Local non-Bayesian social learning with stubborn agents

DANIEL VIAL, University of Michigan, USA
VIJAY SUBRAMANIAN, University of Michigan, USA

In recent years, people have increasingly turned to social networks like Twitter and Facebook for news. In contrast to traditional news sources, these platforms allow users to simultaneously read news articles and share opinions with other users. Among other effects, this has led to the rise of fake news, sometimes spread via bots (automated social media accounts masquerading as real users).

In this work, we devise and analyze a mathematical model describing such platforms. The model includes a large number of agents attempting to learn an underlying true state of the world in an iterative fashion. At each iteration, these agents update their beliefs about the true state based on noisy observations of the true state and the beliefs of a subset of other agents. These subsets may include a special type of agent we call bots, who attempt to convince others of an erroneous true state rather than learn (modeling users spreading fake news). This process continues for a finite number of iterations we call the learning horizon.

Our analysis details three cases for the outcome of this process: agents may learn the true state, mistake the erroneous state promoted by the bots as true, or believe the state falls between the true and erroneous states. Which outcome occurs depends on the relationship between the number of bots and the learning horizon. This leads to several interesting consequences; for example, we show that agents can initially learn the true state but later forget it and believe the erroneous state to be true instead. In short, we argue that varying the learning horizon can lead to dramatically different outcomes. This is in contrast to existing works studying models like ours, which typically fix a finite horizon or only consider an infinite horizon.

1 INTRODUCTION

With the rise of social networks like Twitter and Facebook, people increasingly receive news through non-traditional sources. One recent study shows that two-thirds of American adults have gotten news through social media [19]. Such news sources are fundamentally different than traditional ones like print media and television, in the sense that social media users read and discuss news on the same platform. As a consequence, users turning to these platforms for news receive information not only from major publications but from others users as well; in the words of [4], a user "with no track record or reputation can in some cases reach as many readers as Fox News, CNN, or the New York Times." This phenomenon famously reared its head during the 2016 United States presidential election, when fake news stories were shared tens of millions of times [4].

In this paper, we study a mathematical model describing this situation. The model includes a large number of agents attempting to learn an underlying true state of the world (e.g. which of two candidates is better suited for office) using information from three sources. First, each agent receives noisy observations of the true state, modeling e.g. news stories from major publications. Second, each agent observes the opinions of a subset of other agents, modeling e.g. discussions with other social media users. Third, each agent may observe the opinions of *stubborn agents* or *bots* who aim to persuade others of an erroneous true state, modeling e.g. users spreading fake news.[1] Based on this information, agents iteratively update their beliefs about the true state in a manner similar to the non-Bayesian social learning model of Jadbabaie *et al.* [16]. This iterative process continues for a finite number of iterations that we refer to as the learning horizon.

[1]The term *stubborn agents* has been used in the social learning and consensus literature to describe such agents; the term *bots* is used in reference to automated social media accounts spreading fake news while masquerading as real users [18].

Authors' addresses: Daniel Vial, University of Michigan, Ann Arbor, MI, USA, dvial@umich.edu; Vijay Subramanian, University of Michigan, Ann Arbor, MI, USA, vgsubram@umich.edu.

Under this model, two competing forces emerge as the learning horizon grows. On the one hand, agents receive more observations of the true state, suggesting that they become more likely to learn. On the other hand, the influence of bots gradually propagates through the system, suggesting that agents become increasingly susceptible to this influence and thus less likely to learn. Hence, while the horizon clearly affects the learning outcome, the nature of this effect – namely, whether learning becomes more or less likely as the horizon grows – is less clear.

This effect of the learning horizon has often been ignored in works with models similar to ours. For example, our model is nearly identical to that in the empirical work [5], in which the authors show that polarized beliefs can arise when there are two types of bots with diametrically opposed viewpoints. However, the experiments in [5] simply fix a large learning horizon and do not consider the effect of varying it. Models similar to ours have also been treated analytically; for example, [3, 14, 16] study non-Bayesian learning models similar to ours. However, these works consider a fixed number of agents and an infinite learning horizon and thus also ignore timescale effects.

The main message of this work is that *the learning horizon plays a prominent role in the learning outcome and therefore should not be ignored.* In particular, we show that the learning outcome depends on the relationship between the horizon T_n and a quantity p_n that describes the number of bots in the system, where both quantities may depend on the number of agents n. Mathematically, letting $\theta \in (0,1)$ denote the true state and $\theta_{T_n}(i^*)$ denote the belief about the true state for a uniformly random agent i^* at the horizon T_n, we show (see Theorem 3.1)

$$\theta_{T_n}(i^*) \xrightarrow[n\to\infty]{\mathbb{P}} \begin{cases} \theta, & T_n(1-p_n) \xrightarrow[n\to\infty]{} 0 \\ 0, & T_n(1-p_n) \xrightarrow[n\to\infty]{} \infty \end{cases}. \tag{1}$$

Here p_n is smaller when more bots are present and 0 is the erroneous true state promoted by the bots. Hence, in words, (1) says the following: if there are sufficiently few bots, in the sense that $T_n(1-p_n) \to 0$, i^* learns the true state; if there are sufficiently many bots, in the sense that $T_n(1-p_n) \to \infty$, i^* adopts the extreme belief 0 promoted by the bots. An interesting consequence arises from taking $T_{n,1} = o(1/(1-p_n))$ and $T_{n,2} = \omega(1/(1-p_n))$, so that

$$\theta_{T_{n,1}}(i^*) \xrightarrow[n\to\infty]{\mathbb{P}} \theta, \quad \theta_{T_{n,2}}(i^*) \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

In words, a typical agent initially (at time $T_{n,1}$) learns the true state θ, then later (at time $T_{n,2}$) "forgets" the true state and adopts the extreme opinion 0! Hence, as illustrated by this example, the chosen learning horizon can lead to drastically different outcomes.

In addition to (1), we prove two other results. First, we consider a regime that falls between the two in (1); namely, the case $T_n(1-p_n) \to c \in (0,\infty)$ (note larger c implies more bots). In this case, the learning outcome depends on another parameter, which we denote by $\eta \in (0,1)$ and which dictates the weight agents place on other agents' opinions in their belief updates. Here we show

$$\theta_{T_n}(i^*) \xrightarrow[n\to\infty]{\mathbb{P}} \theta(1-e^{-c\eta})/(c\eta). \tag{2}$$

The limit in (2) is depicted graphically as a function of c in Figure 1, which offers an intuitive interpretation: if an adversary deploys bots in hopes of driving agent opinions to 0, the marginal benefit of deploying additional bots is smaller when c is larger (i.e. when more bots have been deployed). In other words, the adversary experiences "diminishing returns". It is also interesting to fix c and consider the limit in (2) as a function of η. As $\eta \to 0$, agents place less weight on others' opinions, and this limit tends to θ; in other words, when agents ignore the network (and thus the harmful effects of bots), they learn. On the other hand, as $\eta \to 1$, agents have increased exposure to bots and the limit tends to $(1-e^{-c})/c$ (interestingly, there is a discontinuity here, i.e. the limiting

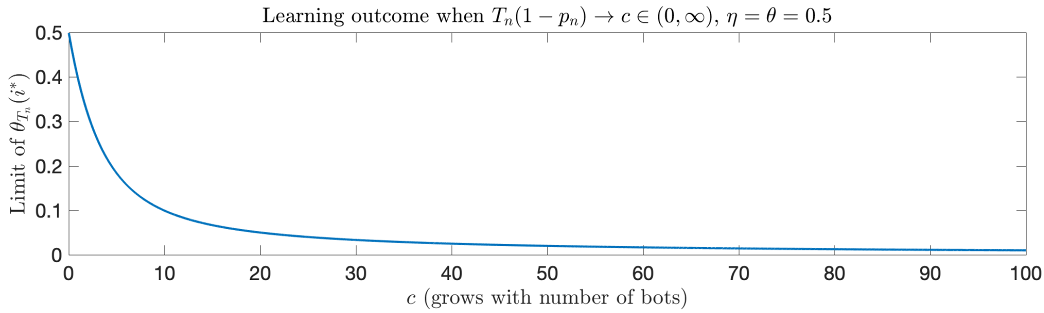

Fig. 1. Graphical depiction of (2).

belief does not fully reach zero as $\eta \to 1$). Finally, in addition to (1) and (2), we show in Theorem 3.2 that all but $O(n^k)$ agents adopt opinion 0 in a certain sub-case of $T_n(1-p_n) \to \infty$, for some $k < 1$. Hence, Theorem 3.2 is stronger than Theorem 3.1 and applies to fewer cases; we also argue empirically that this stronger result likely fails in other cases.

Before proceeding, we note our particular choice of T_n guarantees that at the learning horizon, an agent's belief is only influenced by a vanishing fraction of other agents and bots; specifically, those within the agent's neighborhood in the graph connecting agents and bots (see Section 3.4). This is why the title of our work refers to the learning as "local". More specifically, our choice of T_n is asymptotically dominated by the mixing time of the random walk on this graph. From an analysis perspective, this means we cannot leverage global properties like the stationary distribution of this walk, in contrast to many works on social learning (see Section 4). In fact, as shown in [8], this random walk exhibits *cutoff*, meaning that at our learning horizon, the distribution of this walk can be maximally far from stationarity. Hence, we cannot even use an approximation of the stationary distribution. Instead, we assume the graph is randomly generated in a manner that guarantees a well-behaved local structure; we then show that analyzing beliefs amounts to analyzing hitting probabilities of random walks on this random graph. Fundamentally, it is from three regimes of these hitting probabilities that the three regimes in (1) and (2) arise (see Section 3.2).

The remainder of the paper is organized as follows. In Section 2, we define the model studied throughout the paper. We then state and discuss our theoretical results and provide some empirical results in Section 3. Finally, we discuss related work in Section 4 and conclude in Section 5.

Notational conventions: Most notation is standard or defined as needed, but we note here that the following conventions are used frequently. For $n \in \mathbb{N}$, we let $[n] = \{1, 2, \ldots, n\}$, and for $n, k \in \mathbb{N}$ we let $k+[n] = [n]+k = \{k+1, k+2, \ldots, k+n\}$. All vectors are treated as row vectors. We let e_i denote the vector with 1 in the i-th position and 0 elsewhere. We denote the set of nonnegative integers by $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. We use $1(A)$ for the indicator function, i.e. $1(A) = 1$ if A is true and 0 otherwise. All random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\mathbb{E}[\cdot] = \int_\Omega \cdot \, d\mathbb{P}$ denoting expectation, $\xrightarrow{\mathbb{P}}$ denoting convergence in probability, and $a.s.$ meaning $\mathbb{P}-$almost surely.

2 MODEL

2.1 Learning model

We begin by defining the model of social learning studied throughout the paper. The main ingredients are (1) a true state of the world, represented as a scalar, (2) a social network connecting two sets of nodes, some who aim to learn the true state and some who wish to persuade others of an erroneous true state, and (3) a learning horizon. We discuss each of these ingredients in turn.

The true state of the world is a constant $\theta \in (0, 1)$. For example, in an election between candidates representing two political parties (say, Party 1 and Party 2), $\theta \approx 0$ can be interpreted as the Party 1 candidate being far superior, $\theta \approx 1$ means the Party 2 candidate is far superior, and $\theta \approx 0.5$ implies the candidates are roughly equal. We emphasize that θ is a deterministic constant and does not depend on time, nor on the number of nodes in the system (both of which will vary).

A directed graph $G = (A \cup B, E)$ connects disjoint sets of nodes A and B (details regarding the graph structure are discussed in Section 2.2). We refer to elements of A as *regular agents*, or simply *agents*, and elements of B as *stubborn agents* or *bots*. While agents attempt to learn the true state θ, bots aim to disrupt this learning and convince agents that the true state is instead 0. In the election example, agents represent voters who study the two candidates to learn which is superior, while bots are loyal to Party 1 and aim to convince agents that the corresponding candidate is superior (despite possible evidence to the contrary). Edges in the graph represent connections in a social network over which nodes share opinions in a manner that will be described shortly. An edge from node i to node j, denoted $i \rightarrow j$, will be interpreted to mean that i influences j's opinion.

Agents and bots share opinions until a learning horizon $T \in \mathbb{N}$. We will allow the horizon to depend on the number of agents $n \triangleq |A|$ and will thus denote it by T_n at times. In the election example, T represents the duration of the election season, i.e. the number of time units that agents can learn about the candidates and that bots can attempt to convince agents of the superiority of the Party 1 candidate. Here T_n will be finite for each finite n, and we will let T_n tend to infinity with n. In particular, we will choose T_n such that an agent's opinion at time T_n only depends on the opinions of a vanishing fraction of all agents and bots; namely, those within the agent's T_n-step incoming neighborhood in the social network G (see Section 3.4 for details).

It remains to specify how agents attempt to learn and how bots aim to disrupt this learning. We begin with the agents. Initially, $i \in A$ believes the state to be $\theta_0(i) = \alpha_0(i)/(\alpha_0(i) + \beta_0(i))$, where $\alpha_0(i) \in [0, \bar{\alpha}]$ and $\beta_0(i) \in [0, \bar{\beta}]$ for some $\bar{\alpha}, \bar{\beta} \in (0, \infty)$ that do not depend on n (if $\alpha_0(i) = \beta_0(i) = 0$, we let $\theta_0(i) = 0.5$ by convention). We refer to $\alpha_0(i), \beta_0(i)$ as the *prior parameters* and will not specify them beyond assuming they lie in the aforementioned intervals.[2] In our running example, the initial belief $\theta_0(i)$ can encode i's past opinions regarding the political parties, e.g. $\theta_0(i) < 0.5$ means i historically prefers Party 1 and is predisposed towards the corresponding candidate before the election season begins. At each time $t \in [T]$, $i \in A$ receives a noisy observation of the true state (e.g. i reads a news story regarding the candidates) and modifies its opinion based on this observation and on the opinions of its incoming neighbors in G (e.g. i discusses the election with its social connections). Mathematically, $i \in A$ updates its belief as $\theta_t(i) = \alpha_t(i)/(\alpha_t(i) + \beta_t(i))$, where

$$\alpha_t(i) = (1 - \eta)(\alpha_{t-1}(i) + s_t(i)) + \frac{\eta}{d_{in}(i)} \sum_{j \in N_{in}(i)} \alpha_{t-1}(j), \tag{3}$$

$$\beta_t(i) = (1 - \eta)(\beta_{t-1}(i) + (1 - s_t(i))) + \frac{\eta}{d_{in}(i)} \sum_{j \in N_{in}(i)} \beta_{t-1}(j).$$

Here $s_t(i) \sim \text{Bernoulli}(\theta)$ is the noisy observation of the true state, $N_{in}(i) \subset A \cup B$ is i's incoming neighbor set in G, $d_{in}(i) = |N_{in}(i)|$, and $\eta \in (0, 1)$ is a constant (independent of agent i and time t). We note that, as η grows, the effect of the network becomes stronger (i.e. the opinions of agent i's neighbors have a stronger effect on i's own opinion); this will be reflected in our results. Also, as discussed in Section 2.2, we will assume $d_{in}(i) > 0 \ \forall \ i \in A$, so (3) is well-defined.

Before discussing the bots, we comment further on the belief update (3). First, assuming $\eta = \alpha_0(i) = \beta_0(i) = 0$ temporarily, we simply have $\theta_t(i) = \sum_{\tau=1}^{t} s_\tau(i)/t$, which is an unbiased estimate of the true state θ. Next, if we drop the assumption $\alpha_0(i) = \beta_0(i) = 0$ (but still assume $\eta = 0$), $\theta_t(i)$

[2] Appendix A.1 shows the effect of the prior parameters vanishes when $T_n \rightarrow \infty$ with n, so specifying them is unnecessary.

is no longer an unbiased estimate. Instead, we can view $\theta_t(i)$ as the mean of a beta distribution with parameters $\alpha_t(i), \beta_t(i)$; in this case, (3) is simply a Bayesian update of the prior distribution Beta$(\alpha_{t-1}(i), \beta_{t-1}(i))$ with a Bernoulli(θ) signal. Finally, dropping the assumption $\eta \neq 0$ to obtain the model we actually consider, (3) is no longer a Bayesian update, as alluded to by the title of our work. This non-Bayesian model is closely related to others in the literature; see Section 4.

Having specified the behavior of agents, we turn to the bots. For $i \in B$, we simply set

$$\alpha_t(i) = 0, \quad \beta_t(i) = \bar{\beta} + (1-\eta)t \quad \forall\, t \in [T]. \tag{4}$$

Hence, the opinion of $i \in B$ is $\theta_t(i) = \alpha_t(i)/(\alpha_t(i) + \beta_t(i)) = 0$, e.g. bots believe the candidate from Party 1 is far superior. To explain the precise form of (4), consider a system composed of only agents (i.e. $B = \emptyset$). Since $\beta_0(i) \leq \bar{\beta}, s_t(i) \geq 0\ \forall\ i \in A$, it is easy to show via (3) that $\beta_t(i) \leq \bar{\beta} + (1-\eta)t$ and $\alpha_t(i) \geq 0\ \forall\ i \in A, t \in [T]$. Hence, not only are bots biased towards state 0, but their bias is maximal, in the sense that their parameters $\alpha_t(i), \beta_t(i)$ are as extreme as an agent's can be.

Note that we can define the bot behavior in an alternative way that will be more convenient for our analysis. Specifically, for $i \in B$, we can set $N_{in}(i) = \{i\}$ (i.e. i has a self-loop and no other incoming edges in G), $\alpha_0(i) = 0$, $\beta_0(i) = \bar{\beta}$, and $s_t(i) = 0\ \forall\ t \in [T]$. Then, assuming $i \in B$ updates its parameters via (3), it is straightforward to show (4) holds. This alternative definition will be used for the remainder of the paper. Finally, since all bots $i \in B$ have the same behavior, we assume (without loss of generality) that the outgoing neighbor set of $i \in B$ is $N_{out}(i) = \{i, i'\}$ for some $i' \in A$, i.e. in addition to its self-loop, each bot has a single outgoing neighbor from the agent set.

2.2 Graph model

Having defined our learning model, we next specify how the social network G is constructed. For this, we use a modification of a well-studied random graph model called the *directed configuration model* (DCM) [10]. The DCM is a means of constructing a graph with a specified degree sequence; our modification is needed to account for the distinct node types at hand (agents and bots).

To begin, we realize a random sequence $\{d_{out}(i), d_{in}^A(i), d_{in}^B(i)\}_{i \in A}$ called the *degree sequence* from some distribution; here we let $A = [n]$. In the construction described next, $i \in A$ will have $d_{out}(i)$ outgoing neighbors (i will influence $d_{out}(i)$ other agents), $d_{in}^A(i)$ incoming neighbors from the A (i will be influenced by $d_{in}^A(i)$ agents), and $d_{in}^B(i)$ incoming neighbors from B (i will be influenced by $d_{in}^B(i)$ bots). Here the total in-degree of i is $d_{in}(i) = d_{in}^A(i) + d_{in}^B(i)$ (as used in (4)). We assume

$$d_{out}(i), d_{in}^A(i) \in \mathbb{N}, d_{in}^B(i) \in \mathbb{N}_0\ \forall\ i \in A, \quad \sum_{i \in A} d_{out}(i) = \sum_{i \in A} d_{in}^A(i). \tag{5}$$

In words, the first condition says i influences and is influenced by at least one agent, and may be influenced by one or more bots. The second condition says sum out-degree must equal sum in-degree in the agent sub-graph; this will be necessary to construct a graph with the given degrees.

After realizing the degree sequence, we begin the graph construction.[3] First, we attach $d_{out}(i)$ outgoing half-edges, $d_{in}^A(i)$ incoming half-edges labeled A, and $d_{in}^B(i)$ incoming half-edges labeled B, to each $i \in A$; we will refer to these half-edges as *outstubs*, *A-instubs*, and *B-instubs*, respectively. We let O_A denote the set of these outstubs. We then pair each outstub in O_A with an A-instub to form edges between agents in an iterative, breadth-first-search fashion that proceeds as follows:

- Sample i^* from A uniformly. For each the $d_{in}^A(i^*)$ A-instubs attached to i^*, sample an outstub uniformly from O_A (resampling if the sampled outstub has already been paired), and connect the instub and outstub to form an edge from some agent to i^*.
- Let $A_1 = \{i \in A \setminus \{i^*\} :$ an outstub of i was paired with an A-instub of $i^*\}$. For each $i \in A_1$, pair the $d_{in}^A(i)$ A-instubs attached to i in the same manner the A-instubs of i^* were paired.

[3] This construction is presented more formally as Algorithm 1 in Appendix A.1.

- Continue iteratively until all A-instubs have been paired. In particular, during the l-th iteration, we pair all A-instubs attached to A_l, the set of agents at distance l from i^* (by distance l, we mean a path of length l exists, but no shorter path exists).

At the conclusion of this procedure, we obtain a graph with edges between agents, along with unpaired B-instubs attached to some agents. It remains to attach these B-instubs to bots. For this, we define $B = n + [\sum_{i \in A} d_{in}^B(i)]$ to be the set of bots (hence, the node set is $A \cup B = [n + \sum_{i \in A} d_{in}^B(i)]$). To each $i \in B$ we add a single self-loop and a single unpaired outstub (as described at the end of Section 2.1). This yields $\sum_{i \in A} d_{in}^B(i)$ unpaired outstubs attached to bots. Finally, we pair these outstubs arbitrarily with the $\sum_{i \in A} d_{in}^B(i)$ unpaired B-instubs from above to form edges from bots to agents (note the exact pairing can be arbitrary since all bots behave the same, per Section 2.1).

Before proceeding, we note that the pairing of A-instubs with outstubs from O_A did not prohibit us from forming agent self-loops (i.e. edges $i \to i$ for $i \in A$), nor did it prohibit multiple edges from $i \in A$ to $i' \in A$. This second observation means the set of edges E formed will in general be a multi-set. For this reason, we re-define the parameter update equations (3) as

$$\alpha_t(i) = (1-\eta)(\alpha_{t-1}(i) + s_t(i)) + \eta \sum_{j \in A \cup B} \frac{|\{j' \to i' \in E : j' = j, i' = i\}|}{d_{in}(i)} \alpha_{t-1}(j), \tag{6}$$

$$\beta_t(i) = (1-\eta)(\beta_{t-1}(i) + (1 - s_t(i))) + \eta \sum_{j \in A \cup B} \frac{|\{j' \to i' \in E : j' = j, i' = i\}|}{d_{in}(i)} \beta_{t-1}(j),$$

i.e. we weigh the opinions of i's incoming neighbors proportional to the number of edges pointing to i. We also note that, instead of attaching bots to B-instubs after pairing all A-instubs as described above, we can pair B-instubs iteratively along with the pairing of A-instubs. Finally, we note that in the case $d_{in}^B(i) = 0 \ \forall \ i \in A$, the construction described above reduces to the standard DCM.

3 RESULTS

Having defined our model, we now turn to our results. We begin by defining the required assumptions in Section 3.1. We then state and discuss two theorems, one each in Sections 3.2 and 3.3. Finally, in Section 3.4, we return to comment on our assumptions.

3.1 Assumptions

To define the assumptions needed to prove our results, we require some notation. First, from the given degree sequence $\{d_{out}(i), d_{in}^A(i), d_{in}^B(i)\}_{i \in A = [n]}$, we define

$$f_n^*(i,j,k) = \frac{1}{n} \sum_{a=1}^{n} 1((d_{out}(a), d_{in}^A(a), d_{in}^B(a)) = (i,j,k)) \ \forall \ (i,j,k) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N}_0, \tag{7}$$

$$f_n(i,j,k) = \sum_{a=1}^{n} \frac{d_{out}(a)}{\sum_{a' \in A} d_{out}(a')} 1((d_{out}(a), d_{in}^A(a), d_{in}^B(a)) = (i,j,k)) \ \forall \ (i,j,k) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N}_0.$$

Here f_n^* and f_n, respectively, are the degree distributions for an agent sampled uniformly and sampled proportional to out-degree, respectively. Note that, since the first agent i^* added to the graph is sampled uniformly from A, the degrees of i^* are distributed as f_n^*. Furthermore, recall that, to pair A-instubs, we sample outstubs uniformly from O_A, resampling if the sampled outstub is already paired. It follows that, each time we add a new agent to the graph (besides i^*), its degrees are distributed as f_n. We also note that, because the degree sequence is random, these distributions are random as well. Using these random distributions, we also define the random variables

$$\tilde{p}_n^* = \sum_{j\in\mathbb{N}, k\in\mathbb{N}_0} \frac{j}{j+k} \sum_{i\in\mathbb{N}} f_n^*(i,j,k), \quad \tilde{p}_n = \sum_{j\in\mathbb{N}, k\in\mathbb{N}_0} \frac{j}{j+k} \sum_{i\in\mathbb{N}} f_n(i,j,k), \tag{8}$$
$$\tilde{q}_n = \sum_{j\in\mathbb{N}, k\in\mathbb{N}_0} \frac{j}{j+k} \frac{1}{j+k} \sum_{i\in\mathbb{N}} f_n(i,j,k).$$

Following the discussion above, $\tilde{p}_n^*$ is the expected value (conditioned on the degree sequence) of the ratio of A-instubs to total instubs for i^*; $\tilde{p}_n$ is the expected value of this same ratio, but for new agents added to the graph (besides i^*). The interpretation of $\tilde{q}_n$ is similar, i.e. the expected ratio of A-instubs to the square of total instubs for new agents (besides i^*). At the end of Section 3.2, we discuss in more detail why these random variables arise in our analysis.

We now define the following four assumptions, which are needed to establish our results. Two of these statements require the degree sequence to be well-behaved (with high probability) – specifically, (A1) requires certain moments of the degree sequence to be finite, while (A3) requires $\{\tilde{p}_n\}_{n\in\mathbb{N}}$ to be close to a deterministic sequence $\{p_n\}_{n\in\mathbb{N}}$. The other two statements, (A2) and (A4), impose maximum and minimum rates of growth for the learning horizon T_n. In particular, T_n must be finite for each finite n by (A2) and then grow to infinity with n by (A4), as mentioned in Section 2.1. We defer further discussion of these assumptions to Section 3.4.

(A1) $\lim_{n\to\infty} \mathbb{P}(\Omega_{n,1}) = 1$, where, for some $\nu_1, \nu_2, \nu_3, \gamma > 0$ independent of n with $\nu_3 > \nu_1$,
$$\Omega_{n,1} = \left\{ \left| \frac{\sum_{i=1}^n d_{out}(i)}{n} - \nu_1 \right| < n^{-\gamma}, \left| \frac{\sum_{i=1}^n d_{out}(i)^2}{n} - \nu_2 \right| < n^{-\gamma}, \left| \frac{\sum_{i=1}^n d_{out}(i) d_{in}^A(i)}{n} - \nu_3 \right| < n^{-\gamma} \right\}.$$
(A2) $\exists\, N \in \mathbb{N}$ and $\zeta \in (0, 1/2)$ independent of n s.t. $T_n \leq \zeta \log(n)/\log(\nu_3/\nu_1)\ \forall\, n \geq N$.
(A3) $\lim_{n\to\infty} \mathbb{P}(\Omega_{n,2}) = 1$, where, for some $p_n \in [0,1]$ s.t. $\lim_{n\to\infty} p_n = p \in [0,1]$, some $0 \leq \delta_n = o(1/T_n)$, and some $\xi \in (0,1)$ independent of n,
$$\Omega_{n,2} = \left\{ |p_n - \tilde{p}_n| < \delta_n, \tilde{p}_n^* \geq \tilde{p}_n, \tilde{q}_n < 1 - \xi \right\}.$$
(A4) $\lim_{n\to\infty} T_n = \infty$.

3.2 First result

We can now present our first result, Theorem 3.1. The theorem states that the belief at time T_n of a uniformly random agent converges in probability as $n \to \infty$. Interestingly, the limit depends on the relative asymptotics of the time horizon T_n and the quantity p_n defined in (A3). For example, this limit is θ when $T_n(1-p_n) \to 0$; note that $T_n(1-p_n) \to 0$ requires p_n to quickly approach 1 (since $T_n \to \infty$ by (A4)), which by (A3) and (8) suggests the number of bots is small. Hence, i^* learns the true state when there are sufficiently few bots. (The other cases can be interpreted similarly.)

