Combinatorics, Probability and Computing (2021), **30**, pp. 79–104
doi:10.1017/S0963548320000073

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

A non-increasing tree growth process for recursive trees and applications

Laura Eslava

Universidad Nacional Autonoma Mexico, Instituto de investigaciones en matematicas aplicadas y en sistemas, CDMX 04510, Mexico
Email: laura@sigma.iimas.unam.mx

(Received 18 January 2017; revised 16 December 2019; first published online 19 October 2020)

Abstract

We introduce a non-increasing tree growth process $((T_n, \sigma_n),\ n \geqslant 1)$, where T_n is a rooted labelled tree on n vertices and σ_n is a permutation of the vertex labels. The construction of (T_n, σ_n) from (T_{n-1}, σ_{n-1}) involves rewiring a random (possibly empty) subset of edges in T_{n-1} towards the newly added vertex; as a consequence $T_{n-1} \not\subset T_n$ with positive probability. The key feature of the process is that the shape of T_n has the same law as that of a random recursive tree, while the degree distribution of any given vertex is not monotone in the process.

We present two applications. First, while couplings between Kingman's coalescent and random recursive trees were known for any fixed n, this new process provides a non-standard coupling of all finite Kingman's coalescents. Second, we use the new process and the Chen–Stein method to extend the well-understood properties of degree distribution of random recursive trees to extremal-range cases. Namely, we obtain convergence rates on the number of vertices with degree at least $c \ln n$, $c \in (1, 2)$, in trees with n vertices. Further avenues of research are discussed.

2020 MSC Codes: 60C05, 05C80

1. Introduction

In a paper published in 1970, Na and Rapoport [21] introduced the problem of modelling how the structure of networks (such as sociograms, communication and acquaintance networks) emerge through time. They considered two models of random trees which they called *growing* and *static*, respectively. The 'growing' model is now known as the uniform attachment model, and each instance of it is usually called a (random) recursive tree. These are part of a broad class of tree growth models where vertices are added sequentially and connected to a random vertex in the current tree. The 'static' model was an early description of what are now referred to as coalescent processes, and the term static was motivated by the fact that in this construction the vertex set is fixed throughout the process. More precisely, it starts with n vertices, and $n - 1$ edges are added one by one (without creating cycles). These two seemingly distinct models of growth have been shown to be related for certain coalescent procedures (*e.g.* additive and Kingman's); that is, their resulting trees can also be constructed by a growth process [1, 20, 22]. In particular, finite Kingman's coalescents, which are introduced in Section 2, correspond to recursive trees; see Remark 2.3.

© The Author(s), 2020. Published by Cambridge University Press.

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

Here we present a non-increasing tree growth process $((T_n, \sigma_n),\ n \geqslant 1)$, where T_n is a rooted labelled tree on n vertices and σ_n is a permutation of the vertex labels; precise definitions will be given shortly. The three key features of this new growth process are as follows.

(1) The shape of T_n has the same distribution as that of recursive trees (vertices are labelled uniformly at random).
(2) Adding edges according to the permutation σ_n (in reverse order), recovers Kingman's coalescent.
(3) There is a positive probability that $T_{n-1} \not\subset T_n$.

Formally, we introduce the class $\mathscr{D}_n$ of *decorated trees* on n vertices and a random mapping $\mathrm{H}_n\colon \mathscr{D}_{n-1} \to \mathscr{D}_n$ such that $\mathrm{H}_n(T_{n-1}, \sigma_{n-1}) \stackrel{\text{dist}}{=} (T_n, \sigma_n)$ for all $n > 1$. Our main result, Theorem 1.1, states that recursively applying the mappings H_n to the unique element in $\mathscr{D}_1$ gives uniformly random elements on $\mathscr{D}_n$; from this the properties above are recovered. The fact that we can construct recursive trees in a non-increasing fashion is, to the best of our knowledge, a novel idea and it opens a wide range of further avenues of research. We discuss some of these in the last section.

We give the name *Robin Hood pruning* to the random mapping H_n that builds (T_n, σ_n) from (T_{n-1}, σ_{n-1}); this is the key conceptual contribution of this work and builds on the correspondence between recursive trees and Kingman's coalescent exploited in [2, 10]. It seems that such a connection had rarely been exploited, with the exception of [6, 23], where an equivalent construction was used to study union-find trees and relate them to recursive trees.

Additionally, we provide applications to high-degree vertices of recursive trees and their maximum degree. Kingman's coalescent had already been exploited by Addario-Berry and the author to describe near-maximum degrees in recursive trees [2, 10]. With the new procedure, we are able to extract finer information about extreme degree values in recursive trees. The main underlying technique is the Chen–Stein method for convergence rates to Poisson distributions. Informally, this method approximates the law of a sum W of indicator variables, by understanding how the law of such indicator variables changes when conditioning on one of them being equal to one. In our case the sum W counts the number of vertices with high degree. The perspective of the Robin Hood pruning allows us to understand how the vertex-degree distributions change when we condition on the degree of one of the vertices being large.

Before we continue to precise statements of our results, we introduce basic notation that will be used throughout the paper, as well as the standard construction of recursive trees.

1.1 Notation

For $n \in \mathbb{N}$, we write $[n] = \{1, \dots, n\}$ and $\mathscr{S}_n$ for the set of permutations on $[n]$. We denote natural logarithms by $\ln(\cdot)$ and logarithms with base 2 by $\log(\cdot)$.

Given a rooted labelled tree $T = (V(T), E(T))$, write $|T| = |V(T)|$ and call $|T|$ the *size* of T. We write $\mathscr{T}_n$ for the set of rooted trees T with vertex set $V(T) = [n]$. By convention, we direct all edges toward the root $r(T)$ and write $e = uv$ for an edge with tail u and head v. For $u \in V(T) \setminus \{r(T)\}$, we write $p_T(u)$ for the *parent* of u, that is, the unique vertex v with uv in $E(T)$. Finally, write $\mathrm{d}_T(v)$ for the number of edges directed toward v in T, and call $\mathrm{d}_T(v)$ the *degree* of v. Note that $\mathrm{d}_T(v) = \#\{u\colon p_T(u) = v\}$.

We say $T \in \mathscr{T}_n$ is *increasing* if its vertex labels increase along root-to-leaf paths; in other words, if $T \in \mathscr{T}_n$ and $p_T(v) < v$ for all $v \in [n] \setminus \{r(T)\}$ (in particular, $r(T) = 1$). We write $\mathscr{I}_n \subset \mathscr{T}_n$ for the set of increasing trees of size n. Using induction, it is easy to see that $|\mathscr{I}_n| = (n-1)!$ for all n. Next, a *tree growth process* is a sequence $(T_n,\ n \geqslant 1)$ of trees with $T_n \in \mathscr{T}_n$ for each n. The process is *increasing* if $T_n \subset T_{n+1}$ for all n; this implies that $T_n \in \mathscr{I}_n$ for all n.

