Asynchronous Majority Dynamics in Preferential Attachment Trees

Maryam Bahrani*, Nicole Immorlica†, Divyarthi Mohan‡, S. Matthew Weinberg§

Abstract

We study information aggregation in networks where agents make binary decisions (labeled incorrect or correct). Agents initially form independent private beliefs about the better decision, which is correct with probability $1/2 + \delta$. The dynamics we consider are asynchronous (each round, a single agent updates their announced decision) and non-Bayesian (agents simply copy the majority announcements among their neighbors, tie-breaking in favor of their private signal).

Our main result proves that when the network is a tree formed according to the preferential attachment model [5], with high probability, the process stabilizes in a correct majority. We extend our results to other tree structures, including balanced $M$-ary trees for any $M$. 

*Princeton University, mbahrani@princeton.edu.
†Microsoft Research, nicimm@microsoft.com
‡Princeton University, dm23@princeton.edu
§Princeton University, smweinberg@princeton.edu. Supported by NSF CCF-1717899.
1 Introduction

Individuals form opinions about the world both through private investigation and through discussion with one-another. A citizen, trying to decide which candidate’s economic policies will lead to more jobs, might form an initial belief based on her own employment history. However, her stated opinion might be swayed by the opinions of her friends. The dynamics of this process, together with the social network structure of the individuals, can result in a variety of societal outcomes. Even if individuals are well-informed, i.e., are more likely to have correct than incorrect initial beliefs, certain dynamics and/or network structures can cause large portions of the population to form mistaken opinions.

A substantial body of work exists modeling these dynamics mathematically, which we overview in Section 1.2. This paper focuses on the model of asynchronous majority dynamics. Initially, individuals have private beliefs over a binary state of the world, but no publicly stated opinion. Initial beliefs are independent: Correct with probability \( 1/2 + \delta \), and Incorrect with probability \( 1/2 - \delta \). In each time step, a random individual is selected to announce a public opinion. Each time an individual announces a public opinion, they simply copy the majority of their neighbors’ announcements, tie-breaking in favor of their private belief. This is clearly naive: a true Bayesian would reason about the redundancy of information among the opinions of her friends, for example. Majority (or other non-Bayesian) dynamics are generally considered a more faithful model of agents with bounded rationality (e.g. voters), whereas Bayesian dynamics are generally considered a more faithful model of fully rational actors (e.g. financial traders).

It’s initially tempting to conjecture that these dynamics in a connected network should result in a Correct consensus; after all, the majority is initially Correct (with high probability) by assumption. Nonetheless, it’s well-understood that individuals can fail miserably to learn. Suppose for instance that the individuals form a complete graph. Then in asynchronous majority dynamics, whichever individual is selected to announce first will have their opinion copied by the entire network. As this opinion is Incorrect with constant probability, there’s a good chance that the entire network makes the wrong decision (this is known as an information cascade, and is not unique to asynchronous majority dynamics [4, 7]). So the overarching goal in these works is to understand in which graphs the dynamics stabilize in correctness with high probability.

For most previously studied dynamics (discussed in Section 1.2), “correctness” means a Correct consensus. This is because the models terminate in a consensus with probability 1, and the only question is whether this consensus is correct or not. With majority dynamics, it is certainly possible that the process stabilizes without a consensus. To see this, suppose individuals form a line graph. In this case, two adjacent individuals with the same initial belief are likely to form a “road block” (if both announce before their other neighbors), sticking to their initial beliefs throughout the process. In this case, with high probability a constant fraction of individuals terminate with a Correct opinion, but also a constant fraction terminate with an Incorrect opinion. As consensus is no longer guaranteed, we’re instead interested in understanding network structures for which the dynamics converge, with high probability, to a majority of nodes having the Correct opinion (i.e., if a majority vote were to be taken, would it be correct w.h.p.?).

Prior work shows that it’s sufficient for the social network to be sparse (every individual has only a constant number of neighbors) and expansive (every group of individuals have many friends outside the group) [11], and the tools developed indeed make strong use of both assumptions. Many networks of interest, however, like the hierarchy of employees in a corporation, are neither sparse nor expansive. Therefore, the focus of this paper is to push beyond these assumptions and develop tools for more general graphs.
1.1 Our Results and Techniques

We focus our attention on trees, the simplest graphs outside the reach of prior techniques. In addition to modeling certain types of social networks (including hierarchical ones, or communication networks in which redundancy is expensive), and forming the backbone of many more, trees already present a number of technical challenges whose absence enabled the prior results. Our main result is the following:

**Theorem 1.** Let $G$ be a tree. Then with probability $1 - o(1)$, asynchronous majority dynamics in $G$ stabilizes in a **Correct** majority if:

- $G$ is formed according to the preferential attachment model.$^1$
- $G$ is a balanced, $M$-ary tree of any degree.$^2$

**Beyond Prior Tools.** In prior work [11], the authors have two key ideas. Without yet getting into full details, one key idea crucially invokes sparsity to claim that most pairs of nodes $u, v$ have distance $d(u, v) = \Omega(\ln n / \ln \ln n)$, which allows them to conclude that after $O(n \ln n / \ln \ln n)$ steps, most nodes are announcing a **Correct** opinion.

Still, just the fact that the dynamics hit a **Correct** majority along the way does not imply that the **Correct** majority will hold thru termination. To wrap up, they crucially invoke expansiveness (building off an argument of [21]) to claim that once there is a **Correct** majority, it spreads to a **Correct** consensus with high probability.

Both properties are necessary for prior work, and both properties fail in trees. For instance, the star graph is a tree, and $d(u, v) \leq 2$ for all $u, v$ (precluding their “majority at $O(n \ln n / \ln \ln n)$ argument). Additionally, trees are not expansive. In particular, the line graph discussed earlier is a tree which hits a **Correct** majority at some point (as this tree happens to be sparse), but does not converge to consensus, so there is no hope for an argument like this.

**New Tools.** Our main technical innovation is an approach to reason about majority without going through consensus. Specifically, we show in Sections 5 and 6 for preferential attachment trees, or balanced $M$-ary trees, that with probability $1 - o(1)$, a $1 - o(1)$ fraction of nodes have **finalized** after $O(n \ln n / \ln \ln n)$ steps. That is, after $O(n \ln n / \ln \ln n)$ steps, most (but not all) of the network has stabilized. We postpone further details to Section 5, but just wish to highlight this approach as a fairly significant deviation from prior work.

From here, our task is now reduced to showing that a **Correct** majority exists w.h.p. after $O(n \ln n / \ln \ln n)$ steps. Our main insight here is that most nodes with $d(u, v) = O(\ln n / \ln \ln n)$ must have some high-degree nodes along the path from $u$ to $v$. We prove that such nodes act like a “road block,” causing announcements on either side to be independent with high probability (and all nodes with $d(u, v) = \Omega(\ln n / \ln \ln n)$ can be handled with similar arguments to prior work).

1.2 Related Work

Information aggregation in social networks is an enormous field, and we will not come close to overviewing it in its entirety. Below, we’ll briefly summarize the most related literature, restrict-

---

$^1$That is, $G$ is created by adding nodes one at a time. When a node is added, it attaches a single edge to a random previous node, selected proportional to its degree.

$^2$That is, $G$ can be rooted at some node $v$. All non-leaf nodes have $M$ children, and all leaves have the same distance to $v$. 

2
ing attention to works that consider two states of the world and independent initial beliefs are independently CORRECT with probability $1/2 + \delta$.

**Bayesian Dynamics.** In Bayesian models, agents are fully rational and sequentially perform Bayesian updates to their public opinion based on the public opinions of their neighbors. Seminal works of Banerjee [4] and Bikchandani, Hirshleifer, and Welch [7] first identified the potential of information cascades in this model. Subsequent works consider numerous variations, aiming to understand what assumptions on the underlying network or information structure results in CORRECT consensus [23, 3, 2]. While the high-level goals of these works align with ours, technically they are mostly unrelated as we consider non-Bayesian dynamics.

**Voter and DeGroot Dynamics.** Prior work also considered other non-Bayesian dynamics. In voter dynamics, individuals update by copying a random neighbor [8, 16]. Similar dynamics (such as 3-majority, or $k$-majority) are analyzed by distributed computing perspective with an emphasis on the rate of convergence to consensus [6, 13, 12]. In the DeGroot model, individuals announce an opinion in $[0, 1]$ (as opposed to $\{0, 1\}$), and update by averaging their neighbors [9, 14]. The biggest difference between these works and ours is that consensus is reached with probability 1 in these models on any connected graph, which doesn’t hold for majority dynamics.

**Majority Dynamics.** The works most related to ours consider majority dynamics. Even synchronous majority dynamics may not result in a consensus (consider again the line graph). These works, like ours, therefore seek to understand what graph structures result in a CORRECT majority. Mossel, Neeman, and Tamuz study synchronous majority dynamics and prove that a CORRECT majority arises as long as the underlying graph is sufficiently symmetric, or sufficiently expansive (in the latter case, they prove that the network further reaches consensus) [21]. Feldman et al. study asynchronous majority dynamics and prove that a CORRECT consensus arises when the underlying graph is sparse and expansive [11]. Work of [24] further studies “retention of information,” which asks whether any recovery procedure (not necessarily a majority vote) at stabilization can recover the ground truth with high probability. In connection to these, our work simply pushes the boundary beyond what classes of graphs are understood in prior work.

The key difference between the synchronous and asynchronous models is captured by the complete graph. In asynchronous dynamics, a CORRECT majority occurs only with probability $1/2 + \delta$, whereas in synchronous dynamics a CORRECT consensus occurs with probability $1 - \exp(-\Omega(n))$. This is because in step one, every node simply announces their private belief, and in step two everyone updates to the majority, which is CORRECT with probability $1 - \exp(-\Omega(n))$. So while the models bear some similarity, and some tools are indeed transferable (e.g. the expansiveness lemma of [21] used in [11]), much of the analyses will necessarily diverge.

