/2}$, whereas under 1-hop PAAD,

$$P_D^{PAAD} \leq \frac{2}{(d-1)^{T/2-1} - 1} .$$

Notice from these expressions that P_D^{PAAD} scales as $(d - 1)^{-T/2}$, which achieves perfect obfuscation, whereas regular adaptive diffusion scales as $2^{-T/2}$.

This shows that there exist snapshots where PAAD significantly improves over adaptive diffusion. However, such examples are rare under the random tree model, and there are also examples of snapshots where adaptive diffusion can achieve a better obfuscation than PAAD. To complete the analysis, we would like to show the analog of Theorem 3.3 for

PAAD. However, the observed snapshot is no longer generated by a standard Galton-Watson branching process, due to the preferential attachment. The analysis techniques developed for Theorem 3.3 does not generalize, and significantly new techniques appear to be needed for a technical analysis. This is outside the scope of this manuscript; instead, we show numerical simulations suggesting that PAAD improves over adaptive diffusion.

4.1 Simulation Results

PAAD requires each virtual source to know some information about its local neighborhood on the contact network; in exchange, we observe empirically that it hides the source better than traditional adaptive diffusion. Figure 6 shows the probability of detection over graphs with a degree distribution of support $\boldsymbol{f} = (2, 5)$ with probability $\boldsymbol{p} = (0.5, 0.5)$. The results are averaged over 10,000 realizations of the random graph and the spreading sequence. This plot shows empirically that preferential attachment adaptive diffusion exhibits better hiding properties than regular adaptive diffusion, and that the benefit of preferential attachment increases with the size of the neighborhood considered for preferential attachment (e.g., one-hop vs. two-hop). Notice that our lower bound on probability of detection is $1/|\partial G_T|$ rather than $1/N_T$, as in [13]; this is because we constrain the source to always be at one of the leaves of the graph, so $1/|\partial G_T|$ lower bounds the probability of detection.

Figure 7 computes the ratio of the observed probability of detection to a lower bound on the probability of detection (i.e., $1/|\partial G_T|$), for both adaptive diffusion (AD) and one-hop PAAD. Empirically, we observe that the advantage of PAAD is greater when the degree distribution is more imbalanced (i.e., when $f_{\max} - f_{\min}$ is large).

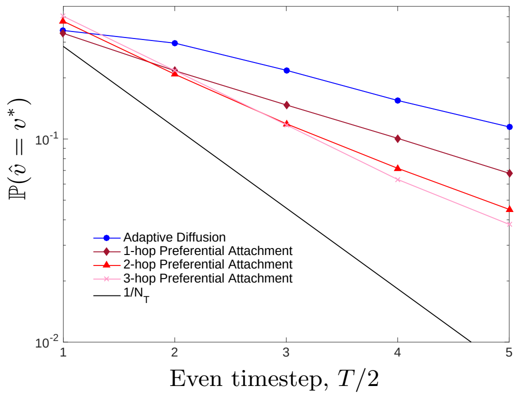

Figure 6: Probability of detection of regular adaptive diffusion compared to 1-, 2-, and 3-hop preferential attachment adaptive diffusion (PAAD).

5. DISCUSSION

We characterize how the probability of detection depends on the degree distribution of the underlying random tree, when messages spread as per adaptive diffusion. This suggests a novel family of protocols, we call preferential attachment adaptive diffusion. We analytically calculate the conditional probability of detection of this family of protocols, and numerical results suggest that this improves over adaptive diffusion.

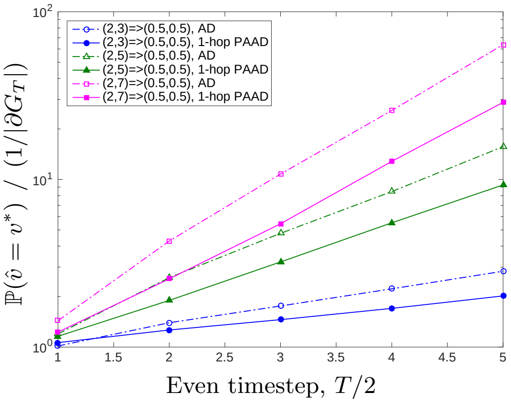

Figure 7: Ratio of observed probability of detection to lower-bound probability of detection, for a range of degree distributions. PAAD has better anonymity properties than regular adaptive diffusion over random, irregular trees.

Ideally, we would like to prove an upper bound on the average probability of detection under adaptive diffusion:

$$\mathbb{P}(\hat{v}^{(T)}_{\text{MAP}} = v^*) \quad = \quad \sum_{\mathcal{G}_T} \mathbb{P}(\mathcal{G}_T)\, \mathbb{P}(\hat{v}^{(T)}_{\text{MAP}} = v^* | \mathcal{G}_T)\ , \quad (16)$$

averaged over all instances of the graphs. We conjecture this has the similar scaling as the concentration result in Theorem 3.3. The main challenge is in proving a sharper concentration of Λ_{G_T}.

Proving a result analogous to Theorem 3.3 for PAAD is an interesting research direction, with new technical challenges due to the fact that snapshot G_T is no longer a Galton-Watson tree. Although the distribution of the snapshot can be modeled by a multi-type branching process, it is an interesting open question how to analyze the typical value of $Q(\mathcal{G}_T, v)$.

Another future research direction is analyzing the general version of adaptive diffusion and PAAD, where the source is not constrained to be at the leaf. This is more practical, in the sense that a natural message spreading should spread to all directions eventually. We believe the same techniques developed in this paper can be naturally generalized to handle this more general adaptive diffusion.

A very relevant question in rumor spreading is, how to hide the relays that were involved in the spreading. How do we hide the fact that you “approved” a particular message on a sensitive topic. Initial investigations have been done in a recent work [18], but there are similar interesting technical and algorithmic questions, such as what is the fundamental tradeoff between the privacy, spreading rate, and the utility lost measured by the increased number of spams in the social filtering process.

It has been recently shown that when there are multiple sources, it is more difficult to locate them [35]. It might be possible to spread the messages faster than the adaptive diffusion and still achieve a perfect obfuscation, by creating multiple pseudo sources. The cost of such an approach is the loss in the social filtering, since we start spreading a message from a remote node who is not necessarily approving the message.

Finally, there has been recent advances in finding the first node to be present in a randomly growing network [6, 21, 23, 24]. Unlike a diffusion, the network itself is progressively growing, for example as a preferential attachment in [6], and one is interested in the fundamental limit on the size of a candidate set in order to ensure that the true *source* or the root that started the random network is in the candidate set with failure probability is at most some positive ε. A natural question of interest is to design a randomized network growth process that hides the identity of the root.