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

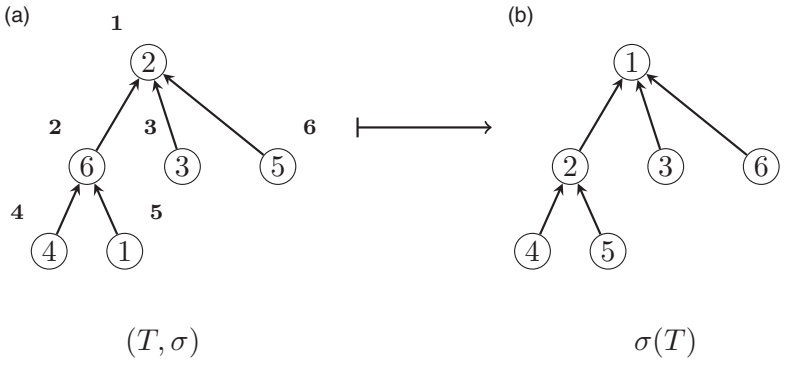


Figure 1. (a) A decorated tree $(T, \sigma) \in \mathscr{D}_6$; the permutation σ is depicted with bold numbers next to the vertices in T (so for example $\sigma(1) = 5$ and $\sigma(6) = 2$). (b) The increasing tree $\sigma(T)$.

Recursive trees on n vertices, which we denote by R_n, are usually constructed as follows. Start with R_1 as a single node with label 1. For each $1 < j \leqslant n$, R_j is obtained from R_{j-1} by adding a new vertex j and connecting it to $v_j \in [j-1]$; the choice of v_j is uniformly random and independent for each $1 < j \leqslant n$. It is readily seen that R_n is a uniformly random tree in $\mathscr{I}_n$. It follows that the process $(R_n,\ n \geqslant 1)$ is a random increasing tree growth process.

1.2 The new growth process

In what follows we extend the concept of increasing trees. If $T \in \mathscr{T}_n$ and $\sigma \in \mathscr{S}_n$, then $\sigma(T)$ is the tree $T' \in \mathscr{T}_n$ with edges $\{\sigma(u)\sigma(v)\colon uv \in E(T)\}$. In words, T' is obtained from T by relabelling the vertices of T according to the permutation σ; see Figure 1 for an example. We say that σ is a *stamp history* for T if $\sigma(T)$ is increasing. If σ is a stamp history for T, then we say that the pair (T, σ) is a *recursively decorated tree* or *decorated tree*, and that vertex v has stamp history $\sigma(v)$. We denote the set of decorated trees of size n by

$$\mathscr{D}_n = \{(T, \sigma)\colon T \in \mathscr{T}_n,\ \sigma \text{ is a stamp history of } T\}.$$

For each $n \geqslant 2$, the Robin Hood pruning $\mathrm{H}_n\colon \mathscr{D}_{n-1} \to \mathscr{D}_n$ is a random mapping that can be applied to any decorated tree. The exact definition of H_n will be given in Section 3. Broadly speaking, $\mathrm{H}_n(T, \sigma)$ is obtained from (T, σ) by pruning some subtrees of T and placing them as subtrees of a new vertex labelled n; additionally, vertex n attaches to a random vertex or becomes the root of the new tree. The stamp history in $\mathrm{H}_n(T, \sigma)$ is adjusted from σ in such a way that vertex n has a uniformly random *time stamp*. Heuristically, the random procedure follows a 'steal from the old to give to the new' scheme; that is, once the time stamp of n has been determined, vertices with an earlier time stamp have larger probability of being reattached to vertex n.

Our main theorem states that, when the input of H_n is uniformly random in $\mathscr{D}_{n-1}$, the output is uniformly random in $\mathscr{D}_n$. For the remainder of the paper, for any $n \geqslant 1$, the pair (T_n, σ_n) denotes a uniformly random element in $\mathscr{D}_n$.

Theorem 1.1 (main coupling)**.** *For each $n \geqslant 2$, the Robin Hood pruning provides a coupling between (T_{n-1}, σ_{n-1}) and (T_n, σ_n) such that $(T_n, \sigma_n) = H_n(T_{n-1}, \sigma_{n-1})$.*

This result boils down to carefully setting up the distribution of the random parameters involved in the Robin Hood pruning. Note that $|\mathscr{D}_1| = 1$, thus the Robin Hood pruning can be unambiguously applied to decorated trees starting from $\mathscr{D}_1$. Theorem 1.1 implies that the tree growth process $((T_n, \sigma_n),\ n \geqslant 1)$ given by $(T_n, \sigma_n) = \mathrm{H}_n((T_{n-1}, \sigma_{n-1}))$ is composed of uniformly random decorated trees, but it yields a non-increasing growth process on trees. This occurs since

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

the rewiring may destroy some subtrees in the previous tree; see Remark 3.4. However, the shape of T_n has the same law as that of R_n; this is proved by a straightforward bijection between $\mathscr{D}_n$ and $\mathscr{I}_n \times \mathscr{S}_n$.

Proposition 1.2. *For each $n \in \mathbb{N}$, $|\mathscr{D}_n| = n!(n-1)!$, and if $(T_n, \sigma_n) \in \mathscr{D}_n$ is chosen uniformly at random then $\sigma_n(T_n) \overset{\text{dist}}{=} R_n$ is a recursive tree of size n and σ_n is a uniformly chosen permutation in $\mathscr{S}_n$.*

Proof. By definition, if $(T, \sigma) \in \mathscr{D}_n$, then $\sigma(T) \in \mathscr{I}_n$. Let $\varphi\colon \mathscr{D}_n \to \mathscr{I}_n \times \mathscr{S}_n$ be defined such that $\varphi(T, \sigma) = (\sigma(T), \sigma)$. For an increasing tree T and $\sigma \in \mathscr{S}_n$, let $T' = \sigma^{-1}(T)$. Then $\varphi(T', \sigma) = (T, \sigma)$; it is also straightforward that φ is injective. Therefore $|\mathscr{D}_n| = |\mathscr{I}_n| \cdot |\mathscr{S}_n| = n!(n-1)!$. The result follows since bijections preserve the uniform measure on finite probability spaces. □

Growth procedures naturally couple families of trees as the size varies. For example, Proposition 1.3 below shows that (T_n, σ_n) is a representation of Kingman's coalescent on $[n]$; informally, the stamp history encodes the addition of edges in the coalescent. Precise definitions are given in Section 2, and for the moment it suffices to say that $\mathbf{C} = (F_n, \dots, F_1)$ denotes a Kingman's coalescent, where the F_j are forests with j trees.

Proposition 1.3. *Let (T_n, σ_n) be uniformly random in $\mathscr{D}_n$ and let $\mathbf{C} = (F_n, \dots, F_1)$ be a Kingman's coalescent. Denote $F_1 = \{T_{\mathbf{C}}\}$; then $T_{\mathbf{C}} \overset{\text{dist}}{=} T_n$ and the forest's evolution is given by σ_n.*

Typically there is no simple coupling of finite n-coalescent processes as n varies. The first application of Theorem 1.1 extends Proposition 1.3; namely, we establish that the Robin Hood pruning, given a Kingman's coalescent on n vertices, produces a Kingman's coalescent on $n+1$ vertices.