Theorem 3.1. *Given (A1), (A2), (A3), and (A4), we have for $i^* \sim A$ uniformly,*

$$\theta_{T_n}(i^*) \xrightarrow[n\to\infty]{\mathbb{P}} \begin{cases} \theta, & T_n(1-p_n) \xrightarrow[n\to\infty]{} 0 \\ \theta(1-e^{-c\eta})/(c\eta), & T_n(1-p_n) \xrightarrow[n\to\infty]{} c \in (0,\infty) \\ 0, & T_n(1-p_n) \xrightarrow[n\to\infty]{} \infty \end{cases}.$$

Before discussing the proof of the theorem, we make several observations:

- Suppose p_n is fixed and consider varying T_n. To be concrete, let $p_n = 1 - (\log n)^{-1/2}$ and define $T_{n,1} = (\log n)^{1/4}$ and $T_{n,2} = (\log n)^{3/4}$ (note $T_{n,1}, T_{n,2}$ satisfy (A2), (A4)). Then $T_{n,1}(1-p_n) \to 0$ and $T_{n,2}(1-p_n) \to \infty$, so by Theorem 3.1, the belief of i^* converges to θ at time $T_{n,1}$ and to 0 at time $T_{n,2}$. In words, i^* initially (at time $(\log n)^{1/4}$) learns the state of the world, then later (at time $(\log n)^{3/4}$) forgets it and adopts the bot opinions!

- Alternatively, suppose T_n is fixed and consider varying p_n. For example, let $p_n = 1 - c/T_n$ for some $c \in (0, \infty)$. Here smaller c implies fewer bots, and Theorem 3.1 says the limiting belief of i^* is a decreasing convex function of c (see Figure 1). One interpretation is that, if an adversary deploys bots in hopes of driving agent beliefs to 0, the marginal benefit of deploying additional bots is smaller when c is larger, i.e. the adversary experiences "diminishing returns". It is also worth noting that, since $(1 - e^{-c\eta})/(c\eta) \to 1$ as $c \to 0$ and $(1 - e^{-c\eta})/(c\eta) \to 0$ as $c \to \infty$, the limiting belief of i^* is continuous as a function of c.
- If $T_n(1 - p_n) \to c \in (0, \infty)$, consider the limiting belief of i^* as a function of η. By Theorem 3.1, this belief tends to θ as $\eta \to 0$ and tends to $(1 - e^{-c})/c$ as $\eta \to 1$. This is expected from (6): when $\eta = 0$, agents ignore the network (and thus avoid exposure to biased bot opinions) and form opinions based only on unbiased signals; when $\eta = 1$, the opposite is true. Interestingly, though, there is an asymmetry here: when $\eta \to 0$, the belief approaches the $T_n(1 - p_n) \to 0$ case, but when $\eta \to 1$, it does *not* approach the $T_n(1 - p_n) \to \infty$ case (since $(1 - e^{-c})/c > 0$).
- If $p_n \to p < 1$, we must have $T_n(1 - p_n) \to \infty$ (since $T_n \to \infty$ by (A4)), and the belief of i^* tends to 0 by Theorem 3.1. Loosely speaking, this says that a necessary condition for learning is that the bots vanish asymptotically (in the sense that $p_n \to 1$).

The proof of Theorem 3.1 is lengthy; for readability, we outline it in Appendix A and defer computational details to Appendix B. However, we next present a short argument to illustrate the fundamental reason why the three cases of the limiting belief arise in Theorem 3.1. (As a disclaimer, this argument is not entirely precise; we refer the reader to the appendices for a rigorous proof.)

At a high level, these three cases arise as follows. First, when $T_n(1 - p_n) \to 0$, the "density" of bots within the T_n-step incoming neighborhood of i^* is small. As a consequence, i^* is not exposed to the biased opinions of bots by time T_n and is able to learn the true state (i.e. $\theta_{T_n}(i^*) \to \theta$ in $\mathbb{P}$). On the other hand, when $T_n(1 - p_n) \to \infty$, this "density" is large; i^* is exposed to bot opinions and thus adopts them (i.e. $\theta_{T_n}(i^*) \to 0$ in $\mathbb{P}$). Finally, when $T_n(1 - p_n) \to c \in (0, \infty)$, the "density" is moderate; i^* does not fully learn, nor does i^* fully adopt bot opinions (i.e. $\theta_{T_n}(i^*) \to \theta(1 - e^{-c\eta})/(c\eta)$ in $\mathbb{P}$).

The explanation of the previous paragraph is not at all surprising; what is more subtle is what precisely *density of bots within the T_n-step incoming neighborhood of i^** means. It turns out that the relevant quantity (and what we mean by this "density") is the probability that a random walker exploring this neighborhood reaches the set of bots.

To illustrate this, we consider a random walk $\{X_l\}_{l \in \mathbb{N}}$ that begins at $X_0 = i^*$ and, for $l \geq 0$, chooses X_{l+1} uniformly from all incoming neighbors of X_l (agents and bots); note here that the walk follows edges in the direction opposite their polarity in the graph. For this walk, it is easy to see that, conditioned on the event $X_l \in A$, the event $X_{l+1} \in A$ occurs with probability

$$\frac{d_{in}^A(X_l)}{d_{in}^A(X_l) + d_{in}^B(X_l)}. \tag{9}$$

Importantly, we can sample this walk and construct the graph simultaneously, by choosing which instub of X_{l-1} to follow *before* actually pairing these instubs. Assuming they are later paired with uniform agent outstubs, and hence connected to agents chosen proportional to out-degree, we can average (9) over the out-degree distribution to obtain that $X_{l+1} \in A$ occurs with probability

$$\sum_{a \in A} \frac{d_{in}^A(a)}{d_{in}^A(a) + d_{in}^B(a)} \frac{d_{out}(a)}{\sum_{a' \in A} d_{out}(a')} = \tilde{p}_n. \tag{10}$$

Now since bots have a self-loop and no other incoming edges, they are absorbing states on this walk. It follows that $X_{T_n} \in A$ if and only if $X_l \in A \; \forall \; l \in [T_n]$; by the argument above, this latter event occurs with probability $\tilde{p}_n^{T_n}$. Since $\tilde{p}_n \approx p_n$ by (A3), we thus obtain that $X_{T_n} \in A$ with probability

$$\tilde{p}_n^{T_n} \approx p_n^{T_n} \approx \left(1 - \frac{\lim_{n\to\infty} T_n(1-p_n)}{T_n}\right)^{T_n} \approx e^{-\lim_{n\to\infty} T_n(1-p_n)}.$$

From this final expression, the three regimes of Theorem 3.1 emerge: when $T_n(1-p_n) \to 0$, the random walker remains in the agent set with probability ≈ 1; this corresponds to i^* avoiding exposure to bot opinions and learning the true state. Similarly, $T_n(1-p_n) \to \infty$ means the walker is absorbed into the bot set with probability ≈ 1, corresponding to i^* adopting bot opinions. Finally, $T_n(1-p_n) \to c \in (0,\infty)$ means the walker stays in the agent set with probability $\approx e^{-c} \in (0,1)$, corresponding to i^* not fully learning nor fully adopting bot opinions.

We note that the actual proof of Theorem 3.1 does not precisely follow the foregoing argument. Instead, we locally approximate the graph construction with a certain branching process; we then study random walks on the tree resulting from this branching process.[4] However, the foregoing argument illustrates the basic reason why the three distinct cases of Theorem 3.1 arise.

Finally, we observe that the argument leading to (10) shows why $\tilde{p}_n$ enters into our analysis. The other random variables defined in (8) enter similarly. Specifically, $\tilde{p}_n^*$ arises in almost the same manner, but pertains only to the first step of the walk; this distinction arises since the walk starts at i^*, the degrees of which relate to $\tilde{p}_n^*$. On the other hand, $\tilde{q}_n$ arises when we analyze the variance of agent beliefs. This is because analyzing the variance involves studying *two* random walks; by an argument similar to (10), the probability of both walks visiting the same agent is

$$\sum_{a\in A} \frac{d_{in}^A(a)}{d_{in}^A(a) + d_{in}^B(a)} \frac{1}{d_{in}^A(a) + d_{in}^B(a)} \frac{d_{out}(a)}{\sum_{a'\in A} d_{out}(a')} = \tilde{q}_n.$$

3.3 Second result

While Theorem 3.1 establishes convergence for the belief of a typical agent, a natural question to ask is how many agents have convergent beliefs. Our second result, Theorem 3.2, provides a partial answer to this question. To prove the result, we require slightly stronger assumptions than those required for Theorem 3.1 (we will return shortly to comment on why these are needed). First, we strengthen (A1) and (A3) to include particular rates of convergence for the probabilities $\mathbb{P}(\Omega_{n,i})$, $i \in \{1,2\}$. Second, we strengthen (A4) with a minimum rate at which $T_n \to \infty$ (specifically, $T_n = \Omega(\log n)$). Third, and perhaps most restrictively, we require $p_n \to p < 1$ in (A1). As a result, Theorem 3.2 only applies to the case $T_n(1-p_n) \to \infty$, for which Theorem 3.1 states the belief of a uniform agent converges to zero. In this setting, Theorem 3.2 provides an upper bound on how many agents' beliefs do *not* converge to zero. In particular, this bound is $O(n^k)$ for some $k < 1$.

Theorem 3.2. *Assume $\exists\, \kappa, \mu > 0$ and $N' \in \mathbb{N}$ independent of n s.t. the following hold:*

- *(A1), with $\mathbb{P}(\Omega_{n,1}) = O(n^{-\kappa})$.*
- *(A2).*
- *(A3), with $\mathbb{P}(\Omega_{n,2}) = O(n^{-\kappa})$ and $p < 1$.*
- *(A4), with $T_n \geq \mu \log n \; \forall \; n \geq N'$.*

Then for any $\epsilon > 0$, $k > 1 - \min\{(1/2) - \zeta, \mu(\epsilon\eta(1-p)/\theta)^2/16, \kappa\}$, and $K > 0$, all independent of n,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\{i \in [n] : \theta_{T_n}(i) > \epsilon\}\right| > Kn^k\right) = 0.$$

[4]This is necessary because the argument leading to (10) assumes instubs are paired with with uniform outstubs, which is not true if resampling of outstubs occurs in the construction from Section 2.2.

We reiterate that $\zeta < 1/2$ by (A2) and $\mu, \kappa > 0$ by the theorem statement. Hence, $\min\{(1/2) - \zeta, \mu(\epsilon\eta(1-p)/\theta)^2/16, \kappa\} > 0$, so one can choose $k < 1$ in Theorem 3.2 to show that the size of the non-convergent set of agents vanishes relative to n. We suspect that such a result is the best one could hope for; in particular, we suspect that showing *all* agent beliefs converge to zero is impossible. This is in part because our assumptions do not preclude the graph from being disconnected. Hence, there may be small connected components composed of agents but no bots; in such components, agent beliefs will converge to θ (not zero). Additionally, while the lower bound for k in Theorem 3.2 is somewhat unwieldy, certain terms are easily interpretable: the bound sharpens as η grows (i.e. as agents place less weight on their unbiased signals), as p decays (i.e. as the number of bots grows), and as θ decays (i.e. as signals are more likely to be zero, pushing beliefs to zero).

As for Theorem 3.1, the proof of Theorem 3.2 is outlined in Appendix A with details provided in Appendix B. The crux of the proof involves obtaining a sufficiently fast rate for the convergence in Theorem 3.1; namely, we show that for some $\gamma > 0$, $\mathbb{P}(\theta_{T_n}(i^*) > \epsilon) = O(n^{-\gamma})$.[5] At a high level, obtaining such a bound requires bounding three probabilities by $O(n^{-\gamma})$, which also helps explain the stronger assumptions of Theorem 3.2:

- As for Theorem 3.1, we first locally approximate the graph construction with a branching process so as to analyze the belief process on a tree. Here strengthening (A1) with $\mathbb{P}(\Omega_{n,1}) = O(n^{-\kappa})$ is necessary to ensure this approximation fails with probability at most $O(n^{-\gamma})$.
- To analyze the belief process on a tree, we first condition on the random tree structure and treat the belief as a weighted sum of i.i.d. signals using an approach similar to Hoeffding's inequality. Namely, we obtain the Hoeffding-like tail $O(e^{-2\epsilon^2 T_n})$; strengthening (A4) with $T_n \geq \mu \log n$ is necessary to show this tail is $O(e^{-2\epsilon^2 \mu \log n}) = O(n^{-2\epsilon^2 \mu}) = O(n^{-\gamma})$.
- Finally, after conditioning on the tree structure, we show this structure is close to its mean. More specifically, letting $\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]$ denote the expected belief for the root node in the tree conditioned on the random tree structure (see Appendix A for details), we show

$$\mathbb{P}(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] > \epsilon) = O\left(n^{-\gamma}\right).$$

  Note the only source of randomness in $\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]$ is the random tree; because this tree is recursively generated, it has a martingale-like structure that can be analyzed using an approach similar to the Azuma-Hoeffding inequality for bounded-difference martingales. Here we require $\mathbb{P}(\Omega_{n,2}) = O(n^{-\kappa})$ to ensure the degree sequence is ill-behaved with probability at most $O(n^{-\gamma})$; we also require $p_n \to p < 1$ in this step (and only in this step).

We now address the most notable difference between Theorems 3.1 and 3.2; namely, that the latter only applies when $p_n \to p < 1$. We believe this reflects a fundamental distinction between the cases $p_n \to p < 1$ and $p_n \to 1$ and is *not* an artifact of our analysis. An intuitive reason for this is that more bots are present in the former case, so fewer random signals are present (recall we model bot signals as being deterministically zero). As a result, $\theta_{T_n}(i^*)$ is "less random", so its concentration around its mean is stronger. Towards a more rigorous explanation, we first note that Appendix A.4.1 provides the following condition for extending Theorem 3.2 to other cases of p_n:

$$\exists\, \gamma' > 0 \ s.t.\ \mathbb{P}(|\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] - L(p_n)| > \epsilon) = O\left(n^{-\gamma'}\right), \tag{11}$$

[5] One may wonder why we derive a separate bound for Theorem 3.2, since we have already bounded $\mathbb{P}(\theta_{T_n}(i^*) > \epsilon)$ to prove Theorem 3.1. The reason for this is that the bound for Theorem 3.1 does not decay quickly enough as $n \to \infty$ to prove Theorem 3.2; on the other hand, the bound for Theorem 3.2 does not decay at all as $n \to \infty$ for the case $T_n(1-p_n) \to [0, \infty)$ and therefore cannot be used for all cases of Theorem 3.1. See Appendix A.4.2 for details.

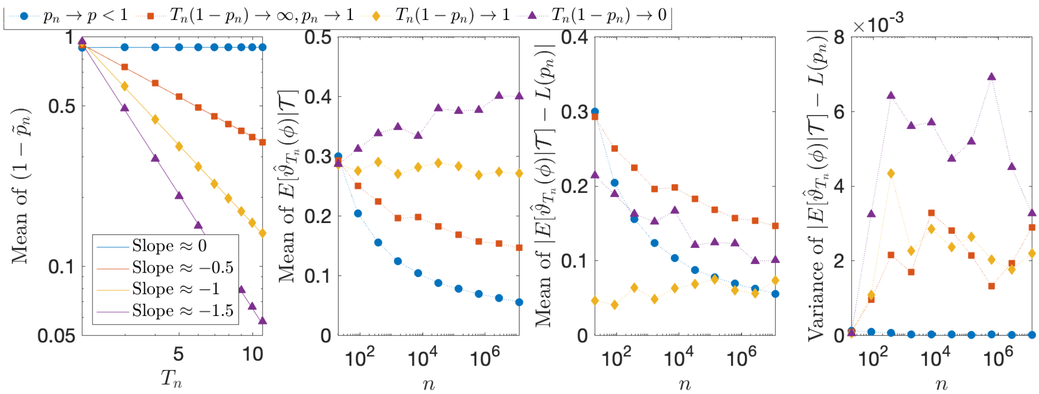

Fig. 2. Empirical comparison of the cases $p_n \longrightarrow p < 1$, $T_n(1-p_n) \longrightarrow \infty$ with $p_n \longrightarrow 1$, $T_n(1-p_n) \longrightarrow 1$, and $T_n(1-p_n) \longrightarrow 0$ (leftmost plot). On average, $\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]$ approaches the corresponding limit from Theorem 3.1 in all cases (second plot from left). However, the error term $|\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] - L(p_n)|$ behaves markedly differently in the case $p_n \longrightarrow p < 1$, with a faster decay on average (second plot from right) and a strikingly lower variance (rightmost plot); we believe this is why Theorem 3.2 only applies in this case.

where $L(p_n)$ is the limit from Theorem 3.1 based on the relative asymptotics of T_n and p_n, i.e.

$$L(p_n) = \begin{cases} \theta, & T_n(1-p_n) \xrightarrow[n\to\infty]{} 0 \\ \theta(1-e^{-c\eta})/(c\eta), & T_n(1-p_n) \xrightarrow[n\to\infty]{} c \in (0,\infty) \\ 0, & T_n(1-p_n) \xrightarrow[n\to\infty]{} \infty \end{cases} .$$

It is the convergence of $|\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] - L(p_n)|$ in (11) that we suspect is fundamentally different in the cases $p_n \longrightarrow p < 1$ and $p_n \longrightarrow 1$. To illustrate this, we provide empirical results in Figure 2. In the leftmost plot, we show $1-\tilde{p}_n$ versus T_n; here the plot is on a log-log scale, so a line with slope m means $(1-\tilde{p}_n) \propto T_n^m$. Hence, we are comparing four cases: $m \approx 0$, so that $p_n \approx p < 1$ (blue circles); $m \approx -0.5$, so that $T_n(1-p_n) \longrightarrow \infty$ and $p_n \longrightarrow 1$ (orange squares); $m \approx -1$, so that $T_n(1-p_n) \longrightarrow 1$ (yellow diamonds); and $m \approx -1.5$, so that $T_n(1-p_n) \longrightarrow 0$ (purple triangles). The second plot reflects the corresponding cases of $L(p_n)$: $\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]$ decays to zero in the first two cases, grows towards $\theta = 0.5$ in the fourth case, and approaches an intermediate limit in the third case. The final two plots illustrate the convergence (or lack thereof) in (11). Here the empirical mean of the error term $|\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] - L(p_n)|$ decays quickly for the first case but decays more slowly (or is even non-monotonic) in the other cases. More strikingly, the empirical variance of this error term is several orders of magnitude smaller in the first case. This suggests that $\mathbb{P}(|\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]-L(p_n)| > \epsilon)$ decays much more rapidly in the case $p_n \longrightarrow p < 1$, which is why we believe this is the only case for which (11) is satisfied. (We point the reader to Appendix C for further details on this experiment.)

3.4 Comments on assumptions

We now return to comment on the assumptions needed to prove our results. First, (A1) states that certain empirical moments of the degree distribution – namely, for $i^* \sim A$ uniformly, the first two moments of $d_{out}(i^*)$ and the correlation between $d_{out}(i^*)$ and $d_{in}^A(i^*)$ – converge to finite limits. Roughly speaking, this says our graph lies in a sparse regime, where typical node degrees do not grow with the number of nodes.[6] We also note $\nu_3 > \nu_1$ in (A1) is minor and simply eliminates an uninteresting case. To see this, first note that when $\Omega_{n,1}$ holds, we have (roughly)

$$\frac{\nu_3}{\nu_1} \approx \frac{\sum_{i=1}^n d_{out}(i) d_{in}^A(i)/n}{\sum_{i=1}^n d_{out}(i)/n} = \sum_{i=1}^n \frac{d_{out}(i)}{\sum_{i'=1}^n d_{out}(i')} d_{in}^A(i) \geq 1, \tag{12}$$

where we have used the assumed inequality $d_{in}^A(i) \geq 1 \ \forall \ i \in [n]$. Hence, $\nu_3 < \nu_1$ cannot occur, so assuming $\nu_3 > \nu_1$ simply eliminates the case $\nu_3 = \nu_1$. This remaining case is uninteresting because ν_3/ν_1 is the limiting number of offspring for each node in the branching process we analyze; thus, if $\nu_3 = \nu_1$, the tree resulting from this process is simply a line graph.

Next, (A2) states $T_n = O(\log n)$. Together with (A1), these assumptions are standard given our analysis approach, which, as discussed previously, locally approximates the graph construction with a branching process. We also note that, with the interpretation of ν_3/ν_1 above, it follows that the number of agents within the T_n-step neighborhood of i^* can be upper bounded by

$$(\nu_3/\nu_1)^{T_n} = O\left((\nu_3/\nu_1)^{\zeta \log(n)/\log(\nu_3/\nu_1)} \right) = O\left(n^{\zeta} \right) = o(n).$$

In words, the size of the aforementioned neighborhood vanishes relative to n. As mentioned in the introduction, this is why our title refers to the learning as "local": only a vanishing fraction of other agents (those within this neighborhood) influence the belief of i^*.

The remaining statements are needed to establish belief convergence on the tree resulting from the branching process. (A4) states $T_n \to \infty$ with n, which is an obvious requirement for convergence. (A3) essentially says that three events occur with high probability. First, $\tilde{p}_n$ should be close to a convergent, deterministic sequence p_n; this is necessary since the asymptotics of p_n play a prominent role in Theorem 3.1. Second, $\tilde{p}_n^* \geq \tilde{p}_n$ essentially says that bots prefer to attach to agents with higher out-degrees, i.e. more influential agents; this is the behavior one would intuitively expect from bots aiming to disrupt learning. Third, $\tilde{q}_n < 1 - \xi \in (0, 1)$ is a minor assumption; for example, if all agents have total in-degree at least 2, $\tilde{q}_n \leq 1/2$.

4 RELATED WORK

Before closing, we discuss some connections between existing work and ours. First, from a modeling perspective, we note our belief update (3) is closely related to the popular non-Bayesian social learning model from [16]. In that model, agent beliefs are distributions over a finite set of possible states of the world (not simply scalars, as in our model), but belief updates are similar. Specifically, at time t agent i updates its belief $\mu_t(i)$ as

$$\mu_t(i) = \eta_{ii} \mathrm{BU}(\mu_{t-1}(i), \omega_t(i)) + \sum_{j \in N_{in}(i)} \eta_{ij} \mu_{t-1}(j),$$

where $\omega_t(i)$ is the signal received by i at t, $\mathrm{BU}(\mu_{t-1}(i), \omega_t(i))$ means a Bayesian update of the prior belief $\mu_{t-1}(i)$ with the observed signal $\omega_t(i)$, and $\sum_{j \in N_{in}(i) \cup \{i\}} \eta_{ij} = 1$. In [16], it is shown that, under certain assumptions, including the graph being fixed and strongly connected, these distributions converge to point masses on the true state as $t \to \infty$ (see Proposition 3 in [16]).

Per the discussion following (3) in Section 2.1, our model can be viewed as a variant of this one, in which all agents have beta beliefs and Bernoulli signals, communicate parameters of distributions instead of the distributions themselves, and average parameters instead of distributions. From this viewpoint, our quantity of interest $\theta_t(i)$ is simply the mean of agent i's belief. However, the crucial assumptions of strong connectedness and an infinite learning horizon from [16] are violated in our

[6] This is analogous to an Erdős-Rényi model with edge formation probability λ/n for some $\lambda > 0$ independent of n, in which degrees converge in distribution to Poisson(λ) random variables that have finite mean and variance.

model (the former since bots have self-loops but no other incoming edges; the latter since we take $T_n = O(\log n)$). This necessitates a different analysis, which in turn requires us to simplify the model from [16] by communicating scalars and by taking a simple form of the weights $\{\eta_{ij}\}_{j \in N_{in}(i) \cup \{i\}}$.

We also note our variant of the model from [16] is quite similar to the model in the working paper [5]. In fact, our belief update and inclusion of bots are both taken from this work (with minor differences to bot behavior). However, this work only includes theoretical results in the case $B = \emptyset$; the case $B \neq \emptyset$ is studied empirically. This allows [5] to use a slightly richer model than ours, including a time-varying graph structure, agent-dependent mixture parameters $\sum_{j \in N_{in}(i) \cup \{i\}} \eta_{ij}$, and three types of nodes (bots, agents susceptible to bot influence, and agents not susceptible to bot influence). Notably, the empirical results from [5] for the case $B \neq \emptyset$ fix a learning horizon and do not investigate the effects of different timescales; in particular, the delicate relationship between timescale and bot prevalence that we describe in Theorem 3.1 is not brought to light in [5].

From an analytical perspective, our approach of analyzing beliefs by studying random walks is not new. Perhaps the most obvious example is the classical deGroot model [11], in which agent i updates its (scalar) belief as $\theta_t(i) = \sum_j \theta_{t-1}(j) W(j,i)$ for some column-stochastic matrix W. Collecting beliefs in vector form yields $\theta_t = \theta_{t-1} W = \cdots = \theta_0 W^t$, where θ_0 is the vector of initial beliefs. From here, it is clear that beliefs relate closely to random walks, since the i-th column of W^t gives the distribution of a t-step random walk from i on the weighted graph defined by W. This observation has been exploited in the literature (see Section 3 of the survey [15], Section 4 of the survey [2] , and the references therein). For example, assuming W is irreducible and aperiodic, and therefore has a well-defined stationary distribution π, [14] establishes conditions for learning using the fact that $\theta_t(i) = \theta_0 W^t e_i^{\mathsf{T}} \approx \theta_0 \pi^{\mathsf{T}}\ \forall\ i$ when t is large. Beyond the deGroot model and deGroot-like models such as ours, random walk interpretations have also been leveraged in Bayesian learning models. For example, [17] considers a model for which agents perform a Bayesian update using their own signal but using the prior of a randomly-chosen neighbor. Exchanging priors with neighbors yields a natural connection to random walks; assuming strong connectedness, the authors exploit the fact that the walk visits every agent infinitely often (i.o.) to derive conditions for learning.

Similar to [16], these works typically assume strong connectedness and long learning horizons so as to leverage properties such as stationary distributions and i.o. visits. This is a fundamental distinction from our work. Indeed, even if we disregard stubborn agents, so that the random walk converges to a stationary distribution, it does *not* converge within our local learning horizon. This is because, as shown in [8], the sparse DCM we consider has mixing time that exceeds

$$\frac{\log n}{\sum_{i \in [n]} \log(d_{in}^A(i)) \frac{d_{out}(i)}{\sum_{i' \in A} d_{out}(i')}} \geq \frac{\log n}{\log(\sum_{i \in [n]} d_{in}^A(i) \frac{d_{out}(i)}{\sum_{i' \in A} d_{out}(i')})} \approx \frac{\log n}{\log(\nu_3/\nu_1)},$$

where the inequality is Jensen's and the approximate equality is (12). The final expression exceeds T_n by (A2), i.e. our learning horizon occurs before the underlying random walk mixes. In fact, [8] shows that the random walk on the DCM exhibits *cutoff*, meaning that the T_n-step distribution of this walk can be maximally far from the stationary distribution (i.e. the total variation distance between these distributions can be 1 for certain starting locations of the walk). Hence, not only can we not use this stationary distribution, we cannot even use an approximation of it. Again, this means our analysis cannot leverage global properties typically used when relating beliefs to random walks and thus requires a different approach. We also note that our idea to simultaneously construct the graph and sample the walk (as discussed in Section 3.2) is taken from [8].

Some other works have considered social learning with stubborn agents. For example, [3] studies a model in which agents meet and either retain their own (scalar) beliefs, adopt the average of their beliefs, or adopt a weighted average; the agent whose belief has a larger weight is called a "forceful" agent. Here the authors show that all agent beliefs converge to a common random variable and study its deviation from the true state. A crucial difference between this work and ours is that [3] assumes even forceful agents are occasionally influenced by other agents. This yields an underlying Markov chain that is irreducible (unlike ours, in which stubborn agents are absorbing states); the analysis then relies on this chain having a well-defined stationary distribution.

Stubborn agents have also been considered in the consensus setting. This setting is similar to the social learning setting we consider, but instead of asking whether agents learn an underlying state, one asks whether agent beliefs converge to a common value, i.e. a consensus. For example, [1] considers a model in which regular agents adopt weighted averages of beliefs upon meeting other agents (regular or stubborn), while stubborn agents always retain their own beliefs. This intuitively prohibits a consensus from forming; indeed, it is shown that agent beliefs fail to converge, and therefore that disagreement can persist indefinitely. Another example is [13], in which an agent's belief at time $t+1$ is a weighted average of their own belief at time 0 and their neighbors' beliefs at time t. In this model, stubborn agents place all weight on their own belief from time 0 and thus do not update their beliefs. The analysis in [13] is similar to ours as it relates agent beliefs to hitting probabilities of the stubborn agent set, but it differs as the learning horizon is infinite in [13].