6. PROOF OF THEOREM 3.3

To facilitate the analysis, we consider an alternative random process that generates unlabeled graphs G'_T according to the same distribution as G_T (i.e., the infected, unlabeled subgraph embedded in $U(G_D^{(T)})$ from the proof of Theorem 3.1). For a given degree distribution D and a stopping time T, the new process is defined as a Galton-Watson process in which the set of offsprings at the first time step is drawn from D and the offsprings at subsequent time steps are drawn from $D-1$. At time $t=0$, a given root node v_T draws its degree d_{v_T} from D, and generates d_{v_T} child nodes. The resulting tree now has depth 1. In each subsequent time step, the process traverses each leaf v of the tree, draws its degree from D, and generates d_v-1 children. The random process continues until the tree has depth $T/2$, since under adaptive diffusion, the infected subgraph at even time T has depth $T/2$. Because the probability of detection in Equation (5) does not depend on the degrees of the leaves of G_T, the random process stops at depth $T/2$ rather than $T/2+1$. We call the output of this random process G'_T. The distribution of G'_T is identical to the distribution as the previous random process imposed on G_T, which follows from the proof of Theorem 3.1. We therefore use G_T to denote the resulting output in the remainder of this proof.

Distribution D is a multinomial distribution with support $\boldsymbol{f}=(f_1,\ldots,f_\eta)$ and probabilities $\boldsymbol{p}=(p_1,\ldots,p_\eta)$. Without loss of generality, we assume $2\leq f_1<\ldots<f_\eta$. Let μ_D denote the mean number of *children* generated by D:

$$\mu_D=\sum_{i=1}^{\eta}p_i(f_i-1).$$

There are two separate classes of distributions, which we deal with as separate cases.

Case 1: When $p_1(f_1-1)>1$, we claim that with high probability, there exists a leaf node v in ∂G_T such that on the unique path from the root v_T to this leaf v, all nodes in this path have the minimum degree f_1, except for a vanishing fraction. To prove this claim, consider a different graph H_T derived from G_T by pruning large degree nodes:

1. For a fixed, positive c, find t_0 such that $T/2=t_0+c\log(t_0)$.
2. Initialize H_T to be identical to G_T.
3. For each node $v\in H_T$, if the hop distance $\delta_H(v,v_T)\leq c\log(t_0)$, do not modify that node.
4. For each node $v\in H_T$, if the hop distance $\delta_H(v,v_T)>c\log(t_0)$ and $d_v>f_1$, prune out all the children of v, as well as all their descendants (Figure 5).

We claim that this pruned process survives with high probability. The branching process that generates H_T is equivalent to a Galton-Watson process that uses distribution $D-1$ for the first $c\log(t_0)$ generations, and a different degree distribution $D'-1$ for the remaining generations; D' has support $\boldsymbol{f'}=(f_1,1)$, probability mass $\boldsymbol{p'}=(p_1,1-p_1)$, and mean number of children $\mu_{D'}=p_1(f_1-1)$.

Note that $f_1\geq 3$ by the assumption that $p_1(f_1-1)>1$. Hence, the inner branching process up to $c\log t_0$ has probability of extinction equal to 0. This means that at a hop distance of t_0 from v_T, there are at least $(f_1-1)^{c\log(t_0)}$ nodes. Each of these nodes can be thought of as the source of an independent Galton-Watson branching process with degree distribution $D'-1$. By the properties of Galton-Watson branching processes ([22], Thm. 6.1), since $\mu_{D'}>1$ by assumption, each independent branching process' asymptotic probability of extinction is the unique solution of $g_{D'}(s)=s$, for $s\in[0,1)$, where $g_{D'}(s)=p_1\,s^{f_1-1}+(1-p_1)$ denotes the probability generating function of the distribution D'. Call this solution $\theta_{D'}$. The probability of any individual Galton-Watson process going extinct in the first generation is exactly $1-p_1$. It is straightforward to show that $g_{D'}(s)$ is convex, and $g_{D'}(1-p_1)>1-p_1$, which implies that the probability of extinction is nondecreasing over successive generations and upper bounded by $\theta_{D'}$. Then for the branching process that generates H_T, the overall probability of extinction (for a given time T) is at most $\theta_{D'}^{(f_1-1)^{c\log t_0}}$. Increasing the constant c therefore decreases the probability of extinction. If there exists at least one leaf at depth T (i.e., extinction did not occur), then there exists at least one path in H_T of length $t_0-c\log t_0$ in which every node (except possibly the final one) has the minimum degree f_1. This gives

$$\begin{aligned}\frac{\log(\Lambda_{H_T})}{T/2} &\leq \frac{t_0\log(f_1-1)+c\log(t_0)\log(f_\eta-1)}{t_0+c\log(t_0)} \qquad (17)\\ &\leq \log(f_1-1)+\frac{c\log t_0}{t_0}\log\frac{f_\eta-1}{f_1-1}\,, \qquad (18)\end{aligned}$$

with probability at least $1-\theta_{D'}^{(f_1-1)^{c\log t_0}}=1-\theta_{D'}^{t_0^{c\log(f_1-1)}}=1-e^{-C_{D'}t_0}$, where $C_{D'}=\log(\theta_{D'})$ and the upper bound in (17) comes from assuming all the interior nodes have maximum degree f_η. Since H_T is a subgraph of a valid snapshot G_T, there exists a path in G_T from the virtual source v_T to a leaf of the tree where the hop distance of the path is exactly $T/2$, and at least t_0 nodes have the minimum degree f_1. Since the second term in (18) is $o(t_0)$, the claim follows. The lower bound $\log(\Lambda_{H_T})/(T/2)\geq\log(f_1-1)$ holds by definition. Therefore, for any $\delta>0$, by setting T (and consequently, t_0) large enough, we can make the second term in (18) arbitrarily small. Thus, for $T\geq C'_{D,\delta}$, where $C'_{D,\delta}$ is a constant that depends only on the degree distribution and δ, the result holds.

Case 2: Consider the case when $p_1(f_1-1)<1$. By the properties of Galton-Watson branching processes ([22], Thm. 6.1), the previous pruned random process that generated graphs H_T goes extinct with probability approaching 1. This implies that with high probability there is no path from the root to a leaf that consists of only minimum degree nodes.

Instead, we introduce a Galton-Watson process with multiple types, derived from the original process. Our approach is to assign a numeric *type* to each node in G_T according to the number of non-minimum-degree nodes in the unique

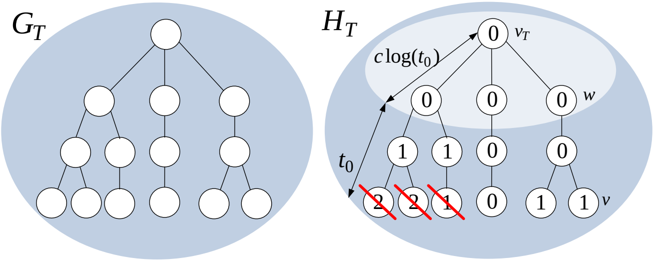

Figure 8: Pruning of a snapshot using multiple types. In this example, the distribution D allows nodes to have degree 2 or 3. We take $t_0 = 2$ and $r = 0.5$, so all descendants of nodes with type $rt_0 = 1$ are pruned.

path between that node and the virtual source. If a node's path to v_T contains too many nodes of high degree, then we prune the node's descendants. The challenge is to choose the smallest pruning threshold that still ensures the pruned tree will survive with high probability. Knowing this threshold allows us to precisely characterize Λ_{G_T} for most of the instances.