Corollary 1.4 (coupling for Kingman's coalescents)**.** *The tree growth process $((T_n, \sigma_n),\ n \geqslant 1)$, coupled as in Theorem 1.1, gives an explicit coupling of all finite Kingman's coalescents.*

The proof of Proposition 1.3 is given in Section 2 and is based on previous connections between recursive trees and Kingman's coalescents; see Remark 2.3.

1.3 High-degree vertices in R_n

In this section we establish a phase change on the number of *high-degree* vertices in recursive trees. Throughout the paper we use log to denote logarithms with base 2.

Phase changes occur on random structures when a class of variables undergo a transition from asymptotic normal limits to asymptotic Poisson limits. The change is marked by the mean of the variables going from infinity to bounded. In recursive trees, for example, the number of fringe trees of a given size undergoes a phase change when the size k of the trees tends to infinite and $k = o(\sqrt{n})$ no longer holds [12]; similar results are given when the fringe trees are required to satisfy any given *property* or *pattern* [5, 16]. For an integer $0 < m \leqslant n$, let us denote the number of high-degree vertices by

$$Z_m^{(n)} = \#\{v \in [n]\colon \mathrm{d}_{R_n}(v) \geqslant m\}$$

and write $\lambda_{n,m} = \mathbb{E}[Z_m^{(n)}]$. The following estimates were implicitly given in [2] and a proof can be found in Appendix A, Proposition A.3. For each $c \in (0, 2)$, there is $\gamma = \gamma(c) > 0$ such that, uniformly over $m < c \ln n$,

$$2^{-m+\log n}(1 - o(n^{-\gamma})) = \mathbb{E}[Z_m^{(n)}] \leqslant 2^{-m+\log n}. \tag{1.1}$$

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

It thus follows that the phase change occurs when $m = m(n) \approx \log n$. Using a Poisson approximation together with (1.1), we obtain the following phase change for the number of high-degree vertices.

Theorem 1.5 (phase change). *For each $c \in (1, \log e)$, there exists $c' \in (1, c)$ such that, if $c' \ln n < m < c \ln n$, then*

$$\frac{Z_m^{(n)} - \lambda_{n,m}}{\sqrt{\lambda_{n,m}}} \xrightarrow{\text{dist}} N(0,1). \tag{1.2}$$

If $m \geqslant \log n$, and (n_j) is a sequence for which $\lambda_{n_j,m} \to \lambda$, we have that

$$Z_m^{(n_j)} \xrightarrow{\text{dist}} \text{Poi}\,(\lambda). \tag{1.3}$$

Remark 1.6. Counting the high-degree vertices is equivalent to counting fringe trees (of all sizes) with a high-degree root. Although it is possible to deduce (1.2), the asymptotic normal distribution of $Z_m^{(n)}$, from [16, Corollary 1.25], the computation of both the mean and variance for the renormalization of the variables is apparently not straightforward.

Remark 1.7. The associated convergence rates for (1.3) given in Theorem 1.8 are strong and novel.

Previous results on the profile of recursive trees consider $X_m^{(n)} = \#\{v \in [n] \colon \mathrm{d}_{R_n}(v) = m\}$, for $m < n$. For fixed values of m, Janson established the joint limiting distribution of $(X_m^{(n)},\ m \geqslant 1)$ in [17]. Addario-Berry and the author addressed the case $m = m(n) \to \infty$, providing all the possible limiting distributions of $(X^{(n)}_{\lfloor \log n \rfloor + k},\ k \in \mathbb{Z})$ and establishing asymptotic normality for $X_m^{(n)}$ when $m = \log n - d$ and $d = d(n)$ slowly tends to infinity [2].

Theorem 1.5 follows from the convergence rates of the next theorem, which in turn, applies the Chen–Stein method to $Z_m^{(n)}$. By changing the perspective of recursive trees to the distribution equivalent T_n we can use the Robin Hood pruning to understand how the variables $(\mathbf{1}_{[\mathrm{d}_{T_n}(v) \geqslant m]},\ v \in [n])$ change when conditioning to $\mathrm{d}_{T_n}(u) \geqslant m$ where u is a vertex chosen uniformly at random. The details of this approach are somewhat delicate, so we defer the discussion to Section 4.

Theorem 1.8 (convergence rates). *Fix $1 < c' < c < 2$. There are constants $\alpha = \alpha(c) \in (0,1)$ and $\beta = \beta(c') > 0$ such that, uniformly for $m = m(n)$ satisfying $c' \ln n < m < c \ln n$,*

$$\mathrm{d}_{\mathrm{TV}}(Z_m^{(n)}, \text{Poi}\,(\lambda_{n,m})) \leqslant O(2^{-m+(1-\alpha)\log n}) + O(n^{-\beta}).$$

The exponent α is determined by the *almost* negative correlation between pairs of vertices in T_n (see Proposition 4.1), while the exponent β depends on an auxiliary coupling based on the Robin Hood pruning (see Proposition 4.4).

Remark 1.9. A detailed but simple track of the conditions on α (see Proposition 4.1) shows that there is a non-empty interval $\mathscr{I} = ((1-\alpha)\log e, c)$ such that, if $c' \in \mathscr{I} \cap (1,2)$, then the bounds in Theorem 1.8 do, in fact, tend to zero.

Remark 1.10. We believe that the constraint $c' > 1$ could be relaxed by obtaining uniform bounds on $\mathbb{P}(\mathrm{d}_{R_n}(i) = m)$ rather than $\mathbb{P}(\mathrm{d}_{R_n}(i) \geqslant m)$.

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

Finally, now consider the maximum degree Δ_n of a recursive tree R_n. Note that $Z_m^{(n)} > 0$ if and only if $\Delta_n \geqslant m$. Therefore, having $\mathbb{E}[Z^{(n)}_{\log n}] \approx 1$ indicates $\Delta_n \approx \log n$. In fact, Devroye and Lu showed that $\Delta_n / \log n \to 1$ a.s. [7]. The first tail bounds on Δ_n were obtained for $\mathbb{P}(\Delta_n < \lfloor \log n \rfloor + i)$ with $i \in \mathbb{Z}$ using singularity analysis of generating functions [13]. The relation between recursive trees and Kingman's coalescent provided simpler proofs of such results, extending them also to $i < 2 \ln n - \log n$ [2]. The bounds in Theorem 1.8 yield broader, tighter bounds.

Corollary 1.11 (max-degree tails)**.** *There exists $C > 0$ such that, uniformly over $0 < i = i(n) < \log e \ln \ln n - C$,*

$$\mathbb{P}(\Delta_n < \lfloor \log n \rfloor - i) = \exp\{-2^{i+\varepsilon_n}\}(1 + o(1)),$$

where $\varepsilon_n = \log n - \lfloor \log n \rfloor$.