Finally, we note that the random graph model we consider was proposed and analyzed in [10]. As mentioned in Appendix A, our analysis uses a branching process approximation of this model that draws from the analysis of [9]. The directed configuration model is a natural extension of the undirected configuration model, the study of which originated in [6, 7, 20].

5 CONCLUSIONS

In this work, we devised and analyzed a model for social learning in the presence of stubborn agents. Our analysis identified a close relationship between the learning horizon, the "density" of stubborn agents, and the learning outcome. Several extensions of our work can be considered. First, it would be useful to generalize our model to allow for agent- and/or time-dependent mixture parameters (i.e. allowing η to vary with i and/or t in (6)). Allowing agent dependence suggests a more heterogeneous model in which some agents place more value on private observations, while others place more value on the opinions of their social connections. Allowing time dependence, and specifically allowing η_t to vanish as t grows, suggests a model in which agents become more "set in their ways" over time. Beyond being a richer model, this may also lead to learning in more cases (even if there are many bots, an agent who becomes sufficiently "set in its ways" may learn before the influence of bots reaches it). Second, and closely related, one could consider adaptive belief updates, rather than assuming agents update beliefs by the same rule at each time step. In particular, agents could place less weight on neighbors whose opinions they deem to be too extreme. This would lead to a much more complicated model, as agents would simultaneously attempt to learn the true state and assess the opinions of their neighbors. On the other hand, if agents could eventually identify extreme beliefs, they may be less susceptible to bots and more likely to learn. Third, one could keep T_n finite for each finite n but allow it to asymptotically dominate our "local" $O(\log n)$ horizon. Here our branching process approximation fails, so this would require a different analysis. However, it would be interesting to see if the three regimes of Theorem 3.1 still hold for such T_n, or if a different phenomenon emerges when global effects of the network take hold.

REFERENCES

[1] Daron Acemoglu, Giacomo Como, Fabio Fagnani, and Asuman Ozdaglar. 2011. Opinion fluctuations and persistent disagreement in social networks. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*. IEEE, 2347–2352.

[2] Daron Acemoglu and Asuman Ozdaglar. 2011. Opinion dynamics and learning in social networks. *Dynamic Games and Applications* 1, 1 (2011), 3–49.

[3] Daron Acemoglu, Asuman Ozdaglar, and Ali ParandehGheibi. 2010. Spread of (mis) information in social networks. *Games and Economic Behavior* 70, 2 (2010), 194–227.

[4] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31, 2 (2017), 211–36.

[5] Marina Azzimonti and Marcos Fernandes. 2018. *Social media networks, fake news, and polarization.* Technical Report. National Bureau of Economic Research.

[6] Edward A Bender and E Rodney Canfield. 1978. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A* 24, 3 (1978), 296–307.

[7] Béla Bollobás. 1980. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics* 1, 4 (1980), 311–316.

[8] Charles Bordenave, Pietro Caputo, and Justin Salez. 2018. Random walk on sparse random digraphs. *Probability Theory and Related Fields* 170, 3-4 (2018), 933–960.

[9] Ningyuan Chen, Nelly Litvak, and Mariana Olvera-Cravioto. 2017. Generalized PageRank on directed configuration networks. *Random Structures & Algorithms* 51, 2 (2017), 237–274.

[10] Ningyuan Chen and Mariana Olvera-Cravioto. 2013. Directed random graphs with given degree distributions. *Stochastic Systems* 3, 1 (2013), 147–186.

[11] Morris H DeGroot. 1974. Reaching a consensus. *J. Amer. Statist. Assoc.* 69, 345 (1974), 118–121.

[12] D.P. Dubhashi and A. Panconesi. 2009. *Concentration of Measure for the Analysis of Randomized Algorithms.* Cambridge University Press.

[13] Javad Ghaderi and R Srikant. 2014. Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. *Automatica* 50, 12 (2014), 3209–3215.

[14] Benjamin Golub and Matthew O Jackson. 2010. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics* 2, 1 (2010), 112–49.

[15] Benjamin Golub and Evan Sadler. 2017. Learning in social networks. (2017).

[16] Ali Jadbabaie, Pooya Molavi, Alvaro Sandroni, and Alireza Tahbaz-Salehi. 2012. Non-Bayesian social learning. *Games and Economic Behavior* 76, 1 (2012), 210–225.

[17] Mohammad Amin Rahimian, Shahin Shahrampour, and Ali Jadbabaie. 2015. Learning without recall by random walks on directed graphs. In *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on.* IEEE, 5538–5543.

[18] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini, and Filippo Menczer. 2017. The spread of fake news by social bots. *arXiv preprint arXiv:1707.07592* (2017), 96–104.

[19] Elisa Shearer and Jeffrey Gottfried. 2017. News use across social media platforms 2017. *Pew Research Center, Journalism and Media* (2017).

[20] Nicholas C Wormald. 1980. Some problems in the enumeration of labelled graphs. *Bulletin of the Australian Mathematical Society* 21, 1 (1980), 159–160.

APPENDIX A PROOF OUTLINES

The proofs of Theorems 3.1 and 3.2 proceed in two steps. First, we show that the graph construction can be locally approximated by a certain branching process. Second, we analyze the beliefs of agents in the graph by instead analyzing the beliefs of agents in the tree resulting from the branching process. We note that studying tree agent beliefs rather than graph agent beliefs is advantageous because the tree has a comparatively simple structure that is more amenable to analysis.

The first step is identical for both theorems, while the second step requires a different analysis for each theorem. In Appendix A.1, we outline the first step, and in Appendices A.2 and A.3, respectively, we outline the second step for Theorems 3.1 and 3.2, respectively. To highlight the key ideas of our analysis, we defer many details to Appendix B; in particular, proofs pertaining to Appendices A.1, A.2, and A.3 , respectively, can be found in Appendices B.1, B.2, and B.3, respectively. Finally, we note that throughout the analysis we use $\mathbb{P}_n$ and $\mathbb{E}_n$, respectively, to denote probability and expectation, respectively, conditioned on the degree sequence $\{d_{out}(i), d_{in}^A(i), d_{in}^B(i)\}_{i\in[n]}$.

A.1 Branching process approximation (Step 1 for proofs of Theorems 3.1 and 3.2)

We first show that the belief of any agent in the graph depends (asymptotically) only on the structure of the agent's local neighborhood and on certain signals realized within this neighborhood. This will facilitate the definition of the branching process with which we will approximate the graph construction. Importantly, the graph agent's belief will *not* depend on the prior parameters α_0, β_0 (asymptotically). This is necessary as we have not specified these priors (beyond assuming they are bounded by some $\bar{\alpha}, \bar{\beta}$ independent of n, as discussed in Section 2.1).

To begin, we require some notation. Let P denote the graph's column-normalized adjacency matrix, i.e. $P(i,j) = |\{i' \to j' \in E : i' = i, j' = j\}|/d_{in}(j)$, and set $Q = (1-\eta)I + \eta P$, where I is the identity matrix of appropriate dimension. (Recall from Section 2.2 that E is in general a multi-set; hence, the numerator in $P(i,j)$ may exceed 1.) Next, for $t \in \mathbb{N}$, let s_t denote the collection of signals $\{s_t(i)\}_{i\in A\cup B}$ in vector form. Finally, for $i \in A$ define

$$\vartheta_{T_n}(i) = \frac{1}{T_n}\sum_{t=0}^{T_n-1} s_{T_n-t} Q^t e_i^{\mathsf{T}}. \tag{13}$$

We note that (13) can be rewritten as

$$\vartheta_{T_n}(i) = \frac{1}{T_n}\sum_{t=0}^{T_n-1}\sum_{j\in A} s_{T_n-t}(j) e_j Q^t e_i^{\mathsf{T}}, \tag{14}$$

where we have used the fact that $s_t(j) = 0 \ \forall\ t \in \mathbb{N}, j \in B$. From this expression, it is clear that $\vartheta_{T_n}(i)$ only depends on the structure of the T_n-step neighborhood into i (since only this sub-graph affects the $e_j Q^t e_i^{\mathsf{T}}$ terms) and on certain signals within this neighborhood, as mentioned above. We can then establish the following.

Lemma A.1. *Given (A4),* $\forall\ \epsilon > 0\ \exists\ N$ *s.t.* $\forall\ n \geq N$, $|\theta_{T_n}(i) - \vartheta_{T_n}(i)| < \epsilon$ *a.s.* $\forall\ i \in A$.

Proof. See Appendix B.1.1. □

Before defining the aforementioned branching process, we formally define the graph construction described in Section 2.2. For this, we will use the following additional notation.

- We let $A_l, l \in \mathbb{N}_0$ denote the set of agents at distance l from the initial agent i^*, i.e. $i \in A_l$ means a path from i to i^* of length l exists, but no shorter path exists (hence, $A_0 = \{i^*\}$, $A_1 = N_{in}(i^*)$, etc.). Similarly, we let $B_l, l \in \mathbb{N}_0$ denote the set of bots at distance l from i^*.

- We let $\{(i, j) : j \in [d_{out}(i)]\}$ denote the set of outstubs belonging to $i \in A$; we let O_A denote the set of all such outstubs.
- For each $(i, j) \in O_A$, we define a label $g((i, j)) \in \{1, 2, 3\}$ as follows:

$$g((i,j)) = \begin{cases} 1, & i \text{ does not yet belong to graph} \\ 2, & i \text{ belongs to graph but } (i,j) \text{ has not been paired} \\ 3, & i \text{ belongs to graph and } (i,j) \text{ has been paired} \end{cases}. \tag{15}$$

  We will explain the utility of these labels shortly.

With this notation in place, we present the formal graph construction as Algorithm 1. We offer some further comments to help explain the algorithm:

- The algorithm takes as input the degree sequence $\{d_{out}(i), d_{in}^A(i), d_{in}^B(i)\}_{i \in A}$, which is used in Line 1 to define O_A. Also in Line 1, we label all outstubs as 1 (since no agents have been added to the graph), and we initialize the set of bots to the empty set.
- In Line 2, we sample the agent i^* from which the graph construction begins. Since i^* then belongs to the graph, we change the labels of its outstubs to 2.
- For the remainder of the algorithm, we proceed in a breadth-first-search fashion, looping over distance l and agents i at distance l from i^*. For each such agent, we do the following:
  - For each of the $d_{in}^A(i)$ instubs of i intended for pairing with agent outstubs, we sample an agent outstub uniformly (Line 7), resampling until an unpaired outstub (i.e. one with label 1 or 2) has been found (Line 9). Upon finding such an outstub, denoted (i', j'), we pair it with i's instub to form an edge from i' to i (Line 10). Note that $g((i', j')) = 1$ implies i' was added to the graph when edge $i' \to i$ was formed; hence, because $i \in A_l$, i' is at distance $l + 1$ from i^* and must be added to A_{l+1} (Line 11). Finally, we update the labels of the outstubs of i' via (15) (Lines 11-12). (Line 8 will be used in the branching process approximation and will be discussed shortly.)
  - For each of the $d_{in}^B(i)$ instubs of i intended for pairing with bot outstubs, we add a new bot with a self-loop and an unpaired outstub to the set of bots, updating B_{l+1} accordingly (Line 15), and then add an edge from the new bot to i (Line 16). Note here that $B = \emptyset$ at the start of the construction; it follows that the k-th bot added to the graph is $n + k + 1$, so $B = n + [\sum_{i \in A} d_{in}^B(i)]$ is the set of bots at the end of the construction.
  - Finally, if all agent outstubs have been paired, the construction terminates (Line 18).

We now return to discuss Line 8 of Algorithm 1. Here τ_n denotes the first iteration at which an outstub with label 2 or 3 is sampled for pairing with an instub. Put differently, $\tau_n > l$ means that for the first l iterations of the construction, only outstubs with label 1 have been sampled. This has two consequences. First, no edges have been added between two nodes both at distance $\leq l$ from i^*, i.e. the l-step incoming neighborhood of i^* is a tree (except for the self-loops attached to bots). Second, no resampling of outstubs has occurred (Line 9); this implies that the outstub (i', j') paired in Line 10 is chosen uniformly from O_A, so the degrees $(d_{out}(i'), d_{in}^A(i'), d_{in}^B(i'))$ of i' are distributed according to the out-degree distribution f_n defined in (7).

These observations motivate a tree construction that we define next. In particular, we will construct a tree (except for bot self-loops) with edges pointing towards the root. Agents will be added to the tree with degrees sampled from f_n, except for the root node, whose degrees are sampled from f_n^* (also defined in (7)), corresponding to the degrees of i^* in the graph construction.

The tree construction requires further notation. First, we let $\hat{A}_l$ ($\hat{B}_l$, respectively) denote agents (bots, respectively) at distance l from the tree's root. We also set $\hat{A} = \cup_{l=0}^{\infty} \hat{A}_l$, $\hat{B} = \cup_{l=0}^{\infty} \hat{B}_l$. (Here and moving forward, we use $\hat{\cdot}$ to distinguish tree-related objects from similarly-defined graph-related ones.) At times, we will use branching process terminology and e.g. refer to $\hat{A}_l$ as the l-th *generation*

ALGORITHM 1: Graph-Construction

Set $O_A = \{(i,j) : i \in A, j \in [d_{out}(i)]\}$, $g((i,j)) = 1 \ \forall \ (i,j) \in O_A$, $B = \emptyset$
Sample i^* uniformly from A; set $g((i^*, j)) = 2 \ \forall \ j \in [d_{out}(i^*)]$; set $A_0 = \{i^*\}$
for $l = 0$ **to** ∞ **do**
    Set $A_{l+1} = B_{l+1} = \emptyset$
    for $i \in A_l$ **do**
        for $j = 1$ **to** $d_{in}^A(i)$ **do**
            Sample (i', j') from O_A uniformly
            if $g((i', j')) \neq 1$ **and** $\tau_n = \infty$ **then** set $\tau_n = l$
            while $g((i', j')) = 3$ **do** sample (i', j') from O_A uniformly
            Add directed edge from i' to i
            if $g((i', j')) = 1$ **then** set $A_{l+1} = A_{l+1} \cup \{i'\}$, $g((i', j')) = 3$, $g((i', j'')) = 2 \ \forall \ j'' \in [d_{out}(i')] \setminus \{j'\}$
            else if $g((i', j')) = 2$ **then** set $g((i', j')) = 3$
        end
        for $j = 1$ **to** $d_{in}^B(i)$ **do**
            Add bot $b = n + |B| + 1$ with self-loop and unpaired outstub, set $B = B \cup \{b\}$, $B_{l+1} = B_{l+1} \cup \{b\}$
            Add directed edge from b to i
        end
        if $g((i', j')) = 3 \ \forall \ (i', j') \in O_A$ **then return**
    end
end

of agents. We let ϕ denote the root node, so that $\hat{A}_0 = \{\phi\}$. We will denote generic node in $\hat{A}_l \cup \hat{B}_l$ as $\mathbf{i} \in \mathbb{N}^l$; here $\mathbf{i} = (i_1, \ldots, i_l)$ encodes the ancestry of $\mathbf{i}$, i.e. $(i_1, \ldots, i_l)$ is the child of $(i_1, \ldots, i_{l-1})$, who is in turn the child of $(i_1, \ldots, i_{l-2})$, etc. Finally, for such $\mathbf{i}$ and for $j \in \mathbb{N}$, $(\mathbf{i}, j) = (i_1, \ldots, i_l, j)$ is the concatenation operation and $\mathbf{i}|j = (i_1, \ldots, i_j)$ denotes $\mathbf{i}$'s ancestor in generation j, with $\mathbf{i}|0 = \phi$ by convention (note also that $\mathbf{i}|l = \mathbf{i}$ for such $\mathbf{i}$).

With this notation in place, we define the tree construction in Algorithm 2. We offer several more explanatory comments:

- Lines 2 and 6-11 define a particular random walk that will be used in Appendix A.2; they do not affect the tree structure and we defer further explanation to Appendix A.2.
- As mentioned above, the root node ϕ has degrees sampled from f_n^* (Line 1), while all other agents have degrees sampled from f_n (Line 13).
- In Line 14, a directed edge is added from $(\mathbf{i}, j)$ to $\mathbf{i}$; the other $d_{out}((\mathbf{i}, j)) - 1$ outstubs of $(\mathbf{i}, j)$ are left unpaired so that the tree structure is preserved (except for bot self-loops).
- At the conclusion of the l-th iteration, $\mathbf{i} \in \hat{A}_l$ has incoming neighbor set (offspring, in the branching process terminology) $\{(\mathbf{i}, j) : j \in [d_{in}^A(\mathbf{i}) + d_{in}^B(\mathbf{i})]\}$. More specifically, the subset $(\mathbf{i}, 1), \ldots, (\mathbf{i}, d_{in}^A(\mathbf{i}))$ of $\mathbf{i}$'s incoming neighbors are agents (Line 14), while the subset $(\mathbf{i}, d_{in}^A(\mathbf{i}) + 1), \ldots, (\mathbf{i}, d_{in}^A(\mathbf{i}) + d_{in}^B(\mathbf{i}))$ of $\mathbf{i}$'s incoming neighbors are bots (Line 17).
- Unlike the graph construction, the tree construction continues indefinitely, yielding an infinite tree (except for bot self-loops) with edges pointing towards the root node ϕ.

Having defined the tree construction, we also define $\hat{\vartheta}_{T_n}(\phi)$ as in (13) but using the tree from Algorithm 2 instead of the graph from Algorithm 1. Specifically, we let

$$\hat{\vartheta}_{T_n}(\phi) = \frac{1}{T_n} \sum_{t=0}^{T_n - 1} \hat{s}_{T_n - t} \hat{Q}^t e_\phi^\mathsf{T} = \frac{1}{T_n} \sum_{t=0}^{T_n - 1} \sum_{\mathbf{i} \in \hat{A}} \hat{s}_{T_n - t}(\mathbf{i}) e_\mathbf{i} \hat{Q}^t e_\phi^\mathsf{T}, \tag{16}$$

ALGORITHM 2: `Tree-Construction`

Define f_n, f_n^* via (7), set $\hat{A}_0 = \{\phi\}$, sample $(d_{out}(\phi), d_{in}^A(\phi), d_{in}^B(\phi))$ from f_n^*
Set $X_0^1 = X_0^2 = \phi$
for $l = 0$ **to** ∞ **do**
  Set $\hat{A}_{l+1} = \hat{B}_{l+1} = \emptyset$
  for $\mathbf{i} \in \hat{A}_l$ **do**
    for $k \in \{1, 2\}$ **do**
      if $X_l^k = \mathbf{i}$ **then**
        Sample j^* from $[d_{in}^A(\mathbf{i}) + d_{in}^B(\mathbf{i})]$ uniformly, set $X_{l+1}^k = (\mathbf{i}, j^*)$
        if $j^* > d_{in}^A(\mathbf{i})$ **then** set $X_{l'}^k = (\mathbf{i}, j^*)\ \forall\ l' \in \{l+2, l+3, \cdots\}$
      end
    end
    for $j = 1$ **to** $d_{in}^A(\mathbf{i})$ **do**
      Sample $(d_{out}((\mathbf{i}, j)), d_{in}^A((\mathbf{i}, j)), d_{in}^B((\mathbf{i}, j)))$ from f_n
      Add directed edge from $(\mathbf{i}, j)$ to $\mathbf{i}$, set $\hat{A}_{l+1} = \hat{A}_{l+1} \cup \{(\mathbf{i}, j)\}$
    end
    for $j = 1$ **to** $d_{in}^B(\mathbf{i})$ **do**
      Add bot $b = (\mathbf{i}, d_{in}^A(\mathbf{i}) + j)$ with self-loop and unpaired outstub, set $\hat{B}_{l+1} = \hat{B}_{l+1} \cup \{b\}$
      Add directed edge from b to i
    end
  end
end

where $\hat{s}_t(\mathbf{i}) \sim \text{Bernoulli}(\theta)\ \forall\ t \in \mathbb{N}, \mathbf{i} \in \hat{A}$; $\hat{s}_t(\mathbf{i}) = 0\ \forall\ t \in \mathbb{N}, \mathbf{i} \in \hat{B}$; $\hat{Q} = (1-\eta)I + \eta\hat{P}$; and $\hat{P}$ is the column-normalized adjacency matrix of the tree from Algorithm 2. We pause to note that

$$0 \leq \hat{\vartheta}_{T_n}(\phi) \leq \frac{1}{T_n} \sum_{t=0}^{T_n - 1} \mathbf{1}\hat{Q}^t e_\phi^\top = 1, \tag{17}$$

where the first inequality holds since (16) is a sum of nonnegative terms, the second follows since $\sum_{\mathbf{i} \in \hat{A}} \hat{s}_{T_n - t}(\mathbf{i}) e_{\mathbf{i}} \leq \mathbf{1}$ component-wise (where $\mathbf{1}$ is the all ones vectors) and since $\hat{Q}^t e_\phi^\top$ is element-wise nonnegative, and the equality holds by column stochasticity of $\hat{Q}$.

We can now state Lemma A.2, which relates the belief of a uniformly random agent in the graph with the belief of the root node in the tree. For the first statement in the lemma, we argue that, conditioned on $\tau_n > T_n$, the T_n-step neighborhood of i^* in the graph and the T_n-step neighborhood of ϕ in the tree are constructed via the same procedure; since the signals are defined in the same manner as well, this implies $\vartheta_{T_n}(i^*)$ and $\hat{\vartheta}_{T_n}(\phi)$ have the same distribution. The second statement of the lemma says that the condition $\tau_n > T_n$ occurs with high probability; it is essentially implied by Lemma 5.4 from [9]. We note that the assumptions (A1) and (A2) are required for this second statement to hold, and are standard assumptions needed to locally approximate a sparse random graph construction with a branching process. Finally, we recall $\zeta < 1/2$ by (A2), which is why the limit shown in Lemma A.2 holds.

Lemma A.2. *Assume (A1) and (A2) hold, and let $\overset{\mathscr{D}}{=}$ denote equality in distribution. Then*

$$\vartheta_{T_n}(i^*) | \{\tau_n > T_n\} \overset{\mathscr{D}}{=} \hat{\vartheta}_{T_n}(\phi), \quad \mathbb{P}(\tau_n \leq T_n | \Omega_{n,1}) = O\left(n^{\zeta - 1/2}\right) \xrightarrow[n \to \infty]{} 0.$$

Proof. See Appendix B.1.2. □

We can now state and prove Lemma A.3, which is the main result for Step 1 of the proofs of the theorems. This result will allow us to analyze convergence of $\theta_{T_n}(i^*)$ (the graph agent belief) by instead analyzing convergence of $\hat{\vartheta}_{T_n}(\phi)$ (the tree agent belief).

Lemma A.3. *Assume (A1), (A2), and (A4) hold. Then* $\forall\, x \in \mathbb{R}$,

$$\mathbb{P}(|\theta_{T_n}(i^*) - x| > \epsilon) \le \mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - x| > \epsilon/2) + \mathbb{P}(\Omega_{n,1}^C) + O\left(n^{\zeta - 1/2}\right).$$

Proof. First, given $\epsilon > 0$, we have for sufficiently large n,

$$\mathbb{P}(|\theta_{T_n}(i^*) - x| > \epsilon) \le \mathbb{P}(|\theta_{T_n}(i^*) - \vartheta_{T_n}(i^*)| + |\vartheta_{T_n}(i^*) - x| > \epsilon) \le \mathbb{P}(|\vartheta_{T_n}(i^*) - x| > \epsilon/2),$$

where the first inequality uses the triangle inequality and in the second we used Lemma A.1 to bound $|\theta_{T_n}(i^*) - \vartheta_{T_n}(i^*)|$ by $\epsilon/2$ *a.s.* Furthermore, by the law of total probability, we have

$$\mathbb{P}(|\vartheta_{T_n}(i^*) - x| > \epsilon/2) \le \mathbb{P}(|\vartheta_{T_n}(i^*) - x| > \epsilon/2 | \tau_n > T_n) + \mathbb{P}(\tau_n \le T_n | \Omega_{n,1}) + \mathbb{P}(\Omega_{n,1}^C).$$

Combining the previous two inequalities and using Lemma A.2 (which applies since (A1), (A2) are assumed to hold), we obtain

$$\mathbb{P}(|\theta_{T_n}(i^*) - x| > \epsilon) \le \mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - x| > \epsilon/2) + O\left(n^{\zeta - 1/2}\right) + \mathbb{P}(\Omega_{n,1}^C),$$

which is what we set out to prove. □

Before proceeding, we state another lemma that will be used in Step 2 of the proofs for both theorems. This lemma uses the fact that each agent in the tree has a unique path to the root. As a result, we can obtain an alternate expression for the terms $e_{\mathbf{i}} \hat{Q}^t e_\phi^\top$ appearing in (16).

Lemma A.4. *For each* $n \in \mathbb{N}$,

$$\hat{\vartheta}_{T_n}(\phi) = \frac{1}{T_n} \sum_{t=0}^{T_n - 1} \sum_{l=0}^{t} \binom{t}{l} \eta^l (1-\eta)^{t-l} \sum_{\mathbf{i} \in \hat{A}_l} \hat{s}_{T_n - t}(\mathbf{i}) \prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1} \ a.s., \tag{18}$$

where by convention $\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1} = 1$ *when* $l = 0$.

Proof. See Appendix B.1.3. □

A.2 Step 2 for proof of Theorem 3.1

Our next goal is to establish convergence of $\hat{\vartheta}_{T_n}(\phi)$, from which convergence of $\theta_{T_n}(i^*)$ will follow via Lemma A.3. For this, we will use Chebyshev's inequality, so we begin with two lemmas describing the limiting behavior of the mean and variance of $\hat{\vartheta}_{T_n}(\phi)$. Here and moving forward, for random variables X and Y we use $\mathrm{Var}_n(X) = \mathbb{E}_n[X^2] - (\mathbb{E}_n[X])^2$ and $\mathrm{Cov}_n(X,Y) = \mathbb{E}_n[XY] - \mathbb{E}_n[X]\mathbb{E}_n[Y]$ to denote variance and covariance conditional on the degree sequence.

Lemma A.5. *Given (A3) and (A4), we have the following:*

$$\lim_{n\to\infty} T_n(1-p_n) = 0 \Rightarrow \lim_{n\to\infty} |\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)] - \theta| 1(\Omega_{n,2}) = 0 \ a.s.$$

$$\lim_{n\to\infty} T_n(1-p_n) = c \in (0,\infty) \Rightarrow \lim_{n\to\infty} \left|\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)] - \theta \frac{1 - e^{-c\eta}}{c\eta}\right| 1(\Omega_{n,2}) = 0 \ a.s.$$

$$\lim_{n\to\infty} T_n(1-p_n) = \infty \Rightarrow \lim_{n\to\infty} |\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)]| 1(\Omega_{n,2}) = 0 \ a.s.$$

Proof. See Appendix B.2.1. □

Lemma A.6. *Given (A3) and (A4),* $\lim_{n\to\infty} \mathit{Var}_n(\hat{\vartheta}_{T_n}(\phi)) 1(\Omega_{n,2}) = 0$ *a.s.*

Proof. See Appendix B.2.2. □

Before proceeding, we briefly describe our approach to proving these lemmas. First, we note that in analyzing the moments of $\hat{\vartheta}_{T_n}(\phi)$, the i.i.d. Bernoulli random variables $\hat{s}_{T_n-t}(\mathbf{i})$ in (18) are easily dealt with; the difficulty arises from the $\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}$ terms. Luckily, there is a simple interpretation of these terms that guides our analysis and that proceeds as follows. First, define a random walk $\{X_l^1\}_{l\in\mathbb{N}_0}$ with $X_0^1 = \phi$ and X_l^1 chosen uniformly from the incoming neighbors of X_{l-1}^1, for each $l \in \mathbb{N}$. Then, as shown in (37) in Appendix B.2.1,

$$\mathbb{E} \sum_{\mathbf{i}\in\hat{A}_l} \prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1} = \mathbb{P}(X_l^1 \in \hat{A}_l).$$

In short, computing the mean of $\hat{\vartheta}_{T_n}(\phi)$ amounts to computing hitting probabilities of the form $\mathbb{P}(X_l^1 \in \hat{A}_l)$. Similarly, to analyze the second moment of $\hat{\vartheta}_{T_n}(\phi)$, we compute hitting probabilities of the form $\mathbb{P}(X_l^1 \in \hat{A}_l, X_l^2 \in \hat{A}_l)$, where X_l^2 is defined in the same manner as X_l^1 and is conditionally independent of X_l^1 given the tree structure. We note that, in principal, the k-th moment of $\hat{\vartheta}_{T_n}(\phi)$ can be computed by analyzing k walks. However, the calculations become exceedingly complex as k grows, and because we only require two moments, we do not study any case $k > 2$.