To simplify the discussion, we start by considering a special case in which D allows nodes to take only two values of degrees, i.e., $\eta = 2$. We subsequently extend the results for $\eta = 2$ to larger, finite values of η. With a slight abuse of a notation, consider a new random process H_T derived from G_T by pruning large degree nodes in the following way:

1. For a fixed, positive c, find t_0 such that $T/2 = t_0 + c\log(t_0)$.

2. Initialize H_T to be identical to G_T.

3. For each node $v \in H_T$, if the hop distance $\delta_H(v, v_T) \leq c\log(t_0)$, do not modify that node, and assign it type 0.

4. For each node $v \in H_T$, if the hop distance $\delta_H(v, v_T) > c\log(t_0)$, assign v a type ξ_v, which is the number of nodes in $\phi(w, v) \setminus \{v\}$ that have the maximum possible degree f_2, where w is the closest node in H_T to v such that $\delta_H(w, v_T) \leq c\log(t_0)$ (Figure 8).

5. Given a threshold $r \in (0,1)$, if a node v has type $\xi_v \geq rt_0$, prune out all the descendants of v. For example, in Figure 8, if $t_0 = 2$ and the threshold is $r = 0.5$, we would prune out all descendants of nodes with $\xi_v \geq 1$.

We show that for an appropriately-chosen threshold r, this pruned tree survives with high probability. By choosing the smallest possible r, we ensure that Λ_{H_T} consists (in all but a vanishing fraction of nodes) of a fraction r nodes with maximum degree, and $(1-r)$ of minimum degree. This allows us to derive the bounds on $\log(\Lambda_{H_T})/(T/2)$ stated in the claim, which hold with high probability.

Let $k \equiv rt_0$. The process that generates H_T is equivalent to a different random branching process that generates nodes in the following manner: set the root's type $\xi_{v_T} = 0$. At time $t = 0$, the root v_T draws a number of children according to distribution D, and generates d_{v_T} children, all type 0. Each leaf generates type 0 children according to child degree distribution $D - 1$ until $c\log(t_0)$ generations have passed. At that point, each leaf v in this branching process (which necessarily has type 0) reproduces as follows: if its type $\xi_v > k$, then v does not reproduce. Otherwise, it either generates $(f_1 - 1)$ children with probability p_1, each with state ξ_v, or it generates $(f_2 - 1)$ children with probability p_2, each with state $\xi_v + 1$. This continues for t_0 generations. Mimicking the notation from Case 1, we use D' to denote the distribution that gives rise to this modified, multi-type random process (in the final t_0 generations); this is a slight abuse of notation since the branching dynamics are multi-type, not defined by realizations of i.i.d. degree random variables.

Lemma 6.1. *Consider a Galton-Watson branching process with child degree distribution $D - 1$, where each node has at least one child with probability 1, and $\mu_{D-1} > 1$. Then the number of leaves in generation t, $Z^{(t)}$, satisfies the following:*

$$Z^{(t)} \geq e^{C_\ell t}$$

with probability at least $1 - e^{C'_\ell t}$, where both C_ℓ and C'_ℓ are constants that depend on the degree distribution.

We omit the proof due to space limitations. The first $c\log(t_0)$ generations ensure that with high probability, we have at least $e^{C_\ell \log t_0}$ independent multi-type Galton-Watson processes originating from the leaves of the inner subgraph; this follows from Lemma 6.1. Here we have encapsulated the constant c from the first $c\log(t_0)$ generations in the constant C_ℓ. For example, in Figure 8, there are 3 independent Galton-Watson processes starting at the leaves of the inner subgraph. We wish to choose r such that the expected number of new leaves generated by *each* of these processes, at *each* time step, is large enough to ensure that extinction occurs with probability less than one. For brevity, let $\alpha \equiv p_1(f_1 - 1)$ and let $\beta \equiv p_2(f_2 - 1)$. Let $x^{(t)}$ denote the $(k+1)$-dimensional vector of the expected number of leaves generated with each type from 0 to k in generation t. This vector evolves according to the following $(k+1) \times (k+1)$ transition matrix M:

$$x^{(t+1)} = x^{(t)} \underbrace{\begin{bmatrix} \alpha & \beta & & & \\ & & \ddots & & \\ & & \ddots & \alpha & \beta \\ & & & & 0 \end{bmatrix}}_{M}.$$

The last row of M is 0 because a node with type k does not reproduce. Since the root of each process always has type 0, we have $x^{(0)} = \boldsymbol{e}_1$, where $\boldsymbol{e}_1$ is the indicator vector with a 1 at index 1 and zeros elsewhere.

Let $Z^{(t)}$ denote the expected number of new leaves created in generation t. This gives

$$\mathbb{E}[Z^{(t)}] = \boldsymbol{e}_1 M^t \mathbb{1}^\intercal_{(k+1)}, \tag{19}$$

where $^\intercal$ denotes a transpose, and $\mathbb{1}_{(k+1)}$ is the $(k+1)$ all-ones vector. When $t < k$, this is a simple binomial expansion of $(\alpha + \beta)^t$. For $t \geq k$, this is a truncated expansion up to k:

$$\mathbb{E}[Z^{(t)}] = \sum_{i=0}^{k} \binom{t}{i} \alpha^{t-i} \beta^i. \tag{20}$$

We seek the necessary and sufficient condition on r for non-extinction, such that $(1/t)\log(\mathbb{E}[Z^{(t)}]) > 0$. Consider a binomial random variable W with parameter $\beta/(\alpha+\beta) = \beta/\mu_D$ and t trials. Equation (20) implies that for large t,

$$\begin{aligned}\mathbb{E}[Z^{(t)}] &= (\alpha+\beta)^t\,\mathbb{P}(W \le k). && (21)\\ &= \mu_D^t\,\exp\Big\{-t\,D_{\mathrm{KL}}\Big(r\,\|\,\frac{\beta}{\mu_D}\Big) + o(t)\Big\}\,, && (22)\end{aligned}$$

by Sanov's theorem [10]. Here $D_{\mathrm{KL}}(r\|\beta/\mu_D)$ denotes the Kullback-Leibler divergence between the two Bernoulli distributions defined by r and β/μ_D, such that $D_{\mathrm{KL}}(r\|\beta/\mu_D) = (1-r)\log((1-r)/(\alpha/\mu_D)) + r\log(r/(\beta/\mu_D))$. We wish to identify the smallest r for which $(1/t)\log(\mathbb{E}[Z^{(t)}])$ is bounded away from zero. Such an r is a sufficient (and necessary) condition for the multi-type Galton-Watson process to have a probability of extinction less than 1. To achieve this, we define the following set of r such that Eq. (22) is strictly positive, for some $\epsilon > 0$:

$$\mathcal{R}_{\alpha,\beta}(\epsilon) \;=\; \big\{\, r \mid \log(\mu_D) \ge D_{\mathrm{KL}}(r\|\beta/\mu_D) + \epsilon \,\big\}\,, \tag{23}$$

Suppose we now choose a threshold $r \in \mathcal{R}_{\alpha,\beta}(\epsilon)$. This is the regime where the modified Galton-Watson process with threshold r has a chance for survival. In other words, the probability of extinction $\theta_{D'}$ is strictly less than one. Precisely, $\theta_{D'}$ is the unique solution to $s = g_{D'}(s)$, where $g_{D'}(s)$ denotes the probability generating function of the described multi-type Galton-Watson process. Using the same argument as in Case 1, we can construct a process where the probability of extinction is asymptotically zero. Precisely, we modify the pruning process such that we do not prune any leaves in the first $c\log(t_0)$ generations. This ensures that with high probability, there are at least $e^{C_\ell \log(t_0)}$ independent multi-type Galton-Watson processes evolving concurrently after time $c\log(t_0)$, each with probability of extinction $\theta_{D'}$. Hence with probability at least $1 - e^{-2C_{D'}t_0}$ (for an appropriate choice of a constant $C_{D'}$ that only depends on the degree distribution D' and the choice of r), the overall process does not go extinct.