The maximum of i.i.d. random variables is, under rather general conditions, distributed in the limit as the Gumbel (or double-exponential) distribution [15]; however, lattice distributions are excluded from this regime. Addressing the case of integer-valued variables, Anderson gives sufficient conditions under which the Gumbel distribution serves as an approximation for their maximum [3]; among those is the geometric distribution. Now, when we randomize the labels in R_n (*e.g.* using the tree T_n instead), vertex degrees become exchangeable and their limiting distributions are geometric. Although the degrees of T_n are not independent, their correlations are weak and the Gumbel-type approximation still arises for the distribution of Δ_n. Goh and Schmutz provide an alternative heuristic based on the fact that $\mathrm{d}_{R_n}(i)$, with $i \to \infty$ slowly, is asymptotically normal [13].

Outline. The paper is divided into two parts. Sections 2 and 3 establish that the Robin Hood pruning is well-defined and provides a new coupling of recursive trees (and Kingman's coalescent), while Sections 4 and 5 apply the Chen–Stein method to such a representation of recursive trees.

We discuss further the connection between recursive trees, Kingman's coalescents and other tree models in Section 2. The precise definition of the Robin Hood pruning H_n, together with the proof of Theorem 1.1, is given in Section 3. The results on high-degree vertices of recursive trees use the Chen–Stein method and the Robin Hood pruning in a non-trivial way; an overview of how we use the Chen–Stein method is given in Section 4. Assuming the existence of an auxiliary coupling (Proposition 4.4), we complete the proofs concerning high-degree vertices in Section 4.1. The auxiliary coupling, based on the Robin Hood pruning, is presented in Section 5. Finally, Section 6 discusses further avenues of research.

2. Kingman's coalescents and recursive trees: distinct representations

Discrete coalescents are processes on partitions of $[n]$ that can be represented by different tree structures. On can encode the coalescent using an n-chain: a sequence of forests where at all times there are n vertices (or *elements*), and $n - 1$ edges are added one by one until a tree is formed. However, there is a more traditional construction using binary search trees (BSTs), where internal nodes correspond to merges and only external nodes correspond to *elements* of the coalescent. In the next section we introduce the representation used in this paper and prove Proposition 1.3. Following that, we discuss the well-known bijection between BSTs and recursive trees and the difference between the two coalescent representations. In addition, we explain the difference between the Yule–Harding model of phylogenetic trees and its uniform model, and highlight the importance of clarifying both the rules applied to the mergings in coalescent processes and their representation as trees.

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

2.1 Recursive trees perspective

Na and Rapoport [21] loosely described this process as the construction of static trees with n vertices:

> Initially, single elements move about at random. Each collision forms a couple. A collision of a couple with a single element forms a triple, a collision of an s-tuple with a t-tuple forms an $(s+t)$-tuple, and so on. At each collision a link is established between an element of one X-tuple and an element of another, the links being rigid so that the elements of the same k-tuple cannot collide. The process goes on until the entire set of n elements has been joined into an n-tuple.

By changing the rule on how to link the elements on the tuples, we obtain distinct coalescent distributions. In the description in [21] there are no restrictions on the elements allowed to be linked during the coalescent.[1] In fact, the discrete multiplicative coalescent arises when any possible link is chosen uniformly at random. It is associated with Kruskal's algorithm for the minimum weighted spanning tree problem [1].

Kingman's coalescent is characterized by the property that the merging probability of any pair of components is independent of the components' sizes. The representation used in this paper uses the fact that at each time the representative of each 'tuple' is the current root of the tree and it is closely related to the 'union-find' algorithm used in computer science (see *e.g.* [24]). We present a formal description following the notation of [2].

A (deterministic) forest f is a set of trees with pairwise disjoint vertex sets. Let $V(f)$ and $E(f)$, respectively, denote the union of the vertex and edge sets of the trees in f. For $n \geqslant 1$, an n-chain is a sequence $C = (f_n, \ldots, f_1)$ of elements of $\mathscr{F}_n = \{f\colon V(f) = [n]\}$ such that, for $1 < i \leqslant n$, f_{i-1} is obtained from f_i by adding a directed edge between the roots of some pair of trees in f_i. In particular, f_n consists of n one-vertex trees and f_1 consists of a single tree on n vertices denoted by t_C. For an example see Figure 2.

Next we introduce the necessary notation to define Kingman's coalescent using n-chains. For an n-chain $(f_n, \ldots, f_1)$ and $1 \leqslant i \leqslant n$, list the trees of f_i in increasing order of their smallest-labelled vertex as $t_1^{(i)}, \ldots, t_i^{(i)}$. Independently for each $1 < i \leqslant n$, let $\{a_i, b_i\} \subset \{\{a, b\}\colon 1 \leqslant a < b \leqslant i\}$ be uniformly chosen at random; in addition, let ξ_i be independent Bernoulli random variables with mean $1/2$.

Definition 2.1. Kingman's n-coalescent is defined as $\mathbf{C} = (F_n, \ldots, F_1)$, constructed as follows. For $1 < i \leqslant n$, F_{i-1} is obtained from F_i by adding an edge between $r(T_{a_i}^{(i)})$ and $r(T_{b_i}^{(i)})$. If $\xi_i = 1$ then direct the edge towards $r(T_{a_i}^{(i)})$; otherwise direct it towards $r(T_{b_i}^{(i)})$. The forest F_{i-1} consists of the new tree and the remaining $i-2$ unaltered trees from F_i.

In other words, if $\mathbf{C} = (F_n, \ldots, F_1)$ is a Kingman's coalescent, then each of the trees of F_i correspond to a set of coalesced elements after $n - i + 1$ steps of the process. At each step, two sets (represented by their roots) coalesce and a new representative is chosen uniformly at random.

To link n-chains with decorated trees, we first define a natural edge labelling that tracks the number of *trees left* in the forest when a given edge comes along. Fix $C = (f_n, \ldots, f_1)$, for each $e \in E(t_C)$, let

$$\rho_C(e) = \max\{i \in [n-1]\colon e \in E(f_i)\}.$$

We next define a vertex labelling $\sigma_C\colon V(t_C) \to [n]$. Let $\sigma_C(r(t_C)) = 1$, and for each $uv \in E(t_C)$, let

$$\sigma_C(u) = \rho_C(uv) + 1.$$

[1] Unfortunately, it was incorrectly presumed in [21] that 'static' trees build uniformly random unrooted labelled trees.