This interpretation explains Lines 2 and 6-11 of Algorithm 2: in Line 2, we begin two random walks at the root node ϕ; each time Lines 6-11 are reached, we advance the random walks one step. Importantly, we simultaneously sample the walks and construct the tree. In particular, the l-th step of the walk is taken at Line 8, *before* the degrees of the corresponding node are realized in Line 13; this is crucial to our computation of the aforementioned hitting probabilities. Finally, we note that in Line 9 of Algorithm 2, the condition $j^* > d_{in}^A(\mathbf{i})$ implies the walk reaches the set of bots $\hat{B}$; since bots have self-loops but no other incoming edges, they act as absorbing states on the walk. This is why the entire future trajectory of the walk can be defined in Line 9.

In Lemmas A.7 and A.8, we compute the hitting probabilities needed for the proofs of Lemmas A.5 and A.6. We note that, in addition to the random variables $\tilde{p}_n, \tilde{p}_n^*, \tilde{q}_n$ defined in (8) in Section 3.1, Lemma A.8 requires the definition of several similar random variables; we define these in (19) (and also recall the definitions of $\tilde{p}_n, \tilde{p}_n^*, \tilde{q}_n$ for convenience). We discuss these in more detail shortly.

$$\begin{aligned}
\tilde{p}_n &= \sum_{j\in\mathbb{N},k\in\mathbb{N}_0} \frac{j}{j+k} \sum_{i\in\mathbb{N}} f_n(i,j,k) & \tilde{p}_n^* &= \sum_{j\in\mathbb{N},k\in\mathbb{N}_0} \frac{j}{j+k} \sum_{i\in\mathbb{N}} f_n^*(i,j,k) \\
\tilde{q}_n &= \sum_{j\in\mathbb{N},k\in\mathbb{N}_0} \frac{j}{j+k} \frac{1}{j+k} \sum_{i\in\mathbb{N}} f_n(i,j,k) & \tilde{q}_n^* &= \sum_{j\in\mathbb{N},k\in\mathbb{N}_0} \frac{j}{j+k} \frac{1}{j+k} \sum_{i\in\mathbb{N}} f_n^*(i,j,k) \\
\tilde{r}_n &= \sum_{j\in\mathbb{N},k\in\mathbb{N}_0} \frac{j}{j+k} \frac{j-1}{j+k} \sum_{i\in\mathbb{N}} f_n(i,j,k) & \tilde{r}_n^* &= \sum_{j\in\mathbb{N},k\in\mathbb{N}_0} \frac{j}{j+k} \frac{j-1}{j+k} \sum_{i\in\mathbb{N}} f_n^*(i,j,k)
\end{aligned} \tag{19}$$

Lemma A.7. *We have*

$$\mathbb{P}_n(X_l^1 \in \hat{A}) = \begin{cases} \tilde{p}_n^* \tilde{p}_n^{l-1}, & l \in \mathbb{N} \\ 1, & l = 0 \end{cases}.$$

Proof. See Appendix B.2.4. □

Lemma A.8. *For $l' > l$, we have*

$$\mathbb{P}_n(X_l^1 \in \hat{A}, X_{l'}^2 \in \hat{A}) = \begin{cases} \mathbb{P}_n(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}) \tilde{p}_n^{l'-l}, & l \in \mathbb{N} \\ \tilde{p}_n^* \tilde{p}_n^{l'-1}, & l = 0 \end{cases}.$$

Furthermore,

$$\mathbb{P}_n(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}) = \begin{cases} \tilde{r}_n^* \tilde{p}_n^{2(l-1)} + \sum_{t=2}^{l} \tilde{q}_n^* \tilde{q}_n^{t-2} \tilde{r}_n \tilde{p}_n^{2(l-t)} + \tilde{q}_n^* \tilde{q}_n^{l-1}, & l \in \{2,3,\dots\} \\ \tilde{r}_n^* + \tilde{q}_n^*, & l = 1 \\ 1, & l = 0 \end{cases} . \tag{20}$$

Proof. See Appendix B.2.5. □

Before proceeding, we comment on the form of (20), which helps explain the definitions in (19). Namely, in (20), $\tilde{r}_n^* \tilde{p}_n^{2(l-1)}$ is the probability of the two random walks visiting different agents on the first step of the walk ($\tilde{r}_n^*$ term), then separately remaining in the agent set for the next $l-1$ steps of the walk ($\tilde{p}_n^{2(l-1)}$ term); similarly, $\tilde{q}_n^* \tilde{q}_n^{t-2} \tilde{r}_n \tilde{p}_n^{2(l-t)}$ is the probability of the walks visiting the same agents for $t-1$ steps ($\tilde{q}_n^* \tilde{q}_n^{t-2}$ term), then visiting a different agent on the t-th step ($\tilde{r}_n$ term), then separately remaining in the agent set for $l-t$ steps ($\tilde{p}_n^{2(l-t)}$ term); finally, $\tilde{q}_n^* \tilde{q}_n^{l-1}$ is the probability of the walks remaining together and in the agent set for l steps. Each of these arguments follows from (19): $\tilde{p}_n$ gives the probability of a single walk proceeding to an agent ($j/(j+k)$ term), $\tilde{q}_n$ gives the probability of two walks proceeding to the same agent ($j/(j+k)$ term for the first walk, $1/(j+k)$ term for the second walk), and $\tilde{r}_n$ gives the probability of two walks proceeding to different agents ($j/(j+k)$ term for the first walk, $(j-1)/(j+k)$ term for the second walk). Similar arguments apply to $\tilde{p}_n^*, \tilde{q}_n^*, \tilde{r}_n^*$, except these pertain to the first steps of the walks.

Equipped with Lemmas A.5 and A.6, we can prove Theorem 3.1. First, suppose $T_n(1-p_n) \to 0$. Given $\epsilon > 0$, we can use Lemma A.3 to obtain (provided the limits exist)

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(|\theta_{T_n}(i^*) - \theta| > \epsilon) &\leq \lim_{n\to\infty} \left(\mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - \theta| > \epsilon/2) + \mathbb{P}(\Omega_{n,1}^C) + O\left(n^{\zeta - 1/2}\right)\right) \\ &= \lim_{n\to\infty} \mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - \theta| > \epsilon/2), \end{aligned} \tag{21}$$

where we have used $\mathbb{P}(\Omega_{n,1}^C) \to 0$ by (A1) and $\zeta < 1/2$ by (A2). Next, using total probability,

$$\mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - \theta| > \epsilon/2) \leq \mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - \theta| > \epsilon/2, \Omega_{n,2}) + \mathbb{P}(\Omega_{n,2}^C). \tag{22}$$

We can further expand the first summand in (22) as

$$\begin{aligned} \mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - \theta| > \epsilon/2, \Omega_{n,2}) &\leq \mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}_n \hat{\vartheta}_{T_n}(\phi)| + |\mathbb{E}_n \hat{\vartheta}_{T_n}(\phi) - \theta| > \epsilon/2, \Omega_{n,2}) \\ &\leq \mathbb{P}\left(|\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}_n \hat{\vartheta}_{T_n}(\phi)| > \frac{\epsilon}{4}, \Omega_{n,2}\right) + \mathbb{P}\left(|\mathbb{E}_n \hat{\vartheta}_{T_n}(\phi) - \theta| > \frac{\epsilon}{4}, \Omega_{n,2}\right), \end{aligned} \tag{23}$$

where we have simply used the triangle inequality and the union bound. Now for the first summand in (23), we have (via total expectation and the conditional form of Chebyshev's inequality)

$$\begin{aligned} \mathbb{P}\left(|\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}_n \hat{\vartheta}_{T_n}(\phi)| > \frac{\epsilon}{4}, \Omega_{n,2}\right) &= \mathbb{E}\left[\mathbb{P}_n\left(|\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}_n \hat{\vartheta}_{T_n}(\phi)| > \frac{\epsilon}{4}\right) 1(\Omega_{n,2})\right] \\ &\leq \frac{16}{\epsilon^2} \mathbb{E}\left[\mathrm{Var}_n(\hat{\vartheta}_{T_n}(\phi)) 1(\Omega_{n,2})\right] \xrightarrow[n\to\infty]{} 0, \end{aligned} \tag{24}$$

where the limit holds by Lemma A.6. For second summand in (23), we write

$$\begin{aligned} \mathbb{P}\left(|\mathbb{E}_n \hat{\vartheta}_{T_n}(\phi) - \theta| > \frac{\epsilon}{4}, \Omega_{n,2}\right) &= \mathbb{E}\left[1\left(|\mathbb{E}_n \hat{\vartheta}_{T_n}(\phi) - \theta| > \frac{\epsilon}{4}\right) 1(\Omega_{n,2})\right] \\ &\leq \frac{4}{\epsilon} \mathbb{E}[|\mathbb{E}_n \hat{\vartheta}_{T_n}(\phi) - \theta| 1(\Omega_{n,2})] \xrightarrow[n\to\infty]{} 0, \end{aligned} \tag{25}$$

where the first two lines use total expectation and the inequality $1(x > y) \leq x/y$ for $x, y > 0$ (which is easily proven by considering the cases $x > y$ and $x \leq y$), and the limit holds by Lemma A.5.

Finally, combining (21), (22), (23), (24), and (25), and recalling that $\mathbb{P}(\Omega_{n,2}^C) \to 0$ by (A3), we obtain

$$0 \le \lim_{n\to\infty} \mathbb{P}(|\theta_{T_n}(i^*) - \theta| > \epsilon) \le \lim_{n\to\infty} \mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - \theta| > \epsilon/2) = 0.$$

Since $\epsilon > 0$ was arbitrary, we conclude that $\theta_{T_n}(i^*)$ converges to θ in probability, completing the proof in the case $T_n(1-p_n) \to 0$. For the cases $T_n(1-p_n) \to c \in (0,\infty)$ and $T_n(1-p_n) \to \infty$, respectively, we can replace θ with $\theta(1-e^{-c\eta})/(c\eta)$ and 0, respectively (the corresponding cases from Lemma A.5), but otherwise follow the same approach.

A.3 Step 2 for proof of Theorem 3.2

Similar to the second step in the proof of Theorem 3.1, we begin by analyzing the limiting behavior of $\hat{\vartheta}_{T_n}(\phi)$. However, we will use a different approach than that used in Theorem 3.1. This approach is made possible by the stronger assumptions of Theorem 3.2, and it will yield a fast rate of convergence that will allow us to prove the theorem.

To explain our approach, we first recall that Lemma A.4 states

$$\hat{\vartheta}_{T_n}(\phi) = \frac{1}{T_n} \sum_{t=0}^{T_n-1} \sum_{l=0}^{t} \binom{t}{l} \eta^l (1-\eta)^{t-l} \sum_{\mathbf{i}\in\hat{A}_l} \hat{s}_{T_n-t}(\mathbf{i}) \prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}.$$

Hence, letting $\mathcal{T}$ denote the collection of random variables defining the tree structure,

$$\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] = \frac{1}{T_n} \sum_{t=0}^{T_n-1} \sum_{l=0}^{t} \binom{t}{l} \eta^l (1-\eta)^{t-l} \sum_{\mathbf{i}\in\hat{A}_l} \mathbb{E}[\hat{s}_{T_n-t}(\mathbf{i})|\mathcal{T}] \prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1} \tag{26}$$

$$= \frac{\theta}{T_n} \sum_{t=0}^{T_n-1} \sum_{l=0}^{t} \binom{t}{l} \eta^l (1-\eta)^{t-l} \sum_{\mathbf{i}\in\hat{A}_l} \prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1},$$

where we have simply used the fact that the signals are i.i.d. Bernoulli(θ) random variables. Our basic approach will now proceed in two steps. First, in Lemma A.9 we condition on the tree structure, so that $\hat{\vartheta}_{T_n}(\phi)$ is simply a weighted sum of i.i.d. Bernoulli(θ) random variables; the lemma shows that this weighted sum is close to its conditional mean $\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]$ with high probability. Second, in Lemma A.10, we show that the conditional mean $\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]$ converges to zero in probability. Before proceeding, we also note that an argument similar to (17) implies

$$0 \le \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] \le \theta \ a.s., \tag{27}$$

which will be used in the proofs of the lemmas in this section.

We now state Lemma A.9. As mentioned, the proof involves analyzing a weighted sum of i.i.d. random variables; hence, our analysis is similar to the derivation of Hoeffding's inequality.

Lemma A.9. *Assume $\exists\, \mu > 0$ and $N' \in \mathbb{N}$ independent of n s.t. the following hold:*

- *(A4), with $T_n \ge \mu \log n \ \forall\ n \ge N'$.*

Then $\forall\ \epsilon > 0$,

$$\mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]| > \epsilon) = O\left(n^{-2\epsilon^2\mu}\right).$$

Proof. See Appendix B.3.1. □

Lemma A.10 states that conditional mean $\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]$ converges to zero in probability. Note that the only source of randomness in $\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]$ is the tree structure. Since the tree structure is generated recursively, $\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]$ has a martingale-like structure; this allows us to use an approach similar to the Azuma-Hoeffding inequality for bounded-difference martingales.

Lemma A.10. *Assume* $\exists\,\kappa, \mu > 0$ *and* $N' \in \mathbb{N}$ *independent of* n *s.t. the following hold:*

- *(A3), with* $P(\Omega_{n,2}) = O(n^{-\kappa})$ *and* $p < 1$.
- *(A4), with* $T_n \geq \mu \log n \;\forall\; n \geq N'$.

Then $\forall\, \epsilon > 0$,

$$\mathbb{P}(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] > \epsilon) = O\left(n^{-\min\{\mu(\epsilon\eta(1-p)/\theta)^2,\kappa\}}\right).$$

Proof. See Appendix B.3.2. □

With Lemmas A.9 and A.10 in place, we can prove Theorem 3.2. First, since $\theta_{T_n}(i^*), \hat{\vartheta}_{T_n}(\phi) \geq 0$, taking $x = 0$ in Lemma A.3 yields

$$\begin{aligned}\mathbb{P}(\theta_{T_n}(i^*) > \epsilon) &\leq \mathbb{P}(\hat{\vartheta}_{T_n}(\phi) > \epsilon/2) + \mathbb{P}(\Omega_{n,1}^C) + O\left(n^{\zeta-1/2}\right) \\ &= \mathbb{P}(\hat{\vartheta}_{T_n}(\phi) > \epsilon/2) + O\left(n^{-\kappa}\right) + O\left(n^{\zeta-1/2}\right),\end{aligned} \tag{28}$$

where the equality is by the theorem assumptions. For the first summand in (28), we write

$$\begin{aligned}\mathbb{P}(\hat{\vartheta}_{T_n}(\phi) > \epsilon/2) &= \mathbb{P}((\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]) + \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] > \epsilon/2) \\ &\leq \mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]| + \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] > \epsilon/2) \\ &\leq \mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]| > \epsilon/4) + \mathbb{P}(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] > \epsilon/4) \\ &= O\left(n^{-\epsilon^2\mu/8} + n^{-\min\{\mu(\epsilon\eta(1-p)/\theta)^2/16,\kappa\}}\right) = O\left(n^{-\min\{\mu(\epsilon\eta(1-p)/\theta)^2/16,\kappa\}}\right),\end{aligned}$$

where the first equality adds and subtracts a term, the first inequality is immediate, the second inequality uses the union bound, the second equality uses Lemmas A.9 and A.10, and the final equality holds since $\eta, p \in (0,1)$ implies $\epsilon^2\mu/8 > \mu(\epsilon\eta(1-p)/\theta)^2/16$. Substituting into (28),

$$\mathbb{P}(\theta_{T_n}(i^*) > \epsilon) = O\left(n^{-\min\{(1/2)-\zeta,\mu(\epsilon\eta(1-p)/\theta)^2/16,\kappa\}}\right). \tag{29}$$

We can then write

$$\begin{aligned}\mathbb{E}\left|\left\{i \in [n] : \theta_{T_n}(i) > \epsilon\right\}\right| &= \sum_{i\in[n]} \mathbb{E}1(\theta_{T_n}(i) > \epsilon) = \sum_{i\in[n]} \mathbb{P}(\theta_{T_n}(i) > \epsilon) \\ &= n\mathbb{P}(\theta_{T_n}(i^*) > \epsilon) = O\left(n^{1-\min\{(1/2)-\zeta,\mu(\epsilon\eta(1-p)/\theta)^2/16,\kappa\}}\right),\end{aligned}$$

where we have used (29). Hence, by Markov's inequality,

$$\begin{aligned}\mathbb{P}\left(\left|\left\{i \in [n] : \theta_{T_n}(i) > \epsilon\right\}\right| > Kn^k\right) &\leq K^{-1}n^{-k}\mathbb{E}\left|\left\{i \in [n] : \theta_{T_n}(i) > \epsilon\right\}\right| \\ &= O\left(n^{-k+(1-\min\{(1/2)-\zeta,\mu(\epsilon\eta(1-p)/\theta)^2/16,\kappa\})}\right) \xrightarrow[n\to\infty]{} 0,\end{aligned}$$

where the limit holds by the assumption on k in the statement of the theorem.

A.4 Other remarks

A.4.1 A sufficient condition for extending Theorem 3.2. Here we show that the condition (11) from Section 3.3 is sufficient to extend Theorem 3.2 to other cases of p_n. Recall this condition is

$$\exists\, \gamma' > 0 \; s.t. \; \mathbb{P}(|\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] - L(p_n)| > \epsilon) = O\left(n^{-\gamma'}\right), \tag{30}$$

where $L(p_n)$ is the limit from Theorem 3.1 based on the relative asymptotics of T_n and p_n, i.e.

$$L(p_n) = \begin{cases} \theta, & T_n(1-p_n) \xrightarrow[n\to\infty]{} 0 \\ \theta(1-e^{-c\eta})/(c\eta), & T_n(1-p_n) \xrightarrow[n\to\infty]{} c \in (0,\infty) \\ 0, & T_n(1-p_n) \xrightarrow[n\to\infty]{} \infty \end{cases}. \tag{31}$$

Suppose (30) holds in the case $T_n(1-p_n) \to 0$, so that $L(p_n) = \theta$. In this case, we have

$$\begin{aligned} \mathbb{P}(|\theta_{T_n}(i^*) - \theta| > \epsilon) &\le \mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - \theta| > \epsilon/2) + O\left(n^{-\min\{\kappa,(1/2)-\zeta\}}\right) \\ &\le \mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]| > \epsilon/4) + \mathbb{P}(|\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] - \theta| > \epsilon/4) + O\left(n^{-\min\{\kappa,(1/2)-\zeta\}}\right) \\ &\le O\left(n^{-\epsilon^2\mu/8}\right) + O\left(n^{-\gamma'}\right) + O\left(n^{-\min\{\kappa,(1/2)-\zeta\}}\right) = O\left(n^{-\min\{\epsilon^2\mu/8,\gamma',\kappa,(1/2)-\zeta\}}\right), \end{aligned}$$

where the first inequality is Lemma A.3 (which holds for all cases of p_n) with $\mathbb{P}(\Omega_{n,1}) = O(n^{-\kappa})$ and the third uses Lemma A.9 (which holds for all cases of p_n) and the sufficient condition (30). Hence, by the argument following (29), we obtain for any $\epsilon > 0, K > 0$, and $k' > 1-\min\{\epsilon^2\mu/8, \gamma', \kappa, (1/2)-\zeta\}$,

$$\mathbb{P}\left(|\{i \in [n] : |\theta_{T_n}(i) - \theta| > \epsilon\}| > Kn^{k'}\right) \xrightarrow[n\to\infty]{} 0,$$

i.e. Theorem 3.2 holds with k replaced by k'. The same argument shows that Theorem 3.2 holds (with only a change of k) in the cases $T_n(1-p_n) \to c \in (0,\infty)$ and $T_n(1-p_n) \to \infty$ with $p_n \to 1$.

A.4.2 Comparing Step 2 for proofs of Theorems 3.1 and 3.2. As shown in Appendices A.2 and A.3, Step 2 for the proofs of both theorems involves bounding $\mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - L(p_n)| > \epsilon/2)$ for the appropriate $L(p_n)$. One may wonder why we have conducted a different analysis for the two theorems. The reason is that, as shown in Appendix B.3.3, the analysis for Step 2 of Theorem 3.2 yields a bound that does not decay with n in the case $T_n(1-p_n) \to c \in [0,\infty)$. Hence, we have derived a bound for Theorem 3.1 that encompasses all cases of $\lim_{n\to\infty} T_n(1-p_n)$. On the other hand, the bound from Theorem 3.1 only states $\mathbb{P}(|\hat{\vartheta}_{T_n}(\phi) - L(p_n)| > \epsilon/2) \to 0$ but does not provide a rate of convergence so cannot be used to prove Theorem 3.2. We also note Appendix B.3.3 shows that, while the bound for Step 2 of Theorem 3.2 *does* decay in n for the case $T_n(1-p_n) \to \infty$ with $p_n \to 1$, it does not decay quickly enough to establish (11).

APPENDIX B PROOF DETAILS

B.1 Branching process approximation (Step 1 for proofs of Theorems 3.1 and 3.2)

B.1.1 Proof of Lemma A.1. For $t \in \mathbb{N}_0$, let α_t, β_t denote the parameters $\{\alpha_t(i)\}_{i\in A\cup B}, \{\beta_t(i)\}_{i\in A\cup B}$ in vector form, and let $\mathbf{1}$ denote the all ones vector. We claim

$$\alpha_t = (1-\eta)\sum_{\tau=1}^{t} s_\tau Q^{t-\tau} + \alpha_0 Q^t, \quad \beta_t = (1-\eta)\sum_{\tau=1}^{t}(\mathbf{1} - s_\tau)Q^{t-\tau} + \beta_0 Q^t \quad \forall\, t \in \mathbb{N}. \tag{32}$$

We prove (32) for α_t; the proof for β_t follows the same approach. First, we use the parameter update equations (6), and the definitions of P and Q from Appendix A.1 (P being the column-normalized adjacency matrix and $Q = (1-\eta)I + \eta P$) to write the parameter update equation in vector form as

$$\alpha_t = (1-\eta)(\alpha_t + s_t) + \eta\alpha_{t-1}P = (1-\eta)s_t + \alpha_{t-1}Q. \tag{33}$$

We next use induction. For $t = 1$, (32) is equivalent to (33). Assuming (32) holds for $t-1$, we have

$$\begin{aligned}\alpha_t &= (1-\eta)s_t + \alpha_{t-1}Q = (1-\eta)s_t + \left((1-\eta)\sum_{\tau=1}^{t-1} s_\tau Q^{(t-1)-\tau} + \alpha_0 Q^{t-1}\right)Q \\ &= (1-\eta)s_t + (1-\eta)\sum_{\tau=1}^{t-1} s_\tau Q^{t-\tau} + \alpha_0 Q^t = (1-\eta)\sum_{\tau=1}^{t} s_\tau Q^{t-\tau} + \alpha_0 Q^t,\end{aligned}$$

which completes the proof. Next, recalling e_i is the vector with 1 in the i-th position and 0 elsewhere,

$$\begin{aligned}\theta_{T_n}(i) &= \frac{\alpha_{T_n}(i)}{\alpha_{T_n}(i)+\beta_{T_n}(i)} = \frac{(1-\eta)\sum_{\tau=1}^{T_n} s_\tau Q^{T_n-\tau}e_i^\mathsf{T} + \alpha_0 Q^{T_n}e_i^\mathsf{T}}{(1-\eta)\sum_{\tau=1}^{T_n} \mathbf{1}Q^{T_n-\tau}e_i^\mathsf{T} + (\alpha_0+\beta_0)Q^{T_n}e_i^\mathsf{T}} \\ &= \frac{(1-\eta)\sum_{\tau=1}^{T_n} s_\tau Q^{T_n-\tau}e_i^\mathsf{T} + \alpha_0 Q^{T_n}e_i^\mathsf{T}}{(1-\eta)T_n + (\alpha_0+\beta_0)Q^{T_n}e_i^\mathsf{T}} = \frac{\frac{1}{T_n}\sum_{\tau=1}^{T_n} s_\tau Q^{T_n-\tau}e_i^\mathsf{T} + \frac{1}{(1-\eta)T_n}\alpha_0 Q^{T_n}e_i^\mathsf{T}}{1 + \frac{1}{(1-\eta)T_n}(\alpha_0+\beta_0)Q^{T_n}e_i^\mathsf{T}},\end{aligned}$$

where the equalities hold by definition, by (32), since the columns of Q sum to 1 by definition, and by multiplying numerator and denominator by $\frac{1}{(1-\eta)T_n}$, respectively. Next, recall from Section 2.1 that $\alpha_0(j) \in [0, \bar{\alpha}]\ \forall\ j \in A \cup B$ for some $\bar{\alpha} > 0$. Hence, α_0 is element-wise upper bounded by $\bar{\alpha}\mathbf{1}$, so $\alpha_0 Q^{T_n}e_i^\mathsf{T} \leq \bar{\alpha}\mathbf{1}Q^{T_n}e_i^\mathsf{T} = \bar{\alpha}$, where we have used column stochasticity of Q. Additionally, $\alpha_0 Q^{T_n}e_i^\mathsf{T} \geq 0$ (since the three terms in the product are elementwise nonnegative). By a similar argument, $0 \leq \beta_0 Q^{T_n}e_i^\mathsf{T} \leq \bar{\beta}$. Taken together, we can use the previous equation to obtain

$$\frac{\frac{1}{T_n}\sum_{\tau=1}^{T_n} s_\tau Q^{T_n-\tau}e_i^\mathsf{T}}{1+\frac{\bar{\alpha}+\bar{\beta}}{(1-\eta)T_n}} \leq \theta_{T_n}(i) \leq \frac{1}{T_n}\sum_{\tau=1}^{T_n} s_\tau Q^{T_n-\tau}e_i^\mathsf{T} + \frac{\bar{\alpha}}{(1-\eta)T_n}.$$

Finally, recall from Section 2.1 that $\bar{\alpha}$ and $\bar{\beta}$ are independent of n. Hence, because $T_n \to \infty$ as $n \to \infty$ (by (A4) in the statement of the lemma), $\bar{\alpha}/T_n, \bar{\beta}/T_n \to 0$ as $n \to \infty$. It follows that, for given $\epsilon > 0$ and n sufficiently large, $|\theta_{T_n}(i) - \frac{1}{T_n}\sum_{\tau=1}^{T_n} s_\tau Q^{T_n-\tau}e_i^\mathsf{T}| < \epsilon$. Finally, by changing the index of summation, it is clear that $\frac{1}{T_n}\sum_{\tau=1}^{T_n} s_\tau Q^{T_n-\tau}e_i^\mathsf{T} = \vartheta_{T_n}(i)$, completing the proof.

B.1.2 Proof of Lemma A.2. We begin by arguing $\vartheta_{T_n}(i^*)|\{\tau_n > T_n\} \overset{\mathscr{D}}{=} \hat{\vartheta}_{T_n}(\phi)$. For this, first consider the sub-graph containing only edges between two agents formed during the first T_n iterations of Algorithm 1. Conditioned on $\tau_n > T_n$, this sub-graph is constructed as follows:

- The initial agent i^* is sampled uniformly from A (Line 2), which implies its degrees ($d_{out}(i^*)$, $d_{in}^A(i^*), d_{in}^B(i^*)$) are distributed as f_n^*. (In fact, this holds even if $\tau_n \leq T_n$.)
- Each time an edge is added to the sub-graph (Line 10), the paired outstub (i', j') is sampled uniformly from O_A (else, $\tau_n > T_n$ is contradicted by Line 8-9), so the degrees ($d_{out}(i'), d_{in}^A(i')$, $d_{in}^B(i')$) of the corresponding agent i' are distributed as f_n.
- The initial agent i^* has no paired outstubs, while all other agents in the sub-graph have one paired outstub (otherwise, an outstub with label 2 was paired within the first T_n iterations, contradicting $\tau_n > T_n$ by Line 8); in particular, the sub-graph has $|\cup_{l=0}^{T_n} A_l|$ nodes and $|\cup_{l=0}^{T_n} A_l| - 1$ edges. Also, every agent in the sub-graph has a path to i^* by the breadth-first-search nature of the construction, so, neglecting edge polarities, we obtain a connected graph with $|\cup_{l=0}^{T_n} A_l|$ nodes and $|\cup_{l=0}^{T_n} A_l| - 1$ edges, i.e. a tree. Finally, since all edges point towards i^* (see Line 10), the sub-graph is a directed tree pointed towards i^*.