Our goal is to find the choice of r with minimum product of degrees $\log(\Lambda_{G_T})/(T/2)$ that survives. We define r_1 as follows:

$$r_1 \equiv \operatorname*{arg\,min}_{r \in \mathcal{R}_{\alpha,\beta}(\epsilon)} \quad (1-r)\log(1-f_1) + r\log(1-f_2).$$

Since $\mathcal{R}_{\alpha,\beta}(\epsilon)$ is just an interval and we are minimizing a linear function with a positive slope, the optimal solution is $r_1 = \inf_{r\in\mathcal{R}_{\alpha,\beta}(\epsilon)} r$. This is a choice that survives with high probability and has the minimum product of degrees. Precisely, with probability at least $1 - e^{-C_{D'}T}$, where $C_{D'}$ depends on D' and ϵ, we have that

$$\frac{\log(\Lambda_{G_T})}{T/2} \le \langle [1-r_1, r_1], \log(f-1)\rangle + \frac{c\log(t_0)}{t_0}\log\left(f_2 - 1\right)$$

where we define the standard inner product $\langle[1-r_1, r_1], \log(\boldsymbol{f}-1)\rangle \triangleq (1-r_1)\log(f_1-1) + r_1\log(f_2-1)$. It follows that

$$\begin{aligned}&\frac{\log(\Lambda_{G_T})}{T/2} - \langle[1-r^*, r^*], \log(\boldsymbol{f}-1)\rangle \le\\ &(r_1 - r^*)\log\left(\frac{f_2-1}{f_1-1}\right) + \frac{c\log(t_0)}{t_0}\log\left(f_2-1\right)\end{aligned} \tag{24}$$

By setting ϵ small enough and t_0 large enough, we can make this as small as we want. For any given $\delta > 0$, there exists a positive $\epsilon > 0$ such that the first term is bounded by $\delta/2$. Further, recall that $T/2 = c\log(t_0) + t_0$. For any choice of ϵ, there exists a $t_{D',\epsilon}$ such that for all $T \ge t_{D',\epsilon}$ the vanishing term in Eq. (22) is smaller than ϵ. For any given $\delta > 0$, there exists a positive $t_{D',\delta}$ such that $T \ge t_{D',\delta}$ implies that the second term is upper bounded by $\delta/2$. Putting everything together (and setting ϵ small enough for the target δ), we get that

$$\mathbb{P}\Big(\frac{\log(\Lambda_{G_T})}{T/2} \ge \langle[1-r^*, r^*], \log(\boldsymbol{f}-1)\rangle + \delta\Big) \;\le\; e^{-C_{D',\delta}T} \tag{25}$$

for all $T \ge C'_{D',\delta}$, where $C_{D',\delta}$ and $C'_{D',\delta}$ are positive constants that only depend on the degree distribution D' and the choice of $\delta > 0$.

For the lower bound, we define the following set of r such that Eq. (22) is strictly negative:

$$\overline{\mathcal{R}}_{\alpha,\beta}(\epsilon) \;=\; \big\{\, r \mid \log(\mu_D) \le D_{\mathrm{KL}}(r\|\beta/\mu_D) - \epsilon \,\big\}\,. \tag{26}$$

Choosing $r \in \overline{\mathcal{R}}_{\alpha,\beta}(\epsilon)$ causes extinction with probability approaching 1. Explicitly, $\mathbb{P}(Z^{(t)} \ne 0)$ is the probability of non-extinction at time t, and $\mathbb{P}(Z^{(t)} \ne 0) \le \mathbb{E}[Z^{(t)}]$. By Equation (22), we have

$$\mathbb{E}[Z^{(t)}] \quad\le\quad e^{t(\log(\mu_D) - D_{\mathrm{KL}}(r\|\beta/\mu_D) + o(t))}$$

where $\log(\mu_D) - D_{\mathrm{KL}}(r\|\beta/\mu_D) \le -\epsilon$. The probability of extinction is therefore at least $1 - \mathbb{E}[Z^{(t)}] \ge 1 - e^{-t(\epsilon + o(t))}$. So defining

$$r_2 \equiv \operatorname*{arg\,max}_{r \in \overline{\mathcal{R}}_{\alpha,\beta}(\epsilon)} \quad (1-r)\log(1-f_1) + r\log(1-f_2),$$

we have

$$\frac{\log(\Lambda_{G_T})}{T/2} \ge \langle[1-r_2, r_2], \log(\boldsymbol{f}-1)\rangle + \frac{c\log(t_0)}{t_0}\log(f_1-1)$$

with probability at least $1 - e^{-C_{D',2}T}$ where $C_{D',2}$ is again a constant that depends on D' and ϵ. It again follows that

$$\begin{aligned}&\frac{\log(\Lambda_{G_T})}{T/2} - \langle[1-r^*, r^*], \log(\boldsymbol{f}-1)\rangle \ge\\ &(r_2 - r^*)\log\left(\frac{f_2-1}{f_1-1}\right) + \frac{c\log(t_0)}{t_0}\log\left(f_1-1\right)\;,\end{aligned} \tag{27}$$

where $r_2 - r^*$ is strictly negative. Again, for any given $\delta > 0$, there exists a positive $\epsilon > 0$ such that the first term is lower bounded by $-\delta/2$, and for any choice of ϵ, there exists a $t_{D',\epsilon}$ such that for all $T \ge t_{D',\epsilon}$ the vanishing term in Eq. (22) is smaller than ϵ. Note that this ϵ might be different from the one used to show the upper bound. We ultimately choose the smaller of the two ϵ values. For any given $\delta > 0$, there exists a positive $t_{D',\delta}$ such that $T \ge t_{D',\delta}$ implies that the second term is lower bounded by $-\delta/2$. Putting everything together (and setting ϵ small enough for the target δ), we get that

$$\mathbb{P}\Big(\frac{\log(\Lambda_{G_T})}{T/2} \le \langle[1-r^*, r^*], \log(\boldsymbol{f}-1)\rangle - \delta\Big) \;\le\; e^{-C_{D',\delta}T} \tag{28}$$

for all $T \ge C'_{D',\delta}$, where $C_{D',\delta}$ and $C'_{D',\delta}$ are positive constants that only depend on the degree distribution D' and the choice of $\delta > 0$. This gives the desired result.