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

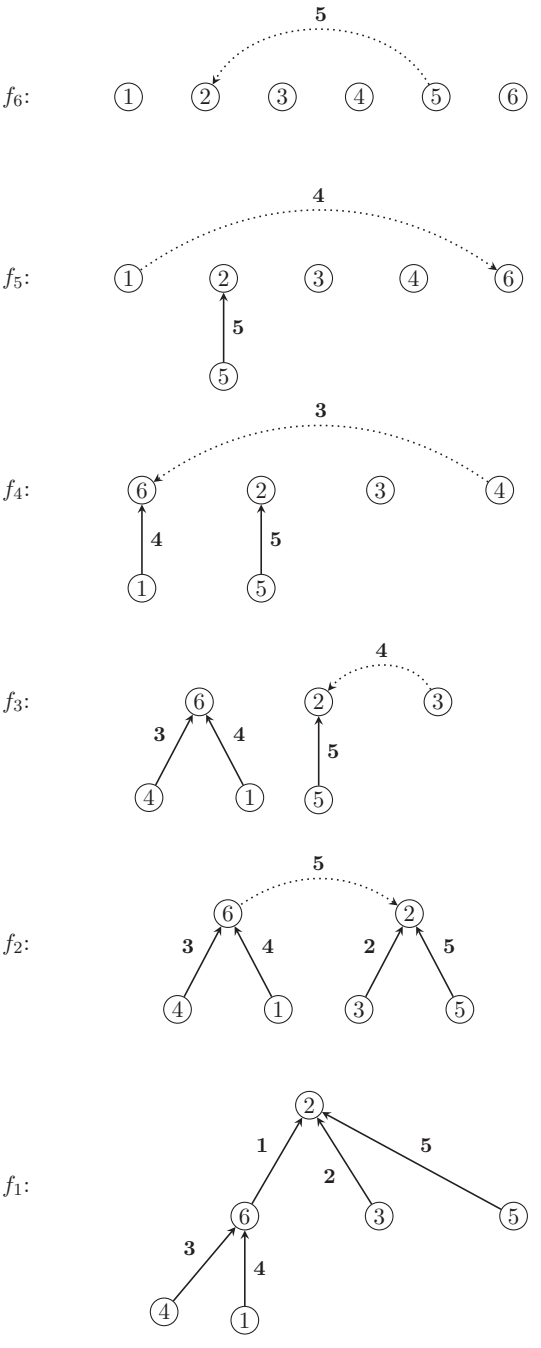


Figure 2. An example of an n-chain with $n = 6$. The edge labelling ρ_n is represented by numbers in bold.

The following proposition shows that the pair $(t_C, \sigma_C) \in \mathscr{D}_n$ contains all the information for recovering the original n-chain C; in other words, if $\mathscr{C}_n$ denotes the set of n-chains, then $\mathscr{D}_n$ and $\mathscr{C}_n$ are in bijection.

Proposition 2.2. *Let $\Upsilon : \mathscr{C}_n \to \mathscr{D}_n$ be defined as follows. For an n-chain $C = (f_n, \ldots, f_1)$, let $\Upsilon(C) = (t_C, \sigma_C)$. Then Υ is a bijection.*

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

Proof. First, we show that $\mathscr{C}_n$ and $\mathscr{D}_n$ have the same cardinality. To count the number of n-chains, consider constructing $(f_n, \dots, f_1)$ by deciding which edge to add from f_k to f_{k-1}. Since there are k trees in f_k, when we have chosen $(f_n, \dots, f_k)$, there are $k(k-1)$ possible directed edges to add. Therefore $|\mathscr{C}_n| = n!(n-1)!$.

Next, let $C = (f_n, \dots, f_1)$ be an n-chain. For each $1 \leqslant i < n$, the new edge in f_i joins the roots of two trees in f_{i+1} and is directed towards the root of the resulting tree. Thus the labels $\{\rho_C(e),\ e \in E(t_C)\}$ decrease along all paths in t_C towards the root $r(t_C)$. Consequently the labels $\{\sigma_C(v),\ v \in [n]\}$ form, indeed, a stamp history of t_C. It follows that Υ is well-defined.

Finally, let $C = (f_n, \dots, f_1)$, $C' = (f'_n, \dots, f'_1)$ be distinct n-chains and write $k = \min\{i\colon f_i \neq f'_i\}$. If $k = 1$ then $t_C \neq t_{C'}$, and clearly $\Upsilon(C) \neq \Upsilon(C')$. Otherwise $t_C = t_{C'}, f_{k-1} = f'_{k-1}$ and the (unique) edges $e \in E(f_{k-1}) \setminus E(f_k)$ and $e' \in E(f'_{k-1}) \setminus E(f'_k)$ are distinct. It follows that $e = uv \in f'_k$ and so $\sigma_C(u) = k > \sigma_{C'}(u)$. This shows that Υ is injective, and so Υ is a bijection between $\mathscr{C}_n$ and $\mathscr{D}_n$. □

Using the bijection of Proposition 2.2, it follows that Proposition 1.3 boils down to showing that $\mathbf{C}$ is uniformly random in $\mathscr{C}_n$.

Proof of Proposition 1.3. Let $\mathbf{C} = (F_n, \dots, F_1)$ be a Kingman's coalescent. For any fixed n-chain $(f_n, \dots, f_1) \in \mathscr{C}_n$,

$$\mathbb{P}((F_n, \dots, F_1) = (f_n, \dots, f_1)) = \prod_{k=1}^{n-1} \mathbb{P}(F_k = f_k \mid (F_n, \dots, F_{k+1}) = (f_n, \dots, f_{k+1})).$$

Among the $k(k+1)$ possible oriented edges connecting roots of f_{k+1}, exactly one of them can be added to f_{k+1} to yield f_k. Thus, regardless of the sequence $(f_n, \dots, f_1)$,

$$\mathbb{P}((F_n, \dots, F_1) = (f_n, \dots, f_1)) = [(n-1)!n!]^{-1}.$$

Recall $F_1 = \{T_{\mathbf{C}}\}$. By Proposition 2.2, $(T_{\mathbf{C}}, \sigma_{\mathbf{C}}) \in \mathscr{D}_n$ and it has a uniform distribution, since the bijection preserves the uniform measure of $\mathbf{C}$. Finally, by Proposition 2.2, it follows that $T_{\mathbf{C}} \overset{\text{dist}}{=} T_n$. The evolution of the forests is given by $\rho_{\mathbf{C}}$; equivalently by $\sigma_{\mathbf{C}}$. □

Remark 2.3. It follows from Propositions 1.2 and 1.3 that, for any fixed n and up to relabelling of vertices, Kingman's coalescent corresponds to recursive trees. See also [1, 2] for direct proofs of this fact.

2.2 The binary search tree connection

Binary search trees have been related to both recursive trees and phylogenetic trees. In this section we briefly discuss these connections and compare them with Kingman's coalescent. Let $\mathscr{B}_n$ be the set containing all plane, rooted, unlabelled binary trees with n external nodes. Trees in $\mathscr{B}_n$ distinguish between left and right subtrees of any given internal vertex. It can be shown that the sizes $|\mathscr{B}_n| = 3 \cdot 5 \cdots (2n-3)$ are given by the Catalan numbers.

Binary search trees are the tree representation of the sorting algorithm Quicksort. Simply described, for each $n \geqslant 1$, the Quicksort algorithm takes a permutation $\sigma \in [n]$ and constructs (step by step) a binary tree with internal vertex labels on $[n]$ as follows. The root is $\sigma(1)$ and vertices $\sigma(2), \dots, \sigma(n)$ are added sequentially so that the final tree satisfies the following property: for any internal node j, all nodes on its left subtree are smaller than j and all nodes on its right tree are larger than j. It follows that, given a shape of a binary tree $B \in \mathscr{B}_{n+1}$, there is exactly one way to label internal vertices.