**Preferential Attachment and Balanced $M$-ary Trees.** There is also substantial prior work studying aggregation dynamics in trees. Here, the most related work is [17, 20], which studies synchronous majority dynamics in balanced $M$-ary trees. Less related are works which study “bottom-up” dynamics in balanced $M$-ary trees [19, 26, 25], $k$-majority dynamics in preferential attachment trees [1], or model cascades themselves as a preferential attachment tree [15]. While these works provide ample motivation for restricting attention to preferential attachment trees, or balanced $M$-ary trees, they bear no technical similarity to ours.
2 Model and Preliminaries

We consider an undirected tree \( G = (V, E) \) with \( |V(G)| = n \) and \( |E(G)| = m \). We denote by \( \deg(v) \) the degree of a node \( v \in V(G) \), \( N(v) \) to be its neighbors \( \{u, (u, v) \in E\} \), and \( d(u, v) \) to be the length of the unique path between \( u \) and \( v \), and let \( P(u, v) \) denote the ordered list of vertices on this path (i.e. starting with \( u \) and ending with \( v \)). We'll also denote by \( D(G) = \max_{u,v}\{d(u, v)\} \) the diameter of \( G \).

Individuals initially have one of two private beliefs, which we'll refer to as \textsc{Correct} (or 1) and \textsc{Incorrect} (or 0). That is, each \( v \in V(G) \) receives an independent private signal \( X(v) \in \{0, 1\} \), and \( \Pr[X(v) = 1] = 1/2 + \delta \), for some constant \( 0 < \delta < 1/2 \).

Individuals also have a \textit{publicly announced} opinion (which we will simply refer to as an announcement). We define \( C^t(v) \in \{\perp, 0, 1\} \) to be the public announcement of \( v \in V(G) \) at time \( t \). Initially, no announcements have been made, i.e. \( C^0(v) = \perp \) for all \( v \). In each subsequent step, a single node \( v^t \) is chosen uniformly at random from \( V(G) \) and updates their announcement (announcements of all other nodes stay the same). \( v^t \) updates her announcement using \textit{majority dynamics}. That is, if \( N_1^t(v) \) denotes the number of \( v \)'s neighbors with a \textsc{Correct} announcement at time \( t \), and \( N_0^t(v) \) denotes the number of \( v \)'s neighbors with an \textsc{Incorrect} announcement, then:

\[
C^t(v) = \begin{cases} 
1 & \text{if } N_1^{t-1}(v) > N_0^{t-1}(v), \text{ and } v = v^t, \\
0 & \text{if } N_1^{t-1}(v) < N_0^{t-1}(v), \text{ and } v = v^t, \\
x(v) & \text{if } N_1^{t-1}(v) = N_0^{t-1}(v), \text{ and } v = v^t, \\
C^{t-1}(v) & \text{if } v \neq v^t.
\end{cases}
\]

Note that we will treat \( \delta \) as an absolute constant. Therefore, the only variable taken inside Big-Oh notation is \( n \), the number of nodes (and, for instance, when we write \( o(1) \) we mean any function of \( n \) that approaches 0 as \( n \) approaches \( \infty \)).

2.1 Concentration Bounds and Tools from Prior Work

Our work indeed makes use of some tools from prior work to get started, which we state below. The concept of a \textit{critical time}, defined below, is implicit in [11].

\textbf{Definition 1.} The critical time from \( u \) to \( u \), \( T(u, u) \), is the first time that node \( u \) announces. The critical time from \( u \) to \( v \), \( T(u, v) \), is recursively defined as the first time that \( v \) announces after the critical time from \( u \) to \( x \), where \( x \) is the neighbor of \( v \) in \( P(u, v) \). We further denote the critical chain from \( u \) to \( v \) as the ordered list of critical times from \( u \) to \( x \) for all \( x \) on \( P(u, v) \).

The following lemma is a formal statement of ideas from prior work (a proof appears in Appendix A). To parse it, it will be helpful to think of the process as first drawing a countably infinite sequence \( S \) of nodes to announce, which then allows each \( C^t(v) \) to be written as a deterministic function of the random variables \( \{X(u), u \in V\} \). Lemma 1 below states that in fact, for early enough \( t \), initial beliefs for only a proper subset of \( V \) suffice.

\textbf{Lemma 1 ([11]).} For all \( t \), and all \( v \), \( C^t(v) \) can be expressed as a function of \( \{X(u), T(u, v) \leq t\} \).

The final theorem we take from prior work is due to Mossel et al., and is used to claim that at minimum the expected number of \textsc{Correct} nodes at termination is at least \( (1/2 + \delta)n \).

\textbf{Theorem 2 ([21]).} Let \( f \) be an odd, monotone Boolean function. Let \( X_1, \ldots, X_n \) be input bits, each sampled i.i.d. from a distribution that is 1 with probability \( p \geq 1/2 \) and 0 otherwise. Then

\[ \mathbb{E}[f(X_1, \ldots, X_n)] \geq p. \]
Note that \( C^t(v) \) is an odd, monotone, Boolean function in variables \( \{X(u), u \in V\} \), and therefore (as long as \( v \) has announced at least once by \( t \)) \( \Pr[C^t(v) = 1] \geq 1/2 + \delta \) for all \( v, t \).

Finally, we’ll make use of the following concentration bound on \( T(u, v) \) repeatedly. It’s proof is a simple application of a Chernoff bound.

**Lemma 2.** \( \Pr[T(u, v) > 8 \cdot \max\{\ln(1/y), d(u, v) + 1\} \cdot n] \leq e^{-y d(u, v)(1-y)^2/3}. \)

### 3 Key Concepts

Before getting into our proofs, we elaborate some key concepts that will be used throughout. In Proposition 1 below, we analyze the connection between critical chains and switches in announcements. Intuitively, Proposition 1 is claiming that every fresh announcement can cause other nodes to switch a previous announcement along critical chains, but that these are the only switches that can occur.

**Definition 2.** Let \( v \) change her announcement at \( t \), and her previous announcement be \( t' > 0 \). We say that node \( u \) is a cause of \( v \) changing her announcement at \( t \) if \( C^t(u) = C^t(v) \), and \( C^t(u) \neq C^t(v) \). Observe that every such change in announcement has a cause.

**Proposition 1.** If \( C^t(v) \neq C^{t-1}(v) \), then \( t = T(u, v) \) for some \( u \). Moreover, if \( u = x_0, x_1, \ldots, x_{|P(u,v)|} = v \) denotes \( P(u,v) \), then \( C^{T(u,x)}(u) = C^t(v) \), and every \( x_i, i > 0 \), has \( C^{T(u,x)}(x_i) = C^{t-1}(v) \) and \( C^{T(u,x)}(x) = C^t(v) \), and \( x_{i-1} \) caused this change.

**Proof.** The proof proceeds by induction on \( t \). Consider \( t = 1 \) as a base case. If \( C^1(v) \neq C^0(v) = \perp \), then it must be because \( v \) announced at time 1, meaning that \( 1 = T(v, v) \) as desired.

Now assume that for all \( v \) and all \( t' < t \) the claim holds, and consider time \( t \). If \( v \) does not announce at time \( t \) then the claim vacuously holds. If \( v \) announces at time \( t \) but does not change their announcement, then again the claim vacuously holds. If \( v \) announces at time \( t \) for the first time, then \( t = T(v, v) \) and the claim holds. The only remaining case is that \( v \) is changing their previous announcement that was made at time \( t' < t \) (and \( v \) did not announce between \( t' \) and \( t \)).

Let’s consider the state of affairs at time \( t' \), when \( v \) announced some opinion \( A \). This means that, at time \( t' \), a majority (tie-breaking for \( X(v) \)) of \( v \)’s neighbors were announcing \( A \). Yet, at time \( t \), a majority (tie-breaking for \( X(v) \)) of \( v \)’s neighbors were announcing \( B = 1 - A \). Therefore, some node adjacent to \( v \) must have switched its announcement to \( B \) at some \( t'' \in (t', t) \), and still announce \( B \) at \( t \) (and caused the change). Call this node \( x \). We now wish to invoke the inductive hypothesis for \( x \) at \( t'' \).

The inductive hypothesis claims there there is some \( u \) (maybe \( u = x \)) such that \( u \) made \( B \) as its first announcement, and then every node \( y \) along the critical chain from \( u \) to \( x \) switched from \( A \) to \( B \) at \( T(u, y) \) (caused by its predecessor), and that \( t'' = T(u, x) \). Let’s first consider the case that \( v \) is not on the path from \( u \) to \( x \) (and therefore \( x \) is on the path from \( u \) to \( v \)), since they are adjacent). Then as \( T(u, x) = t'' \in (t', t) \), and \( v \) does not announce in \((t', t)\), we see that \( T(u, v) = t \) (immediately by definition of critical times). Moreover, as \( P(u, v) \) is simply \( P(u, x) \) concatenated with \( v \), the inductive hypothesis already guarantees that \( u \) announced \( B \) at \( T(u, u) \), and that every node \( y \) on \( P(u, v) \) switched from \( A \) to \( B \) at \( T(u, y) \). So the last step is to show that in fact \( v \) must not be on the path from \( u \) to \( x \), and then the inductive step will be complete.

\(^3\)That is, flipping all \( X(v) \) simultaneously to \( 1 - X(v) \) would cause \( C^t(v) \) to flip (odd), and changing any subset of initial beliefs from 0 to 1 cannot change \( C^t(v) \) from 1 to 0 (monotone).
Finally, we show that we cannot have \( v \) on the path from \( u \) to \( x \), completing the inductive step. Assume for contradiction that \( v \) were on the path from \( u \) to \( x \). Then as \( t'' = T(u, x) \), we would necessarily have \( t' \geq T(u, v) \) (immediately from definition of critical times). However, by hypothesis, \( C^d(v) = A \), contradicting the inductive hypothesis that \( v \) caused \( u \) to change, which implies \( C^{T(u,v)}(v) = C^d(v) = B \) (as \( C^d(v) = C^{t''}(v) \) because \( v \) does not announce in \((t', t'')\)). So \( v \) cannot be on the path from \( u \) to \( x \).

Below, we make use of Proposition 1 to prove that the process terminates quickly (proof in Appendix B). Note that [11] already proves that the process on trees terminates with probability \( 1 - o(1) \) after \( O(n^2) \) steps, so Corollary 1 is a strict improvement when \( D(G) = o(n) \).

**Corollary 1.** Let \( T \) denote the last time that a node changes its announcement. Then with probability \( 1 - o(1) \), \( T \leq 8 \cdot \max\{2 \ln(n), D(G) + 1\} \).