We now address the general case for D with support greater than two. We follow the identical structure of the argument.

The first major difference is that node types are no longer scalar, but tuples. Each node v's type ξ_v is the $(\eta-1)$-tuple listing how many nodes in the path $\phi(w,v)\setminus\{v\}$ had each non-minimum degree from f_2 to f_η, where w is the closest node to v such that $\delta_H(w,v_T)\le c\log(t_0)$. Consequently, the threshold $\boldsymbol{r}=[r_1,\ldots,r_{\eta-1}]$ is no longer a scalar, but a vector-valued, pointwise threshold on each element of ξ_v. We let $\boldsymbol{k}=[k_1=r_1t_0,\ldots,k_{\eta-1}=r_{\eta-1}t_0]$ denote the time-dependent threshold, and we say $\boldsymbol{k}<\xi_v$ if $k_i<(\xi_v)_i$ for $1\le i\le\eta-1$. The matrix M is no longer second-order, but a tensor. Equation (19) still holds, except M is replaced with its tensor representation. For brevity, let $\alpha=p_1(f_1-1)$ and $\beta_i=p_{i+1}(f_{i+1}-1)$. Let $\tilde\beta=\sum_{i=1}^{\eta-1}\beta_i$. Hence, Equation (20) gets modified as

$$\mathbb{E}[Z^{(t)}]=\sum_{i_1=0}^{k_1}\cdots\sum_{i_{\eta-1}=0}^{k_{\eta-1}}\binom{t}{i_1,\ldots,i_{\eta-1}}\alpha^{t-\tilde\beta}\beta_1^{i_1}\cdots\beta_{\eta-1}^{i_{\eta-1}}. \tag{29}$$

Now we consider a *multinomial* variable W with parameters β_i/μ_D for $1\le i\le\eta-1$ and t trials. As before, equation (29) can equivalently be written as

$$\begin{aligned}\mathbb{E}[Z^{(t)}] &= \mu_D^t\,\mathbb{P}(W\le\boldsymbol{k})\\ &= \mu_D^t\,\exp\Big\{-t\,D_{\mathrm{KL}}\Big(\boldsymbol{r}\,\|\,\Big(\frac{\boldsymbol{\beta}}{\mu_D}\Big)\Big)+o(t)\Big\}\end{aligned} \tag{30}$$

where $\boldsymbol{\beta}/\mu_D$ denotes elementwise division. Here $D_{\mathrm{KL}}(\boldsymbol{r}\|\boldsymbol{\beta}/\mu_D)$ denotes the Kullback-Leibler divergence between the two generalized Bernoulli distributions defined by $\boldsymbol{r}$ and $\boldsymbol{\beta}/\mu_D$, such that $D_{\mathrm{KL}}(\boldsymbol{r}\|\boldsymbol{\beta}/\mu_D)=(1-\sum r_i)\log((1-\sum r_i)/(\alpha/\mu_D))+\sum_i r_i\log(r_i/(\beta_i/\mu_D))$. Once again, we wish to obtain bounds on $\mathbb{P}(W\le\boldsymbol{k})$. As before, we define the following set of r such that Eq. (30) is strictly positive, for some $\epsilon>0$:

$$\mathcal{R}_{\alpha,\boldsymbol{\beta}}(\epsilon)\;=\;\Big\{\,\boldsymbol{r}\;\Big|\;\log(\mu_D)\ge D_{\mathrm{KL}}(\boldsymbol{r}\|\Big(\frac{\boldsymbol{\beta}}{\mu_D}\Big))+\epsilon\,\Big\}\,, \tag{31}$$

We now choose a threshold $\boldsymbol{r}\in\mathcal{R}_{\alpha,\beta}(\epsilon)$. Using the same argument as before, we can construct a process where the probability of extinction is asymptotically zero. We again do not prune any leaves in the first $c\log(t_0)$ generations. This ensures that with high probability, there are at least $e^{C_\ell\log(t_0)}$ independent multi-type Galton-Watson processes evolving concurrently after time $c\log(t_0)$, each with probability of extinction $\theta_{D'}$. Hence with probability at least $1-e^{-2C_{D'}t_0}$ (for an appropriate choice of a constant $C_{D'}$ that only depends on the degree distribution D' and the choice of $\boldsymbol{r}$), the overall process does not go extinct.

We define $\boldsymbol{r}_1$ analogously to the $\eta=2$ case:

$$\boldsymbol{r}_1\equiv\mathop{\arg\min}_{\boldsymbol{r}\in\mathcal{R}_{\alpha,\boldsymbol{\beta}}(\epsilon)}\quad\langle\boldsymbol{r},\log(\boldsymbol{f}-1)\rangle\,,$$

where we now define $\langle\boldsymbol{r},\log(\boldsymbol{f}-1)\rangle\equiv(1-\sum_i r_i)\log(f_1-1)+\sum_{j=1}^{\eta-1}r_j\log(f_{j+1}-1)$. Therefore with probability at least $1-e^{-C_{D'}T}$, where $C_{D'}$ depends on D' and ϵ, we have that

$$\frac{\log(\Lambda_{G_T})}{T/2}\le\langle\boldsymbol{r}_1,\log(\boldsymbol{f}-1)\rangle+\frac{c\log(t_0)}{t_0}\log\left(f_\eta-1\right).$$

It follows that

$$\frac{\log(\Lambda_{G_T})}{T/2}-\langle\boldsymbol{r}^*,\log(\boldsymbol{f}-1)\rangle\le$$
$$\sum_{j=1}^{\eta-1}((r_1)_j-r_j^*)\log\left(\frac{f_{j+1}-1}{f_1-1}\right)+\frac{c\log(t_0)}{t_0}\log\left(f_\eta-1\right)\,. \tag{32}$$

By setting ϵ small enough and t_0 large enough, we can make this as small as we want. For any given $\delta>0$, there exists a positive $\epsilon>0$ such that each term in the summation in (32) is bounded by δ/η. Further, recall that $T/2=c\log(t_0)+t_0$. For any choice of ϵ, there exists a $t_{D',\epsilon}$ such that for all $T\ge t_{D',\epsilon}$ the vanishing term in Eq. (22) is smaller than ϵ. For any given $\delta>0$, there exists a positive $t_{D',\delta}$ such that $T\ge t_{D',\delta}$ implies that the second term of (32) is upper bounded by δ/η. Putting everything together (and setting ϵ small enough for the target δ), we get that

$$\mathbb{P}\Big(\frac{\log(\Lambda_{G_T})}{T/2}\ge\langle\boldsymbol{r}^*,\log(\boldsymbol{f}-1)\rangle+\delta\Big)\;\le\;e^{-C_{D',\delta}T} \tag{33}$$

for all $T\ge C'_{D',\delta}$, where $C_{D',\delta}$ and $C'_{D',\delta}$ are positive constants that only depend on the degree distribution D' and the choice of $\delta>0$. Similar analysis proves the lower bound:

$$\mathbb{P}\Big(\frac{\log(\Lambda_{G_T})}{T/2}\le\langle\boldsymbol{r}^*,\log(\boldsymbol{f}-1)\rangle-\delta\Big)\;\le\;e^{-C_{D',\delta}T} \tag{34}$$

for all $T\ge C'_{D',\delta}$.This gives the desired result.