There are $n!$ distinct permutations as input for the quicksort algorithm. Devroye introduced the representation of the binary search tree (process) using time stamps which record the entire insertion process. Using this representation, the rotation correspondence maps (one-to-one)

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

recursive trees (on $n-1$ vertices) and binary search trees (with n external vertices). For a thorough description of the correspondence, see [16, Section 2, Figures 1–2].

On the other hand, phylogenetic trees on n species are represented by elements in $\mathscr{B}_n$. In this case species are assumed to have a common ancestor (the root), internal nodes are also ancestors, and the time elapsed between differentiation of species, *i.e.* the length of the branches, is omitted. Two common distributions on phylogenetic trees are the uniform one, known as the Catalan model, and the Yule–Harding model. The latter is a process that constructs trees starting from the root, by branching a uniformly random external node and replacing it with a cherry (an internal node with two external nodes). Clearly, this construction corresponds one-to-one to the quicksort algorithm. We remark that the Yule–Harding process does not yield uniform phylogenetic trees (as the BST is not uniform in $\mathscr{B}_n$). For further comparison between the two models, see *e.g.* [4, Section 3].

It has been presumed that Kingman's coalescent is the bottom-up construction of the Yule–Harding model; see *e.g.* [5]. However, such correspondence has to be done carefully, as merges in principle are not bound to satisfy planarity constraints. As we can see from the bijections and $n!$-to-1 mappings from Propositions 2.2 and 1.2, there should be a correspondence between BSTs with time stamps and Kingman's coalescent.

The construction of Kingman's coalescent as a binary tree in $\mathscr{B}_n$ with time stamps is as follows. Using the same random variables used in Definition 2.1, add an internal node connecting the two roots of the merging trees, while the coin flip indicates which of the trees is the left child of the new internal vertex; the time stamps indicate the (reversed) order of addition of internal nodes. Note that in this construction, the symmetry breaking of the coin flip is still necessary.

Conversely, we describe how to interpret time stamps of a BST as the merging history of a Kingman's coalescent. To do so, we have to label external nodes uniformly at random (so that there are a total of $n!(n-1)!$ different processes). Now the role of internal vertices is as follows. At step $k \in [n-1]$, the two sets of external vertices in each of the subtrees of the vertex with time stamp $n-k$ are the subsets to be merged in the coalescent.

Kingman's coalescent has uniform distribution when considering all possible merging histories with elements labelled exchangeably. However, considering only the final tree (in either $\mathscr{T}_n$ or $\mathscr{B}_n$) yields a non-uniform distribution: there are $n!(n-1)!$ total ways to merge the subtrees (if we use the symmetry breaking at each merging), but there are only $|\mathscr{T}_n| = n^{n-1}$ and $|\mathscr{B}_n| = 3 \cdot 5 \cdots (2n-3)$ different rooted, labelled trees and phylogenetic trees, respectively.

3. The Robin Hood pruning

The *Robin Hood pruning* $\mathrm{H}_n \colon \mathscr{D}_{n-1} \to \mathscr{D}_n$ is a random procedure based on randomizing the parameters of a deterministic mapping $\mathrm{h}_n \colon \mathscr{D}_{n-1} \times \mathscr{P}_n \to \mathscr{D}_n$ where the set $\mathscr{P}_n$ defines all possible ways to prune a decorated tree on $n-1$ vertices. The distribution on $\mathscr{P}_n$ is tailored so that the Robin Hood pruning in fact yields a coupling of $((T_n, \sigma_n); n \geqslant 1)$.

First we introduce the necessary notation to define $\mathscr{P}_n$, the *deterministic pruning* h_n and verify that, indeed, the mapping h_n is well-defined. We then continue to define the distribution on $\mathscr{P}_n$ used in defining the Robin Hood pruning (Definition 3.5). The proof of Theorem 1.1 requires us to characterize the properties of the uniform distribution in decorated trees. For the characterization in Lemma 3.6 and the proof of Theorem 1.1, we emphasize the difference between deterministic elements (T, π) of $\mathscr{D}_n$ and random elements $(\boldsymbol{T}, \boldsymbol{\sigma})$ using bold notation; the distribution of $(\boldsymbol{T}, \boldsymbol{\sigma})$ is not given *a priori*.

3.1 A deterministic process

Informally, we define all possible ways to prune a decorated tree on $n-1$ vertices using three parameters $(k, l, x) \in \mathscr{P}_n$: the *time stamp* k of the new vertex, its *point of attachment* l, and the

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

vertices to be rewired encoded by time stamp in the sequence $x = (x_1, \dots, x_{n-1})$. Once a stamp history is given to a tree T, $\mathscr{V}_n$ contains the vertices to be pruned and rewired towards the new vertex n.

We now proceed to precise definitions. Let $n \geqslant 2$ and set

$$\mathscr{P}_n = \{(k, l, x)\colon 1 \leqslant l < k \leqslant n,\ x \in \{0,1\}^{n-1}\} \cup \{(1, 0, x)\colon x \in \{0,1\}^{n-1},\ x_1 = 1\}.$$

Additionally, for $(k, l, x) \in \mathscr{P}_n$ and a permutation $\sigma \in \mathscr{S}_{n-1}$, let

$$\mathscr{V}_n(k, l, x, \sigma) = \mathscr{V}_n(k, x, \sigma) = \{v \in [n-1]\colon x_{\sigma(v)} = 1,\ \sigma(v) \geqslant k\}.$$

Remark 3.1. The definition of $\mathscr{P}_n$ is such that $\sigma^{-1}(1) \in \mathscr{V}_n$ if and only if $k = 1$.

The following definition of a deterministic pruning is illustrated in Figure 3.

Definition 3.2. Fix $n \geqslant 2$, $(T, \sigma) \in \mathscr{D}_{n-1}$ and $(k, l, x) \in \mathscr{P}_n$. We define (T', σ') and set

$$\mathrm{h}_n((T, \sigma), (k, l, x)) = (T', \sigma')$$

as follows. First, let $\mathscr{V} = \mathscr{V}_n(k, x, \sigma)$ and construct T' from T. For each $v \in \mathscr{V} \setminus \{r(T)\}$, replace the edge $vp_T(v)$ with an edge connecting v to a new vertex labelled n. Now if $k = 1$ then attach $r(T)$ to n; otherwise attach vertex n to $\sigma^{-1}(l)$. In other words the edges of T' are given by

$$E(T') = \begin{cases} (E(T) \cup \{vn; v \in \mathscr{V}\}) \setminus \{vp_T(v); v \in \mathscr{V}\} & \text{if } k = 1, \\ \{n\sigma^{-1}(l)\} \cup (E(T) \cup \{vn; v \in \mathscr{V}\}) \setminus \{vp_T(v); v \in \mathscr{V}\} & \text{if } k > 1. \end{cases}$$

Second, let $\sigma'\colon [n] \to [n]$ be defined by $\sigma'(n) = k$ and for $v < n$,

$$\sigma'(v) = \sigma(v) + \mathbf{1}_{[\sigma(v) \geqslant k]}.$$

Lemma 3.3. *For any $n \geqslant 2$, $h_n\colon \mathscr{D}_{n-1} \times \mathscr{P}_n \to \mathscr{D}_n$ is well-defined. That is, for any $(T, \sigma) \in \mathscr{D}_{n-1}$ and $(k, l, x) \in \mathscr{C}_n$,*

$$h_n((T, \sigma), (k, l, x)) \in \mathscr{D}_n.$$

Proof. Write $\mathrm{h}_n((T, \sigma), (k, l, x)) = (T', \sigma')$. When $k = 1$, it is clear that T' is a tree. When $k > 1$, let $w = \sigma^{-1}(l)$ be the parent of n in T' and let $(w = v_1, \dots, v_j = r(T))$ be the path from w to the root of T. Since σ is a stamp history of T, $l = \sigma(v_1) > \sigma(v_2) > \cdots > \sigma(v_j) = 1$; moreover, $l < k$. It follows that $v_i \notin \mathscr{V}(k, l, x, \sigma)$ for all $i \in [j]$, and consequently no edges in the path from n to the root in T' close a cycle by connecting to n.

