Electron. Commun. Probab. **26** (2021), article no. 55, 1–13.
https://doi.org/10.1214/21-ECP423
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Sufficient condition for root reconstruction by parsimony on binary trees with general weights*

Sebastien Roch† Kun-Chieh Wang‡

Abstract

We consider the problem of inferring an ancestral state from observations at the leaves of a tree, assuming the state evolves along the tree according to a two-state symmetric Markov process. We establish a general branching rate condition under which maximum parsimony, a common reconstruction method requiring only the knowledge of the tree topology (but not of the substitution rates or other parameters), succeeds better than random guessing uniformly in the depth of the tree. We thereby generalize previous results of [13, 37]. Our results apply to both deterministic and i.i.d. edge weights.

Keywords: Markov model on a tree; reconstruction problem; branching number; parsimony.
MSC2020 subject classifications: 92D15.
Submitted to ECP on December 31, 2020, final version accepted on August 26, 2021.

1 Introduction

Ancestral reconstruction In biology, the inferred evolutionary history of organisms is depicted by a phylogenetic tree, that is, a rooted tree whose branchings indicate past speciation events with the leaves representing living species. The evolution of features, such as the nucleotide at a given position in the genome of a species, the presence or absence of a protein or the number of horns in a lizard, is commonly assumed to follow Markovian dynamics along the tree [35]. That is, on each edge, the state of the feature changes according to a Markov process; at bifurcations, two independent copies of the feature evolve along the outgoing edges starting from the state at the branching point.

Here we consider the problem of inferring an ancestral state from observations of a feature at the leaves of a known phylogenetic tree. We refer to this problem, which has important applications in biology [26, 8, 31], as the ancestral reconstruction problem. Many rigorous results have been obtained in previous work; see, e.g., [20, 4, 19, 36, 24, 9, 25, 27, 5, 13, 37, 33, 28, 29, 3, 21, 10, 23, 11, 15, 1] for a partial list. Typically, one seeks an estimator of the root state which is strictly superior to random guessing—uniformly in the depth of the tree—under a uniform prior on the root [25]. Whether such

*Supported by NSF grants DMS-1248176, DMS-1149312 (CAREER), DMS-1614242, CCF-1740707 (TRIPODS), DMS-1902892, DMS-1916378, DMS-2023239, as well as a Simons Fellowship and a Vilas Associates Award.

†Department of Mathematics, University of Wisconsin–Madison, United States of America. E-mail: roch@wisc.edu

‡Google.

an estimator exists has been shown to depend on a trade-off between the mixing rate of the Markov process and the growth rate of the tree. In some cases, for instance in two-state symmetric Markov chains on binary trees [20, 19] as well as on more general trees [9], sharp thresholds have been established. We focus here on the special case of two-state symmetric chains on binary trees.

The threshold of maximum parsimony Maximum parsimony is an ancestral reconstruction method with a long history in evolutionary biology. See e.g. [35]. Its underlying principle is simple and intuitive: it calls for assigning a state to *each* internal vertex in such a way as to *minimize the total number of changes* along the edges; the resulting state at the *root* (possibly not unique) is the desired reconstructed state. An advantage of maximum parsimony, which accounts partly for its occasional use in practice, is that it only requires knowledge of the tree—not of the substitution probabilities along the edges, which can be difficult to estimate accurately from data and can affect the accuracy of reconstruction [7]. In fact maximum parsimony is equivalent to maximum likelihood when only the tree topology is known [36]. Another root state estimator which only requires knowledge of the tree is recursive majority, studied in [24, 26].

The theoretical properties of maximum parsimony for ancestral state reconstruction have been widely studied [34, 36, 12, 37, 13, 6, 14, 16, 17, 18]. From its very definition, one might expect maximum parsimony to perform well when the probabilities of substitution along the edges are "sufficiently small." This statement was confirmed rigorously for the two-state symmetric model on complete binary trees with constant mutation probability in [37] (as first conjectured in [34]). Below a critical probability, maximum parsimony does indeed beat random guessing. A related result was obtained in [13] under a common random tree model known as the Yule tree. Here we substantially generalize both of these results. We give a general sufficient condition on the branching rate under which maximum parsimony succeeds. Our condition is related to the branching number of the tree, which roughly speaking generalizes the concept of vertex degree and plays an important role in many processes on trees. See e.g. [22]. Other generalized notions of degree have been used in related contexts [32].

Definitions In order to state our results formally, we begin with some definitions. Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ be an infinite complete binary tree rooted at ρ with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$. That is, all vertices of $\mathcal{T}$ have exactly two children; in particular, $\mathcal{T}$ has no leaf. Every edge $e \in \mathcal{E}$ is assigned a weight $\theta_e \in [0, 1]$. If $e = (x, y)$ where x is the parent of y, we also write $\theta_y = \theta_e$. We use the notation $x \leq y$ to indicate that x is an ancestor of y and we write $x < y$ for $x \leq y$ and $x \neq y$. We also let $s(x)$ be the sibling of $x \neq \rho$.

Under the *Cavender-Farris-Neyman (CFN) model* over $\mathcal{T}$ and $\theta = (\theta_y)_{y \in \mathcal{V}}$, also known as Neyman 2-state model,we associate to each vertex $x \in \mathcal{V}$ a state $\sigma_x \in \{0, 1\}$ as follows. The state at the root, σ_ρ, is picked uniformly at random in $\{0, 1\}$. Recursively, if y has parent x, state σ_y is equal to σ_x with probability θ_y, otherwise it is picked uniformly at random in $\{0, 1\}$. We let $p_y = \mathbb{P}_{\mathcal{T},\theta}[\sigma_x \neq \sigma_y] = \frac{1-\theta_y}{2}$ be the *probability of a substitution* on edge $e = (x, y)$. Here we use $\mathbb{P}_{\mathcal{T},\theta}$ to indicate probabilities taken under the CFN model over $\mathcal{T}$ and θ. The CFN model is equivalent to the Fortuin-Kastelyn random cluster representation of the ferromagnetic Ising model on $\mathcal{T}$ with a free boundary. See e.g. [9] and references therein.

Informally, a root state estimator is a map returning a (possibly randomized) guess for the state of the root, based on the knowledge of the states on an observed cutset.

Definition 1.1 (Cutset)**.** *A* cutset *is a minimal, finite set of vertices $\pi \subseteq \mathcal{V}$ such that all infinite self-avoiding paths starting at ρ must visit π. We let $\mathscr{C}(\mathcal{T})$ be the collection of all*

cutsets of $\mathcal{T}$. We denote by $\mathcal{T}^\pi = (\mathcal{V}^\pi, \mathcal{E}^\pi)$ the finite tree obtained from $\mathcal{T}$ after removing all descendants of the vertices in π.

As mentioned above, our focus in this work is on a root state estimator known in phylogenetics as *maximum parsimony*. Fix a cutset π and assume that the states on π are observed. The parsimony principle dictates that one assigns to each vertex x (ancestor to the observed cutset π) a state $\hat{\sigma}_x$ such that the overall number of changes along the edges of $\mathcal{T}^\pi$, namely, $\sum_{(x,y)\in\mathcal{E}^\pi} \mathbf{1}\{\hat{\sigma}_x \neq \hat{\sigma}_y\}$, is minimized, where we let by default $\hat{\sigma}_z = \sigma_z$ for all $z \in \pi$. In case both 0 and 1 can be obtained in this way as root state, a uniformly random value in $\{0,1\}$ is returned. We let $\mathrm{RA}^\pi_{\mathcal{T},\theta}$ be the *reconstruction accuracy* of parsimony, i.e. the probability that it correctly reconstructs the root state.

Main result: deterministic weights In our main result, we give conditions under which the reconstruction accuracy of maximum parsimony is uniformly bounded away from $1/2$.