Finally, we prove one last technical lemma which will be used in future sections regarding the probability that a single node announces \textsc{Correct} throughout the process (the proof appears in Appendix B). Beginning with \( v \)'s first announcement, because the graph is a tree, prior to \( v \)'s first announcement all of \( v \)'s neighbors' announcements are independent. Therefore, it initially seems like we should expect \( v \)'s initial announcement to be \textsc{Correct} except with probability exponentially small in \( \deg(v) \) - indeed, this would hold if the dynamics were synchronous. However, since the dynamics are \textit{asynchronous}, there's a good chance that \( v \) announces before any of its neighbors and simply announces \( X(v) \). That is, the probability that \( v \)'s initial announcement is \textsc{Incorrect} is at least \( \frac{1/2-\delta}{\deg(v)} \), so we cannot hope for such strong guarantees. This observation highlights one (of several) crucial differences between synchronous and asynchronous dynamics. Still, the lemma below shows roughly that the only bad event is \( v \) announcing before many of its neighbors. Below for a set \( S \), we'll use \( C^d_S(v) \) to denote the following modified dynamics: First, set \( C^d_S(v) = \text{Incorrect} \) for all \( v \in S \), and all \( t \). Then, run the asynchronous majority dynamics as normal. In other words, the modified dynamics hard-code an \textsc{Incorrect} announcement for all nodes in \( S \) and otherwise run asynchronous majority dynamics as usual (this extension will be necessary for a later argument).

**Definition 3.** We say that a node \( v \) is safe thru \( T \) if \( C^d(v) \in \{\bot, 1\} \) for all \( t \leq T \).

We further say that a node \( v \) is safe thru \( T \), even against \( S \) if \( C^d_S(v) \in \{\bot, 1\} \) for all \( t \leq T \).

**Proposition 2.** For all \( a \), there exist absolute constants \( b, c, d \) such that for any \( v \), \( S \) with \( |S| = a \), and \( T \leq n \cdot e^{b \deg(v)} \), \( v \) is safe thru \( T \), even against \( S \), with probability at least \( 1 - c - d \ln(\deg(v))/\deg(v) \).

4 Forming an Initial Majority

In this section, we prove that in any tree, a \textsc{Correct} majority forms after a near-linear number of steps (but may later fade). The main idea is to show that the announcements of most pairs of nodes are independent with probability \( 1 - o(1) \), and use Chebyshev's inequality to show that the number of \textsc{Correct} announcements therefore concentrates around its expectation. This latter step is the crux of the proof. To show it, we consider three cases depending on the length and degree sequence of the path between a pair of nodes. If the path is long, then, similar to prior work, there is simply not enough time for the pair to influence each other. If the path is short, but (some of) the intermediate nodes have high degrees, then these effectively block influence because the announcements of these high-degree nodes is effectively independent of what's happening on the path. Finally, if the path is short and the intermediate nodes have low degrees, then the pair
certainly may influence each other. However, a counting argument shows there can only be a vanishingly small fraction of such pairs. The main result of this section is the following:

**Theorem 3.** [Majority in trees] For any $T \leq \frac{\ln n}{32 \ln \ln n}$, after $T_n$ steps, with probability at least $1 - o(1)$, the announcements of at least $(\frac{1}{2} + \frac{\delta}{2} - e^{-T}) \cdot n$ nodes in a tree are **Correct**.

First, we analyze the expected number of **Correct** nodes using Theorem 2 (proof in Appendix C).

**Lemma 3.** The expected number of nodes $v$ with $C^T(v) = \text{Correct}$ is at least $(1/2 + \delta - e^{-T})n$.

From here, we now need to show that the number of **Correct** announcements concentrates around its expectation. To this end, we’ll show that most pairs of nodes can be written as functions of disjoint initial beliefs, and are therefore independent. A lemma of [11] shows that this suffices:

**Lemma 4.** [Restated from [11]] Let $p_{uv}$ denote the probability, over the randomness in order of announcements, that $C^t(u)$ and $C^t(v)$ can be written as functions of disjoint sets of initial beliefs. Then if $\sum_{u,v}(1 - p_{uv}) = o(n^2)$, the number of **Correct** nodes at time $t$ is within $\delta n / 2$ of its expectation with probability $1 - o(1)$.

So our remaining task is to show that $\sum_{u,v}(1 - p_{uv}^T) = o(n^2)$, and this is the point where we diverge from prior work. For a given pair $u, v$, there are three possible cases. Below, case one is most similar to prior work, and cases two/three are fairly distinct.

**Case One: Long Paths.** One possibility is that $d(u, v) \geq \max\{4T, \ln \ln(n)\}$ (the choice of $\ln \ln(n)$ is arbitrary, any super-constant function would suffice). Let’s look at the midpoint $x$, which has distance $d(u, v)/2$ to both $u$ and $v$. Note that if $T(x, u) > T_n$ and $T(x, v) > T_n$, then Lemma 1 implies that $C^{T_n}(u)$ and $C^{T_n}(v)$ can be written as functions of disjoint initial beliefs. A direct application of Lemma 2 with $y = 1/4$ yields that the probability that $T(x, u) < T_n$ is $o(1)$. A union bound (over two events) then lets us conclude by Lemma 1 that $p_{uv}^T = o(1)$.

**Case Two: Short Paths A.** Another possibility is that $d(u, v) < \max\{4T, \ln \ln(n)\}$. Here, there will be two subcases. First, maybe it’s the case that $d(u, v)$ is small and the product of degrees on the path from $u$ to $v$ is small. In this case, it very well could be that $p_{uv}$ is tiny, which is bad. However, we prove that there cannot be many such pairs (and so in total they contribute $o(n^2)$ to the sum). The following lemma shows in fact that even if we remove the restriction that $d(u, v) = O(T)$, there simply cannot be many pairs of nodes such that the product of degrees on $P(u, v)$ is small (proof in Appendix C).

**Lemma 5.** Let $K$ be the set of pairs of nodes $(u, v)$ such that $\prod_{w \in P(u, v)} \deg(w) \leq X$. When $u$ and $v$ are chosen uniformly at random, we have $\Pr((u, v) \in K) \leq X/n$.

**Case Three: Short Paths B.** The final possibility is that $d(u, v) < \max\{4T, \ln \ln(n)\}$, and also that $\prod_{w \in P(u, v)} \deg(w)$ is large. In this case, we will prove that with probability $1 - o(1)$ there is some block in $P(u, v)$ causing $u$’s and $v$’s announcements to be independent (proof in Appendix C).

**Definition 4.** We say that a node $x \in P(u, v)$ cuts $u$ from $v$ thru $T$ if some node $y$ in $P(u, x)$ is safe thru $T$ even against $S_y$, where $S_y$ are $y$’s (at most) two neighbors in $P(u, v)$.

**Lemma 6.** Let $x$ cut $u$ from $v$ thru $T$ and also cut $v$ from $u$ thru $T$. Then $C^T(u)$ and $C^T(v)$ can be written as functions of disjoint sets of initial beliefs.
Next, we wish to show that with good probability there is indeed a node on \( P(u, v) \) that cuts \( u \) from \( v \) and also \( v \) from \( u \) (proofs in Appendix C).

**Lemma 7.** For any \( u, v \in V \) with \( \prod_{w \in P(u, v)} \deg(w) = X \), \( d(u, v) \leq \frac{\ln(X)}{2 \ln \ln(X)} \), and \( T = e^{O(X^{1/(4d(u, v))})} \), with probability \( 1 - o(X^{1/3}) \) there exists an \( x \) that cuts \( u \) from \( v \) thru \( Tn \) and also \( v \) from \( u \) thru \( Tn \).

**Corollary 2.** For any \( u, v \in V \) with \( \prod_{w \in P(u, v)} \deg(w) = X \), \( d(u, v) \leq \frac{\ln(X)}{2 \ln \ln(X)} \), and \( T = e^{O(X^{1/(4d(u, v))})} \), \( 1 - p^T_{u,v} = o(X^{-1/3}) \).

Now, we’ll put together case one, Lemma 6 and Corollary 2 together to prove Theorem 3, which is mostly a matter of setting parameters straight (and appears in Appendix C).

To conclude, at this point we have proven that a majority takes hold after \( n \cdot \frac{\ln n}{\ln \ln n} \) steps for any tree. The remaining work is to prove that it does not disappear.

## 5 Stabilizing Quickly

In this section, we identify properties of a tree which cause it to stabilize quickly. Our main theorem will then follow by proving that both balanced \( M \)-ary trees and preferential attachment trees have this property. The main idea is to consider nodes that are “close” to leaves in the following formal sense:

**Definition 5.** We say that a node \( v \) is an \((X, Y)\)-leaf in \( G \) if there exists a rooting of \( G \) such that \( v \) has \( \leq X \) descendants, and the longest path from \( v \) to one of its descendants is at most \( Y \). Note that leaves are \((0, 0)\)-leaves. When we refer to a node’s parent, children, or descendants, it will be with respect to this rooting.

**Definition 6.** We say that a node \( v \) is:

- finalized at \( T \), if \( C^t(v) = C^T(v) \) for all \( t \geq T \).
- nearly-finalized at \( T \) with respect to \( u \) if there exists a \( t' \geq T \) such that \( v \) is finalized at \( t' \) and for all \( t \in (T, t') \) when \( v \) announces, it either updates \( C^t(v) = C^t(u) \), if \( C^t(u) \neq \bot \), or \( C^t(v) = C^{t-1}(v) \), if \( C^t(u) = \bot \).

Intuitively, a node is finalized if it is done changing its announcement. A node \( v \) is nearly-finalized with respect to \( u \) if \( v \) is not quite finalized, but changes in \( u \) are the only reason why \( v \) would change its announcement (and moreover, \( v \) will copy \( u \) every announcement until \( v \) finalizes).

The main result of this section is as follows:

**Theorem 4.** Let \( v \) be an \((X, Y)\)-leaf. Then with probability \( X e^{-T/Y} \), \( v \) is nearly-finalized at \( Tn \) with respect to its parent.

The main insight for the proof of Theorem 4 will be the following lemmas. Below, Lemma 8 asserts that once all of \( v \)’s children are nearly-finalized with respect to \( v \), any changes in \( v \)’s opinion are to copy its parent, and Lemma 9 builds off this to claim that we can relate the time until \( v \) nearly-finalizes to its critical times. Importantly, Lemma 9 does not require all critical paths to hit \( v \), but only those from its descendents.