In summary, the sub-graph is a directed tree pointing towards an agent with degrees distributed as f_n^*, in which all other nodes have degrees distributed as f_n. This is precisely the procedure used to construct the sub-graph of agents during the first T_n iterations of Algorithm 2. Additionally, Algorithms 1 and 2 add bots in the same manner (Lines 15-16 in Algorithm 1, Lines 17-18 in Algorithm 2). Taken together, we conclude that, conditioned on $\tau_n > T_n$, the T_n-step neighborhood into i^* is constructed in the same manner in Algorithm 1 as the T_n-step neighborhood into ϕ is constructed in Algorithm 2. Furthermore, by (14) and (16), it is clear that $\vartheta_{T_n}(i)$ and $\hat{\vartheta}_{T_n}(\phi)$, respectively, depend only on these respective neighborhoods, and on the signals $s_{T_n-t}(i)$ and $\hat{s}_{T_n-t}(\mathbf{i})$, respectively. Since the signals $s_{T_n-t}(i)$ and $\hat{s}_{T_n-t}(\mathbf{i})$ are also defined in the same manner ($s_{T_n-t}(i), \hat{s}_{T_n-t}(\mathbf{i}) \sim \text{Bernoulli}(\theta)$ for $i \in A, \mathbf{i} \in \hat{A}$; $s_{T_n-t}(i) = \hat{s}_{T_n-t}(\mathbf{i}) = 0$ for $i \in B, \mathbf{i} \in \hat{B}$), we ultimately conclude that $\vartheta_{T_n}(i^*)$ and $\hat{\vartheta}_{T_n}(\phi)$ have the same distribution when $\tau_n > T_n$ holds.

We next argue $\{\tau_n > T_n\}$ occurs with high probability when $\Omega_{n,1}$ holds. For this, we note that Algorithm 1 is nearly identical to the graph construction described in Section 5.2 of [9]. More specifically, the only difference is that the construction in [9] does not include the pairing of agent instubs with bots in Lines 15-16 of Algorithm 1. However, these lines do not affect τ_n. Moreover, when (A1) holds, the assumptions of Lemma 5.4 from [9] are satisfied. This lemma states that, if $t_n < (\log n)/(2\log(\nu_3/\nu_1))$ and $\nu_3 > \nu_1$ (with ν_1, ν_3 defined as in (A1)), then $P(\tau_n \leq t_n|\Omega_{n,1}) = O((\nu_3/\nu_1)^{t_n}/\sqrt{n})$. In particular, by (A2) we have $T_n \leq \zeta \log(n)/\log(\nu_3/\nu_1)$ for n sufficiently large, with $\zeta \in (0, 1/2)$ independent of n; substituting gives

$$\mathbb{P}(\tau_n \leq T_n|\Omega_{n,1}) = O\left(\frac{(\nu_3/\nu_1)^{\zeta \log(n)/\log(\nu_3/\nu_1)}}{\sqrt{n}}\right) = O\left(n^{\zeta-1/2}\right).$$

B.1.3 Proof of Lemma A.4. We first claim that for $l \in \mathbb{N}_0$ and $\mathbf{i} \in \hat{A}_l$,

$$e_{\mathbf{i}}\hat{P}^{l'}e_\phi = \begin{cases} \prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}, & l' = l \\ 0, & l' \in \mathbb{N}_0 \setminus \{l\} \end{cases}. \tag{34}$$

(Recall $\hat{P}$ is the column-normalized adjacency matrix.) We prove (34) separately for $l = 0$ and $l \in \mathbb{N}$. When $l = 0$, the only case is $\mathbf{i} = \phi$ (since $\hat{A}_0 = \{\phi\}$); if $l' = 0$, the left side is clearly 1 and the right side is 1 by convention; if $l' \in \mathbb{N}$, the left side is 0 since $e_\phi \hat{P}^{l'} = 0$ (ϕ has no outgoing neighbors in the tree). Next, we aim to prove (34) for $\mathbf{i} \in \hat{A}_l$ and $l \in \mathbb{N}$. For such $\mathbf{i}$, there is a unique path from $\mathbf{i}$ to ϕ with length l that visits the nodes $\mathbf{i}|l = \mathbf{i}, \mathbf{i}|l-1, \ldots, \mathbf{i}|0 = \phi$. By definition of $\hat{P}$, it follows that

$$e_{\mathbf{i}}\hat{P}^l e_\phi = \hat{P}(\mathbf{i}|l, \mathbf{i}|l-1)\hat{P}(\mathbf{i}|l-1, \mathbf{i}|l-2)\cdots\hat{P}(\mathbf{i}|1, \mathbf{i}|0) = \frac{1}{d_{in}(\mathbf{i}|l-1)}\frac{1}{d_{in}(\mathbf{i}|l-2)}\cdots\frac{1}{d_{in}(\phi)}.$$

On the other hand, if $l' \neq l$, no path of length l' from $\mathbf{i}$ to ϕ exists, so $e_{\mathbf{i}}\hat{P}^{l'}e_\phi = 0$. This proves (34).

Recalling that $\hat{Q} = (1-\eta)I + \eta\hat{P}$, we next claim that $\forall\ t \in \mathbb{N}_0$,

$$\hat{Q}^t = \sum_{l=0}^{t}\binom{t}{l}\eta^l(1-\eta)^{t-l}\hat{P}^l. \tag{35}$$

We prove (35) inductively: both sides equal I when $t = 0$; assuming (35) is true for t, we have

$$\begin{aligned}\hat{Q}^{t+1} &= ((1-\eta)I + \eta\hat{P})\sum_{l=0}^{t}\binom{t}{l}\eta^l(1-\eta)^{t-l}\hat{P}^l \\ &= \sum_{l=0}^{t}\binom{t}{l}\eta^l(1-\eta)^{t+1-l}\hat{P}^l + \sum_{l=1}^{t+1}\binom{t}{l-1}\eta^l(1-\eta)^{t+1-l}\hat{P}^l \\ &= (1-\eta)^{t+1}I + \sum_{l=1}^{t}\left(\binom{t}{l}+\binom{t}{l-1}\right)\eta^l(1-\eta)^{t+1-l}\hat{P}^l + \eta^{t+1}\hat{P}^{t+1} \\ &= (1-\eta)^{t+1}I + \sum_{l=1}^{t}\binom{t+1}{l}\eta^l(1-\eta)^{t+1-l}\hat{P}^l + \eta^{t+1}\hat{P}^{t+1},\end{aligned}$$

where in the first line we have used the definition of $\hat{Q}$ and the inductive hypothesis, the second line simply uses the distributive property, the third rearranges summations, and the fourth uses Pascal's rule ($[t+1]$ has $\binom{t+1}{l}$ subsets of cardinality l; $\binom{t}{l-1}$ that contain 1 and $\binom{t}{l}$ that do not contain 1). This completes the proof of (35).

Having established (35) and (34), we can combine them to obtain $\forall\, t \in \mathbb{N}_0, \mathbf{i} \in \hat{A}_l$,

$$e_{\mathbf{i}}\hat{Q}^t e_\phi = \begin{cases}\binom{t}{l}\eta^l(1-\eta)^{t-l}\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}, & l \le t \\ 0, & l > t\end{cases}.$$

Finally, substituting the previous equation into (16), and recalling $\hat{A} = \cup_{l=0}^{\infty}\hat{A}_l$, we obtain

$$\hat{\vartheta}_{T_n}(\phi) = \frac{1}{T_n}\sum_{t=0}^{T_n-1}\sum_{l=0}^{t}\sum_{\mathbf{i}\in\hat{A}_l}\hat{s}_{T_n-t}(\mathbf{i})\binom{t}{l}\eta^l(1-\eta)^{t-l}\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1},$$

which completes the proof.

B.2 Step 2 for proof of Theorem 3.1

B.2.1 Proof of Lemma A.5. First, letting $\mathcal{D}$ denote the degree sequence and $\mathcal{T}$ denote the set of random variables defining the tree structure, we can use Lemma A.4 to write

$$\begin{aligned}\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)] &= \frac{1}{T_n}\sum_{t=0}^{T_n-1}\sum_{l=0}^{t}\binom{t}{l}\eta^l(1-\eta)^{t-l}\mathbb{E}_n\left[\sum_{\mathbf{i}\in\hat{A}_l}\mathbb{E}[\hat{s}_{T_n-t}(\mathbf{i})|\mathcal{D},\mathcal{T}]\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}\right] \\ &= \frac{\theta}{T_n}\sum_{t=0}^{T_n-1}\sum_{l=0}^{t}\binom{t}{l}\eta^l(1-\eta)^{t-l}\mathbb{E}_n\left[\sum_{\mathbf{i}\in\hat{A}_l}\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}\right] = \frac{\theta}{T_n}\sum_{t=0}^{T_n-1}\sum_{l=0}^{t}\binom{t}{l}\eta^l(1-\eta)^{t-l}\mathbb{P}_n(X_l^1 \in \hat{A})\end{aligned} \tag{36}$$

where the first equality uses the tower property of conditional expectation and the fact that $\hat{A}_l$ and $d(\mathbf{i}|j)^{-1}$ are fixed given the tree structure, the second uses the fact that $\hat{s}_{T_n-t}(\mathbf{i}) \sim$ Bernoulli(θ), and the third holds by the tower property and the definition of X_l^1, i.e.

$$\mathbb{P}_n(X_l^1 \in \hat{A}) = \mathbb{E}_n[\mathbb{P}(X_l^1 \in \hat{A}|\mathcal{D},\mathcal{T})] = \mathbb{E}_n\left[\sum_{\mathbf{i}\in\hat{A}_l}\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}\right]. \tag{37}$$

Here we have also used the fact that $\{X_l^1\}_{l\in\mathbb{N}}$ is a random walk starting at the root of a directed tree; hence, for $\mathbf{i} \in \hat{A}_l$, $\mathbb{P}(X_l^1 = \mathbf{i}|\mathcal{D},\mathcal{T})$ is the probability of the lone path from ϕ to $\mathbf{i}$, which is

$\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}$, and $X_l^1 \in \hat{A} \Leftrightarrow X_l^1 = \mathbf{i}$ for some $\mathbf{i} \in \hat{A}_l$. Next, using (36) and Lemma A.7, we obtain

$$\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)] = \frac{\theta}{T_n} \sum_{t=0}^{T_n-1} \left(\sum_{l=1}^{t} \binom{t}{l} \eta^l (1-\eta)^{t-l} \tilde{p}_n^* \tilde{p}_n^{l-1} + (1-\eta)^t \right), \tag{38}$$

where by convention the summation over l is zero when $t = 0$. Adding and subtracting $(1-\eta)^t \tilde{p}_n^*/\tilde{p}_n$, the previous equation can be rewritten as

$$\begin{aligned}
\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)] &= \frac{\theta}{T_n} \frac{\tilde{p}_n^*}{\tilde{p}_n} \sum_{t=0}^{T_n-1} \sum_{l=0}^{t} \binom{t}{l} (\eta \tilde{p}_n)^l (1-\eta)^{t-l} + \frac{\theta}{T_n} \left(1 - \frac{\tilde{p}_n^*}{\tilde{p}_n}\right) \sum_{t=0}^{T_n-1} (1-\eta)^t \\
&= \frac{\theta}{T_n} \frac{\tilde{p}_n^*}{\tilde{p}_n} \sum_{t=0}^{T_n-1} (1-\eta(1-\tilde{p}_n))^t + \frac{\theta}{T_n} \left(1 - \frac{\tilde{p}_n^*}{\tilde{p}_n}\right) \frac{1-(1-\eta)^{T_n}}{\eta} \\
&= \frac{\theta}{T_n} \frac{\tilde{p}_n^*}{\tilde{p}_n} \frac{1-(1-\eta(1-\tilde{p}_n))^{T_n}}{\eta(1-\tilde{p}_n)} + \frac{\theta}{T_n} \left(1 - \frac{\tilde{p}_n^*}{\tilde{p}_n}\right) \frac{1-(1-\eta)^{T_n}}{\eta},
\end{aligned}$$

where we have simply used the binomial theorem and computed two geometric series.

Next, we assume temporarily that $p_n \to 1$ as $n \to \infty$. By (A3), we have for $\omega \in \Omega_{n,2}$

$$\tilde{p}_n(\omega) \in (p_n - \delta_n, p_n + \delta_n).$$

Hence, by $p_n \to 1$, and since $\delta_n \to 0$ by (A3), we have for $\gamma_1 > 0$, n sufficiently large, and such ω

$$1 - \gamma_1 < \frac{\tilde{p}_n^*(\omega)}{\tilde{p}_n(\omega)} < 1 + \gamma_1,$$

where we have also used the fact that $1 \geq \tilde{p}_n^* \geq \tilde{p}_n$ on $\Omega_{n,2}$ by (A3). Also, by (A4), it is clear that $(1-(1-\eta))^{T_n}/T_n \to 0$, so for given $\gamma_2 > 0$ and n sufficiently large,

$$0 < \frac{\theta}{T_n} \frac{1-(1-\eta)^{T_n}}{\eta} < \gamma_2.$$

Combining the previous four equations implies that for n sufficiently large and $\omega \in \Omega_{n,2}$,

$$\begin{aligned}
\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)](\omega) &< (1+\gamma_1) \frac{\theta}{T_n} \frac{1-(1-\eta(1-p_n-\delta_n))^{T_n}}{\eta(1-p_n-\delta_n)} + \gamma_1\gamma_2, \\
\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)](\omega) &> (1-\gamma_1) \frac{\theta}{T_n} \frac{1-(1-\eta(1-p_n+\delta_n))^{T_n}}{\eta(1-p_n+\delta_n)} - \gamma_1\gamma_2.
\end{aligned} \tag{39}$$

We complete the proof for the case $T_n(1-p_n) \to 0$; the proof for the other two cases is similar. In this case, we can use Lemma B.1 from Appendix B.4 to obtain for any $\gamma_3 > 0$ and for n large enough

$$\begin{aligned}
1 - \gamma_3 &< \frac{1-(1-\eta(1-p_n-\delta_n))^{T_n}}{T_n\eta(1-p_n-\delta_n)} < 1 + \gamma_3, \\
1 - \gamma_3 &< \frac{1-(1-\eta(1-p_n+\delta_n))^{T_n}}{T_n\eta(1-p_n+\delta_n)} < 1 + \gamma_3.
\end{aligned}$$

Combining the previous two equations gives for n large and $\omega \in \Omega_{n,2}$

$$\begin{aligned}
\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)](\omega) &< \theta(1+\gamma_1)(1+\gamma_3) + \gamma_1\gamma_2 = \theta + \theta(\gamma_1+\gamma_3+\gamma_1\gamma_3) + \gamma_1\gamma_2, \\
\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)](\omega) &> \theta(1-\gamma_1)(1-\gamma_3) - \gamma_1\gamma_2 = \theta - \theta(\gamma_1+\gamma_3-\gamma_1\gamma_3) - \gamma_1\gamma_2.
\end{aligned}$$

Hence, for given $\gamma > 0$, we can find $\gamma_1, \gamma_2, \gamma_3$ sufficiently small and n sufficiently large such that, for $\omega \in \Omega_{n,2}$, $|\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)](\omega) - \theta| < \gamma$. This clearly also implies $|\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)](\omega) - \theta| 1(\Omega_{n,2})(\omega) < \gamma$ for such ω. On the other hand, for $\omega \notin \Omega_{n,2}$, it is trivial that $|\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)](\omega) - \theta|1(\Omega_{n,2})(\omega) = 0 < \gamma$. This completes the proof for the case $T_n(1-p_n) \to 0$.

We now return to the case $p_n \to p \in [0,1)$. In this case, it follows from (A4) that $T_n(1-p_n) \to [0,\infty)$ cannot occur, i.e. we need only consider the case $T_n(1-p_n) \to \infty$. First, note that since $p_n \to p < 1$ and $\delta_n \to 0$, we have $p_n + \delta_n < 1-\gamma_1$ for some $\gamma_1 > 0$ and n sufficiently large. For such n, and for $\omega \in \Omega_{n,2}$, we then obtain $\tilde{p}_n(\omega) < 1-\gamma_1$; substituting into (38) (evaluated at ω) gives

$$\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)](\omega) < \frac{\theta}{T_n}\sum_{t=0}^{T_n-1}\left(\sum_{l=1}^{t}\binom{t}{l}\eta^l(1-\eta)^{t-l}(1-\gamma_1)^{l-1} + (1-\eta)^t\right) \tag{40}$$

$$< \frac{\theta}{T_n}\frac{1}{1-\gamma_1}\sum_{t=0}^{T_n-1}\sum_{l=0}^{t}\binom{t}{l}\eta^l(1-\eta)^{t-l}(1-\gamma_1)^l = \frac{\theta}{T_n}\frac{1}{1-\gamma_1}\frac{1-(1-\eta\gamma_1)^{T_n}}{\eta\gamma_1} < \frac{\theta}{T_n}\frac{1}{1-\gamma_1}\frac{1}{\eta\gamma_1}$$

where in the first inequality we used $\tilde{p}_n(\omega) < 1-\gamma_1$ and $\tilde{p}_n^*(\omega) \le 1$, in the second we used $1-\gamma_1 \in (0,1)$ (so that $(1-\eta)^t < (1-\eta)^t/(1-\gamma_1)$), for the equality we used the binomial theorem and computed a geometric series, and the final inequality is immediate. Since θ, η, γ_1 are independent of n, while $T_n \to \infty$ as $n \to \infty$ by (A4), it is clear from this final expression that, for given $\gamma > 0$, n sufficiently large, and $\omega \in \Omega_{n,2}$, $0 \le \mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)](\omega) < \gamma$. It follows that $|\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)]|1(\Omega_{n,2}) \to 0$ *a.s.*, completing the proof.

B.2.2 Proof of Lemma A.6. First, suppose $p_n \to p \in [0,1)$. Then, since $\hat{\vartheta}_{T_n}(\phi) \le 1$ *a.s.* (see (17) and the following argument), $\mathrm{Var}_n(\hat{\vartheta}_{T_n}(\phi)) \le \mathbb{E}_n\hat{\vartheta}_{T_n}(\phi)^2 \le \mathbb{E}_n\hat{\vartheta}_{T_n}(\phi)$. Furthermore, since $T_n \to \infty$ by (A4), the fact that $p_n \to p \in [0,1)$ means only the case $T_n(1-p_n) \to \infty$ can occur. In this case, since $\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)]1(\Omega_{n,2}) \to 0$ *a.s.* by Lemma A.5, we immediately obtain from $\mathrm{Var}_n(\hat{\vartheta}_{T_n}(\phi)) \le \mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)]$ that $\mathrm{Var}_n(\hat{\vartheta}_{T_n}(\phi))1(\Omega_{n,2}) \to 0$ *a.s.* as well. Hence, it only remains to prove the lemma in the case $p_n \to 1$, which we assume to hold for the remainder of the proof.

Towards this end, letting $\mathcal{D}$ denote the degree sequence and $\mathcal{T}$ denote the set of random variables defining the tree structure (as in Appendix B.2.1), we have

$$\mathrm{Var}_n(\hat{\vartheta}_{T_n}(\phi)) = \mathbb{E}_n[\mathrm{Var}(\hat{\vartheta}_{T_n}(\phi)|\mathcal{D},\mathcal{T})] + \mathrm{Var}_n(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{D},\mathcal{T}]). \tag{41}$$

We next consider the two summands in (41) in turn. In particular, we aim to show that each summand multiplied by $1(\Omega_{n,2})$ tends to zero *a.s.* as n tends to infinity.

For the first summand in (41), we use the fact that the signals are i.i.d. Bernoulli(θ) given the tree structure, as well as Lemma A.4, to write

$$\begin{aligned}\mathrm{Var}(\hat{\vartheta}_{T_n}(\phi)|\mathcal{D},\mathcal{T}) &= \frac{1}{T_n^2}\sum_{t=0}^{T_n-1}\sum_{l=0}^{t}\sum_{\mathbf{i}\in\hat{A}_l}\mathrm{Var}(\hat{s}_{T_n-t}(\mathbf{i})|\mathcal{D},\mathcal{T})\left(\binom{t}{l}\eta^l(1-\eta)^{t-l}\prod_{j=0}^{l-1}d_{in}(\mathbf{i}|j)^{-1}\right)^2\\ &= \frac{1}{T_n^2}\sum_{t=0}^{T_n-1}\sum_{l=0}^{t}\sum_{\mathbf{i}\in\hat{A}_l}\theta(1-\theta)\left(\binom{t}{l}\eta^l(1-\eta)^{t-l}\prod_{j=0}^{l-1}d_{in}(\mathbf{i}|j)^{-1}\right)^2\\ &\le \frac{1}{T_n^2}\sum_{t=0}^{T_n-1}\sum_{l=0}^{t}\binom{t}{l}\eta^l(1-\eta)^{t-l}\sum_{\mathbf{i}\in\hat{A}_l}\prod_{j=0}^{l-1}d_{in}(\mathbf{i}|j)^{-1} \le \frac{1}{T_n},\end{aligned}$$

where in the final step we have used $\sum_{\mathbf{i}\in\hat{A}_l}\prod_{j=0}^{l-1}d_{in}(\mathbf{i}|j)^{-1} \le 1$ and $\sum_{l=0}^{t}\binom{t}{l}\eta^l(1-\eta)^{t-l} = 1$. It immediately follows that $0 \le \mathbb{E}_n[\mathrm{Var}(\hat{\vartheta}_{T_n}(\phi)|\mathcal{D},\mathcal{T})]1(\Omega_{n,2}) \le 1/T_n$ *a.s.* Hence, because $T_n \to \infty$ as $n \to \infty$ by (A4), analysis of the first summand in (41) is complete.

For the second summand in (41), we first use the argument of (36) to write

$$\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{D},\mathcal{T}] = \frac{\theta}{T_n}\sum_{t=0}^{T_n-1}\sum_{l=0}^{t}\binom{t}{l}\eta^l(1-\eta)^{t-l}\sum_{\mathbf{i}\in\hat{A}_l}\prod_{j=0}^{l-1}d_{in}(\mathbf{i}|j)^{-1}$$
$$= \frac{\theta}{T_n}\sum_{l=0}^{T_n-1}\sum_{\mathbf{i}\in\hat{A}_l}\prod_{j=0}^{l-1}d_{in}(\mathbf{i}|j)^{-1}\sum_{t=l}^{T_n-1}\binom{t}{l}\eta^l(1-\eta)^{t-l} \triangleq \frac{\theta}{T_n}\sum_{l=0}^{T_n-1}Y_l u_{T_n,l},$$

where we have defined $Y_l = \sum_{\mathbf{i}\in\hat{A}_l}\prod_{j=0}^{l-1}d_{in}(\mathbf{i}|j)^{-1}$ and $u_{T_n,l} = \sum_{t=l}^{T_n-1}\binom{t}{l}\eta^l(1-\eta)^{t-l}$. Therefore,

$$\mathrm{Var}_n(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{D},\mathcal{T}]) = \frac{\theta^2}{T_n^2}\left(\sum_{l=0}^{T_n-1}u_{T_n,l}^2\mathrm{Var}_n(Y_l) + 2\sum_{l=0}^{T_n-1}u_{T_n,l}\sum_{l'=l+1}^{T_n-1}u_{T_n,l'}\mathrm{Cov}_n(Y_l,Y_{l'})\right). \quad (42)$$

It remains to compute the variance and covariance terms in (42). First, for any $l,l'\in\mathbb{N}$, we note

$$\mathbb{E}_n[Y_lY_{l'}] = \mathbb{E}_n\left[\mathbb{P}(X_l^1\in\hat{A}|\mathcal{D},\mathcal{T})\mathbb{P}(X_{l'}^2\in\hat{A}|\mathcal{D},\mathcal{T})\right] \quad (43)$$
$$= \mathbb{E}_n\left[\mathbb{P}(X_l^1\in\hat{A},X_{l'}^2\in\hat{A}|\mathcal{D},\mathcal{T})\right] = \mathbb{P}_n(X_l^1\in\hat{A},X_{l'}^2\in\hat{A}),$$

where we have used the argument of (37) and the fact that $\{X_i^1\}_{i=1}^\infty$ and $\{X_i^2\}_{i=1}^\infty$ are independent random walks given the tree structure. By a similar argument, $\mathbb{E}_n[Y_l] = \mathbb{P}_n(X_l^1\in\hat{A})$. Hence, using Lemmas A.7 and A.8, and assuming for the moment that $l>1$, we have

$$\begin{aligned}\mathrm{Var}_n(Y_l) &= \mathbb{P}_n(X_l^1\in\hat{A},X_l^2\in\hat{A}) - (\mathbb{P}_n(X_l^1\in\hat{A}))^2\\
&= \tilde{r}_n^*\tilde{p}_n^{2(l-1)} + \sum_{t=2}^{l}\tilde{q}_n^*\tilde{q}_n^{t-2}\tilde{r}_n\tilde{p}_n^{2(l-t)} + \tilde{q}_n^*\tilde{q}_n^{l-1} - (\tilde{p}_n^*\tilde{p}_n^{l-1})^2\\
&= \frac{\tilde{r}_n^*}{\tilde{p}_n^2}\tilde{p}_n^{2l} + \frac{\tilde{q}_n^*\tilde{r}_n}{\tilde{p}_n^4}\tilde{p}_n^{2l}\sum_{l=0}^{l-2}\left(\frac{\tilde{q}_n}{\tilde{p}_n^2}\right)^t + \frac{\tilde{q}_n^*}{\tilde{q}_n}\tilde{q}_n^l - \frac{(\tilde{p}_n^*)^2}{\tilde{p}_n^2}\tilde{p}_n^{2l}\\
&= \frac{\tilde{r}_n^*}{\tilde{p}_n^2}\tilde{p}_n^{2l} + \frac{\tilde{q}_n^*\tilde{r}_n}{\tilde{p}_n^4}\tilde{p}_n^{2l}\frac{1-(\tilde{q}_n/\tilde{p}_n^2)^{l-1}}{1-(\tilde{q}_n/\tilde{p}_n^2)} + \frac{\tilde{q}_n^*}{\tilde{q}_n}\tilde{q}_n^l - \frac{(\tilde{p}_n^*)^2}{\tilde{p}_n^2}\tilde{p}_n^{2l}\\
&= \frac{\tilde{r}_n^*}{\tilde{p}_n^2}\tilde{p}_n^{2l} + \frac{\tilde{q}_n^*\tilde{r}_n}{\tilde{p}_n^2}\tilde{p}_n^{2l}\frac{1-(\tilde{q}_n/\tilde{p}_n^2)^{l-1}}{\tilde{p}_n^2-\tilde{q}_n} + \frac{\tilde{q}_n^*}{\tilde{q}_n}\tilde{q}_n^l - \frac{(\tilde{p}_n^*)^2}{\tilde{p}_n^2}\tilde{p}_n^{2l}\\
&= \frac{\tilde{r}_n^*}{\tilde{p}_n^2}\tilde{p}_n^{2l} + \frac{\tilde{q}_n^*\tilde{r}_n}{\tilde{p}_n^2(\tilde{p}_n^2-\tilde{q}_n)}\tilde{p}_n^{2l} - \frac{\tilde{q}_n^*\tilde{r}_n}{\tilde{q}_n(\tilde{p}_n^2-\tilde{q}_n)}\tilde{q}_n^l + \frac{\tilde{q}_n^*}{\tilde{q}_n}\tilde{q}_n^l - \frac{(\tilde{p}_n^*)^2}{\tilde{p}_n^2}\tilde{p}_n^{2l}\\
&= \frac{1}{\tilde{p}_n^2}\left(\tilde{r}_n^* + \frac{\tilde{q}_n^*\tilde{r}_n}{\tilde{p}_n^2-\tilde{q}_n} - (\tilde{p}_n^*)^2\right)\tilde{p}_n^{2l} + \frac{\tilde{q}_n^*}{\tilde{q}_n}\left(1-\frac{\tilde{r}_n}{\tilde{p}_n^2-\tilde{q}_n}\right)\tilde{q}_n^l. \quad (44)\end{aligned}$$