Theorem 1.2 (Reconstruction accuracy of parsimony: deterministic weights)**.** *Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ be an infinite complete binary tree with edge weights $\theta = (\theta_z)_{z\in\mathcal{V}}$ satisfying $\theta_* := \inf_{z\in\mathcal{V}} \theta_z > 0$ and*

$$\sup\left\{\kappa > 0 : \inf_{\pi\in\mathscr{C}(\mathcal{T})} \sum_{x\in\pi} \prod_{\rho\neq z\leq x} \kappa^{-1}\theta_z > 0\right\} > \frac{3}{2}. \tag{1.1}$$

Then $\inf_{\pi\in\mathscr{C}(\mathcal{T})} \mathrm{RA}^\pi_{\mathcal{T},\theta} > \frac{1}{2}$, i.e., the reconstruction accuracy of parsimony on $\mathcal{T}$ is away from $1/2$.

Condition (1.1) involves the *branching number*, a generalized notion of branching rate which plays a key role in the analysis of many stochastic processes on trees and tree-like graphs. See e.g. [22]. The following example provides some intuition in a special case. See Lemma 4.1 for another illustration.

Example 1.3 (Fixed edge weights)**.** As a simple illustration, observe that, when all weights are equal to $\theta_* \in (0,1]$, the supremum in (1.1) is attained for $\kappa = 2\theta_*$. Indeed, the sum in (1.1) when $\kappa = 2\theta_*$ simplifies to

$$\sum_{x\in\pi} \prod_{\rho\neq z\leq x} (2\theta_*)^{-1}\theta_z = \sum_{x\in\pi} \prod_{\rho\neq z\leq x} \frac{1}{2} = 1 > 0$$

for any cutset π, where the equality can be proved by induction on the graph distance from the root to the furthest vertex in π. On the other hand, letting π_n be the cutset of all vertices at graph distance n from ρ, for any $\varepsilon > 0$ it holds that

$$\sum_{x\in\pi_n} \prod_{\rho\neq z\leq x} ((2+\varepsilon)\theta_*)^{-1}\theta_z = (2+\varepsilon)^{-n}\cdot 2^n \to 0,$$

as $n \to +\infty$. Hence, in this case, condition (1.1) reduces to $2\theta_* > 3/2$, that is, $\theta_* > 3/4$. In terms of substitution probability, this is $p_* = \frac{1-\theta_*}{2} < 1/8$.

The argument in the example above leads to the following corollary.

Corollary 1.4 (All substitution probabilities below the threshold)**.** *Let the substitution probabilities $(p_z)_{z\in\mathcal{V}}$ satisfy $p_z \in [0, p_*]$ for all $z \in \mathcal{V}$, for some $p_* < 1/8$. Then the reconstruction accuracy of maximum parsimony on $\mathcal{T}$ is bounded away from $1/2$.*

In general, Theorem 1.2 cannot be improved in the following sense. It was shown in [37, Theorem 4.1] that when $p_z = p_* > 1/8$, that is, $\theta_z = 1 - 2p_z = \theta_* < 3/4$ for all z then $\inf_n \mathrm{RA}^{\pi_n}_{\mathcal{T},\theta} = \frac{1}{2}$, where π_n is defined in Example 1.3. On the other hand, it is not

known whether the reconstruction accuracy necessarily converges to $1/2$ if (1.1) is not satisfied. We leave this as an open problem.

Zhang et al. [37] also established the special case of Theorem 1.2 when $p_z = p_* < 1/8$ for all z and $\pi = \pi_n$. Their proof proceeds through a careful analysis of the limit of a recurrence for $\mathrm{RA}^{\pi_n}_{\mathcal{T},\theta}$ first derived in [34]. Our more general result follows from a softer argument which relies on the instability of a fixed point of this recurrence corresponding to asymptotic reconstruction accuracy $1/2$. A more detailed proof sketch is given in Section 2 following some preliminaries. We note that our proof method may be of more general interest, e.g., to extend the results beyond the two-state case where the higher dimensionality of the system may complicate significantly the derivation of an explicit limit even when edge weights are constant.

Main result: i.i.d. weights We also obtain a related result in the case of edge weights that are i.i.d. No lower bound on the weights is needed in this case, unlike Theorem 1.2.

Theorem 1.5 (Reconstruction accuracy of parsimony: i.i.d. weights)**.** *Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ be an infinite complete binary tree with edge weights $\theta = (\theta_z)_{z\in\mathcal{V}}$ drawn i.i.d. from a distribution Θ over $(0, 1]$. Let μ be the mean of Θ and assume that $\mu > 3/4$. Then, for any $\delta > 0$, there is $\varepsilon > 0$ such that $\inf_{\pi\in\mathscr{C}(\mathcal{T})} \mathrm{RA}^{\pi}_{\mathcal{T},\theta} > \frac{1}{2}(1 + \varepsilon)$, with probability at least $1 - \delta$.*

The previous theorem covers in particular the case of the pure birth process, or Yule tree, which is a popular random model of phylogenetic trees. See e.g. [35]. In that case, $\theta_z = e^{-2T_z}$, where T_z is an exponential with rate λ. To derive the corresponding threshold, we note that

$$\mu = \mathbb{E}[\theta_z] = \int_0^\infty e^{-2t}\lambda e^{-\lambda t}\,\mathrm{d}t = \frac{\lambda}{\lambda + 2}.$$

Then $\mu > \frac{3}{4} \iff \lambda > 6$, which is consistent with the results of [13, Theorem 2.3].

2 Preliminaries

Computing parsimony Our proofs are based on a recurrence for the reconstruction accuracy. Maximum parsimony can be computed efficiently by dynamic programming, which is referred to as the *Fitch method*.

Definition 2.1 (Parsimony recursion)**.** *The Fitch method recursively constructs a set $\widehat{S}^{\pi}_z$ of possible states for each vertex $z \in \mathcal{V}^{\pi}$, starting from π, as follows. If $z \in \pi$, $\widehat{S}^{\pi}_z = \{\sigma_z\}$. If $z \notin \pi$ and has children x and y,*

$$\widehat{S}^{\pi}_z = \begin{cases} \widehat{S}^{\pi}_x \cap \widehat{S}^{\pi}_y, & \text{if } \widehat{S}^{\pi}_x \cap \widehat{S}^{\pi}_y \neq \emptyset \\ \widehat{S}^{\pi}_x \cup \widehat{S}^{\pi}_y, & \text{o.w.} \end{cases}$$

The method returns the maximum parsimony estimator $\hat{\sigma}_\rho$ which is equal to the unique state in $\widehat{S}^{\pi}_\rho$ if $|\widehat{S}^{\pi}_\rho| = 1$, and otherwise returns a uniformly random value in $\{0, 1\}$.