**Lemma 8.** Let all of \( v \)’s children be nearly-finalized with respect to \( v \) at \( T \), and let \( u \) be \( v \)’s parent. Let also \( t > t' \geq T \) be two timesteps during which \( v \) announced. Then if \( C^t(v) \neq C^{t-1}(v) \), we must have \( C^t(v) = C^t(u) \).
Lemma 9. Let $T_v := \max\{T(x,v), x \text{ is a descendant of } v\}$. Then $v$ is nearly-finalized at $T_v$ with respect to its parent.

These above lemmas suffice to prove Theorem 4. We will also need the following implications of Theorem 4. Below, Lemma 10 will be helpful in proving Corollary 3. Corollary 3 lets us claim that while nearly-finalized nodes are not themselves finalized, their existence implies the existence of other finalized nodes. This will be helpful in wrapping up in the following section, since the process only terminates once nodes are finalized.

Lemma 10. For any $t > T_v$, if a child of $v$ changes their announcement at $t$, $v$ becomes finalized at $t$.

Corollary 3. For any $T$, with probability $1 - e^{-T}$, $v$ has $\lfloor (\deg(v) - 1)/2 \rfloor$ children who are finalized at $T_v + Tn$.

Moreover, if $v$ is an $(X,Y)$-leaf and finalized at $t \geq T_v$, then with probability $1 - Xe^{-T/Y}$, all of $v$’s descendants are finalized at $t + nT$.

6 Wrapping Up: Preferential Attachment and Balanced $M$-ary Trees

In this section, we show how to make use of Theorem 4 to conclude that a $1 - o(1)$ fraction of nodes are finalized by $n \ln n$. Proofs for the two cases follow different paths, but both get most of their mileage from the developments in Section 5.

6.1 Preferential Attachment Trees Stablize Quickly

Let’s first be clear what we mean by a preferential attachment tree.\(^4\)

Definition 7 (Preferential Attachment Tree). $n$ nodes arrive sequentially, attaching a single edge to a pre-existing node at random proportional to its degree. Specifically:

- Let $v_i$ denote the $i^{th}$ node to arrive.
- Let $\deg_t(v_i)$ denote the degree of node $v_i$ after a total of $t$ nodes have arrived.
- There is a special node $v_0$, which only $v_1$ connects to upon arrival, and no future nodes.
- When $v_{i+1}$ arrives, $v_{i+1}$ attaches a single edge to a previous node, choosing node $v_j$, $j \in [1,i]$, with probability $\frac{\deg_t(v_i)}{2i-1}$.

Our main argument for preferential attachment trees is that most nodes are in a “good” subtree, defined below. All subsequent proofs are in Appendix E. At a high level the plan is as follows: first, we prove that because most nodes are $(X,Y)$-leaves for small $X,Y$, these nodes quickly become nearly-finalized. Next, we prove that most such nodes are part of a small subtree whose parent is likely to be safe thru the entire process. Therefore, the parent of this subtree is finalized early, and once the subtree becomes nearly-finalized, it finalizes quickly as well.

\(^4\)Note that this is the standard definition of preferential attachment used for heuristic arguments. Most prior rigorous work uses a slightly modified definition that produces a forest instead of a tree in order to rigorously analyze (say) the degree distribution. As we are only interested in (fairly loose) bounds on the degrees, our results are rigorous in the original model.
Definition 8. Say that a subtree rooted at $v$ is good if:

- $v$ is a $(X,Y)$-leaf, for $X = \ln^{O(1)} n$ and $Y = O(\ln \ln n)$.
- $v$’s parent has degree at least $\ln^{O(1)} n$.

Proposition 3. Let the subtree rooted at $v$ be good, and let the diameter of the entire graph be $O(\ln n)$. Then with probability $1 - o(1)$, the entire subtree rooted at $v$ is finalized by $n^{\ln n / 32 \ln \ln n}$.

Proposition 4. For a tree built according to the preferential attachment model, the following simultaneously hold with probability $1 - o(1)$.

- $n - o(n)$ nodes are in good subtrees.
- The diameter of the entire graph is $O(\ln n)$.

Theorem 5. A tree built according to the preferential attachment model stabilizes in a correct majority with probability $1 - o(1)$.

6.2 Balanced $M$-ary Trees Stabilize Quickly

Let’s first be clear what we mean by a balanced $M$-ary tree.

Definition 9. We say a tree is a balanced $M$-ary tree if there is a root $v$ such that all non-leaf nodes have exactly $M$ children, and all root-leaf paths have the same length.

Our plan of attack is as follows (all proofs are in Appendix E). First, the case for large $M$ (say, $M > \ln n$) is actually fairly straight-forward as a result of Proposition 2. This is because every pair of nodes has a high-degree block on their path, meaning that the “Case Three” argument used in Section 4 actually applies all the way until the process terminates. The $M \leq \ln n$ case is more interesting, and requires the tools developed in Section 5.

Here, the plan is as follows. Corollary 3 roughly lets us claim that all nearly-finalized nodes must have a decent number of finalized children, and moreover that all these finalized children have finalized descendents. Iterating this counting inductively through children, we see that actually most descendents of nearly-finalized nodes of sufficient height must themselves be finalized.

Formally, the approach is to first get a bound on the height for which we can claim that nodes are indeed nearly-finalized with high probability (Corollary 4, immediately from Theorem 4). $\ln \ln n$ turns out to be a good choice.

Corollary 4. Let $v$ be distance $h$ from a leaf. Then $v$ is an $(2^M, h)$-leaf, and therefore $v$ is nearly-finalized with respect to its parent at $n^{\ln n / 32 \ln \ln n}$ with probability $2^{Mh} \cdot e^{-\frac{\ln n}{\ln n}} = e^{-\frac{\ln n}{\ln n} + h \ln(2^M)}$.

In particular, if $h = o(\sqrt{\ln n})$, then $v$ is nearly-finalized with respect to its parent at $n^{\ln n / 32 \ln \ln n}$ with probability $1 - o(1)$.

Proposition 5. Let $v$ have height $h = \ln \ln n$ in a balanced $M$-ary tree for $M \leq \ln n$. Then with probability $1 - o(1)$, at most $2^{Mh} \cdot (2/3)^h = o(M^h)$ descendents of $v$ are not finalized by $n^{\ln n / 32 \ln \ln n}$.

Theorem 6. Any $M$-ary tree stabilizes in a correct majority with probability $1 - o(1)$.
References


A Omitted Proofs From Section 2

Proof of Lemma 1. The proof proceeds by induction on $t$. The claim trivially holds for a base case of $t = 0$, as all $C^0(v)$ are deterministically $\perp$. Assume for inductive hypothesis that the claim holds for $t - 1$, and now consider $t$. First, consider all $v$ that are not selected to announce. By inductive hypothesis, $C^{t-1}(v)$ can be expressed as a function of $\{X(u), t - 1 \geq T(u,v)\}$. As $C^t(v) = C^{t-1}(v)$, and $t \geq t - 1$, the inductive step holds for such $v$.

Next, consider the $v$ selected to announce. Clearly, we can write $C^t(v)$ as a function of $\{C^{t-1}(x), (x,v) \in E\}$. By inductive hypothesis, each $C^{t-1}(x)$ can be written as a function of $\{X(u), t - 1 \geq T(u,x)\}$. Finally, we observe immediately from the definition of critical times that if $T(u,x) \leq t - 1$, and $x$ is a neighbor of $v$, and $v$ announces at $t$, then $T(u,v) \leq t$. Therefore, for all $x$ adjacent to $v$, and all $u$ such that $t - 1 \geq T(u,x)$, we also have $t \geq T(u,v)$, meaning that we can indeed write $C^t(v)$ as a function of $\{X(u), t \geq T(u,v)\}$.

\footnote{Could remove this since it’s in prior work, although I don’t think we explained it this clearly. Only reason to keep it is because it gives good intuition for the process.}
Proof of Lemma 2. Let’s first analyze the random variable $T(u,v)$. Note that the random variable $T(u,u)$ is just a geometric random variable of rate $1/n$ (because we are waiting for $u$ to be selected to announce). Moreover, once we hit $T(u,u)$, $T(u,x)$ is just an independent geometric random variable of rate $1/n$, where $x$ is the next node on $P(u,v)$. Following the same reasoning, we see that the random variable $T(u,v)$ is the sum of $d(u,v)+1$ i.i.d. geometric random variables of rate $1/n$ (also called a negative binomial distribution with parameters $(d(u,v)+1)/n$).

Moreover, we can couple the event that $T(u,v) > Tn$ (resp. $< Tn$) with the event that the sum of $Tn$ independent Bernoulli’s with rate $1/n$ exceeds $d(u,v)+1$ (resp., does not exceed). By the Chernoff bound, this is upper bounded by $e^{-(d(u,v)+1-1)2T/3}$.

Plugging in for $T = 8 \cdot \max\{\ln(1/y), d(u,v)+1\}$, we get the first statement. Plugging in for $T = y \cdot (d(u,v)+1)$, we get the second.

B Omitted Proofs From Section 3

Proof of Corollary 1. We’ll only appeal to Proposition 1 and the randomness over order of announcements, and observe that our bound holds even for worst-case initial private beliefs. Lemma 2 guarantees that for a single $(u, v)$ path, $T(u,v) \leq 8 \cdot \max\{2 \ln(n), D(G)+1\}$ with probability $1-1/n^4$.

Taking a union bound over all pairs $(u, v)$, we see that with probability $1-o(1)$, each $T(u,v) \leq \max\{2 \ln n, D(G) + 1\}$. Combining with Proposition 1 immediately proves the corollary.

Proof of Proposition 2. We’ll consider several possible points of failure, and show that each happens with probability at most $O(\ln(\deg(v)) / \deg(v))$. The proposition will then follow from a union bound.

Failure One. First, we’ll consider it to be a failure if fewer than $k = \omega(1)$ (exact value to be set later) of $v$’s neighbors announce before $v$’s initial announcement. Observe that the order of initial announcements among $v$ and its neighbors is uniformly at random, so this probability is exactly $k/(\deg(v)+1) < k/\deg(v)$. Call the event $Z$ that at least $k$ of $v$’s neighbors have announced by $v$’s initial announcement.