Next, using (19) and Jensen's inequality, we have

$$\tilde{r}_n = \sum_{j\in\mathbb{N},k\in\mathbb{N}_0}\left(\frac{j}{j+k}\right)^2\sum_{i\in\mathbb{N}}f_n(i,j,k) - \tilde{q}_n \geq \left(\sum_{j\in\mathbb{N},k\in\mathbb{N}_0}\frac{j}{j+k}\sum_{i\in\mathbb{N}}f_n(i,j,k)\right)^2 - \tilde{q}_n = \tilde{p}_n^2 - \tilde{q}_n, \quad (45)$$

and so $1-\tilde{r}_n/(\tilde{p}_n^2-\tilde{q}_n)\leq 0$, i.e. the second term in (44) is non-positive, so $\forall\ l>1$,

$$\mathrm{Var}_n(Y_l) \leq \frac{1}{\tilde{p}_n^2}\left(\tilde{r}_n^* + \frac{\tilde{q}_n^*\tilde{r}_n}{\tilde{p}_n^2-\tilde{q}_n} - (\tilde{p}_n^*)^2\right)\tilde{p}_n^{2l}. \quad (46)$$

In the case $l = 1$, we have (again by Lemmas A.7 and A.8)

$$\mathrm{Var}_n(Y_l) = (\tilde{r}_n^* + \tilde{q}_n^*) - \tilde{p}_n^* \leq \tilde{r}_n^* + \frac{\tilde{q}_n^* \tilde{r}_n}{\tilde{p}_n^2 - \tilde{q}_n} - (\tilde{p}_n^*)^2 = \frac{1}{\tilde{p}_n^2}\left(\tilde{r}_n^* + \frac{\tilde{q}_n^* \tilde{r}_n}{\tilde{p}_n^2 - \tilde{q}_n} - (\tilde{p}_n^*)^2\right)\tilde{p}_n^{2l},$$

where the inequality is (45) and $\tilde{p}_n^* \leq 1$; hence, (46) holds for $l = 1$ as well. Finally, since $Y_0 = 1$ *a.s.*, it is immediate that (46) also holds for $l = 0$. We next analyze the covariance terms in (42). First, if $l' > l > 0$, we can use (43) and Lemmas A.7 and A.8 to obtain

$$\begin{aligned}
\mathbb{E}_n[Y_l Y_{l'}] &= \mathbb{P}_n(X_l^1 \in \hat{A}, X_{l'}^2 \in \hat{A}) = \tilde{p}_n^{l'-l}\mathbb{P}_n(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}) = \tilde{p}_n^{l'-l}\mathbb{E}_n[Y_l^2], \\
\mathbb{E}_n[Y_{l'}] &= \mathbb{P}(X_{l'}^2 \in \hat{A}) = \tilde{p}_n^* \tilde{p}_n^{l'-1} = \tilde{p}_n^* \tilde{p}_n^{l-1}\tilde{p}_n^{l'-l} = \mathbb{P}(X_l^1 \in \hat{A})\tilde{p}_n^{l'-l} = \mathbb{E}_n[Y_l]\tilde{p}_n^{l'-l}, \\
&\Rightarrow \mathrm{Cov}_n(Y_l, Y_{l'}) = \tilde{p}_n^{l'-l}\left(\mathbb{E}_n[Y_l^2] - (\mathbb{E}_n[Y_l])^2\right) = \tilde{p}_n^{l'-l}\mathrm{Var}_n(Y_l).
\end{aligned}$$

On the other hand, if $l' > l = 0$, we have $Y_l = 1$ *a.s.*, so $\mathrm{Cov}_n(Y_l, Y_{l'}) = 0 = \tilde{p}_n^{l'}\mathrm{Var}_n(Y_0)$. Hence, combined with (46), we have argued

$$\mathrm{Cov}_n(Y_l, Y_{l'}) = \tilde{p}_n^{l'-l}\mathrm{Var}_n(Y_l) \leq \frac{1}{\tilde{p}_n^2}\left(\tilde{r}_n^* + \frac{\tilde{q}_n^* \tilde{r}_n}{\tilde{p}_n^2 - \tilde{q}_n} - (\tilde{p}_n^*)^2\right)\tilde{p}_n^{l+l'} \ \forall\, l \in \mathbb{N}_0, l' > l. \tag{47}$$

Hence, combining (42), (46), and (47), we obtain

$$\begin{aligned}
&\mathrm{Var}_n(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{D}, \mathcal{T}]) \\
&\quad\leq \frac{1}{\tilde{p}_n^2}\left(\tilde{r}_n^* + \frac{\tilde{q}_n^* \tilde{r}_n}{\tilde{p}_n^2 - \tilde{q}_n} - (\tilde{p}_n^*)^2\right)\frac{\theta^2}{T_n^2}\left(\sum_{l=0}^{T_n-1} u_{T_n,l}^2 \tilde{p}_n^{2l} + 2\sum_{l=0}^{T_n-1} u_{T_n,l}\sum_{l'=l+1}^{T_n-1} u_{T_n,l'}\tilde{p}_n^{l+l'}\right) \\
&\quad\leq \frac{1}{\tilde{p}_n^2}\left(\tilde{r}_n^* + \frac{\tilde{q}_n^* \tilde{r}_n}{\tilde{p}_n^2 - \tilde{q}_n} - (\tilde{p}_n^*)^2\right)\frac{1}{T_n^2}\left(\sum_{l=0}^{T_n-1} u_{T_n,l}^2 + 2\sum_{l=0}^{T_n-1} u_{T_n,l}\sum_{l'=l+1}^{T_n-1} u_{T_n,l'}\right) \\
&\quad= \frac{1}{\tilde{p}_n^2}\left(\tilde{r}_n^* + \frac{\tilde{q}_n^* \tilde{r}_n}{\tilde{p}_n^2 - \tilde{q}_n} - (\tilde{p}_n^*)^2\right)\left(\frac{1}{T_n}\sum_{l=0}^{T_n-1} u_{T_n,l}\right)^2 = \frac{1}{\tilde{p}_n^2}\left(\tilde{r}_n^* + \frac{\tilde{q}_n^* \tilde{r}_n}{\tilde{p}_n^2 - \tilde{q}_n} - (\tilde{p}_n^*)^2\right),
\end{aligned}$$

where the second inequality is simply $\theta, \tilde{p}_n \leq 1$, the first equality is immediate, and the second equality holds by definition of $u_{T_n,l}$. It clearly follows that

$$\mathrm{Var}_n(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{D}, \mathcal{T}])1(\Omega_{n,2}) \leq \frac{1}{\tilde{p}_n^2}\left(\tilde{r}_n^* + \frac{\tilde{q}_n^* \tilde{r}_n}{\tilde{p}_n^2 - \tilde{q}_n} - (\tilde{p}_n^*)^2\right)1(\Omega_{n,2}), \tag{48}$$

and so we can complete the proof by showing the right side of (48) tends to zero *a.s.* Clearly, the right side is zero if $\omega \notin \Omega_{n,2}$; we aim to also show that, given $\gamma > 0$, $\exists\, N$ s.t. for $n > N$ and $\omega \in \Omega_{n,2}$,

$$\frac{1}{\tilde{p}_n(\omega)^2}\left(\tilde{r}_n^*(\omega) + \frac{\tilde{q}_n^*(\omega)\tilde{r}_n(\omega)}{\tilde{p}_n(\omega)^2 - \tilde{q}_n(\omega)} - \tilde{p}_n^*(\omega)^2\right) < \gamma. \tag{49}$$

To prove (49), we first recall that by (A3), we have for $\omega \in \Omega_{n,2}$, $\tilde{p}_n^*(\omega) \geq \tilde{p}_n(\omega) > p_n - \delta_n$. Hence, since we are assuming $p_n \to 1$, and since $\delta_n \to 0$ by (A3), we have for $\gamma' > 0$, n sufficiently large, and such ω, $\tilde{p}_n(\omega)^2, \tilde{p}_n^*(\omega)^2 > 1 - \gamma'$. We thus obtain for n large and $\omega \in \Omega_{n,2}$,

$$\frac{1}{\tilde{p}_n(\omega)^2}\left(\tilde{r}_n^*(\omega) + \frac{\tilde{q}_n^*(\omega)\tilde{r}_n(\omega)}{\tilde{p}_n(\omega)^2 - \tilde{q}_n(\omega)} - \tilde{p}_n^*(\omega)^2\right) < \frac{1}{1-\gamma'}\left(\tilde{r}_n^*(\omega) + \frac{\tilde{q}_n^*(\omega)\tilde{r}_n(\omega)}{1 - \gamma' - \tilde{q}_n(\omega)} - (1-\gamma')\right). \tag{50}$$

To further upper bound the right side of (50), we note $\tilde{r}_n \leq 1 - \tilde{q}_n$ *a.s.* by the first equality in (45). The same argument gives $\tilde{r}_n^* \leq 1 - \tilde{q}_n^*$ *a.s.* Note, however, that to use the second bound, we must ensure $1-\gamma'-\tilde{q}_n(\omega)>0$. To this end, recall that $\tilde{q}_n(\omega)<1-\xi$ for $\omega\in\Omega_{n,2}$ by (A3). Hence, assuming we choose $\gamma'<\xi$, we obtain $1-\gamma'-\tilde{q}_n(\omega)>0$ for such ω. Thus,

$$\begin{aligned}
&\frac{1}{\tilde{p}_n(\omega)^2}\left(\tilde{r}_n^*(\omega)+\frac{\tilde{q}_n^*(\omega)\tilde{r}_n(\omega)}{\tilde{p}_n(\omega)^2-\tilde{q}_n(\omega)}-\tilde{p}_n^*(\omega)^2\right)\\
&\quad<\frac{1}{1-\gamma'}\left((1-\tilde{q}_n^*(\omega))+\frac{\tilde{q}_n^*(\omega)(1-\tilde{q}_n(\omega))}{1-\gamma'-\tilde{q}_n(\omega)}-(1-\gamma')\right)\\
&\quad=\frac{1}{1-\gamma'}\left(\tilde{q}_n^*(\omega)\left(\frac{1-\tilde{q}_n(\omega)}{1-\gamma'-\tilde{q}_n(\omega)}-1\right)+\gamma'\right)=\frac{1}{1-\gamma'}\left(\tilde{q}_n^*(\omega)\left(\frac{\gamma'}{1-\gamma'-\tilde{q}_n(\omega)}\right)+\gamma'\right)\\
&\quad=\frac{\gamma'}{1-\gamma'}\left(\frac{\tilde{q}_n^*(\omega)}{1-\gamma'-\tilde{q}_n(\omega)}+1\right)<\frac{\gamma'}{1-\gamma'}\left(\frac{\tilde{q}_n^*(\omega)}{\xi-\gamma'}+1\right)\leq\frac{\gamma'}{1-\gamma'}\left(\frac{1}{\xi-\gamma'}+1\right),
\end{aligned}\tag{51}$$

where the first inequality uses (50) and the bounds from the previous paragraph, the equalities are straightforward, the second inequality uses $\tilde{q}_n(\omega)<1-\xi$ for $\omega\in\Omega_{n,2}$ by (A3), and the third uses $\tilde{q}_n^*(\omega)\leq 1$ (recall we have chosen $\gamma'<\xi$). Finally, it is straightforward to see the final bound in (51) tends to zero with γ'. Hence, for sufficiently small γ', (49) follows, completing the proof.

B.2.3 Notation for proofs of Lemmas A.7 and A.8. In the next two subsections, we prove Lemmas A.7 and A.8. For these proofs, we let $\mathcal{D}$ denote the degree sequence $\{d_{out}(i),d_{in}^A(i),d_{in}^B(i)\}_{i\in[n]}$, and we let D denote a realization of this set. Note that the random variables defined in (19) are all functions of $\mathcal{D}$; for a realization D of $\mathcal{D}$, we let e.g. $\tilde{p}_{n,D}$ denote the realization of $\tilde{p}_n$. We similarly define $f_{n,D},f_{n,D}^*$ for realizations of f_n,f_n^*, defined in (7). Finally, letting $g(D)=\mathbb{P}(\cdot|\mathcal{D}=D)$, we have $\mathbb{P}_n(\cdot)=g(\mathcal{D})$ by definition of $\mathbb{P}_n$. Hence, to prove Lemma A.7, it suffices to show

$$\mathbb{P}(X_l\in\hat{A}|\mathcal{D}=D)=\begin{cases}\tilde{p}_{n,D}^*\tilde{p}_{n,D}^{l-1}, & l\in\mathbb{N}\\ 1, & l=0\end{cases}.$$

while to prove Lemma A.8, it suffices to show

$$\mathbb{P}(X_l^1\in\hat{A},X_{l'}^2\in\hat{A}|\mathcal{D}=D)=\begin{cases}\mathbb{P}(X_l^1\in\hat{A},X_l^2\in\hat{A}|\mathcal{D}=D)\tilde{p}_{n,D}^{l'-l}, & l\in\mathbb{N}\\ \tilde{p}_{n,D}^*\tilde{p}_{n,D}^{l'-1}, & l=0\end{cases},\tag{52}$$

$$\mathbb{P}(X_l^1\in\hat{A},X_l^2\in\hat{A}|\mathcal{D}=D)=\begin{cases}\tilde{r}_{n,D}^*\tilde{p}_{n,D}^{2(l-1)}+\sum_{t=2}^l\tilde{q}_{n,D}^*\tilde{q}_{n,D}^{t-2}\tilde{r}_{n,D}\tilde{p}_{n,D}^{2(l-t)}+\tilde{q}_{n,D}^*\tilde{q}_{n,D}^{l-1}, & l\in\{2,3,\ldots\}\\ \tilde{r}_{n,D}^*+\tilde{q}_{n,D}^*, & l=1\\ 1, & l=0\end{cases}.\tag{53}$$

B.2.4 Proof of Lemma A.7. The $l=0$ case is trivial, since $X_0^1=\phi\in\hat{A}$, so we assume $l\in\mathbb{N}$ moving forward. First, since $\hat{A}^C=\hat{B}$ is an absorbing set, we have $X_l^1\in\hat{A}\Rightarrow X_{l-1}^1\in\hat{A}$, so

$$\mathbb{P}(X_l^1\in\hat{A}|\mathcal{D}=D)=\mathbb{P}(X_l^1\in\hat{A}|X_{l-1}^1\in\hat{A},\mathcal{D}=D)\mathbb{P}(X_{l-1}^1\in\hat{A}|\mathcal{D}=D).\tag{54}$$

For the first term in (54), we have

$$\begin{aligned}
&\mathbb{P}(X_l^1 \in \hat{A} | X_{l-1}^1 \in \hat{A}, \mathcal{D} = D) \qquad (55)\\
&= \sum_{j \in \mathbb{N}, k \in \mathbb{N}_0} \mathbb{P}(X_l^1 \in \hat{A} | d_{in}^A(X_{l-1}^1) = j, d_{in}^B(X_{l-1}^1) = k, X_{l-1}^1 \in \hat{A}, \mathcal{D} = D)\\
&\qquad\qquad \times \mathbb{P}(d_{in}^A(X_{l-1}^1) = j, d_{in}^B(X_{l-1}^1) = k | X_{l-1}^1 \in \hat{A}, \mathcal{D} = D)\\
&= \begin{cases} \sum_{j \in \mathbb{N}, k \in \mathbb{N}_0} \frac{j}{j+k} \sum_{i \in \mathbb{N}} f_{n,D}(i,j,k) = \tilde{p}_{n,D}, & l \in \{2,3,\ldots\} \\ \sum_{j \in \mathbb{N}, k \in \mathbb{N}_0} \frac{j}{j+k} \sum_{i \in \mathbb{N}} f_{n,D}^*(i,j,k) = \tilde{p}_{n,D}^*, & l = 1 \end{cases},
\end{aligned}$$

where the second equality holds by Algorithm 2. More specifically, for $l > 1$, the degrees of X_{l-1}^1 are sampled from $f_{n,D}$ (Line 13 in Algorithm 2) after realizing X_{l-1}^1 (Line 8), yielding the $\sum_{i \in \mathbb{N}} f_{n,D}(i,j,k)$ term; further, X_l^1 is chosen uniformly from the incoming neighbors of X_{l-1}^1 (Line 8) after realizing the degrees of X_{l-1}^1, yielding the $j/(j+k)$ term (the $l = 1$ case is similarly justified). Combining (54) and (55), and using the fact that $X_0^1 = \phi \in \hat{A}$ by definition, completes the proof in the case $l = 1$. For $l > 1$, we again use (54) and (55) to obtain

$$\mathbb{P}(X_l^1 \in \hat{A} | \mathcal{D} = D) = \tilde{p}_{n,D} \mathbb{P}(X_{l-1}^1 \in \hat{A} | \mathcal{D} = D) = \cdots = \tilde{p}_{n,D}^{l-1} \mathbb{P}(X_1^1 \in \hat{A} | \mathcal{D} = D) = \tilde{p}_{n,D}^{l-1} \tilde{p}_{n,D}^*,$$

which completes the proof.

B.2.5 Proof of Lemma A.8. We begin by proving the first statement in the lemma, i.e. (52). First, we note that for the $l = 0$ case, $X_0 = \phi \in \hat{A}$ by definition, so $\mathbb{P}(X_l^1 \in \hat{A}, X_{l'}^2 \in \hat{A} | \mathcal{D} = D) = \mathbb{P}(X_{l'}^2 \in \hat{A} | \mathcal{D} = D)$, and the statement holds by Lemma A.7. For the $l \in \mathbb{N}$ case, we first write

$$\begin{aligned}
\mathbb{P}(X_l^1 \in \hat{A}, X_{l'}^2 \in \hat{A} | \mathcal{D} = D) &= \mathbb{P}(X_l^1 \in \hat{A}, X_{l'-1}^2 \in \hat{A}, X_{l'}^2 \in \hat{A} | \mathcal{D} = D)\\
&= \mathbb{P}(X_{l'}^2 \in \hat{A} | X_l^1 \in \hat{A}, X_{l'-1}^2 \in \hat{A}, \mathcal{D} = D) \mathbb{P}(X_l^1 \in \hat{A}, X_{l'-1}^2 \in \hat{A} | \mathcal{D} = D),
\end{aligned}$$

where the first equality holds since $\hat{A}^C = \hat{B}$ is an absorbing set (i.e. $X_{l'}^2 \in \hat{A} \Rightarrow X_{l'-1}^2 \in \hat{A}$) and the second simply rewrites a conditional probability. Next, by the same argument as (55),

$$\mathbb{P}(X_{l'}^2 \in \hat{A} | X_l^1 \in \hat{A}, X_{l'-1}^2 \in \hat{A}, \mathcal{D} = D) = \tilde{p}_{n,D},$$

where we have used the $l' > 1$ case of (55), since $l' > l \geq 1$. Hence, the previous two equations give

$$\begin{aligned}
\mathbb{P}(X_l^1 \in \hat{A}, X_{l'}^2 \in \hat{A} | \mathcal{D} = D) &= \tilde{p}_{n,D} \mathbb{P}(X_l^1 \in \hat{A}, X_{l'-1}^2 \in \hat{A} | \mathcal{D} = D)\\
&= \cdots = \tilde{p}_{n,D}^{l'-l} \mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A} | \mathcal{D} = D).
\end{aligned}$$

This completes the proof of (52). For the second statement, i.e. (53), the $l = 0$ case is trivial, since $X_0^1 = X_0^2 = \phi \in \hat{A}$ by definition, so we assume $l \in \mathbb{N}$ for the remainder of the proof. First, let $\tau = \inf\{t \in \mathbb{N}_0 : X_t^1 \neq X_t^2\}$ denote the first step at which the two walks diverge. Note that $X_0^1 = X_0^2 = \phi$ by definition, so $\tau \in \mathbb{N}$ *a.s.*; also, due to the tree structure, the walks remain apart forever after diverging, i.e. $X_{\tau+1}^1 \neq X_{\tau+1}^2, X_{\tau+2}^1 \neq X_{\tau+1}^2, \ldots$ *a.s.* Next, for $l \in \mathbb{N}$, we write

$$\begin{aligned}
&\mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A} | \mathcal{D} = D) \qquad (56)\\
&= \sum_{t=1}^{l} \mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, \tau = t | \mathcal{D} = D) + \mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, \tau > l | \mathcal{D} = D)
\end{aligned}$$

We begin by computing the second term in (56). Here we have

$$
\begin{aligned}
&\mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, \tau > l | \mathcal{D} = D) \qquad (57)\\
&= \mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, X_l^1 = X_l^2, X_{l-1}^1 = X_{l-1}^2, \dots, X_1^1 = X_1^2 | \mathcal{D} = D)\\
&= \mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, X_l^1 = X_l^2, X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, X_{l-1}^1 = X_{l-1}^2, \dots, X_1^1 = X_1^2 | \mathcal{D} = D)\\
&= \mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, X_l^1 = X_l^2 | X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, X_{l-1}^1 = X_{l-1}^2, \dots, X_1^1 = X_1^2, \mathcal{D} = D)\\
&\quad \times \mathbb{P}(X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, X_{l-1}^1 = X_{l-1}^2, \dots, X_1^1 = X_1^2 | \mathcal{D} = D)\\
&= \mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, X_l^1 = X_l^2 | X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, X_{l-1}^1 = X_{l-1}^2, \dots, X_1^1 = X_1^2, \mathcal{D} = D)\\
&\quad \times \mathbb{P}(X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, \tau > l-1 | \mathcal{D} = D),
\end{aligned}
$$

where the first and last equalities hold by definition of τ and the second holds since $\hat{A}^C = \hat{B}$ is an absorbing set. Now for $l > 1$, we obtain

$$
\begin{aligned}
&\mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, X_l^1 = X_l^2 | X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, X_{l-1}^1 = X_{l-1}^2, \dots, X_1^1 = X_1^2, \mathcal{D} = D) \qquad (58)\\
&= \mathbb{P}(X_l^1 \in \hat{A}, X_l^1 = X_l^2 | X_{l-1}^1 \in \hat{A}, X_{l-1}^1 = X_{l-1}^2, \mathcal{D} = D)\\
&= \sum_{j \in \mathbb{N}, k \in \mathbb{N}_0} \mathbb{P}(X_l^1 \in \hat{A}, X_l^1 = X_l^2 | d_{in}^A(X_{l-1}^1) = j, d_{in}^B(X_{l-1}) = k, X_{l-1}^1 \in \hat{A}, X_{l-1}^1 = X_{l-1}^2, \mathcal{D} = D)\\
&\qquad \times \mathbb{P}(d_{in}^A(X_{l-1}^1) = j, d_{in}^B(X_{l-1}) = k | X_{l-1}^1 \in \hat{A}, X_{l-1}^1 = X_{l-1}^2, \mathcal{D} = D)\\
&= \sum_{j \in \mathbb{N}, k \in \mathbb{N}_0} \frac{j}{j+k} \frac{1}{j+k} \sum_{i \in \mathbb{N}} f_{n,D}(i,j,k) = \tilde{q}_{n,D},
\end{aligned}
$$

where the first equality uses independence and eliminates repetitive events, and the third follows an argument similar to that following (55). Combining (57) and (58),

$$
\begin{aligned}
\mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, \tau > l | \mathcal{D} = D) &= \tilde{q}_{n,D} \mathbb{P}(X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, \tau > l-1 | \mathcal{D} = D) \qquad (59)\\
&= \cdots = \tilde{q}_{n,D}^{l-1} \mathbb{P}(X_1^1 \in \hat{A}, X_1^2 \in \hat{A}, \tau > 1 | \mathcal{D} = D).
\end{aligned}
$$

Finally, by an argument similar to (58), we have

$$
\begin{aligned}
&\mathbb{P}(X_1^1 \in \hat{A}, X_1^2 \in \hat{A}, \tau > 1 | \mathcal{D} = D) = \mathbb{P}(X_1^1 \in \hat{A}, X_1^1 = X_1^2 | \mathcal{D} = D) \qquad (60)\\
&= \sum_{j \in \mathbb{N}, k \in \mathbb{N}_0} \mathbb{P}(X_1^1 \in \hat{A}, X_1^1 = X_1^2 | d_{in}^A(\phi) = j, d_{in}^B(\phi) = k, \mathcal{D} = D) \mathbb{P}(d_{in}^A(\phi) = j, d_{in}^B(\phi) = k | \mathcal{D} = D)\\
&= \sum_{j \in \mathbb{N}, k \in \mathbb{N}_0} \frac{j}{j+k} \frac{1}{j+k} \sum_{i \in \mathbb{N}} f_{n,D}^*(i,j,k) = \tilde{q}_{n,D}^*.
\end{aligned}
$$

Hence, combining (59) and (60) gives

$$
\mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, \tau > l | \mathcal{D} = D) = \tilde{q}_{n,D}^* \tilde{q}_{n,D}^{l-1} \ \forall \ l \in \mathbb{N}. \qquad (61)
$$

For the first term in (56), we first consider the $t = l$ summand. For $l > 1$, similar to (58),

$$\begin{aligned}
&\mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, \tau = l | \mathcal{D} = D) \\
&\quad = \mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, X_l^1 \neq X_l^2, X_{l-1}^1 = X_{l-1}^2, \ldots, X_1^1 = X_1^2 | \mathcal{D} = D) \\
&\quad = \mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, X_l^1 \neq X_l^2, X_{l-1}^1 \in \hat{A}, X_{l-1}^1 = X_{l-1}^2, \ldots, X_1^1 = X_1^2 | \mathcal{D} = D) \\
&\quad = \mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, X_l^1 \neq X_l^2 | X_{l-1}^1 \in \hat{A}, X_{l-1}^1 = X_{l-1}^2, \ldots, X_1^1 = X_1^2, \mathcal{D} = D) \\
&\qquad \times \mathbb{P}(X_{l-1}^1 \in \hat{A}, X_{l-1}^1 = X_{l-1}^2, \ldots, X_1^1 = X_1^2 | \mathcal{D} = D) \\
&\quad = \sum_{j \in \mathbb{N}, k \in \mathbb{N}_0} \frac{j}{j+k} \frac{j-1}{j+k} \sum_{i \in \mathbb{N}} f_{n,D}(i,j,k) \mathbb{P}(X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, \tau > l-1 | \mathcal{D} = D) = \tilde{r}_{n,D} \tilde{q}_{n,D}^{l-2} \tilde{q}_{n,D}^*,
\end{aligned}$$

where in the final step we have also used (61). Similarly, for $l = 1$,

$$\begin{aligned}
\mathbb{P}(X_1^1 \in \hat{A}, X_1^2 \in \hat{A}, \tau = 1 | \mathcal{D} = D) &= \mathbb{P}(X_1^1 \in \hat{A}, X_1^2 \in \hat{A}, X_1^1 \neq X_1^2 | \mathcal{D} = D) \\
&= \sum_{j \in \mathbb{N}, k \in \mathbb{N}_0} \frac{j}{j+k} \frac{j-1}{j+k} \sum_{i \in \mathbb{N}} f_{n,D}^*(i,j,k) = \tilde{r}_{n,D}^*.
\end{aligned}$$

To summarize, we have shown

$$\mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, \tau = l | \mathcal{D} = D) = \begin{cases} \tilde{q}_{n,D}^* \tilde{q}_{n,D}^{l-2} \tilde{r}_{n,D}, & l \in \{2, 3, \ldots\} \\ \tilde{r}_{n,D}^*, & l = 1 \end{cases}. \tag{62}$$

Next, we consider the $t < l$ summands in (56) (such summands are present only for $l > 1$). We have

$$\begin{aligned}
&\mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, \tau = t | \mathcal{D} = D) \\
&\quad = \mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, X_{l-1}^1 \neq X_{l-1}^2, \tau = t | \mathcal{D} = D) \\
&\quad = \mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A} | X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, X_{l-1}^1 \neq X_{l-1}^2, \tau = t, \mathcal{D} = D) \\
&\qquad \times \mathbb{P}(X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, X_{l-1}^1 \neq X_{l-1}^2, \tau = t | \mathcal{D} = D) \\
&\quad = \prod_{h=1}^{2} \mathbb{P}(X_l^h \in \hat{A} | X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, X_{l-1}^1 \neq X_{l-1}^2, \tau = t, \mathcal{D} = D) \\
&\qquad \times \mathbb{P}(X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, \tau = t | \mathcal{D} = D),
\end{aligned}$$

where in the first equality we used the fact that $\hat{A}^C = \hat{B}$ is an absorbing set and the fact that once the walks diverge they remain apart; in the second equality we used the fact that X_l^1 and X_l^2 are conditionally independent given the event $X_{l-1}^1 \neq X_{l-1}^2$. Further, for $h \in \{1, 2\}$,

$$\mathbb{P}(X_l^h \in \hat{A} | X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, X_{l-1}^1 \neq X_{l-1}^2, \tau = t, \mathcal{D} = D) = \sum_{j \in \mathbb{N}, k \in \mathbb{N}_0} \frac{j}{j+k} \sum_{i \in \mathbb{N}} f_{n,D}(i,j,k) = \tilde{p}_{n,D},$$

and so, combining the previous two equations and applying recursively yields

$$\begin{aligned}
\mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A}, \tau = t | \mathcal{D} = D) &= \tilde{p}_{n,D}^2 \mathbb{P}(X_{l-1}^1 \in \hat{A}, X_{l-1}^2 \in \hat{A}, \tau = t | \mathcal{D} = D) \\
&= \cdots = \tilde{p}_{n,D}^{2(l-t)} \mathbb{P}(X_t^1 \in \hat{A}, X_t^2 \in \hat{A}, \tau = t | \mathcal{D} = D) \\
&= \begin{cases} \tilde{q}_{n,D}^* \tilde{q}_{n,D}^{t-2} \tilde{r}_{n,D} \tilde{p}_{n,D}^{2(l-t)}, & t \in \{2, 3, \ldots, l-1\} \\ \tilde{r}_{n,D}^* \tilde{p}_{n,D}^{2(l-1)}, & t = 1 \end{cases} \quad \forall\, l \in \{2, 3, \ldots\}.
\end{aligned} \tag{63}$$

where the final equality uses (62). Finally, combining (56), (61), (62), and (63) yields

$$\mathbb{P}(X_l^1 \in \hat{A}, X_l^2 \in \hat{A} | \mathcal{D} = D) = \begin{cases} \tilde{r}_{n,D}^* \tilde{p}_{n,D}^{2(l-1)} + \sum_{t=2}^{l} \tilde{q}_{n,D}^* \tilde{q}_{n,D}^{t-2} \tilde{r}_{n,D} \tilde{p}_{n,D}^{2(l-t)} + \tilde{q}_{n,D}^* \tilde{q}_{n,D}^{l-1}, & l \in \{2,3,\dots\} \\ \tilde{r}_{n,D}^* + \tilde{q}_{n,D}^*, & l = 1 \end{cases},$$

which is what we set out to prove.