What is described above is the bottom-up phase of the Fitch method. (A top-down phase, which we will not require here, then assigns a state to each vertex in $\mathcal{V}^{\pi}$ consistent with a maximum parsimony solution; e.g. [35].) Let π be an arbitrary cutset on $\mathcal{T}$ with states σ_u, $u \in \pi$, and let $\widehat{S}^{\pi}_z$, $z \in \mathcal{V}^{\pi}$, be the corresponding reconstructed sets under the Fitch method. We define

$$\alpha^{\pi}_z = \mathbb{P}_{\mathcal{T},\theta}\left[\widehat{S}^{\pi}_z = \{\sigma_z\}\right], \qquad \beta^{\pi}_z = \mathbb{P}_{\mathcal{T},\theta}\left[\widehat{S}^{\pi}_z = \{1 - \sigma_z\}\right].$$

Under our randomization rule, the reconstruction accuracy of maximum parsimony $\mathrm{RA}^{\pi}_{\mathcal{T},\theta}$ is given by

$$\mathbb{P}_{\mathcal{T},\theta}\left[\widehat{S}^{\pi}_\rho = \{\sigma_\rho\}\right] + \frac{1}{2}\mathbb{P}_{\mathcal{T},\theta}\left[\widehat{S}^{\pi}_\rho = \{0, 1\}\right] = \alpha^{\pi}_\rho + \frac{1}{2}(1 - \alpha^{\pi}_\rho - \beta^{\pi}_\rho) = \frac{1}{2} + \frac{1}{2}(\alpha^{\pi}_\rho - \beta^{\pi}_\rho). \quad (2.1)$$

Proof sketch Fix θ satisfying the assumptions of Theorem 1.2 and fix a cutset π. To analyze $(\alpha_z^\pi, \beta_z^\pi)$, it is natural to take advantage of the recursive nature of $\mathcal{T}$.Let x and y be the children of z. The event $\widehat{S}_z^\pi = \{\sigma_z\}$ occurs when either (i) $\widehat{S}_x^\pi = \{\sigma_z\}$ and $\widehat{S}_y^\pi = \{\sigma_z\}$, or (ii) $\widehat{S}_x^\pi = \{\sigma_z\}$ and $\widehat{S}_y^\pi = \{0,1\}$ or vice versa. By the Markov property of the CFN model, the random variables $\widehat{S}_x^\pi$ and $\widehat{S}_y^\pi$, which are functions only of the states of π below x and y respectively, are conditionally independent given σ_z. Hence, letting $q_u = 1 - p_u$ for $u = x, y$ and taking into account the possibility of a mutation along the edges (z, x) and (z, y), it follows as first derived in [34, Lemma 7.20] that

$$\begin{aligned}\alpha_z^\pi &= (q_x\alpha_x^\pi + p_x\beta_x^\pi)(q_y\alpha_y^\pi + p_y\beta_y^\pi) \\ &\quad +(q_x\alpha_x^\pi + p_x\beta_x^\pi)(1-\alpha_y^\pi-\beta_y^\pi) + (1-\alpha_x^\pi-\beta_x^\pi)(q_y\alpha_y^\pi + p_y\beta_y^\pi), \qquad (2.2)\end{aligned}$$

where the first and second lines on the r.h.s. correspond respectively to cases (i) and (ii) above. Similarly,

$$\begin{aligned}\beta_z^\pi &= (p_x\alpha_x^\pi + q_x\beta_x^\pi)(p_y\alpha_y^\pi + q_y\beta_y^\pi) \\ &\quad +(p_x\alpha_x^\pi + q_x\beta_x^\pi)(1-\alpha_y^\pi-\beta_y^\pi) + (1-\alpha_x^\pi-\beta_x^\pi)(p_y\alpha_y^\pi + q_y\beta_y^\pi). \qquad (2.3)\end{aligned}$$

In the case that $p_u = p$ for all u, a fixed point analysis was performed in [34, Theorem 7.22]. It was found that, if $p \geq 1/8$, there is a single fixed point $(1/3, 1/3)$ which corresponds informally to "having no information about the root." While if $p < 1/8$, there is an additional fixed point $(\alpha_p^\infty, \beta_p^\infty)$ with $\alpha_p^\infty > \beta_p^\infty$. Convergence to $(1/3, 1/3)$ in the first case and $(\alpha_p^\infty, \beta_p^\infty)$ in the second case was established rigorously in [37]. One step in [37] involved the derivation of a new recurrence in terms of the quantities $\alpha_z^{\pi_n} - \beta_z^{\pi_n}$ and $1 - (\alpha_z^{\pi_n} + \beta_z^{\pi_n})$, which facilitates the analysis of the limit in the fixed edge weight case.

Going back to binary trees with general weights, as our starting point we further modify the recurrence of [37]. For all $z \in \mathcal{V}$, we define

$$\mathbf{d}_z^\pi = \alpha_z^\pi - \beta_z^\pi, \qquad \text{and} \qquad \mathbf{u}_z^\pi = 3(\alpha_z^\pi + \beta_z^\pi) - 2. \qquad (2.4)$$

We show in Proposition 2.2 below that $(\mathbf{d}_z^\pi, \mathbf{u}_z^\pi)$ satisfies the following recurrence

$$\mathbf{d}_z^\pi = \left(\frac{4 - \mathbf{u}_y^\pi}{6}\right)\theta_x\mathbf{d}_x^\pi + \left(\frac{4 - \mathbf{u}_x^\pi}{6}\right)\theta_y\mathbf{d}_y^\pi, \qquad (2.5)$$

$$\mathbf{u}_z^\pi = \frac{3}{2}\theta_x\theta_y\mathbf{d}_x^\pi\mathbf{d}_y^\pi - \frac{1}{2}\mathbf{u}_x^\pi\mathbf{u}_y^\pi, \qquad (2.6)$$

for $z \in \mathcal{V}^\pi - \pi$ with children x, y, as well as the inequalities $0 \leq \mathbf{d}_z^\pi \leq 1$ and $-1/2 \leq \mathbf{u}_z^\pi \leq 1$ for all $z \in \mathcal{V}^\pi$ and the boundary conditions $\mathbf{d}_z^\pi = \mathbf{u}_z^\pi = 1$ for all $z \in \pi$. Our choice of parametrization is motivated in part by the fact that the "no information" fixed point is now at $(0, 0)$ and that $|\mathbf{d}_z^\pi|, |\mathbf{u}_z^\pi| \leq 1$. At a high level we show that, under the branching rate condition (1.1), the fixed point $(0, 0)$ is "unstable" and that $\mathbf{d}_z^\pi$ in particular stays bounded away form 0. That in terms implies a lower bound on the reconstruction accuracy as, by (2.1), we have

$$\mathrm{RA}_{\mathcal{T},\theta}^\pi = \frac{1}{2} + \frac{1}{2}\mathbf{d}_\rho^\pi.$$

The link between stability and the weighted branching rate in (1.1) can be seen from (2.5). Consider first the simpler special case where all weights are equal to θ_* and $\pi = \pi_n$: by symmetry, $\mathbf{d}_x^\pi$ takes the same value for all x at the same graph distance from ρ; and moreover assuming that we are close to the fixed point $(0, 0)$, that is, $(\mathbf{u}_x^\pi, \mathbf{d}_x^\pi) \approx (0, 0)$, (2.5) implies

$$\mathbf{d}_z^\pi = \left(\frac{4 - \mathbf{u}_y^\pi}{6}\right)\theta_x\mathbf{d}_x^\pi + \left(\frac{4 - \mathbf{u}_x^\pi}{6}\right)\theta_y\mathbf{d}_y^\pi \approx 2\frac{4}{6}\theta_*\mathbf{d}_x^\pi = \frac{4}{3}\theta_*\mathbf{d}_x^\pi.$$

Hence, in that case, the solution is expected to grow when $\frac{4}{3}\theta_* > 1$, corresponding to the condition derived in Example 1.3. More generally, we use (2.5) to relate the $\mathbf{d}$-value at the root to the $\mathbf{d}$-values on a cutset (see Lemma 3.1). To deal with the nonlinear nature of (2.5) and (2.6), we control the $\mathbf{u}$-values thanks to the quadratic form of (2.6) which implies a quick decay towards 0 (see Lemma 3.3).

Recurrence Before proceeding to the proof of our main results, we first establish a basic recurrence which follows from the work of [37]. We give a short proof for completeness.