Acknowledgement

We would like to thank the anonymous reviewers for their helpful comments. This work has been supported by NSF NSF CISE award CCF-1422278 and SaTC award CNS-1527754.

7. REFERENCES

[1] Whisper, 2012. http://whisper.sh.

[2] Yik yak, 2013. http://www.yikyakapp.com/.

[3] Secret, 2014. https://www.secret.ly.

[4] Team blind, 2015. http://us.teamblind.com/.

[5] D. Acemoglu, A. Makhdoumi, A. Malekian, and A. E. Ozdaglar. Privacy-constrained network formation. MIT Department of Economics Working Paper, 2015.

[6] S. Bubeck, L. Devroye, and G. Lugosi. Finding adam in random growing trees. *arXiv preprint arXiv:1411.3317*, 2014.

[7] D. Chaum. The dining cryptographers problem: Unconditional sender and recipient untraceability. *Journal of cryptology*, 1(1), 1988.

[8] I. Clarke, O. Sandberg, B. Wiley, and T.W. Hong. Freenet: A distributed anonymous information storage and retrieval system. In *Designing Privacy Enhancing Technologies*, 2001.

[9] H. Corrigan-Gibbs and B. Ford. Dissent: accountable anonymous group messaging. In *CCS*. ACM, 2010.

[10] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[11] R. Dingledine, M.J. Freedman, and D. Molnar. The free haven project: Distributed anonymous storage service. In *Designing Privacy Enhancing Technologies*, 2001.

[12] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. Technical report, DTIC Document, 2004.

[13] G. Fanti, P. Kairouz, S. Oh, and P. Viswanath. Spy vs. spy: Rumor source obfuscation. In *SIGMETRICS Perform. Eval. Rev.*, volume 43, pages 271–284, 2015.

[14] Giulia Fanti, Peter Kairouz, Sewoong Oh, K. Ramchandran, and P. Viswanath. Hiding the rumor source. *arXiv preprint arXiv:1509.02849*, 2015.

[15] S. Feizi, K. Duffy, M. Kellis, and M. Medard. Network infusion to infer information sources in networks. 2014.

[16] V. Fioriti and M. Chinnici. Predicting the sources of an outbreak with a spectral technique. *arXiv:1211.2333*, 2012.

[17] M.J. Freedman and R. Morris. Tarzan: A peer-to-peer anonymizing network layer. In *Proc. CCS*. ACM, 2002.

[18] G. Giakkoupis, R. Guerraoui, A. Jégou, A. Kermarrec, and N. Mittal. Privacy-conscious information diffusion in social networks. pages 480–496, 2015.

[19] S. Goel, M. Robson, M. Polte, and E. Sirer. Herbivore: A scalable and efficient protocol for anonymous communication. Technical report, 2003.

[20] P. Golle and A. Juels. Dining cryptographers revisited. In *Advances in Cryptology-Eurocrypt 2004*, 2004.

[21] E. Gwynne and S. Bubeck. Asymptotic behavior of the eden model with positively homogeneous edge weights. *arXiv preprint arXiv:1508.05140*, 2015.

[22] T. E. Harris. *The theory of branching processes*. Courier Corporation, 2002.

[23] V. Jog and P. Loh. Persistence of centrality in random growing trees. *arXiv preprint arXiv:1511.01975*, 2015.

[24] V. Jog and P. Loh. Analysis of centrality in sublinear preferential attachment trees via the cmj branching process. *arXiv preprint arXiv:1601.06448*, 2016.

[25] W. Luo, W. Tay, and M. Leng. How to identify an infection source with limited observations. 2013.

[26] W. Luo, W. P. Tay, and M. Leng. Rumor spreading and source identification: A hide and seek game. *arXiv preprint arXiv:1504.04796*, 2015.

[27] E. A. Meirom, C. Milling, C. Caramanis, S. Mannor, A. Orda, and S. Shakkottai. Localized epidemic detection in networks with overwhelming noise. 2014.

[28] C. Milling, C. Caramanis, S. Mannor, and S. Shakkottai. Network forensics: Random infection vs spreading epidemic. In *SIGMETRICS*. ACM, 2012.

[29] C. Milling, C. Caramanis, S. Mannor, and S. Shakkottai. Detecting epidemics using highly noisy data. In *MobiHoc*, pages 177–186, 2013.

[30] P. C. Pinto, P. Thiran, and M. Vetterli. Locating the source of diffusion in large-scale networks. *Physical review letters*, 109(6):068702, 2012.

[31] B. A. Prakash, J. Vreeken, and C. Faloutsos. Spotting culprits in epidemics: How many and which ones? In *ICDM*, volume 12, pages 11–20, 2012.

[32] D. Shah and T. Zaman. Detecting sources of computer viruses in networks: theory and experiment. In *ACM SIGMETRICS Performance Evaluation Review*, volume 38, pages 203–214. ACM, 2010.

[33] D. Shah and T. Zaman. Finding rumor sources on random graphs. *arXiv preprint arXiv:1110.6230*, 2011.

[34] D. Shah and T. Zaman. Rumors in a network: Who's the culprit? *Information Theory, IEEE Transactions on*, 57:5163–5181, Aug 2011.

[35] S. Spencer and R. Srikant. On the impossibility of localizing multiple rumor sources in a line graph. *ACM SIGMETRICS Performance Evaluation Review*, 43(2):66–68, 2015.

[36] J. van den Hooff, D. Lazar, M. Zaharia, and N. Zeldovich. Vuvuzela: Scalable private messaging resistant to traffic analysis. In *Proceedings of the 25th Symposium on Operating Systems Principles*, pages 137–152. ACM, 2015.

[37] L. von Ahn, A. Bortz, and N.J. Hopper. K-anonymous message transmission. In *Proc. CCS*. ACM, 2003.

[38] Z. Wang, W. Dong, W. Zhang, and C.W. Tan. Rumor source detection with multiple observations: Fundamental limits and algorithms. In *ACM SIGMETRICS*, 2014.

[39] K. Zhu and L. Ying. A robust information source estimator with sparse observations. *arXiv preprint arXiv:1309.4846*, 2013.