Failure Two. Next, we’ll consider it a failure if any of $v$’s neighbors haven’t announced by $v$’s $\ell$th announcement. Observe that for any specific neighbor $w$ of $v$, the probability that $w$ announces before $v$’s next announcement is exactly $1/2$, independent of whether $w$ announced before any of $v$’s previous announcements. Therefore, the probability that a $v$ announces $\ell$ times before $w$ announces once is exactly $2^{-\ell}$. Call the event $W$ that all of $v$’s neighbors have announced by $v$’s $\ell$th announcement.

Failure Three. Next, we’ll consider it a failure if $C^*_S(v)$ would be INCORRECT during any of $v$’s first $\ell$ announcements, conditioned on $Z$. To this end, consider the (further) modified dynamics where each neighbor $w$ of $v$ ignores $v$’s announcement when updating their own (i.e. pretends it is $\perp$). Observe in these modified dynamics that essentially the graph has been disconnected by removing $v$. Therefore, at any time $t$, the announcements of $v$’s neighbors at time $t$ are independent (but of course, $w$’s announcement at $t$ may still be correlated with $w$’s own past announcements).

In particular, for a particular $i$th announcement of $v$, observe that each of $v$’s neighbors either announce $\perp$, or announce 1 with probability at least $1/2 + \delta$. Therefore, the expected fraction of $v$’s announced neighbors who announce 1 is at least $1/2 + \delta$, and these are all independent. By the Chernoff bound, and conditioned on $Z$, the probability that at least a $1/2 + \delta/2$ fraction of such neighbors announce 1 is at least $1 - e^{-\Omega(k\delta^2)}$. As $k = \omega(1)$ and $|S| = O(1)$, this indeed suffices in order for the modified dynamics to result in a CORRECT majority among $v$’s neighbors, even
against $S$ (to see why this suffices, observe that each of the “bad nodes” in $S$ can affect at most one of $v$’s neighbors. So even if we hardcode all of these neighbors to INCORRECT, the rest are independent). So by taking a union bound over all $\ell$ announcements, we get that the probability that $C_2^T(v)$ would be INCORRECT during any of the first $\ell$ announcements in the further modified dynamics is at most $\ell e^{-\Omega(k\delta^2)}$.

Now, we need to reason about the original “against $S$” dynamics. We wish to claim that for all orders of announcements, and all initial beliefs, if the further modified dynamics ever result in an INCORRECT majority during one of $v$’s first $\ell$ announcements, then so do the original against $S$ dynamics. To see this, assume for contradiction that $v$’s $i$th announcement was the first INCORRECT announcement in the original against $S$ dynamics, while all of $v$’s first $i$ “announcements” were CORRECT in the further modified dynamics. Then it must be the case that one of $v$’s neighbors announce INCORRECT in the original dynamics, but CORRECT in the modified dynamics, despite $v$ announcing CORRECT during each of its first $i-1$ announcements. This is a contradiction, as the inclusion of $v$’s CORRECT announcements cannot cause $v$’s announcement to switch from CORRECT to INCORRECT. Therefore, we’ve now shown that Failure Three occurs with probability at most $\ell e^{-\Omega(k\delta^2)}$.

**Failure Four.** Next, we’ll consider it a failure if any of $v$’s remaining announcements are INCORRECT conditioned on $W$. Conditioned on this, and by exactly the same reasoning in the previous case, we see that if $v$ makes a total of $L$ announcements during the first $T$ steps, then the probability that any of the last $L-\ell$ of them are INCORRECT is at most $(L - \ell) \cdot e^{-\Omega(deg(v)\delta^2)}$.

**Failure Five.** Finally, we’ll consider it a failure if $v$ announces more than $L$ times during the first $T$ steps. Note that the number of announcements is simply the sum of $T$ i.i.d. Bernoulli random variables that are 1 with probability $1/n$ (so a binomial distribution with parameters $T$ and $1/n$). So the probability that $v$ announces more than $L$ times is at most $e^{-\Omega(Ln/T)}$, for $L \geq 2T/n$, by the multiplicative Chernoff bound. Call the event that Failure Five doesn’t occur $X$.

**Wrapping Up.** First, Failure One occurs with probability at most $k/\deg(v)$. Conditioned on Failure One not occurring (i.e., $Z$ occurring), Failure Three occurs with probability at most $\ell e^{-\Omega(k\delta^2)}$. Failure Five occurs with probability at most $e^{-\Omega(Ln/T)}$, and Failure Two occurs with probability at most $2^{-\ell}$. Conditioned on $X$ and $W$, Failure Four occurs with probability at most $(L - \ell) \cdot e^{-\Omega(deg(v)\delta^2)}$. Therefore, the probability that any of the five failures occur is at most:

$$k/\deg(v) + \ell e^{-\Omega(k\delta^2)} + e^{-\Omega(Ln/T)} + 2^{-\ell} + (L - \ell) e^{-\Omega(deg(v)\delta^2)}.$$  

Setting $k = \Theta(\ln(deg(v))/\delta^2)$, $\ell = \Theta(\ln(deg(v)))$, $L = \Theta(\ln(deg(v))\cdot T/n)$, we get that the total probability of failure is $O(\ln(deg(v))/\deg(v) + \ln(deg(v))\cdot T/n) e^{-\Omega(deg(v)\delta^2)}$. The proposition then follows by observing that the first term dominates the second whenever $T/n = e^{O(deg(v))}$.

As a final sanity check, observe that if none of the five failures occur, then indeed $v$ is safe thru $T$ even against $S$.

\[\square\]

### C Omitted Proofs From Section 4

**Proof of Lemma 3.** Let ANNONCED be the event that a given node $v$ announces by time $Tn$. For such a node $v$, we apply Theorem 2, together with the observation that $C^T(v)$ is an odd monotone boolean function of the random variables $\{X(u), u \in V\}$, to conclude that $v$ is CORRECT with probability at least $1/2 + \delta$ (not independently of other nodes). Then, we observe that $v$ announces by $Tn$ with probability at least $1 - (1 - 1/n)^n Tn \geq 1 - e^{-T}$ (because each of $Tn$ announcements are made by a uniformly random one of the $n$ nodes). Therefore, the probability a given node is
Correct is $\Pr[\text{Correct}] \Pr[\text{Announced}] \geq (1/2 + \delta)(1 - e^{-T}) \geq 1/2 + \delta - e^{-T}$.
The lemma follows by linearity of expectation. \hfill \square

**Proof of Lemma 5.** For any $u \in V$ and $x > 0$, define

$$K_\ell(u, x) = \left\{ v \in V(G) : \prod_{w \in P(u, v)} \deg(w) < x \text{ and } 1 \leq d(u, v) \leq \ell \right\}.$$  

Fix a vertex $u$ and consider rooting the tree at $u$. We claim that for all $\ell \in [n]$ and $x > 0$, we have $|K_\ell(u, x)| \leq x$, with equality only if all of $u$’s neighbors are in $K_\ell(u, x)$. Therefore, once we pick a random $u$, the probability that we pick $v$ such that $(u, v) \in K$ is at most $x/n$. We prove this by induction on the path length $\ell$.

**Base case.** Suppose $\ell = 1$. The vertices at distance exactly 1 from $v$ are its neighbors. Fix any $x > 0$. We have two possibilities:

- $\deg(u) > x$: Note that for all $v \in V(G), \prod_{w \in P(u, v)} \deg(w) \geq \deg(u) \geq x$, implying $|K_1(u, x)| = 0 \leq x$.

- $\deg(u) \leq x$: In this case, $K_1(u, x) \subseteq N(u)$ implies $|K_1(u, x)| \leq \deg(u) \leq x$. Here, notice that equality is possible only if all of $u$’s neighbors are in $K_1(u, x)$.

**Inductive hypothesis.** Assume for all nodes $u \in V(G)$ and $x > 0$, we have $|K_{\ell-1}(u, x)| < x$.

**Inductive step.** Fix any node $u$. For each $u' \in N(u)$, let us consider the remaining subset of $K_\ell(u, x)$ rooted at $u'$. We claim that the following inequality holds for all $x > 0$ and $\ell \geq 2$:

$$|K_\ell(u, x)| \leq \sum_{w \in N(u)} \left|K_{\ell-1} \left(w, \frac{x}{\deg(u)} \right)\right|.$$  

We observe that for any node $v$ satisfying $d(u, v) \in \{1, 2, \ldots, \ell\}$, there must be a neighbor $u' \in N(u)$ such that $d(u', v) \in \{0, 1, \ldots, \ell-1\}$. Furthermore, for any path $P(u, v)$ satisfying $\prod_{w \in P(u, v)} \deg(w) < x$, removing $u$ from $P(u, v)$ gives a new path $P(u', v)$, where $u'$ is a neighbor of $u$, and $\prod_{w \in P(u', v)} \deg(w) < \frac{x}{\deg(u)}$.

The right-hand-side is both over-counting and under-counting the size of $K_\ell(u, x)$. It is over-counting in the following two senses.

- It may be the case that $K_{\ell-1} \left(u', \frac{x}{\deg(u)} \right) \cap K_{\ell-1} \left(u'', \frac{x}{\deg(u)} \right) \neq \emptyset$ for two neighbors $u', u'' \in N(u)$, so some nodes can be counted several times.

- It also may be the case that $u \in K_{\ell-1} \left(u', \frac{x}{\deg(u)} \right)$ for any $u' \in N(u)$. In this case, therefore, the node $u$ — which satisfies $d(u, u) = 0$ and is thus not in $K_{\ell}(u, x)$ — is counted.

The under-counting is due to the fact that the definition of $K_\ell$ excludes the root of the tree. In other words, for each $u' \in N(u)$, $u' \notin K_{\ell-1} \left(u', \frac{u}{\deg(u)} \right)$. Therefore, none of the neighbors of $u$ are counted in the right-hand-side, so the right-hand-side is off by exactly $\deg(u)$. However, this cancels the second case of over-counting listed above: for all $u'$, either $u \in K_{\ell-1} \left(u', \frac{x}{\deg(u)} \right)$, in
which case $u$ is overcounted. Or $u \notin K_{\ell-1}(u', \frac{x}{\deg(u)})$, in which case by the inductive hypothesis the bound is strict, and we can subtract one. So in either case, we get to cancel this undercounting.