Proposition 2.2 (Recurrence and basic properties)**.** *The following hold:*

- [Boundary conditions] *For all* $z \in \pi$, $\mathbf{d}_z^\pi = 1$ *and* $\mathbf{u}_z^\pi = 1$.
- [Recurrence] *For all* $z \in \mathcal{V}^\pi - \pi$, *if* x, y *are the children of* z, *the system* (2.5) *and* (2.6) *holds.*
- [Bounds] *For all* $z \in \mathcal{V}^\pi$, *we have* $0 \leq \mathbf{d}_z^\pi \leq 1$ *and* $-1/2 \leq \mathbf{u}_z^\pi \leq 1$.

Proof. We start with the boundary conditions. By Definition 2.1(a) and the definitions of α_z^π and β_z^π, we have for all $z \in \pi$ that $\alpha_z^\pi = \mathbb{P}_{\mathcal{T},\theta}[\widehat{S}_z^\pi = \{\sigma_z\}] = 1$ and $\beta_z^\pi = \mathbb{P}_{\mathcal{T},\theta}[\widehat{S}_z^\pi = \{1-\sigma_z\}] = 0$. So $\mathbf{u}_z^\pi = 3(\alpha_z^\pi + \beta_z^\pi) - 2 = 1$ and $\mathbf{d}_z^\pi = \alpha_z^\pi - \beta_z^\pi = 1$.

The second statement is merely a change of variables. We briefly expand on the first equation (the other one being similar). Let $z \in \mathcal{V}^\pi - \pi$ with children x, y. We define $\Sigma_u = \alpha_u^\pi + \beta_u^\pi$ and $\Delta_u = \alpha_u^\pi - \beta_u^\pi$. By the definitions of p_x and q_x, note that

$$q_x\alpha_x^\pi + p_x\beta_x^\pi = \left(\frac{1+\theta_x}{2}\right)\alpha_x^\pi + \left(\frac{1-\theta_x}{2}\right)\beta_x^\pi = \frac{1}{2}\Sigma_x + \frac{1}{2}\theta_x\Delta_x,$$

and, similarly, $p_x\alpha_x^\pi + q_x\beta_x^\pi = \frac{1}{2}\Sigma_x - \frac{1}{2}\theta_x\Delta_x$. Hence, by (2.2) and (2.3),

$$\alpha_z^\pi = \frac{1}{4}(\Sigma_x + \theta_x\Delta_x)(\Sigma_y + \theta_y\Delta_y) + \frac{1}{2}(\Sigma_x + \theta_x\Delta_x)(1-\Sigma_y) + \frac{1}{2}(1-\Sigma_x)(\Sigma_y + \theta_y\Delta_y),$$

and

$$\beta_z^\pi = \frac{1}{4}(\Sigma_x - \theta_x\Delta_x)(\Sigma_y - \theta_y\Delta_y) + \frac{1}{2}(\Sigma_x - \theta_x\Delta_x)(1-\Sigma_y) + \frac{1}{2}(1-\Sigma_x)(\Sigma_y - \theta_y\Delta_y).$$

Subtracting the above two equations, we get

$$\begin{aligned}\Delta_z &= \frac{1}{2}\theta_x\Delta_x\Sigma_y + \frac{1}{2}\theta_y\Delta_y\Sigma_x + \theta_x\Delta_x(1-\Sigma_y) + \theta_y\Delta_y(1-\Sigma_x)\\ &= \left(1 - \frac{1}{2}\Sigma_y\right)\theta_x\Delta_x + \left(1 - \frac{1}{2}\Sigma_x\right)\theta_y\Delta_y,\end{aligned}$$

which after plugging in (2.4) gives (2.5).

Because α_z^π and β_z^π are probabilities and further $\alpha_z^\pi + \beta_z^\pi \leq 1$, we have that $\mathbf{d}_z^\pi \leq 1$ and $\mathbf{u}_z^\pi \leq 1$, for all z. Moreover, that together with the boundary conditions and (2.5) implies that $\mathbf{d}_z^\pi \geq 0$ for all z by induction. In turn, that together with the boundary conditions and (2.6) implies that $\mathbf{u}_z^\pi \geq -1/2$ for all z. □

3 Deterministic weights

Before proceeding with the proof of Theorem 1.2, we first prove some lemmas.

3.1 Controlling $\mathbf{d}$- and $\mathbf{u}$-values

In the first lemma, we express the $\mathbf{d}$-value at the root as a function of the $\mathbf{d}$- and $\mathbf{u}$-values above an arbitrary cutset. Recall that $s(z)$ is the sibling of z.

Lemma 3.1 (Controlling the root with a cutset)**.** *For any cutset π' in $\mathcal{T}^\pi$, it holds that*

$$\mathbf{d}_\rho^\pi = \sum_{x\in\pi'} \mathbf{d}_x^\pi \prod_{\rho\neq z\leq x} \left\{\frac{4-\mathbf{u}_{s(z)}^\pi}{6}\right\}\theta_z. \tag{3.1}$$

Proof. The result follows by recursively applying (2.5) from the root down to π'. We implicitly use the fact that, by definition, a cutset is minimal. □

Our second lemma shows that $\mathbf{d}$-values cannot grow too fast down the tree. This fact will be useful to proving the next key lemma. We will need the lower bound $\theta_* = \inf_{z\in\mathcal{V}} \theta_z > 0$, on the θ-values. For $v, w \in \mathcal{V}^\pi$, we let $\gamma(v, w)$ be the graph distance between v and w in $\mathcal{T}^\pi$. Recall that $\theta_* \leq 1 < 2$.

Lemma 3.2 (Growth of $\mathbf{d}$-values)**.** *Fix any $v \in \mathcal{V}^\pi$. For all $\varepsilon' > 0$ and all descendants w of v in $\mathcal{V}^\pi$,*

$$\mathbf{d}_v^\pi \leq \varepsilon' \quad \textit{implies} \quad \mathbf{d}_w^\pi \leq \varepsilon' \left(\frac{2}{\theta_*}\right)^{\gamma(v,w)}.$$

Proof. Let $z \in \mathcal{V}^\pi$, not on π, have children x and y. (Note that, in the case where z is the parent of a vertex on the cutset π, z itself cannot be on the cutset by minimality and therefore both its children are in $\mathcal{V}^\pi$.) By Proposition 2.2, we have $\mathbf{u}_x^\pi, \mathbf{u}_y^\pi \leq 1$ and $\mathbf{d}_x^\pi, \mathbf{d}_y^\pi \geq 0$, which implies that both terms on the r.h.s. of (2.5) are non-negative. Hence, using $\theta_x \geq \theta_*$, (2.5) gives $\mathbf{d}_z^\pi \geq \frac{1}{2}\theta_* \mathbf{d}_x^\pi$. In particular, $\mathbf{d}_z^\pi \leq \varepsilon'$ implies that $\mathbf{d}_x^\pi \leq \varepsilon'(2/\theta_*)$. Recursing gives the claim. □

Our final lemma controls $\mathbf{u}$-values at the root of a subtree where $\mathbf{d}$-values are uniformly small.

Lemma 3.3 (Controlling $\mathbf{u}$ when $\mathbf{d}$ is small)**.** *Fix any $0 < \phi \leq 1/9$ and $v \in \mathcal{V}^\pi$. Then there exists $\varepsilon' > 0$ depending only on θ_* and ϕ such that: $\mathbf{d}_v^\pi \leq \varepsilon'(2/\theta_*)$ implies $|\mathbf{u}_v^\pi| \leq 4\phi$.*

Proof. Let H be the smallest non-negative integer such that

$$\left(\frac{1}{2}\right)^{-1+2^H} \leq \phi. \tag{3.2}$$

Define $\varepsilon' > 0$ to be the largest positive real such that

$$\varepsilon' \left(\frac{2}{\theta_*}\right)^{H+1} \leq 0.99 \qquad \text{and} \qquad \frac{3}{2}\left[\varepsilon' \left(\frac{2}{\theta_*}\right)^{H+1}\right]^2 \leq \phi. \tag{3.3}$$

The rest of the proof proceeds in two steps: we derive a simplified recurrence for $\mathbf{u}$-values and solve it.