APPENDIX

Protocol 1 Adaptive Diffusion [13]

```
Input: contact network $G = (V, E)$, source $v^*$, time $T$, degree $d$
Output: set of infected nodes $V_T$
$V_0 \leftarrow \{v^*\}$, $h \leftarrow 0$, $v_0 \leftarrow v^*$
$v^*$ selects one of its neighbors $u$ at random
$V_1 \leftarrow V_0 \cup \{u\}$, $h \leftarrow 1$, $v_1 \leftarrow u$
let $\mathcal{N}(u)$ represent $u$'s neighbors
$V_2 \leftarrow V_1 \cup \mathcal{N}(u) \setminus \{v^*\}$, $v_2 \leftarrow v_1$
$t \leftarrow 3$
for $t \leq T$ do
    $v_{t-1}$ randomly selects $u \in \mathcal{N}(v_{t-1}) \setminus \{v_{t-2}\}$
    $h \leftarrow h + 1$
    $v_t \leftarrow u$
    for all $v \in \mathcal{N}(v_t) \setminus \{v_{t-1}\}$ do
        Infection Message($G$,$v_t$,$v$,$V_t$)
        if $t + 1 > T$ then
            break
        Infection Message($G$,$v_t$,$v$,$V_t$)
    $t \leftarrow t + 2$
procedure Infection Message($G$,$u$,$v$,$V_t$)
    if $v \in V_t$ then
        for all $w \in \mathcal{N}(v) \setminus \{u\}$ do
            Infection Message($G$,$v$,$w$,$G_t$)
    else
        $V_t \leftarrow V_{t-2} \cup \{v\}$
```

A. PROOF OF THEOREM 3.1

We first analyze the probability of detection for any given estimator (see Eq. (39)); we then show that the estimator in (4) is a MAP estimator, maximizing this probability of detection. Finally, we show that using the MAP estimator in (4) gives the probability of detection in Eq. (5).

We begin with some definitions. Consider the following random process, in which we fix a source v^* and generate a

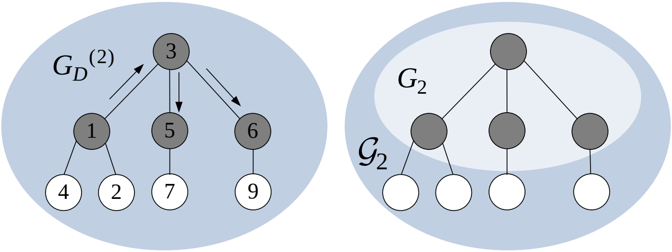

Figure 9: One realization of the random, irregular-tree branching process. Although each realization of the random process $G_D^{(t)}$ yields a labelled graph, the adversary observes G_T and $\mathcal{G}_T$, which are *unlabelled*. White nodes are uninfected, grey nodes are infected.

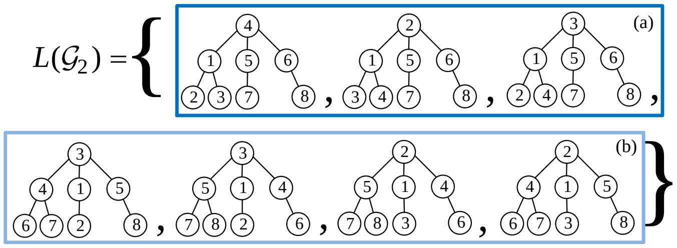

Figure 10: $L(\mathcal{G}_2)$ for the snapshot $\mathcal{G}_2$ illustrated in Figure 9. Boxes (a) and (b) illustrate the two families partitioning $L(\mathcal{G}_2)$.

(random) labelled tree $G_D^{(t)}$ for each time t and for a given degree distribution D. At time $t = 0$, $G_D^{(t)}$ consists of a single node v^*, which is given a label 1. The source v^* draws a degree d_1 from D, and generates d_1 child nodes, labelled in order of creation (i.e., 2 through $d_1 + 1$). At the next time step, $t = 1$, the source picks one of these neighbors uniformly at random to be the new virtual source and infects that neighbor. According to Protocol 1, each time a node v is infected, v draws its degree d_v from D, then generates $d_v - 1$ labelled child nodes. So at the end of time $t = 1$, $G_D^{(1)}$ contains the source and its uninfected neighbors, as well as the new virtual source and its uninfected neighbors. An example of $G_D^{(2)}$ is shown in Figure 9 (left panel) with $d_1 = 3$ and virtual source at node 3. Grey nodes are infected and white nodes are uninfected neighbors. Note that the node labelled 1 is always exactly one hop from a leaf of $G_D^{(t)}$ for all $t > 0$; also, nodes infect their neighbors in ascending order of their labels. The leaves of $G_D^{(t)}$ represent the uninfected neighbors of infected leaves in standard adaptive diffusion spreading over a given graph. Define $\Omega_{(t,D)}$ as the set of all labelled trees generated at time t according to this random process.

At some time T, the adversary observes the snapshot of infected subgraph G_T. Notice that we do not need to generate the entire contact network, since G_T is conditionally independent of the rest of the contact network given its one-hop neighbors. Hence, we only need to generate (and consider) the one hop neighbors of G_T at any given T. We use $\mathcal{G}_T$ to denote this random graph that includes G_T and its one hop neighbors as generated according to the previously explained random process. Notice that the adversary only observes $\mathcal{G}_T$, which is an *unlabelled* snapshot of the infection and its one hop neighbors (see Figure 9, right panel). We refer to the leaves of G_T as 'infected leaves', denoted by ∂G_T, and the leaves of $\mathcal{G}_T$ as 'uninfected leaves' denoted by $\partial \mathcal{G}_T$. Define

$$L(\mathcal{G}_T) \equiv \{\tilde{G} \in \Omega_{(T,D)} \mid U(\tilde{G}) = \mathcal{G}_T\},$$

i.e., the set of all labelled graphs (generated according to the described random process) whose unlabelled representation $U(\tilde{G})$ is equal to the snapshot $\mathcal{G}_T$. Figure 10 illustrates $L(\mathcal{G}_T)$ for the graph $\mathcal{G}_2$ in Figure 9.

We define a *family* $C_{\mathcal{G}_T,v} \subseteq L(\mathcal{G}_T)$ as the set of all labelled graphs whose labeling could have been generated by a breadth-first labeling of $\mathcal{G}_T$ starting at node $v \in \partial G_T$. Here, a breadth-first labeling is a valid order of traversal for a breadth-first search of $\mathcal{G}_T$ starting at node v. We restrict v to be a valid source for an adaptive diffusion spread—that is, it is an infected leaf in ∂G_T. Note that a BFS labeling starting from two different nodes on the unlabelled tree can yield the same labelled graph. In Figure 10, boxes (a) and (b) illustrate the two families contained in $L(\mathcal{G}_2)$.

Let $\mathbb{P}(C_{\mathcal{G},v}) \equiv \mathbb{P}(G_D^{(T)} \in C_{\mathcal{G},v})$ denote the probability that the labelled graph $G_D^{(T)}$ with snapshot $\mathcal{G}_T$ is generated from a node v. From the definition of the random process for generating labelled graphs, we get

$$\mathbb{P}(C_{\mathcal{G}_T,v}) = \underbrace{\left(\prod_{w \in G_T} \mathbb{P}_D(d_w)\right)}_{\text{degrees of } G} \underbrace{Q(\mathcal{G}_T, v)}_{\text{virtual sources}} \underbrace{|C_{\mathcal{G}_T,v}|}_{\substack{\text{count of} \\ \text{isomorphisms}}} \qquad (35)$$

where $\mathbb{P}_D(d)$ is the probability of observing degree d under degree distribution D, and

$$Q(\mathcal{G}_T, v) = \frac{\mathbb{1}_{v \in \partial G_T}}{d_v \prod_{w \in \Phi_{v,v_T} \setminus \{v, v_T\}} (d_w - 1)}$$

is the probability of passing the virtual source from v to the virtual source v_T given the structure of $\mathcal{G}_T$, where Φ_{v,v_T} is the unique path from v to v_T in $\mathcal{G}_T$. Eq. (35) holds because for all instances in $C_{\mathcal{G}_T,v}$, the probability of the degrees of the nodes and the probability of the path of the virtual source remain the same.