The proof of the claim follows from applying the inductive hypothesis to each term in the sum on the right-hand-side.

$$|K_{\ell}(u, x)| \leq \sum_{w \in N(u)} \left| K_{\ell-1} \left( w, \frac{x}{\deg(u)} \right) \right|$$

$$< \sum_{\substack{w \in N(u) \setminus \{y\}}} \frac{x}{\deg(u)} = x.$$ 

\[\square\]

**Proof of Lemma 6.** First, define an auxiliary process that replaces $C^t(x)$ with \textsc{Correct} for all $t \geq T(x, x)$ (and keeps it as $\bot$ before $T(x, x)$), and otherwise runs asynchronous majority dynamics. Under this modified dynamics, it’s clear that all announcements of $u$ and $v$ are independent.

Next, as $x$ cuts $u$ from $v$ thru $T$, there is some node $y$ on $P(u, x)$ that is safe thru $T$ even against $S_y$. If $y$ happens to be $x$ itself, then $C^t(u)$ (in the true asynchronous majority dynamics) is accurately computed by the above modified process.

If $y$ happens to not be $x$, this means that $C^t(y)$ is known to be \textsc{Correct} (or $\bot$, before its first announcement) without inspecting $C^{t-1}(z)$ for $z \in S_y$. Therefore, $C^t(y)$ can be written as a function of initial beliefs only of nodes it can reach without going through $S_y$, for all $t \leq T$.

More formally, let $y_u$ and $y_v$ with $x \in P(y_u, y_v)$ be nodes safe thru $T$ even against $S_{y_u}$ and $S_{y_v}$, respectively (such nodes are guaranteed by the lemma predicate). Let $A_u$ be the set of nodes that can be reached from $u$ without going through $y_u$, i.e., $\{w \mid w \neq P(u, w)\}$. Similarly, let $A_{y_u}$ denote the set of nodes that can be reached from $y_u$ without going through $S_{y_u}$. Then, for all $t \leq T$, $C^t(y)$ can be written as a function of initial beliefs $\{X(w) \mid w \in A_{y_u}\}$, and therefore $C^t(u)$, and in particular $C^T(u)$, can be written as a function of initial beliefs $\{X(w) \mid w \in A_{y_u} \cup A_u\}$. By a similar argument, $C^T(v)$, can be written as a function of initial beliefs $\{X(w) \mid w \in A_{y_v} \cup A_v\}$, where $A_{y_v}$ are nodes that can be reached from $y_v$ without going through $S_{y_v}$. As $y_u \neq y_v$ and the path $P(y_u, y_v)$ intersects $S_{y_u} \cup S_{y_v}$, the sets $A_{y_u} \cup A_u$ and $A_{y_v} \cup A_v$ are disjoint as required.

\[\square\]

**Proof of Lemma 7.** First, observe that whether or not a node $w$ is safe thru $T_n$, even against $S_w$, are independent events. This is because whether or not $w$ is safe thru $T_n$, even against $S_w$, can be determined as a function of $A_w$, and the sets $A_w, w \in P(u, v)$ are disjoint.

Now, observe that there certainly exists a node $x$ such that $\prod_{w \in P(u, x)} \deg(w) \geq X$, and $\prod_{w \in P(x, v)} \deg(w) \geq \sqrt{X}$ (this is simply because the product of both terms is at least $X$). Let’s further restrict attention to $P^*(u, x) = \{w \mid P(u, x), \deg(w) \geq X^{1/(4|P(u, x)|)}\}$. It’s further clear that we still have $\prod_{w \in P^*(u, x)} \deg(w) \geq X^{1/4}$. Now, let’s analyze the probability that no node $y$ on $P^*(u, x)$ is safe thru $T_n$, even against $S_y$. By independence, and using Proposition 2, this is just $\prod_{w \in P^*(u, x)} O(\ln(\deg(w)) \times \deg(w))$ (note that the application of Proposition 2 is valid by our assumptions on $T$). Below, let $\ell = 4|P(u, x)|$.

We already know that $\prod_{w \in P^*(u, x)} \deg(w) \geq X^{1/4}$, so we wish to upper bound $\prod_{w \in P^*(u, x)} O(\ln(\deg(w)))$ conditioned on this. Note that $\prod_{w \in P^*(u, x)} \deg(w) = X^{1/4} \Rightarrow \sum_{w \in P^*(u, x)} \ln(\deg(w)) = \ln(X)/4$. So we are interested in upper bounding $\sum_{w \in P^*(u, x)} \ln(c \cdot \ln(\deg(w)))$ (for the $c$ implied by the big-Oh), given that $\sum_{w \in P^*(u, x)} \ln(\deg(w)) = \ln(X)/4$. As $\ln(\cdot)$ is concave, this is maximized by setting all $\ln(\deg(w)) = \ln(X)/\ell$. 

16
So we get that the total probability that there exists no node on \( P^*(u, x) \) (and hence, \( P(u, x) \)) that is safe thru \( T_n \), even against \( S_y \) is at most \( \left( \frac{\ell}{4} \ln(X) \right)^{\ell/4} / X \). So now we just wish to see what conditions on \( \ell, X \) ensure that this is \( o(1) \). Note that \( \ell/4 \leq d(u, v) \), so this term is upper bounded by:

\[
\left( c \ln(X) \right)^{d(u, v)} \cdot \ln(X) / X.
\]

Further, as \( d(u, v) \leq \frac{\ln(X)}{2 \ln \ln(X)} \), this term is upper bounded by:

\[
e^{\ln(c) \frac{\ln(X)}{2 \ln \ln(X)}} \cdot \ln(X) / X
\]

\[
= O(X^{2/3}) / X = O(X^{-1/3}).
\]

**Proof of Theorem 3.** Lemma 3 states that the expected number of \text{Correct} nodes is at least \( (1/2 + \delta - e^{-T})n \). Combined with Lemma 4, if we can show that \( \sum_{u, v} 1 - p^t_{uv} = o(n^2) \), then with probability \( 1 - o(1) \), the realized number of \text{Correct} nodes at time \( T_n \) is \( (1/2 + \delta/2 - e^{-T})n \), proving the theorem.

Let \( A \) denote the set of all \((u, v)\) such that \( d(u, v) \geq \frac{\ln(n)}{8 \ln \ln(n)} \). Let further \( B \) denote the set of all \((u, v) \notin A\) such that \( \prod_{w \in P(u, v)} \deg(w) \leq X \). Finally, let \( C \) denote all remaining pairs.

Observe that as \( T \leq \frac{\ln(n)}{32 \ln \ln(n)} \), pairs in \( A \) satisfy the hypotheses of case one. Pairs in \( B \) satisfy the hypotheses of Lemma 5 with \( X = \sqrt{n} \), and therefore there are at most \( n^{3/2} \) such pairs. Finally, let’s confirm that pairs in \( C \) satisfy the hypotheses of Lemma 7 for \( X = \sqrt{n} \).

It’s easy to see that \( d(u, v) \leq \frac{\ln(n)}{8 \ln \ln(n)} \leq \frac{\ln(\sqrt{n})}{2 \ln \ln(\sqrt{n})} \), so the constraints on \( d(u, v) \) are satisfied. Moreover, observe that as \( d(u, v) \leq \frac{\ln(n)}{8 \ln \ln(n)} \), our bound becomes:

\[
T = e^{O(X^{2 \ln \ln(n) / \ln(n)})}
\]

\[
= e^{O(e^{\ln(n)})}
\]

\[
= e^{O(\ln(n))}
\]

Which is certainly satisfied by any \( T \leq \frac{\ln(n)}{8 \ln \ln(n)} \). So now we can write:

\[
\sum_{u, v} 1 - p^t_{uv} = \sum_{(u, v) \in A} 1 - p^t_{uv} + \sum_{(u, v) \in B} 1 - p^t_{uv} + \sum_{(u, v) \in C} 1 - p^T_{uv}
\]

\[
\leq |A| \cdot o(1) + |B| + |C| \cdot o(1)
\]

\[
\leq |A| \cdot o(1) + n^{3/2} + |C| \cdot o(1) = o(n^2).
\]

\[\square\]
D Omitted Proofs From Section 5

Proof of Lemma 8. Assume that \( v \) changes her opinion at \( t > T \) from \( A \) to \( B \not= A \), and her previous announcement (of \( A \)) was at time \( t'' \). Note that \( t'' \ge t' > T \) by assumption. By Proposition 1, there must have been some node \( x \) that had \( C^{t''}(x) \not= B \), but \( C^t(x) = B \). However, during this entire window, \( v \)'s announcement is \( A \), and during this entire window all of \( v \)'s children are nearly-finalized with respect to \( v \). Therefore, any child of \( v \) that changes her announcement during this window necessarily changes it to \( A \). Therefore, if \( v \) changes opinion to \( B \) at \( t \), it must be because \( x = u \), and therefore we indeed have \( C^t(v) = C^t(u) \) as desired.

Proof of Lemma 9. The proof proceeds by induction. As a base case, consider when \( v \) is a leaf. Then every time that \( v \) announces it will copy its parent, as long as the announcement of its parent is not \( \perp \). If its announcement is \( \perp \), then \( v \) will simply announce its initial belief. Therefore, \( v \) is nearly-finalized with respect to its parent as soon as \( v \) announces for the first time, which occurs at \( T(v,v) \). So the base case holds.

Now assume that the lemma holds for all children of \( v \). Observe first that \( T_v > T_x \) for all \( x \) that are children of \( v \) (this is simply because any descendant of \( x \) is also a descendant of \( v \), and for such nodes \( y \), \( T(y,x) < T(y,v) \)). Therefore, the inductive hypothesis claims that all children of \( v \) are nearly-finalized with respect to \( v \) at \( T_v \). Moreover, observe that \( v \) necessarily announces at \( T_v \) (because \( T_v = T(y,v) \) for some \( y \), and \( v \) announces at each \( T(y,v) \)).

Lemma 8 then claims that for any \( t > T_v \) during which \( v \) changes her announcement, \( v \) copies her parent, \( u \). We claim that this implies that \( v \) is nearly-finalized with respect to \( u \). That is, until \( v \) finalizes, \( v \) always copies her parent.

Indeed, assume for contradiction that \( v \) is not nearly-finalized with respect to \( u \) at \( T_v \). Then there must exist some time \( t > T_v \) where \( v \) changes their announcement (because \( v \) must not be finalized at \( T_v \)). Let \( t \) be the latest such time. Moreover, there must exist some time \( t' \in (T_v,t) \) where \( v \) announced such that either \( C^t(v) \not= C^t(u) \not= \perp \), or \( C^t(v) \not= C^{t-1}(v) \) and \( C^t(u) = \perp \).