1. Simplified recurrence: Assume that $\mathbf{d}_v^\pi \leq \varepsilon'(2/\theta_*)$. Let w be a descendant of v with graph distance $\gamma(v, w) \leq H$. Then, by Lemma 3.2, $\mathbf{d}_w^\pi \leq \varepsilon' \left(2/\theta_*\right)^{H+1} < 1$, where we used (3.3). This show, in particular, that all descendants of v in $\mathcal{V}$ within graph distance H are in fact strictly above π, because $\mathbf{d}$-values are 1 on the cutset π. Moreover, by the recurrence (2.6) and the inequality (3.3), for any descendant w_0 of v with children w_1, w_2 in $\mathcal{V}^\pi$ that are within graph distance H of v, we have

$$\left|\mathbf{u}_{w_0}^\pi\right| \leq \frac{3}{2}\theta_{w_1}\theta_{w_2}\mathbf{d}_{w_1}^\pi \mathbf{d}_{w_2}^\pi + \frac{1}{2}\left|\mathbf{u}_{w_1}^\pi\right|\left|\mathbf{u}_{w_2}^\pi\right| \leq \phi + \frac{1}{2}\left|\mathbf{u}_{w_1}^\pi\right|\left|\mathbf{u}_{w_2}^\pi\right|, \tag{3.4}$$

where we used that θ-values are ≤ 1 and (3.3). Define

$$\mathbf{U}_h = \sup\left\{|\mathbf{u}^\pi_w| \,:\, w \in \mathcal{V}^\pi,\, v \leq w \text{ and } \gamma(v,w) = h\right\},$$

and $\mathbf{U}_0 = |\mathbf{u}^\pi_v|$. By the remark above, the set in the previous display is non-empty for all $h = 0, \ldots, H-1$. Taking a supremum on both sides of (3.4) gives the recurrence

$$\mathbf{U}_h \leq \phi + \frac{1}{2}\mathbf{U}^2_{h+1}, \qquad 0 \leq h \leq H-1. \tag{3.5}$$

2. Solution: We show by induction on h (backwards from $H-1$) that

$$\mathbf{U}_h \leq 3\phi + \left(\frac{1}{2}\right)^{-1+2^{H-h}}. \tag{3.6}$$

For the base of the induction $h = H-1$, we have indeed that

$$\mathbf{U}_{H-1} \leq \phi + \frac{1}{2} \leq 3\phi + \left(\frac{1}{2}\right)^{-1+2^{H-(H-1)}},$$

where we used (3.5) and the fact that $\mathbf{u}$-values are ≤ 1 in absolute value. Assume the induction claim (3.6) holds for all $h'+1 \leq h \leq H-1$. We show it then holds for $h = h'$. Indeed, by (3.5) again,

$$\begin{aligned}
\mathbf{U}_{h'} &\leq \phi + \frac{1}{2}\mathbf{U}^2_{h'+1} \leq \phi + \frac{1}{2}\left[3\phi + \left(\frac{1}{2}\right)^{-1+2^{H-(h'+1)}}\right]^2 \\
&\leq \phi\left[1 + \frac{9\phi}{2} + 3\left(\frac{1}{2}\right)^{-1+2^{H-(H-1)}}\right] + \left(\frac{1}{2}\right)^{2\{-1+2^{H-(h'+1)}\}+1} \\
&\leq \phi\left[\frac{5}{2} + \frac{9\phi}{2}\right] + \left(\frac{1}{2}\right)^{-1+2^{H-h'}}.
\end{aligned}$$

Because by assumption $\phi \leq 1/9$, the square bracket above is ≤ 3. That concludes the induction.

By our choice of H in (3.2), that implies $|\mathbf{u}^\pi_v| = \mathbf{U}_0 \leq 3\phi + (1/2)^{-1+2^H} \leq 4\phi$. □

3.2 Proof of main theorem

Proof of Theorem 1.2. Fix $\pi \in \mathscr{C}(\mathcal{T})$ and assume that $\theta_* > 0$ and that (1.1) holds. Then there is $0 < \phi \leq 1/9$ and $0 < \zeta < 1$ such that

$$\sum_{x \in \pi'} \prod_{\rho \neq z \leq x} \left\{\frac{2}{3}(1-\phi)\right\}\theta_z \geq \zeta, \tag{3.7}$$

for all cutsets $\pi' \in \mathscr{C}(\mathcal{T})$. For this value of ϕ, let ε' be as in Lemma 3.3 and define $\varepsilon = \varepsilon'\zeta < \varepsilon'$. The proof proceeds by contradiction. Assume that $\mathbf{d}^\pi_\rho \leq \varepsilon$. Let π' be the cutset of those nodes closest to the root where the $\mathbf{d}$-values first cross above ε', i.e., formally $\pi' = \{x \in \mathcal{V}^\pi \,:\, \mathbf{d}^\pi_x > \varepsilon' \text{ and } \mathbf{d}^\pi_z \leq \varepsilon',\ \forall z \leq x\}$. Such a cutset (which is necessarily minimal) exists because $\mathbf{d}^\pi_v = 1$ for all $v \in \pi$ and $\varepsilon' > \varepsilon$. By Lemma 3.2, for all z on or above π', i.e. such that $z \leq x$ for some $x \in \pi'$, we have $\mathbf{d}^\pi_z \leq \varepsilon'\frac{2}{\theta_*}$ and $\mathbf{d}^\pi_{s(z)} \leq \varepsilon'\frac{2}{\theta_*}$, since the immediate parent of z (and $s(z)$) has $\mathbf{d}$-value $\leq \varepsilon'$ by definition of π'. By Lemma 3.3, we then have

$$|\mathbf{u}^\pi_z| \leq 4\phi \quad \text{and} \quad \left|\mathbf{u}^\pi_{s(z)}\right| \leq 4\phi. \tag{3.8}$$

By Lemma 3.1, (3.7) and (3.8), summing over π'

$$\mathbf{d}_\rho^\pi \;=\; \sum_{x\in\pi'} \mathbf{d}_x^\pi \prod_{\rho\neq z\le x} \left\{\frac{4-\mathbf{u}_{s(z)}^\pi}{6}\right\} \theta_z > \sum_{x\in\pi'} \varepsilon' \prod_{\rho\neq z\le x} \left\{\frac{2}{3}(1-\phi)\right\} \theta_z \ge \varepsilon'\zeta = \varepsilon,$$

which is a contradiction. □

4 I.i.d. weights

In this section, we prove our main result in the i.i.d. weight case. Because there is no lower bound on the weights, Theorem 1.2 cannot be applied directly to this case. In particular, the absence of a lower bound makes controlling the $\mathbf{u}$-values more challenging. Here we identify a subtree of $\mathcal{T}$ where $\mathbf{u}$-values are well-behaved. The existence of such a subtree is established with a coupling to a percolation process, where open edges roughly indicate that weights are uniformly bounded in a properly defined neighborhood.