Now we show that σ' is a stamp history for T'. It is clear that σ' is a permutation of $[n]$, so it suffices to prove that $\sigma'(v) > \sigma'(p_{T'}(v))$ for all $v \in V(T) \setminus \{r(T')\}$. First, for vertices v with $p_{T'}(v) = n$ we have $\sigma(v) \geqslant k$, and consequently $\sigma'(v) = \sigma(v) + 1 > k = \sigma'(n)$.

Second, consider $v, w < n$ with $p_{T'}(v) = w$. It follows that $vw \in E(T)$ and thus $\sigma(v) > \sigma(w)$. Consequently $\mathbf{1}_{[\sigma(v) \geqslant k]} \geqslant \mathbf{1}_{[\sigma(w) \geqslant k]}$ and so $\sigma'(v) > \sigma'(w)$. The last case occurs when $k > 1$ and $p_{T'}(n) = w = \sigma^{-1}(l)$. We then have $\sigma'(n) = k > l = \sigma(w) = \sigma'(w)$. □

Remark 3.4. Whenever $(k, l, x) \in \mathscr{P}_n$ has $x_j = 1$ for some $j \geqslant k$, setting

$$(T', \sigma') = \mathrm{h}_n((T, \sigma), (k, l, x)) \quad \text{and} \quad v = \sigma^{-1}(j)$$

yields $n = p_{T'}(v) \neq p_T(v) \in [n-1]$. This implies that $T \not\subset T'$.

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

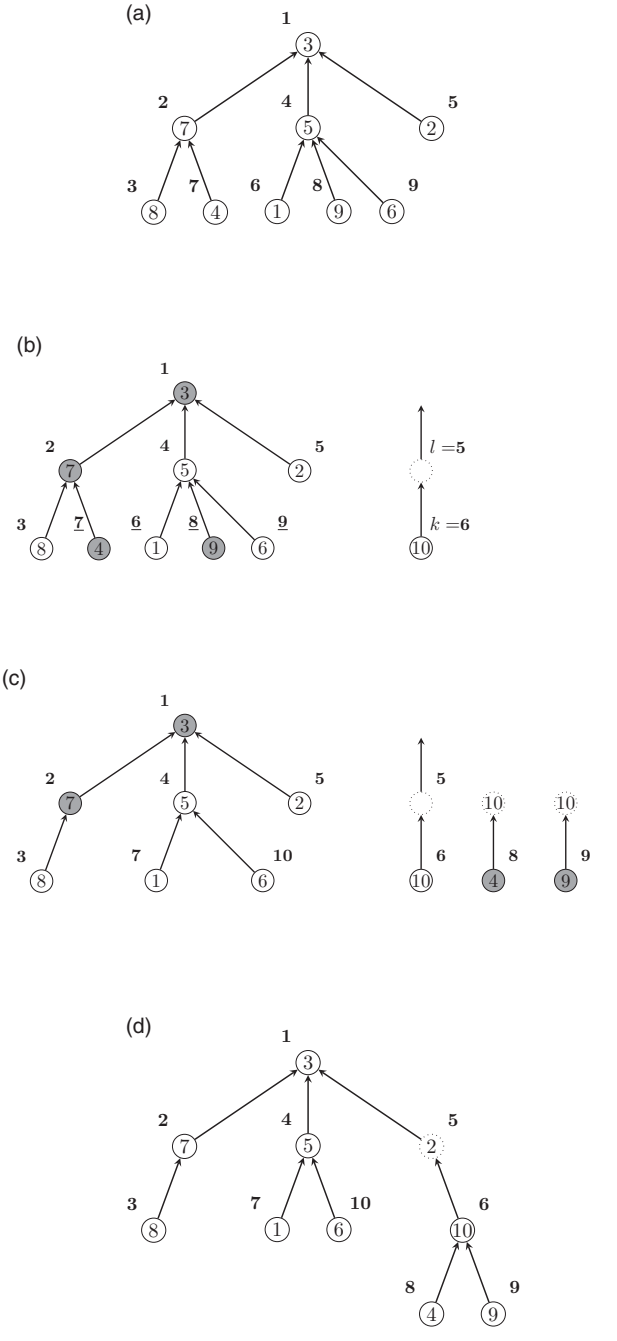


Figure 3. An example of the Robin Hood pruning for (T, σ) with $k = 6$, $l = 5$ and $x_1 = x_2 = x_7 = x_8 = 1$; all other $x_i = 0$. (a) A tree (T, σ) in $\mathscr{D}_9$. The permutation σ is depicted with bold numbers. (b) Vertices in grey satisfy $X_{\sigma(v)} = 1$ and underlined numbers indicate time stamps $\sigma(i) \geqslant k$. (c) Nodes i with $\sigma(i) \geqslant k$ and $x_i = 1$ have been pruned and time stamps have been adjusted. (d) The resulting tree $\mathrm{h}_{10}((T, \sigma), (k, l, x)) \in \mathscr{D}_{10}$.

3.2 The random process

The H_n-set is a sample of $\mathscr{P}_n$ according to the following distribution.

Definition 3.5. Fix $n \geqslant 1$. Let $K \stackrel{\text{dist}}{=} \mathrm{Unif}(1, 2, \ldots, n)$; if $K = 1$ let $L = 0$, and if $K > 1$ let $L = \mathrm{Unif}(1, 2, \ldots K - 1)$. Independently, let $X = (X_1, \ldots, X_{n-1})$, where $X_i \stackrel{\text{dist}}{=} \mathrm{Bernoulli}(1/i)$

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

are independent variables. An H_n-set is a triple of random variables with the same law as $(K, L, X) \in \mathscr{P}_n$.

We are ready to define the Robin Hood pruning. For each $n \geqslant 2$, let $(K, L, X) \in \mathscr{P}_n$ be an H_n-set and define

$$\mathrm{H}_n(T, \sigma) = \mathrm{h}_n((T, \sigma), (K, L, X)).$$

The law of $\mathrm{H}_n(T, \sigma)$ depends on the initial input (T, σ); however, the distribution of the H_n-set is tailored so that $\mathrm{H}_n(T_{n-1}, \sigma_{n-1})$ preserves the uniform measure on decorated trees. In order to prove Theorem 1.1, we start with a characterization of (T_n, σ_n).