The probability of observing a given snapshot $\mathcal{G}_T$ is precisely $\mathbb{P}(G_D^{(T)} \in L(\mathcal{G}_T))$. Notice that $C_{\mathcal{G}_T,v}$ partitions $L(\mathcal{G}_T)$ in to family of labelled trees that are generated from the same source. This give the following decomposition:

$$\mathbb{P}(G_D^{(T)} \in L(G_T)) = \sum_{v \in \mathcal{C}_{\mathcal{G}_T}} \mathbb{P}(C_{\mathcal{G}_T,v}), \qquad (36)$$

where we define $\mathcal{C}_{\mathcal{G}_T}$ as the set of possible candidates of the source that generate distinct labelled trees, i.e.

$$\mathcal{C}_{\mathcal{G}_T} \equiv \{v \in G_T \,|\, C_{\mathcal{G}_T,v} \neq C_{\mathcal{G}_T,v'} \ \forall \ v' \in \mathcal{C}_{\mathcal{G}_T}, \ v' \neq v\} \,. \qquad (37)$$

Notice that this set is not unique, since there can be multiple nodes that represent the same family $C_{\mathcal{G}_T,v}$. We pick one of such node v to represent the class of nodes that can generate the same family of labelled trees. We use this v to index these families and not to denote any particular node in ∂G_T.

Consider an estimate of the source $\hat{v}(\mathcal{G}_T)$. In general, $\hat{v}(\mathcal{G}_T)$ is a random variable, potentially selected from a set of candidates. We define detection ($\overline{D}$) as the event in which

$\hat{v}(\mathcal{G}_T) = v_1(G_D^{(T)})$; i.e., the estimator outputs the node that started the random process. We can partition the set of candidate nodes ∂G_T, by grouping together those nodes that are indistinguishable to the estimator into *classes*. Precisely, we define a subset of nodes indexed by $v \in \mathcal{C}_{\mathcal{G}_T}$,

$$\chi_{\mathcal{G}_T,v} \equiv \{v' \in \partial G_T \,|\, C_{\mathcal{G}_T,v} = C_{\mathcal{G}_T,v'}\}\,. \tag{38}$$

For a given snapshot, there are as many classes as there are families. In Figure 10, the class associated with family (a) has one element—namely, the node labeled '1' in family (a). The class associated with family (b) contains two nodes: the node labeled '1' in family (b), and the node labeled '5' in the rightmost graph of family (b), since both nodes give rise to the same family.

We consider, without loss of generality, an estimator that selects a node in a given class with probability $\mathbb{P}(\hat{v}(\mathcal{G}_T) \in \chi_{\mathcal{G}_T,v})$. Notice that $|\chi_{\mathcal{G}_T,v}|$ denotes the number of (indistinguishable) source candidates in this class. From Eq. (36), the probability of detection given a snapshot is

$$\begin{aligned}\mathbb{P}(\overline{D}|\mathcal{G}_T) &= \frac{\mathbb{P}\left(G_D^{(T)} \in L(\mathcal{G}_T) \wedge \overline{D}\right)}{\mathbb{P}(G_D^{(T)} \in L(\mathcal{G}_T))}. \qquad (39)\\ &= \frac{\sum\limits_{v \in \mathcal{C}_{\mathcal{G}_T}} \mathbb{P}(C_{\mathcal{G}_T,v})\mathbb{P}\left(\overline{D}\,\middle|\, G_D^{(T)} \in C_{\mathcal{G}_T,v}\right)}{\sum_{v \in \mathcal{C}_{\mathcal{G}_T}} \mathbb{P}(C_{\mathcal{G}_T,v})} \qquad (40)\end{aligned}$$

where $\mathbb{P}(\overline{D}|G_D^{(T)} \in C_{\mathcal{G}_T,v}) = \mathbb{P}(\hat{v}(\mathcal{G}_T) \in \chi_{\mathcal{G}_T,v})/|\chi_{\mathcal{G}_T,v}|$. We use the following observation:

Lemma A.1.

$$\frac{\mathbb{P}(C_{\mathcal{G}_T,v})/|\chi_{\mathcal{G}_T,v}|}{\sum_{v' \in \mathcal{C}_{\mathcal{G}_T}} \mathbb{P}(C_{\mathcal{G}_T,v'})} = \frac{1}{d_{v_T} \prod\limits_{\substack{w \in \phi(v,v_T)\\ \setminus\{v,v_T\}}} (d_w - 1)}. \tag{41}$$

Proof of this lemma is omitted due to space limitations. We refer to a longer version of this paper for the proof of this lemma [14]. The above main technical lemma is derived from combinatorial counting of all possible graph relabelings that can result from a particular source. The counting involves counting the number of possible ways to run breadth first search, with different orderings of the offsprings of each node. Substituting Equation (41) into Equation (40), we get that

$$\mathbb{P}(\overline{D}|\mathcal{G}_T) \quad = \quad \sum_{v \in \mathcal{C}_{\mathcal{G}_T}} \frac{\mathbb{P}(\hat{v}(\mathcal{G}_T) \in \chi_{\mathcal{G}_T,v})}{d_{v_T} \prod\limits_{\substack{w \in \phi(v,v_T)\setminus\\ \{v,v_T\}}} (d_w - 1)}.$$

Since each term of this summation is bounded by

$$\frac{\mathbb{P}(\hat{v}(\mathcal{G}_T) \in \chi_{\mathcal{G}_T,v})}{d_{v_T} \prod\limits_{\substack{w \in \phi(v,v_T)\setminus\\ \{v,v_T\}}} (d_w - 1)} \le \frac{1}{\min\limits_{v \in \mathcal{C}_{\mathcal{G}_T}} d_{v_T} \prod\limits_{\substack{w \in \phi(v,v_T)\\ \setminus\{v,v_T\}}} (d_w - 1)},$$

and $\sum_{v \in \mathcal{C}_{\mathcal{G}_T}} \mathbb{P}(\hat{v}(\mathcal{G}_T) \in \chi_{\mathcal{G}_T,v}) = 1$, it must hold that

$$\mathbb{P}(\overline{D}|\mathcal{G}_T) \quad \le \quad \frac{1}{\min\limits_{v \in \mathcal{C}_{\mathcal{G}_T}} d_{v_T} \prod\limits_{\substack{w \in \phi(v,v_T)\\ \setminus\{v,v_T\}}} (d_w - 1)}.$$

This upper bound on the detection probability is achieved exactly if we choose weight $\mathbb{P}(\hat{v}(\mathcal{G}_T) \in \chi_{\mathcal{G}_T,v}) = 1$ for the class(es) minimizing the product $\prod_{w \in \phi(v,v_T)\setminus\{v,v_T\}}(d_w - 1)$, i.e.,

$$\hat{v}(G_T) = \arg\min_{v \in \partial G_T} \prod_{\substack{w \in \phi(v,v_T)\\ \setminus\{v,v_T\}}} (d_w - 1).$$