Proof of Theorem 1.5. First, we need the following percolation result. To each edge $e=(x,y)$ of $\mathcal{T}$, where x is the parent of y, we assign an independent random weight $\tilde\theta_y$ drawn from a distribution $\widetilde\Theta$ over $(0,1]$. We also pick an independent indicator variable $\tilde J_y$, which is 1 with probability $\tilde q\in[0,1]$ and 0 otherwise. Let $\tilde{\mathcal{T}}=(\tilde{\mathcal{V}},\tilde{\mathcal{E}})$ be the subtree of $\mathcal{T}$ whose vertices x satisfy $\prod_{\rho\neq z\le x}\tilde J_z=1$ and whose edges are those with endvertices satisfying that condition. We let $\mathcal{N}_{\mathrm{ext}}$ be the event of non-extinction, i.e., the event that $\tilde{\mathcal{T}}$ is infinite. By standard branching process arguments [2], the extinction probability $\tilde\varphi$ satisfies $\tilde\varphi=\tilde q^2\tilde\varphi^2+2(1-\tilde q)\tilde q\tilde\varphi+(1-\tilde q)^2$, i.e.,

$$\mathbb{P}[\mathcal{N}_{\mathrm{ext}}]=\left(\frac{1-\tilde q}{\tilde q}\right)^2. \tag{4.1}$$

Lemma 4.1 (Branching condition: random edge weights on open cluster)**.** *Fix $\tilde q\in[0,1/2)$ and assume that $\widetilde\Theta$ has mean $\tilde\mu\in(0,1)$. Then, conditioned on $\mathcal{N}_{\mathrm{ext}}$, almost surely*

$$\kappa(\tilde{\mathcal{T}},\tilde\theta):=\sup\left\{\kappa>0:\inf_{\pi\in\mathscr{C}(\tilde{\mathcal{T}})}\sum_{x\in\pi}\prod_{\rho\neq z\le x}\kappa^{-1}\tilde\theta_z>0\right\}=2\tilde q\tilde\mu. \tag{4.2}$$

Proof. This result can be proved along the lines of Proposition 3.2, Proposition 5.1 and Corollary 5.2 in [30]. We sketch a proof.

For any $c>0$ and any rooted, locally finite tree $(\mathcal{T}',\theta')$ with positive edge weights, it follows from the definition of $\kappa(\mathcal{T}',\theta')$ that the property $\{\mathcal{T}'$ is finite or $\kappa(\mathcal{T}',\theta')\le c\}$ is inherited, in the sense that every finite weighted tree $(\mathcal{T}',\theta')$ has this property and that the descendant subtrees of the root of $(\mathcal{T}',\theta')$ have the property whenever $(\mathcal{T}',\theta')$ itself has it. From an argument similar to Proposition 3.2 in [30] (and using the independence of the weights), $\kappa(\tilde{\mathcal{T}},\tilde\theta)$ is almost surely constant conditioned on $\mathcal{N}_{\mathrm{ext}}$.

Let $\tilde\pi_n\in\mathscr{C}(\tilde{\mathcal{T}})$ be the set of vertices of $\tilde{\mathcal{T}}$ at graph distance n from the root. Letting $\widetilde{\mathcal{F}}_n=\sigma\left(\{\tilde J_z,\tilde\theta_z:z\in\pi_m,m\le n\}\right)$, it can be checked that the sequence $\widetilde{\mathcal{M}}_n:=\sum_{x\in\pi_n}\prod_{\rho\neq z\le x}(2\tilde q\tilde\mu)^{-1}\tilde\theta_z$ is a non-negative martingale. It therefore converges to a finite limit almost surely as $n\to+\infty$. That implies that $\kappa(\tilde{\mathcal{T}},\tilde\theta)\le 2\tilde q\tilde\mu$ almost surely (whether or not extinction occurs).

By the previous two paragraphs it suffices to show that, conditioned on $\mathcal{N}_{\mathrm{ext}}$, $\kappa(\tilde{\mathcal{T}},\tilde\theta)\ge 2\tilde q\tilde\mu$ with positive probability. Suppose we perform the following percolation process on $(\tilde{\mathcal{T}},\tilde\theta)$ with parameter $p\in[0,1]$: each edge $e=(x,y)\in\tilde{\mathcal{E}}$ is open independently with probability $p\,\tilde\theta_y$; let $\widetilde{\mathcal{O}}_p$ be the event that there is then an infinite open path originating from the root on $\tilde{\mathcal{T}}$. By a union bound, the inequality $\mathbb{P}\left[\widetilde{\mathcal{O}}_p\,\middle|\,(\tilde{\mathcal{T}},\tilde\theta)\right]\le\sum_{x\in\pi}\prod_{\rho\neq z\le x}p\,\tilde\theta_z$

holds for any $\pi \in \mathscr{C}(\tilde{\mathcal{T}})$. On the other hand, $\mathbb{E}\left[\mathbb{P}\left[\widetilde{\mathcal{O}}_p \,\middle|\, (\tilde{\mathcal{T}}, \tilde{\theta})\right]\right] > 0$ whenever $\tilde{q}\tilde{\mu}p > 1/2$. Indeed, combining the process generating $(\tilde{\mathcal{T}}, \tilde{\theta})$ with the subsequent percolation process on $(\tilde{\mathcal{T}}, \tilde{\theta})$ is equivalent to a percolation process on the original binary tree $\mathcal{T}$ with parameter $\tilde{q}\tilde{\mu}p$. The results follows. □

As discussed briefly above, we use a coupling argument. In order to describe the coupling, we first need to define some constants (not depending on θ). Recall that μ is the mean of the edge weight distribution Θ and that δ is the desired failure probability. Let $q \in (0,1)$ be close enough to 1 that

$$\left(\frac{1-q}{q}\right)^2 \leq \frac{\delta}{3}, \tag{4.3}$$

and $2q\mu > 3/2$. Let then $0 < \phi \leq 1/9$ be such that

$$2q\mu > \frac{3}{2(1-\phi)}. \tag{4.4}$$

Let H be the smallest non-negative integer such that $(1/2)^{-1+2^H} \leq \phi$. Because $\mathbb{P}_{\theta\sim\Theta}[\theta \leq \theta_*] \to 0$ as $\theta_* \to 0$, we can take $\theta_* \in (0,1)$ small enough that $\tau = \mathbb{P}[\theta \leq \theta_*]$, satisfies

$$(1-\tau)^{2^{H+1}} > q, \qquad \text{and} \qquad 1-(1-\tau)^{2^{H+2}} \leq \frac{\delta}{3}. \tag{4.5}$$

We are now ready to define the coupled process $(\tilde{\mathcal{T}}, \tilde{\theta})$. We use the following notation: Ber(ψ) is a Bernoulli random variable with success probability ψ and $Z|\mathcal{A}$ is the random variable Z conditioned on the event $\mathcal{A}$. For each z, we generate $\theta_z = (1-I_z)\theta_{z,0} + I_z\theta_{z,1}$, where $I_z \sim \text{Ber}(\mathbb{P}[\theta_z > \theta_*])$, $\theta_{z,0} \sim \theta_z|\{\theta_z \leq \theta_*\}$, and $\theta_{z,1} \sim \theta_z|\{\theta_z > \theta_*\}$ are independent. For $z \in \mathcal{V}$, let $\mathcal{D}_h(z)$ be the descendants of z lying exactly h levels below it, i.e., $\mathcal{D}_h(z) = \{w \in \mathcal{V} : z \leq w \text{ and } \gamma(z,w) = h\}$. Define $\tilde{J}_z = \prod_{w\in\mathcal{D}_{H+1}(z)} I_w$, and $\tilde{\theta}_z = \theta_{z,1}$. Note that by construction the random variables $\tilde{J}_z \sim \text{Ber}((1-\tau)^{2^{H+1}})$ and $\tilde{\theta}_z$, $z \in \mathcal{V}$, are independent. Let $\tilde{\mathcal{T}}$ be defined as in Lemma 4.1, let $\tilde{q} = (1-\tau)^{2^{H+1}}$ and let $\tilde{\mu}$ be the mean of Θ conditioned on being larger than θ_*. By (4.1), (4.3), and (4.5), $\mathbb{P}[\mathcal{N}_{\text{ext}}] \leq \frac{\delta}{3}$.