A direct application of Lemma 8 in fact immediately rules the second case out. So let \( A = C^t(v) \), and \( B = C^t(u) \not= A \). Note also that both \( A \) and \( B \) are not \( \perp \). But now recall that \( v \) announced during time \( t' \) and updated to a majority of her neighbors (tie-breaking for \( X(v) \)). All of her neighbors except for \( u \) are nearly finalized with respect to \( v \) at \( t' \). So there cannot be some time \( t'' > t' \) where a majority of \( v \)'s neighbors (tie-breaking for \( X(v) \)) are \( B \)! This is because the first such \( t'' \) would have required some neighbor to flip from \( A \) to \( B \). It cannot be a child of \( v \) because they can only flip to \( A \). It cannot be \( u \) because \( u \) is already announcing \( B \). So in fact, \( v \) must be finalized at \( t' \) (contradicting that \( v \) changed their announcement at \( t > t' \)).

Proof of Theorem 4. With Lemma 9, we just need to get a bound on \( T_v \). We do this by partitioning the \( Tn \) steps into \( Y \) epochs of length \( Tn/Y \). We’ll assign descendants that are at distance \( i \) from \( v \) to epoch \( Y - i \).

We’ll now observe the following: if it is the case that all descendants of \( v \) make at least one announcement during their assigned epoch, then \( T_v \le Tn \). This is because for any descendant \( u \) of \( v \), we have an ordered sequence of announcements along the path from \( u \) to \( v \) terminating by \( T_v \). So we just want to bound the probability that any descendant doesn’t announce during its assigned epoch.

For a single node, the probability that it doesn’t announce during a window of length \( Tn/Y \) is exactly \( (1 - 1/n)^{Tn/Y} \le e^{-T/Y} \). Taking a union bound over each of the \( X \) nodes yields that the probability that any node fails is at most \( Xe^{-T/Y} \).
Proof of Lemma 10. Let $t' \in [T_v, t)$ be the most recent announcement of $v$ (we know that such a $t'$ exists, because $v$ announces at $T_v$), and say that $v$ announced $A$. Then $v$ copies the majority of its neighbors at $t'$, so a majority of its neighbors announced $A$. In between $t'$ and $t$, $v$ does not announce again (by definition). Other children of $v$ might announce, but because they nearly-finalized with respect to $v$, they will only announce $A$. $v$'s parent may announce, and could announce either $A$ or $B$. But we know that one of $v$'s children changes their announcement at $t$, and because they are all nearly-finalized with respect to $v$, this announcement will be $A$, pushing the majority of $v$'s neighbors further towards $A$. Maybe $v$'s parent will indeed change their announcement from $A$ to $B$, but all this does is cancel out the child’s switch to $A$, maintaining an $A$ majority. No other children of $v$ can possibly switch to $B$ without $v$ switching first, so $v$ will stay $A$ forever. □

Proof of Corollary 3. At any step $t > T_v$, let’s look at the current states of $v$’s children. We know that every child is nearly-finalized with respect to $v$, so they are either finalized, or copying $v$ with every announcement. If $\lceil (\deg(v) - 1)/2 \rceil$ of $v$’s children are finalized, then the corollary statement is satisfied.

If not, then $v$ has at least $\lceil (\deg(v) - 1)/2 \rceil + 1 > \deg(v)/2$ non-finalized children. If all of these children agree with $v$’s current announcement, then in fact $v$ is finalized. This is because each of these $\lceil (\deg(v) - 1)/2 \rceil$ children will not change their announcement unless prompted by $v$, and $v$ will not change her announcement unless one of them change (because this constitutes a strict majority of $v$’s neighbors). If $v$ is finalized, then each of these children that agree with $v$ are certainly finalized, and the corollary is again satisfied.

So if $v$ is not finalized, and $v$ has fewer than $\lceil (\deg(v) - 1)/2 \rceil$ finalized children, $v$ must have a non-finalized child that disagrees with $v$. Whoever this child is, it is nearly-finalized with respect to $v$. So if it is selected to announce, it will change its opinion to match $v$, causing $v$ to finalize by Lemma 10, and have the desired number of finalized children by the reasoning above. So if $v$ does not yet have the desired number of finalized children by $t$, there is at least one child of $v$ such that if they are selected to announce, $v$ will then have the desired number of finalized children. The probability that this child is not selected to announce for all $t \in (T_v, T]$ is at most $e^{-T}$.

For the “Moreover…” part of the statement, we’ll use a similar proof to that of Theorem 4. We’ll again break down the $Tn$ steps into epochs of length $nT/Y$. This time, we’ll put descendants at distance $i$ from $v$ into epoch $i$. By the reasoning further above (which claims that once $v$ is finalized and a child $x$ of $v$ announces, $x$ is for sure finalized), we again conclude that as long as every descendant makes an announcement during its assigned epoch, all descendants of $v$ are finalized. The probability that any descendant fails to announce during its assigned epoch is at most $e^{-T/Y}$, so a union bound gives that the total failure probability is $Xe^{-T/Y}$. □

E Omitted Proofs From Section 6

Proof of Proposition 3. First, observe that $v$’s parent is safe thru $O(n \ln n)$ with probability $1 - o(1)$. By Corollary 1, the entire process actually terminates by $O(n \ln n)$. Therefore, as soon as $v$’s parent announces, they are finalized (except with probability $o(1)$).

So, the probability that $v$’s parent hasn’t announced by $n^{\ln n/64 \ln \ln n}$ is $o(1)$. By Theorem 4, $v$ is nearly-finalized with respect to its parent by $n^{\ln n/64 \ln \ln n}$ with probability $1 - o(1)$ (this can be observed by plugging in $T = \ln n/64 \ln \ln n$, $X = \ln^{O(1)}(n)$, $Y = O(\ln \ln n)$). When both events happen, this in fact means that $v$ is finalized. Finally, Corollary 3 implies that with probability $1 - o(1)$, the entire subtree rooted at $v$ is finalized by an additional $\ln n/64 \ln \ln n$ steps, proving the proposition. □
Lemma 11. For a tree built according to the preferential attachment model, the following simultaneously hold with probability $1 - o(1)$.

- $n - o(n)$ nodes are in good subtrees.
- The diameter of the entire graph is $O(\ln n)$.

Proof. We want to claim that most subtrees rooted at nodes $v$ that arrive after $n/\log n$ and is a child of one of the first $n/\log n$ are good. We will call the first $n/\log n$ nodes to arrive, the early nodes. For any node $v$, let $s(v)$ be the size of the subtree $H_v$ rooted at $v$ at the end of the process. Consider the process after all the early nodes arrive. Every time an early node $v$ succeeds and gets a child $u$, we call $u$ a critical node and the subtree $H_u$ a critical subtree, and $v$ is the source of $H_u$.

To make the proofs simpler we will consider the preferential attachment process as a bunch of poisson processes, as introduced by Pittel [22]. In this model, there is a node $v_1$ arrives with an edge to the special node $v_0$. We consider a continuous process where the remaining nodes arrive in the following way. Let $G(k)$ be the graph (excluding $v_0$) after $k$ nodes have arrived. For each node $v \in G(k)$, let $\deg_k(v)$ denote its current degree. For each node $u \in G(k)$ start a poisson process independently with rate $\deg_k(u)$. So initially, $v_1$ starts a poisson process with rate 1 (its current degree). When one of them succeeds, we say that a new node $v_{k+1}$ arrives and attaches to the node $u$ that succeeded. We then restart the processes for all nodes in $G_{k+1}$. We run the process till $n$ nodes have arrived. Let $Y_k(u)$ be waiting time for $u$ to succeed. $Y_k(u)$ is an exponential random variable with rate $\deg_k(u)$. Observe that $Y_k = \min_{u \in G(k)} Y_k(u)$ is also exponential with rate $\sum_{u \in G(k)} \deg_k(u) = 2k - 1$, the waiting time for the node $v_{k+1}$ to arrive. Let $\tau_i = \sum_{k=1}^{i-1} Y_k$ denote the time at which the $i^{th}$ node arrives. Recall that the expectation of an exponential random variable with rate $= \lambda$ is $1/\lambda$. Therefore we get, $E(\tau_i) = \sum_{k=1}^{i-1} \frac{1}{2k-1} = \Theta(\log k)$.

Let $v$ be the $i^{th}$ node to arrive. Observe that the growth the subtree doesn’t depend on where $v$ attaches itself to. This is simply because of the recursive nature of the preferential attachment model. That is, conditioned on a node $u$ attaching to $v$’s subtree, $u$ attaches to a pre-existing node in the subtree at random proportional to its degree. So we will now bound the size and diameter of the subtree without worrying about the parent of $v$.

After $v$ arrives, we run the process till we get $n - i$ more nodes. Let $R_v = \sum_{k=1}^{n-i} Y_k$ denote the remaining amount of time needed to finish. The expected waiting time for the process to finish is $E(R_v) = \sum_{k=1}^{n-i} \frac{1}{2k-1} = \Theta(\log(n/i)) \leq \log(n/i)$. To bound the concentration probability we observe that $R_v$ is a sum of independent exponential random variables with rate at least $2i$ and use the following theorem by Janson [18].

Theorem 7 ([18]). Let $X = \sum_{i=1}^{n} X_i$ with $X_i \sim \text{Exp}(a_i)$ independent, $\mu = E(X)$ and $a = \min_i a_i$.

- For any $\lambda \geq 1$, $\Pr(X \geq \lambda \mu) \leq \lambda^{-1} e^{-a \mu (\lambda - 1 - \ln \lambda)}$
- For any $\lambda \leq 1$, $\Pr(X \leq \lambda \mu) \leq e^{-a \mu (\lambda - 1 - \ln \lambda)}$

For $\lambda = 2$, we have $\Pr(R_v \geq 2 \log(n/i)) \leq e^{-\Omega(2i \log(n/i))}$. For any $i \geq n/\log n$, we get that the waiting time for the remaining $n - i$ nodes to arrive is more than $4 \log \log n$ with probability at most $e^{-\Omega(n \log \log n)}$.