B.3 Step 2 for proof of Theorem 3.2

B.3.1 Proof of Lemma A.9. We first write

$$\begin{aligned} \mathbb{P}\left(\left|\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]\right| > \epsilon\right) &= \mathbb{E}\left[\mathbb{P}\left(\left|\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]\right| > \epsilon \middle| \mathcal{T}\right)\right] \\ &= \mathbb{E}\left[\mathbb{P}\left(\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] > \epsilon \middle| \mathcal{T}\right) + \mathbb{P}\left(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] - \hat{\vartheta}_{T_n}(\phi) > \epsilon \middle| \mathcal{T}\right)\right] \end{aligned} \tag{64}$$

where the first equality uses the law of total expectation and the second is immediate. For the first summand in the expectation in (64), we fix $\lambda > 0$ and write

$$\begin{aligned} &\mathbb{P}\left(\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] > \epsilon \middle| \mathcal{T}\right) = \mathbb{P}\left(\exp(\lambda(\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}])) > e^{-\lambda\epsilon} \middle| \mathcal{T}\right) \\ &\le e^{-\lambda\epsilon}\mathbb{E}\left[\exp(\lambda(\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]) \middle| \mathcal{T}\right] \\ &= e^{-\lambda\epsilon} \prod_{t=0}^{T_n-1} \prod_{l=0}^{t} \prod_{\mathbf{i}\in\hat{A}_l} \mathbb{E}\left[\exp\left(\frac{\lambda}{T_n}\binom{t}{l}\eta^l(1-\eta)^{t-l}\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}(\hat{s}_{T_n-t}(\mathbf{i}) - \theta)\right)\middle| \mathcal{T}\right] \\ &\le e^{-\lambda\epsilon} \prod_{t=0}^{T_n-1} \prod_{l=0}^{t} \prod_{\mathbf{i}\in\hat{A}_l} \exp\left(\frac{1}{8}\left(\frac{\lambda}{T_n}\binom{t}{l}\eta^l(1-\eta)^{t-l}\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}\right)^2\right) \\ &\le e^{-\lambda\epsilon} \prod_{t=0}^{T_n-1} \prod_{l=0}^{t} \prod_{\mathbf{i}\in\hat{A}_l} \exp\left(\frac{\lambda^2}{8T_n^2}\binom{t}{l}\eta^l(1-\eta)^{t-l}\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}\right) \\ &= \exp\left(-\lambda\epsilon + \frac{\lambda^2}{8T_n}\frac{1}{T_n}\sum_{t=0}^{T_n-1}\sum_{l=0}^{t}\sum_{\mathbf{i}\in\hat{A}_l}\binom{t}{l}\eta^l(1-\eta)^{t-l}\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}\right) \\ &= \exp\left(-\lambda\epsilon + \frac{\lambda^2}{8T_n\theta}\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]\right) \le \exp\left(-\lambda\epsilon + \frac{\lambda^2}{8T_n}\right). \end{aligned} \tag{65}$$

Here the first equality holds by monotonicity of $x \mapsto e^{\lambda x}$, the first inequality is Markov's, the second equality holds by (26), the second inequality uses Lemma B.3 from Appendix B.4, the third inequality uses $\binom{t}{l}\eta^l(1-\eta)^{t-l}, \prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1} \le 1$, the third equality is immediate, the fourth equality again uses (26), and the fourth inequality uses (27). Since the preceding argument holds $\forall\, \lambda > 0$, we choose $\lambda = 4\epsilon T_n$ to minimize the bound. Upon substituting into (65), we obtain $e^{-2\epsilon^2 T_n}$. The same argument holds for the second summand in the expectation of (64). We also note that the bound $e^{-2\epsilon^2 T_n}$ is non-random, so we may discard the expectation. In summary, we have shown

$$\mathbb{P}\left(\left|\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]\right| > \epsilon\right) \le 2e^{-2\epsilon^2 T_n}.$$

Hence, for n sufficiently large, we have by assumption on T_n

$$\mathbb{P}\left(\left|\hat{\vartheta}_{T_n}(\phi) - \mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]\right| > \epsilon\right) \le 2e^{-2\epsilon^2\mu\log n} = 2n^{-2\epsilon^2\mu} = O\left(n^{-2\epsilon^2\mu}\right),$$

which is what we set out to prove.

B.3.2 Proof of Lemma A.10. We begin by deriving a bound conditioned on the degree sequence. First, we fix $\tilde{\lambda} > 0$ and use monotonicity of $x \mapsto e^{\tilde{\lambda} x}$ and Markov's inequality to write

$$\mathbb{P}_n(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] > \epsilon) \leq e^{-\tilde{\lambda}\epsilon} \mathbb{E}_n \exp(\tilde{\lambda}\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]). \tag{66}$$

The bulk of the proof will involve bounding the expectation term. For this, we first note

$$\mathbb{E}_n \exp(\tilde{\lambda}\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]) = \mathbb{E}_n \exp\left(\tilde{\lambda}\frac{\theta}{T_n}\sum_{t=0}^{T_n-1}\sum_{l=0}^{t}\binom{t}{l}\eta^l(1-\eta)^{t-l}\sum_{\mathbf{i}\in\hat{A}_l}\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}\right)$$

$$= \mathbb{E}_n \exp\left(\frac{\tilde{\lambda}\theta}{T_n}\sum_{l=0}^{T_n-1}\left(\sum_{t=l}^{T_n-1}\binom{t}{l}\eta^l(1-\eta)^{t-l}\right)\left(\sum_{\mathbf{i}\in\hat{A}_l}\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}\right)\right) = \mathbb{E}_n \prod_{l=0}^{T_n-1}\exp(\lambda u_{T_n,l} Y_l),$$

where the first equality holds by (26), the second rearranges summations, and in the third we have defined $\lambda = \tilde{\lambda}\theta/T_n$, $u_{T_n,l} = \sum_{t=l}^{T_n-1}\binom{t}{l}\eta^l(1-\eta)^{t-l}$, and $Y_l = \sum_{\mathbf{i}\in\hat{A}_l}\prod_{j=0}^{l-1} d_{in}(\mathbf{i}|j)^{-1}$. For the remainder of the proof, we use $\mathbb{E}_{n,l}$ to denote conditional expectation with respect to the degree sequence *and* the set of random variables realized during the first l iterations of Algorithm 2 (i.e. the random variables defining the first l generations of the tree). Using this notation, we have

$$\begin{aligned}
\mathbb{E}_n\Big[\prod_{l=0}^{T_n-1}\exp(\lambda u_{T_n,l}Y_l)\Big] &= \mathbb{E}_n\Big[\mathbb{E}_{n,T_n-2}\Big[\prod_{l=0}^{T_n-1}\exp(\lambda u_{T_n,l}Y_l)\Big]\Big] \qquad (67)\\
&= \mathbb{E}_n\Big[\prod_{l=0}^{T_n-2}\exp(\lambda u_{T_n,l}Y_l)\mathbb{E}_{n,T_n-2}\Big[\exp(\lambda u_{T_n,T_n-1}Y_{T_n-1})\Big]\Big]\\
&= \mathbb{E}_n\Big[\prod_{l=0}^{T_n-3}\exp(\lambda u_{T_n,l}Y_l)\exp(\lambda(u_{T_n,T_n-2}+u_{T_n,T_n-1}\tilde{p}_n)Y_{T_n-2})\\
&\qquad\times\mathbb{E}_{n,T_n-2}\Big[\exp(\lambda u_{T_n,T_n-1}(Y_{T_n-1}-\tilde{p}_n Y_{T_n-2}))\Big]\Big],
\end{aligned}$$

where in the third equality we have multiplied and divided $\exp(\lambda u_{T_n,T_n-1}\tilde{p}_n Y_{T_n-2})$. Next, we note

$$\begin{aligned}
Y_{T_n-1} &= \sum_{\mathbf{i}'\in\hat{A}_{T_n-2}}\sum_{\mathbf{i}\in\hat{A}_{T_n-1}:\mathbf{i}|(T_n-2)=\mathbf{i}'}\prod_{j=0}^{T_n-2} d_{in}(\mathbf{i}|j)^{-1} = \sum_{\mathbf{i}'\in\hat{A}_{T_n-2}}\sum_{\mathbf{i}\in\hat{A}_{T_n-1}:\mathbf{i}|(T_n-2)=\mathbf{i}'}\prod_{j=0}^{T_n-2} d_{in}(\mathbf{i}'|j)^{-1} \qquad (68)\\
&= \sum_{\mathbf{i}'\in\hat{A}_{T_n-2}}\prod_{j=0}^{T_n-2} d_{in}(\mathbf{i}'|j)^{-1}|\{\mathbf{i}\in\hat{A}_{T_n-1}:\mathbf{i}|(T_n-2)=\mathbf{i}'\}| = \sum_{\mathbf{i}'\in\hat{A}_{T_n-2}}\prod_{j=0}^{T_n-3} d_{in}(\mathbf{i}'|j)^{-1}d_{in}(\mathbf{i}')^{-1}d_{in}^A(\mathbf{i}'),
\end{aligned}$$

where in the first equality we rewrote the sum based on the construction of $\hat{A}_{T_n-1}$ in Algorithm 2, in the second we have used the fact that $\mathbf{i}|j = \mathbf{i}'|j$ for $j \in \{0,\ldots,T_n-2\}$ by Algorithm 2 (in words, $\mathbf{i}$ and $\mathbf{i}'$ share the same ancestry in the tree), in the third we have recognized that the $\mathbf{i}$-th summand does not depend on $\mathbf{i}$, and in the fourth we have used $\mathbf{i}'|(T_n-2) = \mathbf{i}'$ (since $\mathbf{i}' \in \hat{A}_{T_n-2}$) and the construction of the agent offspring of $\mathbf{i}'$ in Algorithm 2. It follows that

$$\mathbb{E}_{n,T_n-2}Y_{T_n-1} = \sum_{\mathbf{i}'\in\hat{A}_{T_n-2}}\prod_{j=0}^{T_n-3} d_{in}(\mathbf{i}'|j)^{-1}\mathbb{E}_{n,T_n-2}(d_{in}^A(\mathbf{i}')/d_{in}(\mathbf{i}')) = \prod_{j=0}^{T_n-3} d_{in}(\mathbf{i}'|j)^{-1}\tilde{p}_n = Y_{T_n-2}\tilde{p}_n,$$

where $\mathbb{E}_{n,T_n-2}(d_{in}^A(\mathbf{i}')/d_{in}(\mathbf{i}')) = \tilde{p}_n$ holds by definition of $d_{in}^A(\mathbf{i}'), d_{in}(\mathbf{i}')$ in Algorithm 2 and of $\tilde{p}_n$ from (19). In summary, we have argued $\mathbb{E}_{n,T_n-2}(Y_{T_n-1} - Y_{T_n-2}\tilde{p}_n) = 0$. On the other hand, we note $0 \leq Y_{T_n-1} \leq Y_{T_n-2} \leq \cdots \leq Y_0 = 1$, where the first inequality holds since Y_{T_n-1} is a sum of nonnegative terms and the second holds by (68) (using $d_{in}(\mathbf{i}') = d_{in}^A(\mathbf{i}') + d_{in}^B(\mathbf{i}') \geq d_{in}^A(\mathbf{i}')$), and where $Y_0 = 1$ by definition. Hence, we can use Lemma B.3 from Appendix B.4 to obtain

$$\mathbb{E}_{n,T_n-2} \exp(\lambda u_{T_n,T_n-1}(Y_{T_n-1} - \tilde{p}_n Y_{T_n-2})) \leq e^{\lambda^2 u_{T_n,T_n-1}^2/8}. \tag{69}$$

Substituting into (67) then yields

$$\begin{aligned}
&\mathbb{E}_n\Big[\prod_{l=0}^{T_n-1} \exp(\lambda u_{T_n,l} Y_l)\Big] &(70)\\
&\leq \mathbb{E}_n\Big[\prod_{l=0}^{T_n-3} \exp(\lambda u_{T_n,l} Y_l) \exp(\lambda(u_{T_n,T_n-2} + u_{T_n,T_n-1}\tilde{p}_n)Y_{T_n-2})\Big] \exp\left(\frac{\lambda^2}{8} u_{T_n,T_n-1}^2\right).
\end{aligned}$$

We can then iteratively apply the preceding argument. Namely, we have

$$\begin{aligned}
&\mathbb{E}_n\Big[\prod_{l=0}^{T_n-3} \exp(\lambda u_{T_n,l} Y_l) \exp(\lambda(u_{T_n,T_n-2} + u_{T_n,T_n-1}\tilde{p}_n)Y_{T_n-2})\Big] \exp\left(\frac{\lambda^2}{8} u_{T_n,T_n-1}^2\right)\\
&= \mathbb{E}_n\Big[\prod_{l=0}^{T_n-4} \exp(\lambda u_{T_n,l} Y_l) \exp(\lambda(u_{T_n,T_n-3} + u_{T_n,T_n-2}\tilde{p}_n + u_{T_n,T_n-1}\tilde{p}_n^2)Y_{T_n-3})\\
&\quad \times \mathbb{E}_{n,T_n-3}\Big[\exp(\lambda(u_{T_n,T_n-2} + u_{T_n,T_n-1}\tilde{p}_n)(Y_{T_n-2} - \tilde{p}_n Y_{T_n-3}))\Big]\Big] \exp\left(\frac{\lambda^2}{8} u_{T_n,T_n-1}^2\right)\\
&\leq \mathbb{E}_n\Big[\prod_{l=0}^{T_n-4} \exp(\lambda u_{T_n,l} Y_l) \exp(\lambda(u_{T_n,T_n-3} + u_{T_n,T_n-2}\tilde{p}_n + u_{T_n,T_n-1}\tilde{p}_n^2)Y_{T_n-3})\Big] &(71)\\
&\quad \times \exp\left(\frac{\lambda^2}{8}\left((u_{T_n,T_n-2} + u_{T_n,T_n-1}\tilde{p}_n)^2 + u_{T_n,T_n-1}^2\right)\right) &(72)\\
&\leq \cdots \leq \mathbb{E}_n\Big[\exp(\lambda u_{T_n,0} Y_0) \exp\left(\lambda \sum_{l=1}^{T_n-1} u_{T_n,l}\tilde{p}_n^{l-1} Y_1\right)\Big] \exp\left(\frac{\lambda^2}{8} \sum_{l=2}^{T_n-1} \left(\sum_{l'=l}^{T_n-1} u_{T_n,l'}\tilde{p}_n^{l'-l}\right)^2\right). &(73)
\end{aligned}$$

(The precise form of the summations in (73) can be verified by considering the case $T_n = 4$ in (71) and (72).) Note that the final step of the iteration is slightly different; this is because the root node has degrees sampled from f_n^* (the uniform distribution) instead of f_n (the size-biased distribution) in Algorithm 2. Nevertheless, a similar argument holds: here we have $\mathbb{E}_{n,0} Y_1 = \tilde{p}_n^* Y_0$ and $Y_1 \in [0, 1]$ *a.s.*, so by an argument similar to that leading to (69),

$$\begin{aligned}
&\mathbb{E}_n\Big[\exp(\lambda u_{T_n,0} Y_0) \exp\left(\lambda \sum_{l=1}^{T_n-1} u_{T_n,l}\tilde{p}_n^{l-1} Y_1\right)\Big]\\
&= \mathbb{E}_n\Big[\exp\left(\lambda\left(u_{T_n,0} + \tilde{p}_n^* \sum_{l=1}^{T_n-1} u_{T_n,l}\tilde{p}_n^{l-1}\right) Y_0\right) \mathbb{E}_{n,0}\Big[\exp\left(\lambda \sum_{l=1}^{T_n-1} u_{T_n,l}\tilde{p}_n^{l-1}(Y_1 - \tilde{p}_n^* Y_0)\right)\Big]\Big]\\
&\leq \mathbb{E}_n\Big[\exp\left(\lambda\left(u_{T_n,0} + \tilde{p}_n^* \sum_{l=1}^{T_n-1} u_{T_n,l}\tilde{p}_n^{l-1}\right) Y_0\right)\Big] \exp\left(\frac{\lambda^2}{8}\left(\sum_{l=1}^{T_n-1} u_{T_n,l}\tilde{p}_n^{l-1}\right)^2\right).
\end{aligned}$$

Combining the previous inequality with (70) and (73) then yields

$$\mathbb{E}_n\Big[\prod_{l=0}^{T_n-1}\exp(\lambda u_{T_n,l}Y_l)\Big]$$
$$\leq \mathbb{E}_n\Big[\exp\left(\lambda\left(u_{T_n,0}+\tilde{p}_n^*\sum_{l=1}^{T_n-1}u_{T_n,l}\tilde{p}_n^{l-1}\right)Y_0\right)\Big]\exp\left(\frac{\lambda^2}{8}\sum_{l=1}^{T_n-1}\left(\sum_{l'=l}^{T_n-1}u_{T_n,l'}\tilde{p}_n^{l'-l}\right)^2\right)$$

Next, we recall $Y_0 = 1$ by definition. Additionally, we have

$$\begin{aligned} u_{T_n,0}+\tilde{p}_n^*\sum_{l=1}^{T_n-1}u_{T_n,l}\tilde{p}_n^{l-1} &= \sum_{t=0}^{T_n-1}(1-\eta)^t+\tilde{p}_n^*\sum_{l=1}^{T_n-1}\sum_{t=l}^{T_n-1}\binom{t}{l}\eta^l(1-\eta)^{t-l}\tilde{p}_n^{l-1} \\ &= \sum_{t=0}^{T_n-1}\left(\sum_{t=1}^{l}\binom{t}{l}\eta^l(1-\eta)^{t-l}\tilde{p}_n^*\tilde{p}_n^{l-1}+(1-\eta)^t\right)=\frac{T_n}{\theta}\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)], \end{aligned}$$

where the first equality uses the definition of $u_{T_n,l}$, the second rearranges summations, and the third uses (38). Combining the previous two equations therefore yields

$$\mathbb{E}_n\Big[\prod_{l=0}^{T_n-1}\exp(\lambda u_{T_n,l}Y_l)\Big]\leq\exp\left(\lambda\frac{T_n}{\theta}\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)]+\frac{\lambda^2}{8}\sum_{l=1}^{T_n-1}\left(\sum_{l'=l}^{T_n-1}u_{T_n,l'}\tilde{p}_n^{l'-l}\right)^2\right).$$

Hence, recalling that $\lambda=\tilde{\lambda}\theta/T_n$, and substituting into (66), we have shown

$$\mathbb{P}_n(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]>\epsilon)\leq\exp\left(-\tilde{\lambda}\epsilon+\tilde{\lambda}\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)]+\frac{\tilde{\lambda}^2\theta^2}{8T_n^2}\sum_{l=1}^{T_n-1}\left(\sum_{l'=l}^{T_n-1}u_{T_n,l'}\tilde{p}_n^{l'-l}\right)^2\right). \tag{74}$$

Clearly, this inequality still holds if we multiply both sides by $1(\Omega_{n,2})$. Additionally, by (A3), $\tilde{p}_n(\omega)<p_n+\delta_n$ for $\omega\in\Omega_{n,2}$, where $p_n\to p$ and $\delta_n\to 0$; since we additionally assume $p<1$ in the statement of the lemma, we conclude $\tilde{p}_n(\omega)<p_n+\delta_n<1$ for $\omega\in\Omega_{n,2}$ and n sufficiently large. For such n, we can therefore write

$$\begin{aligned} &\mathbb{P}_n(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]>\epsilon)1(\Omega_{n,2}) \\ &\quad\leq\exp\left(-\tilde{\lambda}\epsilon+\tilde{\lambda}\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)]+\frac{\tilde{\lambda}^2\theta^2}{8T_n^2}\sum_{l=1}^{T_n-1}\left(\sum_{l'=l}^{T_n-1}u_{T_n,l'}(p_n+\delta_n)^{l'-l}\right)^2\right)1(\Omega_{n,2}) \\ &\quad\leq\exp\left(-\tilde{\lambda}\epsilon+\tilde{\lambda}\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)]+\frac{\tilde{\lambda}^2\theta^2}{8T_n\eta^2(1-(p_n+\delta_n))^2}\right)1(\Omega_{n,2}), \end{aligned}$$

where the second inequality uses Lemma B.2 from Appendix B.4. Additionally, since $p_n\to p<1$, we can use the argument leading to (40) to obtain $\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)](\omega)<c/T_n$ (for some c independent of n) whenever $\omega\in\Omega_{n,2}$ and n is sufficiently large. For such n, we obtain

$$\mathbb{P}_n(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]>\epsilon)1(\Omega_{n,2})\leq\exp\left(-\tilde{\lambda}\epsilon+\frac{\tilde{\lambda}c}{T_n}+\frac{\tilde{\lambda}^2\theta^2}{8T_n\eta^2(1-(p_n+\delta_n))^2}\right)1(\Omega_{n,2}), \tag{75}$$

Now since $\tilde{\lambda} > 0$ was arbitrary, we can choose $\tilde{\lambda} = 4T_n\epsilon\eta^2(1-(p_n+\delta_n))^2/\theta^2$. Upon substituting into the exponent in the previous equation, this exponent becomes

$$
\begin{aligned}
&-\tilde{\lambda}\epsilon + \frac{\tilde{\lambda}^2}{8T_n\eta^2(1-(p_n+\delta_n))^2} + \frac{\tilde{\lambda}c}{T_n} = -2T_n\epsilon^2\eta^2(1-(p_n+\delta_n))^2/\theta^2 + 4c\epsilon\eta^2(1-(p_n+\delta_n))^2/\theta^2 \\
&= -2T_n\epsilon^2\eta^2\left((1-p_n)^2 - 2(1-p_n)\delta_n + \delta_n^2\right)/\theta^2 + 4c\epsilon\eta^2(1-(p_n+\delta_n))^2/\theta^2 \\
&= -2T_n\epsilon^2\eta^2(1-p_n)^2/\theta^2 + 2T_n\epsilon^2\eta^2\delta_n(2(1-p_n)-\delta_n)/\theta^2 + 4c\epsilon\eta^2(1-(p_n+\delta_n))^2/\theta^2 \\
&\leq -2T_n\epsilon^2\eta^2(1-p_n)^2/\theta^2 + 4T_n\epsilon^2\eta^2\delta_n/\theta^2 + 4c\epsilon\eta^2/\theta^2,
\end{aligned} \tag{76}
$$

where the inequality simply uses $p_n, \delta_n > 0$ and $p_n + \delta_n \in (0,1)$ (for large n). Now note that since $p_n \to p$, we have (for example) $(1-p_n)^2 > (1-p)^2/2$ for n sufficiently large. Additionally, since $\delta_n = o(1/T_n)$, we have (for example) $T_n\delta_n < c/\epsilon$ for n sufficiently large. Combining these observations, we can upper bound (76) as

$$
-2\epsilon^2T_n\eta^2(1-p_n)^2/\theta^2 + 4\eta^2\epsilon^2T_n\delta_n/\theta^2 + 4\eta^2\epsilon c/\theta^2 \leq -(\epsilon\eta(1-p))^2T_n/\theta^2 + 8c\epsilon\eta^2/\theta^2.
$$

Hence, substituting into (75) gives

$$
\mathbb{P}_n(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] > \epsilon)1(\Omega_{n,2}) \leq \exp(8c\epsilon\eta^2/\theta^2)\exp(-(\epsilon\eta(1-p)/\theta)^2T_n)1(\Omega_{n,2}). \tag{77}
$$

Finally, we write

$$
\begin{aligned}
\mathbb{P}(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] > \epsilon) &= \mathbb{E}[\mathbb{P}_n(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] > \epsilon)1(\Omega_{n,2}) + \mathbb{P}_n(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] > \epsilon)1(\Omega_{n,2}^C)] \\
&\leq O\left(e^{-(\epsilon\eta(1-p)/\theta)^2T_n}\right) + \mathbb{P}(\Omega_{n,2}^C) = O\left(e^{-(\epsilon\eta(1-p)/\theta)^2\mu\log n} + n^{-\kappa}\right),
\end{aligned}
$$

where the first equality is the law of total expectation, the inequality uses (77) and upper bounds a probability by 1, and the second equality uses the assumptions in the statement of the lemma.