We apply Lemma 4.1 to $(\tilde{\mathcal{T}}, \tilde{\theta})$ to obtain a branching rate condition similar to (3.7).

Lemma 4.2 (Towards controlling $\mathbf{d}$ at the root)**.** *There is a deterministic (i.e., not depending on θ) $0 < \zeta < 1$ such that, on the event of non-extinction, with probability at least $1 - \delta/3$*

$$\inf_{\tilde{\pi}'\in\mathscr{C}(\tilde{\mathcal{T}})} \sum_{x\in\tilde{\pi}'} \prod_{\rho\neq z\leq x} \left\{\frac{2}{3}(1-\phi)\right\} \tilde{\theta}_z \geq \zeta. \tag{4.6}$$

Proof. On the event of non-extinction, we have almost surely that

$$\sup\left\{\kappa > 0 : \inf_{\pi\in\mathscr{C}(\tilde{\mathcal{T}})} \sum_{x\in\pi} \prod_{\rho\neq z\leq x} \kappa^{-1}\tilde{\theta}_z > 0\right\} = 2\tilde{q}\tilde{\mu} > 2q\mu > \frac{3}{2(1-\phi)},$$

where we used the fact that $\tilde{\mu} > \mu$ as well as (4.4) and (4.5). In particular, we can choose a deterministic (i.e., not depending on θ) $0 < \zeta < 1$ such that, on the event of non-extinction, with probability at least $1 - \delta/3$ the inequality in (4.6) holds. □

The purpose of the coupling is to show that the argument used in Lemma 3.3 to control the $\mathbf{u}$-values can be applied to the vertices in $\tilde{\mathcal{T}}$. This is stated in the next lemma.

For the choice of ϕ, H and θ_* above, let ε' be chosen as in the proof of Lemma 3.3, i.e., the largest positive real such that

$$\frac{3}{2}\left[\varepsilon'\left(\frac{2}{\theta_*}\right)^{H+1}\right]^2 \le \phi, \qquad \text{and} \qquad \varepsilon'\left(\frac{2}{\theta_*}\right)^{H+1} \le 0.99.$$

Lemma 4.3 (Controlling $\mathbf{u}$ on $\tilde{\mathcal{T}}$)**.** *For all $z \in \mathcal{V} \cap \tilde{\mathcal{V}}$*

$$\mathbf{d}_z^\pi \le \varepsilon' \frac{2}{\theta_*} \textit{ implies } |\mathbf{u}_z^\pi| \le 4\phi \quad \textit{and} \quad \mathbf{d}_{s(z)}^\pi \le \varepsilon' \frac{2}{\theta_*} \textit{ implies } \left|\mathbf{u}_{s(z)}^\pi\right| \le 4\phi. \tag{4.7}$$

Proof. First, observe that for any $z \in \mathcal{V} \cap \tilde{\mathcal{V}}$ it must be that $\theta_w > \theta_*$, $\forall w \in \cup_{h \le H+1} \mathcal{D}_h(z)$. Indeed, note that the unique v such that $v \le z$ and $\gamma(v,z) = h$ satisfies $v \in \tilde{\mathcal{T}}$ and $\mathcal{D}_{H+1-h}(z) \subseteq \mathcal{D}_{H+1}(v)$.

As a result, repeating the proof of Lemma 3.3, it follows that for all $z \in \mathcal{V} \cap \tilde{\mathcal{V}}$

$$\mathbf{d}_z^\pi \le \varepsilon' \frac{2}{\theta_*} \implies |\mathbf{u}_z^\pi| \le 4\phi \text{ and } \mathbf{d}_{s(z)}^\pi \le \varepsilon' \frac{2}{\theta_*} \implies \left|\mathbf{u}_{s(z)}^\pi\right| \le 4\phi,$$

where we used the fact that the parent of z is in $\tilde{\mathcal{T}}$ and therefore has all its descendants within $H+1$ level with θ-values above θ_*. The latter in turn implies that $s(z)$ itself has all its descendants within H level with θ-values above θ_*. That concludes the proof of the lemma. □

The rest of the proof of Theorem 1.5 is similar to that of Theorem 1.2, except that we restrict the argument to $\tilde{\mathcal{T}}$. Define $\varepsilon = \varepsilon'\zeta < \varepsilon'$. Let $\mathcal{H}$ be the event that $\theta_w > \theta_*$ for all w such that $\gamma(\rho, w) \le H$. By (4.5), $\mathbb{P}[\mathcal{H}] \ge 1 - \delta/3$. Condition on $\mathcal{H}$, $\mathcal{N}_{\text{ext}}$ and (4.6), which jointly occur with probability at least $1-\delta$. Assume by contradiction that $\mathbf{d}_\rho^\pi \le \varepsilon$ and let $\pi' = \{x \in \mathcal{V}^\pi \,:\, \mathbf{d}_x^\pi > \varepsilon' \text{ and } \mathbf{d}_z^\pi \le \varepsilon',\ \forall z \le x\}$. By Lemma 3.2, for all $z \in \mathcal{V} \cap \tilde{\mathcal{V}}$ such that $z \le x$ for some $x \in \pi'$, we have $\mathbf{d}_z^\pi \le \varepsilon' \frac{2}{\theta_*}$ and $\mathbf{d}_{s(z)}^\pi \le \varepsilon' \frac{2}{\theta_*}$, since the immediate parent of z (and $s(z)$) has $\mathbf{d}$-value $\le \varepsilon'$ by definition of π' and θ-values on $\tilde{\mathcal{T}}$ are above θ_*. By (4.7), we then have

$$|\mathbf{u}_z^\pi| \le 4\phi \quad \text{and} \quad \left|\mathbf{u}_{s(z)}^\pi\right| \le 4\phi, \tag{4.8}$$

for all $z \in \mathcal{V} \cap \tilde{\mathcal{V}}$ such that $z \le x$ for some $x \in \pi'$. Similarly to Theorem 1.2, by Lemma 3.1, (4.6) and (4.8),

$$\mathbf{d}_\rho^\pi \;=\; \sum_{x \in \pi'} \mathbf{d}_x^\pi \prod_{\rho \ne z \le x} \left\{\frac{4 - \mathbf{u}_{s(z)}^\pi}{6}\right\} \theta_z \ge \sum_{x \in \pi' \cap \tilde{\mathcal{V}}} \mathbf{d}_x^\pi \prod_{\rho \ne z \le x} \left\{\frac{4 - \mathbf{u}_{s(z)}^\pi}{6}\right\} \tilde{\theta}_z \ge \varepsilon'\zeta = \varepsilon,$$

where we used again that θ-values on $\tilde{\mathcal{T}}$ are above θ_* as well as the fact that all terms in the sum are non-negative by the general bounds on the $\mathbf{d}$- and $\mathbf{u}$-values. This is a contradiction. □

References

[1] Louigi Addario-Berry, Luc Devroye, Gabor Lugosi, and Vasiliki Velona, *Broadcasting on random recursive trees*, arXiv:2006.11787 [math] (2021), arXiv:2006.11787.

[2] Krishna B. Athreya and Peter E. Ney, *Branching processes*, Die Grundlehren der mathematischen Wissenschaften, Band 196, Springer-Verlag, New York-Heidelberg, 1972. MR-0373040

[3] Nayantara Bhatnagar, Allan Sly, and Prasad Tetali, *Decay of correlations for the hardcore model on the d-regular random graph*, Electron. J. Probab. **21** (2016), Paper No. 9, 42. MR-3485351

[4] P. M. Bleher, J. Ruiz, and V. A. Zagrebnov, *On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice*, J. Statist. Phys. **79** (1995), no. 1-2, 473–482. MR-1325591

[5] Christian Borgs, Jennifer T. Chayes, Elchanan Mossel, and Sébastien Roch, *The kesten-stigum reconstruction bound is tight for roughly symmetric binary channels*, 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), 21-24 October 2006, Berkeley, California, USA, Proceedings, IEEE Computer Society, 2006, pp. 518–530.