Similarly we will get a lower bound for $D_s$, the time taken for $v$ to get $s$ many descendants. Let $H_v(l)$ denote the subtree rooted at $v$ after $v$ gets $l - 1$ descendants, and let $d_l(u)$ denote the current degree of a node $u \in H_v(l)$. Let $W_l(u)$ denote the waiting time for $u \in H_v(l)$ to succeed. $W_l(u)$ is an exponential random variable with rate $d_l(u)$. So if $v$ has $l - 1$ descendants, then the waiting time for a new node attach to the subtree is $W_l = \min_{u \in H_v} W_l(u)$. The random variable $W_l$
is exponential with rate $\sum_{u \in H_v} d_i(u) = 2l - 1$. Thus $D_s = \sum_{i=1}^s W_i \geq \sum_{i=\sqrt{s}}^s W_i$. To use theorem 7, we note that $E(D_s) \geq \Theta(\log \sqrt{s}) \geq \frac{\log s}{n}$ and the min rate is at least $2\sqrt{s}$. Therefore we get, for some constant $\lambda \leq 1$, $Pr(D_s < \lambda \log \sqrt{s}) \leq e^{-\Omega(\sqrt{s} \log s)}$. Let $k = 5$ and $\lambda < 4/5$, for $s \geq \log^{2k} n$ we have $Pr(D_s \leq \log log n) \leq e^{-\Omega(\log^k n)}$.

By taking a union bound, we get that the probability that some node $v$ that arrives after $n/\log n$ nodes has $R_v \geq C \log \log n$ or has $D_s \leq C \log \log n$ is at most $n \left( e^{-\Omega(\log^k n)} + e^{-\Omega(\log^k n)} \right) < o(1)$. For a node to have a subtree of size $s$, $D_s \leq R_v$. Therefore, with probability $1 - o(1)$, all nodes that arrive after $n/\log n$ nodes have subtrees of size at most $\log^{10} n$.

Let us now bound the height of a subtree of size $s$. The diameter in preferential attachment models are well understood and in particular for trees we have tight bounds as shown by Dommers et. al in [10]. A lower bound for this random tree model was first proven in [22] and later a matching upper bound was given in [10]. The following is a combination of two theorems from [10] restricted to our model.

**Theorem 8 ([10]).** Let $G(t)$ be the tree formed by the BA model with $t$ nodes. Then, with high probability, the diameter of $G(t)$ is $\Theta(\log t)$.

We know that the subtree $H$ is a by preferential attachment subtree. Therefore, we get that for any subtree $H$ of size $s$ the diameter of $H$ is at most $c \log s$, with probability $1 - o(1)$. Moreover, with probability $1 - o(1)$, all nodes that arrives after $n/\log n$ have subtrees of size at most $\log^1 0 n$. Let $v$ be the $i^\text{th}$ node to arrive, with $i > n/\log n$. Then, with probability $1 - o(1)$, the diameter of the subtree rooted at $v$ is at most $O(\log \log n)$. This implies that $v$ is a $(X, Y)$-leaf, for $X = \log^{10} n$ and $Y = c \log \log n$, with probability $1 - o(1)$.

We now want to prove that most early nodes have high degree. Once all the early nodes have arrived, that is, the graph has $n/\log n$ many nodes, we want to bound the number times any early node gets a child. Note that, the process runs until we get $n - n/\log n$ many more nodes. By our previous calculation we know that the remaining time $R$ is sum of independent exponential variables with rate at least $n/\log n$ and $E(R)$ is $\Theta(\log \log n) \geq \frac{\log \log n}{2}$. We use theorem 7 to get, the probability of $R$ is less than $\frac{\log \log n}{2}$ is at most $e^{-\Omega(\log n)}$.

Let $Z_d$ denote that time required to get $d$ critical nodes. It is easy to see that the waiting between the $(k-1)^{\text{th}}$ and $k^{\text{th}}$ critical nodes is an exponential variable with rate $\frac{2n}{\log n} + k - 1$. Therefore $Z_d$ is sum of $d$ independent exponential variables with rate at least $\frac{2n}{\log n} - 1$. We have $E(Z_d) = \Theta(\log(2x+d/2i)) \leq \log(d/x+1)$, where $x = n/\log n$. We again use the concentration bound from theorem 7 to get that the probability that $Z_d$ is more than $\log(d/x+1)$ with probability at most $e^{-\Omega(\log n)}$. Therefore for $d \leq \Theta(n/\sqrt{\log n})$, we have that with probability $1 - o(1)$, $R > \frac{\log \log n}{2}$ and $Z_d < \frac{\log \log n}{2}$. That is, there are $\Theta(n/\sqrt{\log n})$ many critical nodes with probability $1 - o(1)$.

Therefore the average degree of an early node is at least $\sqrt{\log n}$. Therefore we can show that with probability $1 - o(1)$, most early nodes have degree at least $(\log n)^{1/4}$. Thus, most critical nodes have a high degree parent. We also know that with high probability all critical nodes have size at most $\log^{10} n$. Therefore at most $o(n)$ many nodes are not part of a good subtree.

**Proof of Theorem 5.** Simply combine Proposition 3 and Lemma 11. Together, they say that with probability $1 - o(1)$, only $o(n)$ nodes are not in good subtrees. Moreover, the expected number of nodes that are in good subtrees but do not finalize by $n_{\frac{\ln n}{2 \ln \ln n}}$ is also $o(n)$. Therefore, Markov’s inequality alone suffices to claim that with probability $1 - o(1)$, $n - o(n)$ nodes have finalized by $n_{\frac{\ln n}{2 \ln \ln n}}$. 

21
Theorem 3 claims that with probability $1 - o(1)$, $n/2 + \delta n/4$ nodes have a CORRECT announcement. When both of these conditions hold, at most $o(n)$ nodes can possibly change their future announcement from this point, and therefore the CORRECT majority holds until termination.

Proof of Theorem 6. For $M \leq \ln n$ there’s not much left to wrap up. We just need to count the number of nodes that we’ve just claimed are finalized by $n \frac{\ln n}{\ln \ln \ln n}$. Observe first that there are $\geq M^D$ nodes of height $\leq \ln n$, and at most $2M^{D-o(1)}$ nodes of height $\geq \ln \ln n$ (where $D$ denotes the height of the root). Therefore, a $1-o(1)$ fraction of all nodes are of height $\leq \ln \ln n$. Moreover, Proposition 5 proves that a $1-o(1)$ fraction of such nodes are finalized by $n \frac{\ln n}{\ln \ln \ln n}$. Therefore, we conclude that only $o(n)$ nodes in the entire graph are not finalized by $n \frac{\ln n}{\ln \ln \ln n}$, except with probability $o(1)$. Theorem 3 claims that with probability $1-o(1)$, $n/2 + \delta n/4$ nodes have a CORRECT announcement. When both of these conditions hold, at most $o(n)$ nodes can possibly change their future announcement from this point, and therefore the CORRECT majority holds until termination.

For $M \geq \ln n$, the argument is actually simpler: all pairs $(u,v)$ that contain some node $x$ on $P(u,v)$ that has degree at least $\ln n$. By Proposition 2, $x$ is safe thru $n \ln^2 n$ even ignoring $S_x$ with probability $1-o(1)$. Therefore, Lemma 6 guarantees that $1 - p_{uv}^{n \ln^2 n} = o(1)$, and Lemma 1 guarantees that there’s a CORRECT majority with probability $1-o(1)$ at $n \ln^2 n$. However, the diameter is $O(\ln n)$, and therefore Corollary 1 guarantees that the entire process stabilizes by $n \ln^2 n$ with probability $1-o(1)$. So taking a union bound, we see that with probability $1-o(1)$ the process has stabilized in a CORRECT majority by $n \ln^2 n$.

Proof of Proposition 5. Corollary 4 allows us to conclude that such nodes are nearly-finalized with respect to their parent at $n \frac{\ln n}{64 \ln \ln \ln n}$ with probability $1-o(1)$. Next, we want to claim that most of the descendents of such nodes are finalized by $n \frac{\ln n}{64 \ln \ln \ln n}$.

We’ll again consider epochs of time passing from the root. We already know that with probability $1-o(1)$, $T_v \leq n \frac{\ln n}{64 \ln \ln \ln n}$, so assume that this holds. We then want to take $T = \frac{\ln n}{64(\ln \ln n)^2}$ and apply Corollary 3. This immediately lets us conclude that with probability $1 - e^{-\frac{\ln n}{64(\ln \ln n)^2}}$, $v$ has $\lfloor M/2 \rfloor \geq M/3$ finalized children.

Now, for each of $v$’s non-finalized children, we’ll apply Corollary 3 again with the same choice of $T$. We’ll continue this process recursively until we reach the bottom.

So, the probability that a single non-finalized descendent does not have the desired number of finalized children by the end of its prescribed epoch is at most $e^{-\frac{\ln n}{64(\ln \ln n)^2}}$. Taking a union bound over all $2M^h$ nodes gives that with probability at most $M^h e^{-\frac{\ln n}{64(\ln \ln n)^2}}$, there is any failure, and as $M \leq \ln n$, $h = \ln \ln n$, this entire bound is $o(1)$.

Now, we want to further use Corollary 3 to conclude that once a node is finalized, all of its nodes are finalized by the end of the epochs. Let now $x$ be some node that is finalized during epoch $i$. Then $x$ is of height $h-i$ and is a $(2M^{h-i}, i)$-leaf. Corollary 3 therefore immediately implies that the probability that any of $x$’s descendents are not finalized by the end of an additional $(h-i)\frac{\ln n}{64(\ln \ln n)^2}$ steps is at most $2M^{h-i}e^{-\frac{\ln n}{64(\ln \ln n)^2}}$.

Taking a further union bound over all nodes, we see that this entire failure probability is again $o(1)$.

So let’s recap what we have now. First, we have an $o(1)$ failure probability that $v$ is not nearly-finalized by $n \frac{\ln n}{64 \ln \ln \ln n}$. Next, we have an $o(1)$ failure probability that any non-finalized node does not have at least $M/3$ finalized children by the end of its prescribed epoch. Finally, we have an $o(1)$ failure probability that any node with an ancestor who was finalized during their prescribed
epoch is not finalized by the end of the entire process (which is exactly $n^{\frac{\ln n}{32 \ln \ln n}}$). So except with probability $o(1)$, at least an $1/3$ fraction of remaining descendents get a finalized ancestor in each epoch, and is itself finalized by the end.

The final step is just a simple counting: if the number of remaining descendents without a finalized ancestor shrinks by a $2/3$ factor, and there are $h$ epochs, then the number of unfinalized nodes at the end is at most a $(2/3)^h$ fraction of the initial set. \qed