B.3.3 Where the proof fails in the case $p_n \to 1$. As shown in Appendix A.4.1, extending Theorem 3.2 to the case $p_n \to 1$ amounts to showing that for some $\gamma' > 0$,

$$
\mathbb{P}(|\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] - L(p_n)| > \epsilon) = O\left(n^{-\gamma'}\right), \tag{78}
$$

where $L(p_n)$ is the appropriate limit from (31). Here we show (roughly) why the approach from the preceding proof fails to establish (78) in the case $p_n \to 1$. To begin, we note we first used the assumption $p_n \to p < 1$ following (74). Hence, in the case $p_n \to 1$, we can still follow the approach leading to (74) to obtain the (one-sided) bound

$$
\begin{aligned}
&\mathbb{P}(\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] - L(p_n) > \epsilon)1(\Omega_{n,2}) \leq \exp(-\tilde{\lambda}(\epsilon + L(p_n))\mathbb{E}\exp(\tilde{\lambda}\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}])1(\Omega_{n,2}) \\
&\leq \exp\left(-\tilde{\lambda}\epsilon + \tilde{\lambda}\left(-L(p_n) + \mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)]\right) + \frac{\tilde{\lambda}^2\theta^2}{8T_n^2}\sum_{l=1}^{T_n-1}\left(\sum_{l'=l}^{T_n-1}u_{T_n,l'}\tilde{p}_n^{l'-l}\right)^2\right)1(\Omega_{n,2}) \\
&\approx \exp\left(-\tilde{\lambda}\epsilon + \frac{\tilde{\lambda}^2\theta^2}{8T_n^2}\sum_{l=1}^{T_n-1}\left(\sum_{l'=l}^{T_n-1}u_{T_n,l'}\tilde{p}_n^{l'-l}\right)^2\right)1(\Omega_{n,2}),
\end{aligned} \tag{79}
$$

where the approximate equality uses $\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)] \approx L(p_n)$ on $\Omega_{n,2}$ by Lemma A.5. We next note

$$\sum_{l=1}^{T_n-1}\left(\sum_{l'=l}^{T_n-1} u_{T_n,l'}\tilde{p}_n^{l'-l}\right)^2 \geq \left(\sum_{l'=1}^{T_n-1} u_{T_n,l'}\tilde{p}_n^{l'-1}\right)^2 = \left(\sum_{l'=1}^{T_n-1}\left(\sum_{t=l'}^{T_n-1}\binom{t}{l'}\eta^{l'}(1-\eta)^{t-l'}\right)\tilde{p}_n^{l'-1}\right)^2$$
$$= (\tilde{p}_n^*)^{-2}\left(\sum_{t=1}^{T_n-1}\sum_{l'=1}^{t}\binom{t}{l'}\eta^{l'}(1-\eta)^{t-l'}\tilde{p}_n^*\tilde{p}_n^{l'-1}\right)^2 = (\tilde{p}_n^*)^{-2}\left(\frac{T_n}{\theta}\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)] - \frac{1-(1-\eta)^{T_n}}{\eta}\right)^2,$$

where the inequality discards nonnegative terms, the first equality is by definition of $u_{T_n,l'}$, the second rearranges summations and multiplies/divides by $(\tilde{p}_n^*)^2$, and the third uses (38). Hence, we have shown (79) is (roughly) lower bounded by

$$\exp\left(-\tilde{\lambda}\epsilon + \frac{\tilde{\lambda}^2}{8}\left(\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)] - \frac{\theta(1-(1-\eta)^{T_n})}{T_n\eta}\right)^2\right)1(\Omega_{n,2}),$$

where we have also used $\tilde{p}_n^* \approx 1$ for large n on $\Omega_{n,2}$ when $p_n \to 1$ by (A3). Now we consider three cases for the exponent in the previous expression:

- $T_n(1-p_n) \to 0$: Here Lemma A.5 states $\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)] \approx \theta$ for large n on $\Omega_{n,2}$; for such n, the exponent is roughly

$$-\tilde{\lambda}\epsilon + \frac{\tilde{\lambda}^2\theta^2}{8}\left(1 - \frac{\theta(1-(1-\eta)^{T_n})}{T_n\eta}\right)^2 \geq -\tilde{\lambda}\epsilon + \frac{\tilde{\lambda}^2\theta^2}{16} = -\frac{4\epsilon^2}{\theta^2},$$

  where the inequality holds for large n (so that $\theta(1-(1-\eta)^{T_n})/(T_n\eta) < 1 - 1/\sqrt{2}$, which holds since $T_n \to \infty$) and the equality holds by choosing the minimizing $\tilde{\lambda}$ (namely, $\tilde{\lambda} = 8\epsilon/\theta^2$). Since this lower bound is constant in n, (79) does not decay as n grows.
- $T_n(1-p_n) \to c \in (0,\infty)$: Here Lemma A.5 states $\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)] \approx \theta(1-e^{-c\eta})/(c\eta)$ for large n on $\Omega_{n,2}$. An argument similar to the previous case shows (79) does not decay as n grows.
- $T_n(1-p_n) \to \infty$ with $p_n \to 1$: Here we consider an example to show (79) does not decay sufficiently quickly for the general case. In particular, we assume $T_n = \bar{c}\log n$ for some constant $\bar{c}$ that satisfies the theorem assumptions and we set $p_n = 1-(\log n)^{-0.9}$. Then since $\delta_n = o((\log n)^{-1})$ per (A3), we have e.g. $1 - p_n + \delta_n < (1-p_n)/2$ for large n. Hence,

$$\mathbb{E}_n[\hat{\vartheta}_{T_n}(\phi)] \gtrsim \frac{\theta(1-(1-\eta(1-p_n+\delta_n))^{T_n})}{\eta T_n(1-p_n+\delta_n)}$$
$$> \frac{\theta(1-(1-(\eta/2)(\log n)^{-0.9})^{\bar{c}\log n})}{(\bar{c}\eta/2)(\log n)^{0.1}} > \frac{\tilde{c}}{(\log n)^{0.1}},$$

  where the first inequality holds by (39) in Appendix B.2.1 (where γ_1, γ_2 are arbitrarily small, hence the approximate inequality), the second holds for our chosen T_n, p_n, δ_n, and the third holds for some constant $\tilde{c}$ and for large n. Hence, the exponent is (roughly) lower bounded by

$$-\tilde{\lambda}\epsilon + \frac{\tilde{\lambda}^2}{8}\frac{\tilde{c}^2}{(\log n)^{0.2}} = -\frac{2\epsilon^2}{\tilde{c}^2}(\log n)^{0.2},$$

  where the equality holds for the minimizer $\tilde{\lambda} = (4\epsilon/\tilde{c}^2)(\log n)^{0.2}$. From here it follows that (79) cannot be $O(n^{-\gamma'})$: if it is, we have for all large n and for some constant $\tilde{C}$,

$$\exp\left(-\frac{2\epsilon^2}{\tilde{c}^2}(\log n)^{0.2}\right) < \tilde{C}n^{-\gamma'} \Rightarrow \exp\left(-\frac{2\epsilon^2}{\tilde{c}^2}(\log n)^{0.2} + \gamma'\log n\right) < \tilde{C}.$$

  The final inequality is a contradiction, since $-(2\epsilon^2/\tilde{c}^2)(\log n)^{0.2} + \gamma'\log n \to \infty$ as $n \to \infty$.

B.4 Auxiliary results

In this appendix, we collect several auxiliary results used in other proofs. (These results are either cited from other sources, or their proofs are computationally heavy but elementary, so we collect them here to avoid cluttering other parts of our analysis.)

Lemma B.1. *For $T_n \to \infty$, $p_n \to 1$, and $\delta_n \to 0$ s.t. $\delta_n = o(1/T_n)$, we have*

$$\frac{1-(1-\eta(1-p_n-\delta_n))^{T_n}}{\eta T_n(1-p_n-\delta_n)} \xrightarrow[n\to\infty]{} \begin{cases} 1, & T_n(1-p_n) \xrightarrow[n\to\infty]{} 0 \\ (1-e^{-c\eta})/(c\eta), & T_n(1-p_n) \xrightarrow[n\to\infty]{} c \in (0,\infty) \\ 0, & T_n(1-p_n) \xrightarrow[n\to\infty]{} \infty \end{cases}, \tag{80}$$

$$\frac{1-(1-\eta(1-p_n+\delta_n))^{T_n}}{\eta T_n(1-p_n+\delta_n)} \xrightarrow[n\to\infty]{} \begin{cases} 1, & T_n(1-p_n) \xrightarrow[n\to\infty]{} 0 \\ (1-e^{-c\eta})/(c\eta), & T_n(1-p_n) \xrightarrow[n\to\infty]{} c \in (0,\infty) \\ 0, & T_n(1-p_n) \xrightarrow[n\to\infty]{} \infty \end{cases}. \tag{81}$$

Proof. We consider the three cases of (80) in turn; the proof of (81) follows the same approach.

First, suppose $\lim_{n\to\infty} T_n(1-p_n) = \infty$. Then since $T_n\delta_n \to 0$ and $T_n(1-p_n) \to \infty$, we have $T_n\delta_n < 1 < T_n(1-p_n)$ for sufficiently large n, which implies $(1-p_n-\delta_n) > 0$ for such n. Clearly, we also have $(1-p_n-\delta_n) < 1$ for all n. Taken together, it follows that $1-(1-\eta(1-p_n-\delta_n))^{T_n} \in (0,1)$ for n large. For such n, we can then write

$$0 < \frac{1-(1-\eta(1-p_n-\delta_n))^{T_n}}{\eta T_n(1-p_n-\delta_n)} < \frac{1}{\eta T_n(1-p_n-\delta_n)},$$

where we used $(1-p_n-\delta_n) > 0$ in the denominator. Now since $T_n(1-p_n) \to \infty$ and $T_n\delta_n \to 0$, $T_n(1-p_n-\delta_n) \to \infty$, so taking $n \to \infty$ in the above inequality gives the result.

Next, suppose $\lim_{n\to\infty} T_n(1-p_n) = c \in (0,\infty)$. Since $\eta T_n(1-p_n-\delta_n) \to \eta c$ by $T_n(1-p_n) \to c$ and $T_n\delta_n \to 0$, it suffices to show $(1-\eta(1-p_n-\delta_n))^{T_n} \to e^{-\eta c}$ as $n \to \infty$. First, since $T_n(1-p_n) \to c$, $\forall\, \epsilon_1 > 0\ \exists\, N_1$ s.t. $c-\epsilon_1 < T_n(1-p_n) < c+\epsilon_1\ \forall\, n \geq N_1$. Further, since $T_n\delta_n \to 0$, $\forall\, \epsilon_2 > 0\ \exists\, N_2$ s.t. $-\epsilon_2 < T_n\delta_n < \epsilon_2\ \forall\, n \geq N_2$. Hence, $\forall\, n \geq \max\{N_1, N_2\}$,

$$\left(1 - \frac{\eta(c+\epsilon_1+\epsilon_2)}{T_n}\right)^{T_n} < (1-\eta(1-p_n-\delta_n))^{T_n} < \left(1 - \frac{\eta(c-\epsilon_1-\epsilon_2)}{T_n}\right)^{T_n}.$$

Next, we note

$$\lim_{n\to\infty}\left(1 - \frac{\eta(c+\epsilon_1+\epsilon_2)}{T_n}\right)^{T_n} = e^{-\eta(c+\epsilon_1+\epsilon_2)}, \lim_{n\to\infty}\left(1 - \frac{\eta(c-\epsilon_1-\epsilon_2)}{T_n}\right)^{T_n} = e^{-\eta(c-\epsilon_1-\epsilon_2)}.$$

Hence, $\forall\, \epsilon_3 > 0\ \exists\, N_3$ s.t. $\forall\, n \geq N_3$,

$$e^{-\eta(c+\epsilon_1+\epsilon_2)} - \epsilon_3 < \left(1 - \frac{\eta(c+\epsilon_1+\epsilon_2)}{T_n}\right)^{T_n}, \left(1 - \frac{\eta(c-\epsilon_1)}{T_n}\right)^{T_n} < e^{-\eta(c-\epsilon_1-\epsilon_2)} + \epsilon_3.$$

Combining these arguments, we obtain $\forall\, n \geq \max\{N_1, N_2, N_3\}$

$$e^{-\eta(c+\epsilon_1+\epsilon_2)} - \epsilon_3 < (1-\eta(1-p_n-\delta_n))^{T_n} < e^{-\eta(c-\epsilon_1-\epsilon_2)} + \epsilon_3.$$

Since both bounds converge to $e^{-\eta c}$ as $\epsilon_1, \epsilon_2, \epsilon_3 \to 0$, $(1-\eta(1-p_n-\delta_n))^{T_n} \to e^{-\eta c}$ follows.

Finally, suppose $\lim_{n\to\infty} T_n(1-p_n) = 0$. First, we observe

$$\frac{1-(1-\eta(1-p_n-\delta_n))^{T_n}}{\eta T_n(1-p_n-\delta_n)} = \frac{1}{T_n}\sum_{t=0}^{T_n-1}(1-\eta(1-p_n-\delta_n))^t \leq 1, \tag{82}$$

where the inequality holds for n s.t. $(1-p_n-\delta_n)>0$ (which indeed occurs for large n; see proof of $T_n(1-p_n)\to\infty$ case), since then the sum is over T_n terms, each upper bounded by 1. On the other hand, we can use the binomial theorem to write

$$\frac{1-(1-\eta(1-p_n-\delta_n))^{T_n}}{\eta T_n(1-p_n-\delta_n)} = \frac{1-\sum_{t=0}^{T_n}\binom{T_n}{t}(-\eta(1-p_n-\delta_n))^t}{\eta T_n(1-p_n-\delta_n)} \tag{83}$$
$$= 1-\sum_{t=2}^{T_n}\frac{(T_n-1)\cdots(T_n-t+1)(-1)^t(\eta(1-p_n-\delta_n))^{t-1}}{t!}.$$

Next, we observe (assuming $(1-p_n-\delta_n)>0$) as above)

$$\sum_{t=2}^{T_n}\frac{(T_n-1)\cdots(T_n-t+1)(-1)^t(\eta(1-p_n-\delta_n))^{t-1}}{t!} \tag{84}$$
$$< \sum_{t=2}^{T_n}\frac{(T_n-1)\cdots(T_n-t+1)(\eta(1-p_n-\delta_n))^{t-1}}{t!} < \sum_{t=2}^{T_n}\frac{(T_n(1-p_n-\delta_n))^{t-1}}{(t-2)!}$$
$$= T_n(1-p_n-\delta_n)\sum_{t=0}^{T_n-2}\frac{(T_n(1-p_n-\delta_n))^t}{t!} < T_n(1-p_n-\delta_n)e^{T_n(1-p_n-\delta_n)},$$

where the first inequality replaces negative terms with positive ones; the second inequality uses $\eta<1$, $(t-2)!<t!$, and $(T_n-j)<T_n$ for $j>0$; and the third inequality upper bounds the summation by replacing its upper limit with infinity. Hence, (82), (83), and (84) yield

$$1 \geq \frac{1-(1-\eta(1-p_n-\delta_n))^{T_n}}{\eta T_n(1-p_n-\delta_n)} > 1-T_n(1-p_n-\delta_n)e^{T_n(1-p_n-\delta_n)}$$
$$\Rightarrow 1 \geq \lim_{n\to\infty}\frac{1-(1-\eta(1-p_n-\delta_n))^{T_n}}{\eta T_n(1-p_n-\delta_n)} \geq 1-\lim_{n\to\infty}T_n(1-p_n-\delta_n)e^{T_n(1-p_n-\delta_n)} = 1,$$

where the final equality holds since $T_n(1-p_n), T_n\delta_n \to 0$ by assumption. □

Lemma B.2. *Let* $u_{T_n,l}=\sum_{t=l}^{T_n-1}\binom{t}{l}\eta^l(1-\eta)^{t-l}$. *Then for any* $x\in(0,1)$,

$$\sum_{l=1}^{T_n-1}\left(\sum_{l'=l}^{T_n-1}u_{T_n,l'}x^{l'-l}\right)^2 \leq \frac{T_n}{\eta^2(1-x)^2}.$$

Proof. For $l\in\mathbb{N}_0$, define $w_l=\sum_{l'=l}^{T_n-1}u_{T_n,l'}x^{l'-l}$. Then

$$w_l = u_{T_n,l}+x\sum_{l'=l+1}^{T_n-1}u_{T_n,l'}x^{l'-(l+1)} = u_{T_n,l}+xw_{l+1}.$$

Assuming temporarily that $u_{T_n,l'}\geq u_{T_n,l''}$ whenever $l'\leq l''$ (which we will return to prove),

$$w_{l+1}\leq u_{T_n,l}\sum_{l'=l+1}^{T_n-1}x^{l'-(l+1)} = u_{T_n,l}\sum_{l'=0}^{T_n-l-2}x^{l'} \leq u_{T_n,l}\sum_{l'=0}^{\infty}x^{l'} = \frac{u_{T_n,l}}{1-x}.$$

Hence, using the previous two equations, we obtain $w_{l+1}-w_l=(1-x)w_{l+1}-u_{T_n-l}\leq 0$, i.e. the sequence $\{w_l\}$ decreases in l. It is also clearly nonnegative. Therefore,

$$\sum_{l=1}^{T_n-1}\left(\sum_{l'=l}^{T_n-1}u_{T_n,l'}x^{l'-l}\right)^2 = \sum_{l=1}^{T_n-1}w_l^2 \leq T_n w_0^2.$$

To further bound the right hand side, we note

$$\begin{aligned} w_0 &= \sum_{l'=0}^{T_n-1}\left(\sum_{t=l'}^{T_n-1}\binom{t}{l'}\eta^{l'}(1-\eta)^{t-l'}\right)x^{l'} = \sum_{t=0}^{T_n-1}\sum_{l'=0}^{t}\binom{t}{l'}(\eta x)^{l'}(1-\eta)^{t-l'} \\ &= \sum_{t=0}^{T_n-1}(\eta x+(1-\eta))^t = \sum_{t=0}^{T_n-1}(1-\eta(1-x))^t \leq \sum_{t=0}^{\infty}(1-\eta(1-x))^t = \frac{1}{\eta(1-x)}, \end{aligned}$$

where the first equality uses the definition of $u_{T_n,l'}$, the second rearranges summations, the third uses the binomial theorem, the fourth is immediate, the inequality is immediate, and the final equality computes a geometric series. Combining the previous two inequalities proves the lemma.

We return to prove $u_{T_n,l'} \geq u_{T_n,l''}$ whenever $l' \leq l''$. For this, we first claim

$$\sum_{t=l}^{t^*}\binom{t}{l}\eta^l(1-\eta)^{t-l} - \sum_{t=l+1}^{t^*+1}\binom{t}{l+1}\eta^{l+1}(1-\eta)^{t-(l+1)} = \binom{t^*+1}{l+1}\eta^l(1-\eta)^{t^*+1-l}\ \forall\, t^*\in\mathbb{N}, l\in\{1,\ldots,t^*\}. \tag{85}$$

We prove (85) by induction on t^*. First, when $t^*=1$, the only case to prove is $l=1$; when $t^*=l=1$, it is immediate that both sides of (85) equal $\eta(1-\eta)$. Next, assume (85) holds for t^*-1. If $l=t^*$, both sides of (85) equal $\eta^{t^*}(1-\eta)$. If $l\in\{1,\ldots,t^*-1\}$, we write

$$\begin{aligned} &\sum_{t=l}^{t^*}\binom{t}{l}\eta^l(1-\eta)^{t-l} - \sum_{t=l+1}^{t^*+1}\binom{t}{l+1}\eta^{l+1}(1-\eta)^{t-(l+1)} \\ &\quad= \left(\sum_{t=l}^{t^*-1}\binom{t}{l}\eta^l(1-\eta)^{t-l} - \sum_{t=l+1}^{t^*}\binom{t}{l+1}\eta^{l+1}(1-\eta)^{t-(l+1)}\right) \\ &\qquad+\binom{t^*}{l}\eta^l(1-\eta)^{t^*-l} - \binom{t^*+1}{l+1}\eta^{l+1}(1-\eta)^{t^*+1-(l+1)} \\ &\quad= \left(\binom{t^*}{l+1}\eta^l(1-\eta)^{t^*-l}\right) + \binom{t^*}{l}\eta^l(1-\eta)^{t^*-l} - \binom{t^*+1}{l+1}\eta^{l+1}(1-\eta)^{t^*-l} \\ &\quad= \eta^l(1-\eta)^{t^*-l}\left(\binom{t^*}{l+1}+\binom{t^*}{l}-\eta\binom{t^*+1}{l+1}\right) \\ &\quad= \eta^l(1-\eta)^{t^*-l}\left(\binom{t^*+1}{l+1}-\eta\binom{t^*+1}{l+1}\right) = \binom{t^*+1}{l+1}\eta^l(1-\eta)^{t^*+1-l}, \end{aligned}$$

where the first equality simply writes the final summands separately, the second uses the inductive hypothesis on the term in parentheses, the third is immediate, the fourth uses Pascal's rule ($[t^*+1]$ has $\binom{t^*+1}{l+1}$ subsets of cardinality $l+1$; $\binom{t^*}{l}$ that contain 1 and $\binom{t^*}{l+1}$ that do not contain 1), and the fifth is immediate. This establishes (85). We then write

$$\begin{aligned} u_{T_n,l'} - u_{T_n,l'+1} &= \sum_{t=l'}^{T_n-1}\binom{t}{l'}\eta^{l'}(1-\eta)^{t-l'} - \sum_{t=l'+1}^{T_n-1}\binom{t}{l'+1}\eta^{l'+1}(1-\eta)^{t-(l'+1)} \\ &= \sum_{t=l'}^{T_n-1}\binom{t}{l'}\eta^{l'}(1-\eta)^{t-l'} - \sum_{t=l'+1}^{T_n}\binom{t}{l'+1}\eta^{l'+1}(1-\eta)^{t-(l'+1)} + \binom{T_n}{l'+1}\eta^{l'+1}(1-\eta)^{T_n-(l'+1)} \\ &= \binom{T_n}{l'+1}\eta^{l'}(1-\eta)^{T_n-l'} + \binom{T_n}{l'+1}\eta^{l'+1}(1-\eta)^{T_n-(l'+1)} = \binom{T_n}{l'+1}\eta^{l'}(1-\eta)^{T_n-(l'+1)} \geq 0, \end{aligned}$$

where the first equality holds by definition of $u_{T_n,l'}$, the second adds and subtracts a term, and the third uses (85). This shows $u_{T_n,l'} \geq u_{T_n,l'+1}$, iterating gives $u_{T_n,l'} \geq u_{T_n,l''}$ whenever $l' \leq l''$. □

LEMMA B.3. *Let Z be a random variable satisfying $\mathbb{E}Z = 0$ and $Z \in [a,b]$ a.s., and let $\lambda > 0$. Then*

$$\mathbb{E}e^{\lambda Z} \leq e^{\lambda^2 (b-a)^2/8}.$$

PROOF. See e.g. Lemma 5.1 in [12]. □

APPENDIX C EXPERIMENTAL DETAILS

The basic workflow of our experiment proceeded as follows:

- Choose a sequence of time horizons T_n that increase linearly, then set n accordingly.
- Realize the degrees $\{d_{out}(i), d_{in}^A(i), d_{in}^B(i)\}_{i\in[n]}$ after selecting n.
- Define the empirical distributions f_n, f_n^* using the degrees as in (7).
- Evaluate quantity of interest $\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]$ empirically via (26) using f_n, f_n^*.

We repeated this experiment 400 times to obtain 400 samples of $\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}]$; the plots in Figure 2 show empirical means and variances across these 400 samples. We used the following parameters:

- We set $\eta = 0.9$ to emphasize the effect of the network.
- We let $d_{in}^A(i) = 1 + \text{Poisson}(\lambda_A - 1)\ \forall\ i \in [n]$, so that $\mathbb{E}[d_{in}^A(i)] = \lambda_A$; we choose λ_A independent of n so that $\mathbb{E}[d_{in}^A(i)] = O(1)$, as required by (A1). In particular, we choose $\lambda_A = 2.1$.
- After realizing $\{d_{in}^A(i)\}_{i\in[n]}$, we assign one outgoing edge to each $i \in [n]$, then assign each of the remaining $\sum_{i\in[A]} d_{in}^A(i) - n$ outgoing edges independently and uniformly at random. Note that this implies $d_{in}^A(i), d_{out}(i) > 0$ and $\sum_{i\in[n]} d_{in}^A(i) = \sum_{i\in[n]} d_{out}(i)$, as required by (5).
- We let $d_{in}^B(i) = \text{Poisson}(\lambda_B)$, with $\lambda_B = \lambda_A(1-p_n)/p_n$, so that

$$\frac{\mathbb{E}d_{in}^A(i)}{\mathbb{E}d_{in}^A(i) + \mathbb{E}d_{in}^B(i)} = \frac{\lambda_A}{\lambda_A + \lambda_B} = \frac{1}{1-(1-p_n)/p_n} = p_n.$$

  (This is not precisely what we desire, since (A3) assumes $p_n \approx \tilde{p}_n = \mathbb{E}_n[\frac{d_{in}^A(v^*)}{d_{in}^A(v^*)+d_{in}^B(v^*)}]$ for v^* sampled proportional to out-degree; however, as shown in the second plot in Figure 2, this empirically yields distinct cases rates of convergence for $(1-p_n) \longrightarrow 0$.)
- We compare four cases of p_n: $p_n = p$ and $p_n = 1 - c_i T_n^{(-i+1)/2}$ for $i \in \{2,3,4\}$, with p and c_i independent of n. Note that the three latter cases satisfy

$$(1-p_n) \propto T_n^{(-i+1)/2} \in \left\{T_n^{-1/2}, T_n^{-1}, T_n^{-3/2}\right\},$$

  as shown in Figure 2. Here p and c_i were chosen via trial-and-error so that all four cases behaved roughly the same at the smallest value of n (as in Figure 2). In particular, we chose

$$p = 0.9, \quad c_2 = 1.3, \quad c_3 = 1.9, \quad c_4 = 2.7.$$

- We let $T_n \in \{2,3,\ldots,11\}$; here the minimum of 2 was chosen since $T_n = 1$ is a trivial case and the maximum of 11 was chosen due to computational limitations.
- Given T_n, we let $n = \lceil \lambda_A^{2T_n} \rceil$. Note that this implies $T_n \approx (\log n)/(2\log \lambda_A)$, roughly the upper bound in (A2). With our choice of T_n and λ_A, n ranged from 20 to (roughly) 12 million.

In addition to the summary statistics shown in Figure 2, we also show histograms of error term $|\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] - L(p_n)|$ across the 400 trials in Figure 3. As discussed in Section 3.3, this term must converge to zero (in probability) at a sufficiently fast rate to prove Theorem 3.2. In Figure 3, these histograms appear to converge quickly to a point mass at zero in the case $p_n \longrightarrow p < 1$; in other cases, such behavior does *not* occur, further suggesting a fundamental difference between the cases.

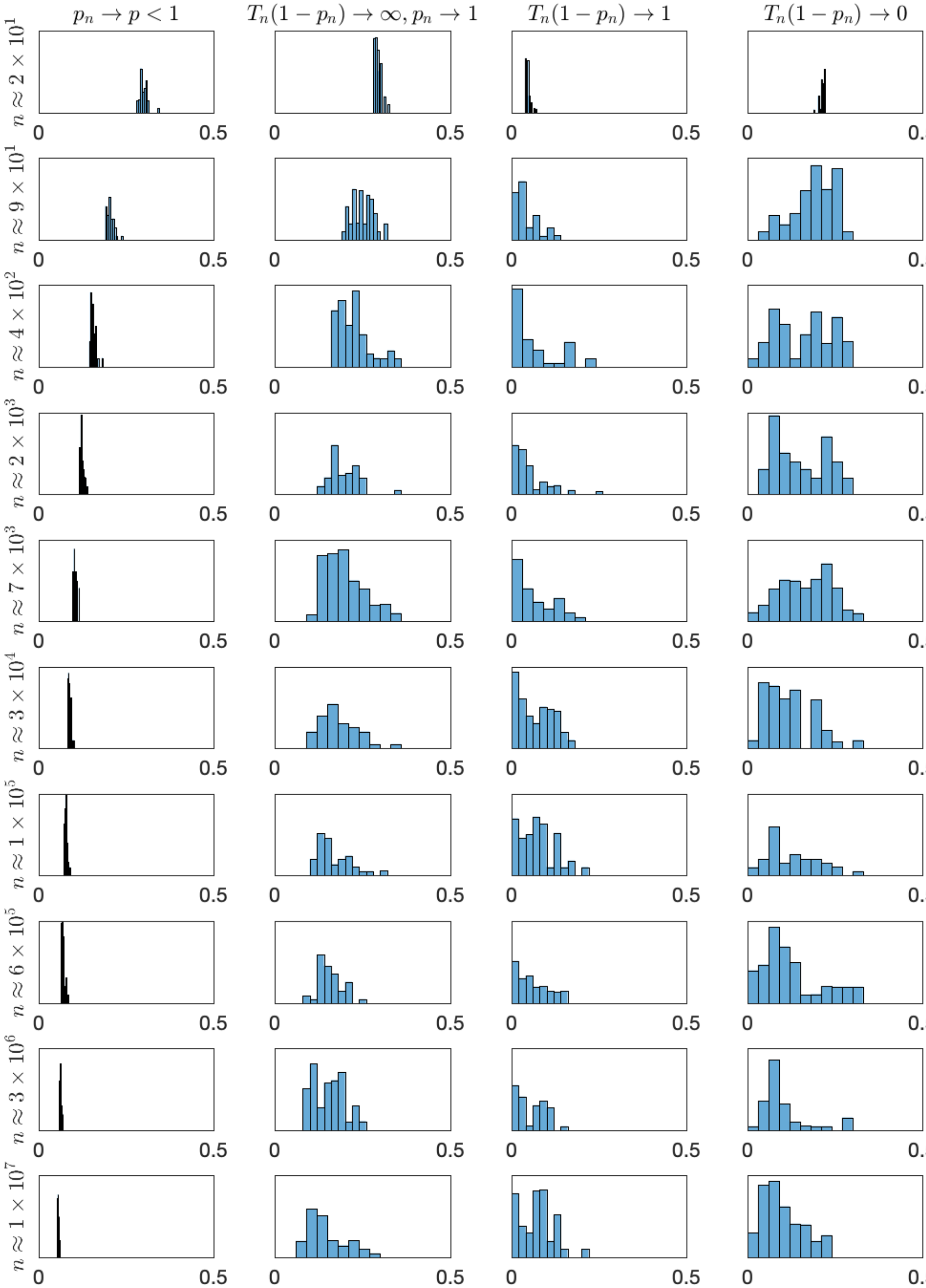

Fig. 3. Histograms of 400 samples of the error term $|\mathbb{E}[\hat{\vartheta}_{T_n}(\phi)|\mathcal{T}] - L(p_n)|$. When $p_n \to p < 1$, the histogram appears to decay quickly to a point mass on zero; in other cases, this does not occur.