[6] Juanjuan Chai and Elizabeth Ann Housworth, *On the number of binary characters needed to recover a phylogeny using maximum parsimony*, Bull. Math. Biol. **73** (2011), no. 6, 1398–1411. MR-2799944

[7] Natalie Cusimano and Susanne S. Renner, *Ultrametric trees or phylograms for ancestral state reconstruction: Does it matter?*, Taxon **63** (2014), no. 4, 721–726, Publisher: International Association for Plant Taxonomy (IAPT).

[8] Constantinos Daskalakis, Elchanan Mossel, and Sébastien Roch, *Evolutionary trees and the Ising model on the Bethe lattice: a proof of Steel's conjecture*, Probab. Theory Related Fields **149** (2011), no. 1-2, 149–189. MR-2773028

[9] William Evans, Claire Kenyon, Yuval Peres, and Leonard J. Schulman, *Broadcasting on trees and the Ising model*, Ann. Appl. Probab. **10** (2000), no. 2, 410–433. MR-1768240

[10] Wai-Tong Fan and Sebastien Roch, *Necessary and sufficient conditions for consistent root reconstruction in Markov models on trees*, Electron. J. Probab. **23** (2018), Paper No. 47, 24. MR-3814241

[11] Wai-Tong Fan and Sebastien Roch, *Statistically consistent and computationally efficient inference of ancestral DNA sequences in the TKF91 model under dense taxon sampling*, Bull. Math. Biol. **82** (2020), no. 2, Paper No. 21, 32. MR-4056716

[12] Mareike Fischer and Bhalchandra D. Thatte, *Maximum parsimony on subsets of taxa*, J. Theoret. Biol. **260** (2009), no. 2, 290–293. MR-2973084

[13] Olivier Gascuel and Mike Steel, *Inferring ancestral sequences in taxon-rich phylogenies*, Math. Biosci. **227** (2010), no. 2, 125–135. MR-2732449

[14] Olivier Gascuel and Mike Steel, *Predicting the Ancestral Character Changes in a Tree is Typically Easier than Predicting the Root State*, Systematic Biology **63** (2014), no. 3, 421–435.

[15] Yuzhou Gu, Hajir Roozbehani, and Yury Polyanskiy, *Broadcasting on trees near criticality*, arXiv:2005.07801 [cs, math] (2020), arXiv:2005.07801.

[16] Lina Herbst and Mareike Fischer, *Ancestral sequence reconstruction with maximum parsimony*, Bull. Math. Biol. **79** (2017), no. 12, 2865–2886. MR-3736430

[17] Lina Herbst and Mareike Fischer, *On the accuracy of ancestral sequence reconstruction for ultrametric trees with parsimony*, Bull. Math. Biol. **80** (2018), no. 4, 864–879. MR-3778384

[18] Lina Herbst, Heyang Li, and Mike Steel, *Quantifying the accuracy of ancestral state prediction in a phylogenetic tree under maximum parsimony*, J. Math. Biol. **78** (2019), no. 6, 1953–1979. MR-3968985

[19] Dmitry Ioffe, *On the extremality of the disordered state for the Ising model on the Bethe lattice*, Lett. Math. Phys. **37** (1996), no. 2, 137–143. MR-1391195

[20] H. Kesten and B. P. Stigum, *Additional limit theorems for indecomposable multidimensional Galton-Watson processes*, Ann. Math. Statist. **37** (1966), 1463–1481. MR-200979

[21] Wenjian Liu, Sreenivasa Rao Jammalamadaka, and Ning Ning, *The tightness of the Kesten-Stigum reconstruction bound of symmetric model with multiple mutations*, J. Stat. Phys. **170** (2018), no. 3, 617–641. MR-3757594

[22] Russell Lyons and Yuval Peres, *Probability on trees and networks*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 42, Cambridge University Press, New York, 2016. MR-3616205

[23] Anuran Makur, Elchanan Mossel, and Yury Polyanskiy, *Broadcasting on random directed acyclic graphs*, IEEE Trans. Inform. Theory **66** (2020), no. 2, 780–812. MR-4069694

[24] Elchanan Mossel, *Recursive reconstruction on periodic trees*, Random Structures Algorithms **13** (1998), no. 1, 81–97. MR-1634344

[25] Elchanan Mossel, *Reconstruction on trees: beating the second eigenvalue*, Ann. Appl. Probab. **11** (2001), no. 1, 285–300. MR-1825467

[26] Elchanan Mossel, *Phase transitions in phylogeny*, Trans. Amer. Math. Soc. **356** (2004), no. 6, 2379–2404. MR-2048522

[27] Elchanan Mossel and Yuval Peres, *Information flow on trees*, Ann. Appl. Probab. **13** (2003), no. 3, 817–844. MR-1994038

[28] Elchanan Mossel, Sébastien Roch, and Allan Sly, *Robust estimation of latent tree graphical models: inferring hidden states with inexact parameters*, IEEE Trans. Inform. Theory **59** (2013), no. 7, 4357–4373. MR-3071334

[29] Elchanan Mossel and Mike Steel, *Majority rule has transition ratio 4 on Yule trees under a 2-state symmetric model*, Journal of Theoretical Biology **360** (2014), 315–318 (en).

[30] Yuval Peres, *Probability on trees: an introductory climb*, Lectures on probability theory and statistics (Saint-Flour, 1997), Lecture Notes in Math., vol. 1717, Springer, Berlin, 1999, pp. 193–280. MR-1746302

[31] Sebastien Roch and Allan Sly, *Phase transition in the sample complexity of likelihood-based phylogeny inference*, Probab. Theory Related Fields **169** (2017), no. 1-2, 3–62. MR-3704765

[32] Alistair Sinclair, Piyush Srivastava, Daniel Štefankovič, and Yitong Yin, *Spatial mixing and the connective constant: optimal bounds*, Probab. Theory Related Fields **168** (2017), no. 1-2, 153–197. MR-3651051

[33] Allan Sly, *Reconstruction for the Potts model*, Ann. Probab. **39** (2011), no. 4, 1365–1406. MR-2857243

[34] M. Steel, *Distributions on bicoloured evolutionary trees*, Ph.D. thesis, Massey University, 1989.

[35] Mike Steel, *Phylogeny—discrete and random processes in evolution*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 89, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2016. MR-3601108

[36] Chris Tuffley and Mike Steel, *Links between maximum likelihood and maximum parsimony under a simple model of site substitution*, Bulletin of Mathematical Biology **59** (1997), no. 3, 581–607 (en).

[37] Louxin Zhang, Jian Shen, Jialiang Yang, and Guoliang Li, *Analyzing the Fitch method for reconstructing ancestral states of ultrametric phylogenetic trees*, Bull. Math. Biol. **72** (2010), no. 7, 1760–1782. MR-2728004

Acknowledgments. We thank Mike Steel for helpful discussions and an anonymous reviewer for suggested improvements to a previous version of the manuscript. Please send all queries to `roch@wisc.edu`.

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS[1])
- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1]LOCKSS: Lots of Copies Keep Stuff Safe `http://www.lockss.org/`
[2]EJMS: Electronic Journal Management System `http://www.vtex.lt/en/ejms.html`
[3]IMS: Institute of Mathematical Statistics `http://www.imstat.org/`
[4]BS: Bernoulli Society `http://www.bernoulli-society.org/`
[5]Project Euclid: `https://projecteuclid.org/`
[6]IMS Open Access Fund: `http://www.imstat.org/publications/open.htm`