Lemma 3.6. *Let $n \geqslant 1$ be an integer. A random decorated tree $(\boldsymbol{T}, \boldsymbol{\sigma}) \in \mathscr{D}_n$ is uniformly random if and only if the following properties are satisfied.*

(i) The permutation $\boldsymbol{\sigma}$ is uniformly random on $\mathscr{S}_n$.
(ii) Conditional on $\boldsymbol{\sigma}$, the vertices $(p_{\boldsymbol{\sigma}(\boldsymbol{T})}(\boldsymbol{\sigma}^{-1}(v)),\ v \in V(\boldsymbol{T}) \setminus \{r(\boldsymbol{T})\})$ are independent.
(iii) For all vertices $v, w \in [n]$ and indices $i, j \in [n]$,

$$\mathbb{P}(p_{\boldsymbol{T}}(v) = w,\ \boldsymbol{\sigma}(v) = j,\ \boldsymbol{\sigma}(w) = i) = \frac{1}{n(n-1)(j-1)} \mathbf{1}_{[j>i]}. \tag{3.1}$$

Proof. Let $(\boldsymbol{T}, \boldsymbol{\sigma}) = (T_n, \sigma_n)$ be uniformly random on $\mathscr{D}_n$. Condition (i) follows directly from Proposition 1.2, which states that $\boldsymbol{\sigma}$ is a uniformly random permutation and $\boldsymbol{\sigma}(\boldsymbol{T})$ has the law of a recursive tree R_n. In addition, Proposition 1.2 implies

$$(p_{\boldsymbol{\sigma}(\boldsymbol{T})}(\boldsymbol{\sigma}^{-1}(v)),\ v \in V(\boldsymbol{T}) \setminus \{r(\boldsymbol{T})\}) \overset{\text{dist}}{=} \{p_{R_n}(j),\ 1 < j \leqslant n\},$$

from which conditions (ii) and (iii) immediately follow: parents in recursive trees are chosen independently for each of the vertices, and for all $v, w, i, j \in [n]$,

$$\begin{aligned}
\mathbb{P}(p_{\boldsymbol{T}}(v) = w,\ \boldsymbol{\sigma}(v) = j,\ \boldsymbol{\sigma}(w) = i) &= \frac{1}{n(n-1)} \mathbb{P}(p_{\boldsymbol{T}}(v) = w \mid \boldsymbol{\sigma}(v) = j,\ \boldsymbol{\sigma}(w) = i) \\
&= \frac{1}{n(n-1)} \mathbb{P}(p_{\boldsymbol{\sigma}(\boldsymbol{T})}(j) = i) \\
&= \frac{1}{n(n-1)(j-1)} \mathbf{1}_{[j>i]}.
\end{aligned}$$

Now consider a random decorated tree $(\boldsymbol{T}, \boldsymbol{\sigma}) \in \mathscr{D}_n$ satisfying conditions (i)–(iii). Fix a decorated tree $(T, \pi) \in \mathscr{D}_n$, and for $v \in V(T) \setminus \{r(T)\}$, let $w_v = p_T(v)$. Condition (ii) on the conditional independence of parents gives, for $v \neq r(T)$,

$$\begin{aligned}
\mathbb{P}(p_{\boldsymbol{T}}(v) = w_v \mid \boldsymbol{\sigma} = \pi) &= \mathbb{P}(p_{\boldsymbol{\sigma}(\boldsymbol{T})}(\boldsymbol{\sigma}(v)) = \pi(w_v) \mid \boldsymbol{\sigma}(v) = \pi(v),\ \boldsymbol{\sigma}(w_v) = \pi(w_v)) \\
&= \frac{\mathbb{P}(p_{\boldsymbol{T}}(v) = w_v,\ \boldsymbol{\sigma}(v) = \pi(v),\ \boldsymbol{\sigma}(w_v) = \pi(w_v))}{\mathbb{P}(\boldsymbol{\sigma}(v) = \pi(v),\ \boldsymbol{\sigma}(w_v) = \pi(w_v))}.
\end{aligned}$$

Using the fact that π is a stamp history for T, so $\pi(v) > \pi(w_v)$, and that $\boldsymbol{\sigma}$ is uniformly random, it follows from (3.1) that

$$\mathbb{P}(p_{\boldsymbol{T}}(v) = w_v \mid \boldsymbol{\sigma} = \pi) = \frac{1}{\pi(v) - 1}. \tag{3.2}$$

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

Any increasing tree $T' \in \mathscr{I}_n$ is determined by the set of parents $\{p_{T'}(v),\ 1 < v \leqslant n\}$. Using the fact that $\pi(T) \in \mathscr{I}_n$ and the conditional independence from condition (ii), we get

$$\begin{aligned}\mathbb{P}(\boldsymbol{\sigma}(\boldsymbol{T}) = \pi(T) \mid \boldsymbol{\sigma} = \pi) = &\quad \mathbb{P}(p_{\boldsymbol{\sigma}(\boldsymbol{T})}(j) = p_{\pi(T)}(j),\ 1 < j \leqslant n \mid \boldsymbol{\sigma} = \pi)\\ = &\quad \mathbb{P}(p_{\boldsymbol{T}}(v) = p_T(v),\ v \in V(t) \setminus \{r(T)\} \mid \boldsymbol{\sigma} = \pi)\\ = &\prod_{v \in V(T) \setminus r(T)} \mathbb{P}(p_{\boldsymbol{T}}(v) = p_T(v) \mid \boldsymbol{\sigma} = \pi)\\ = &\quad [(n-1)!]^{-1},\end{aligned}$$

where the last equality holds by (3.2) and the fact that $\{\pi(v),\ v \in V(T) \setminus \{r(T)\}\} = \{2, \dots, n\}$. Finally, using the equation above and the fact that $\boldsymbol{\sigma}$ is uniformly random, we have

$$\begin{aligned}\mathbb{P}((\boldsymbol{T}, \boldsymbol{\sigma}) = (T, \pi)) &= \mathbb{P}(\boldsymbol{\sigma}(\boldsymbol{T}) = \pi(T) \mid \boldsymbol{\sigma} = \pi)\, \mathbb{P}(\boldsymbol{\sigma} = \pi)\\ &= \frac{1}{n!}\mathbb{P}(\boldsymbol{\sigma}(\boldsymbol{T}) = \pi(T) \mid \boldsymbol{\sigma} = \pi)\\ &= [n!(n-1)!]^{-1}.\end{aligned}$$

This holds regardless of the choice of (T, π), so $(\boldsymbol{T}, \boldsymbol{\sigma})$ is uniformly random in $\mathscr{D}_n$. □

We are now ready to prove Theorem 1.1.

Proof of Theorem 1.1. Let $(T_{n-1}, \sigma_{n-1}) \in \mathscr{D}_{n-1}$ be a uniformly random decorated tree. Let (K, L, X) be an H_n-set and let $(\boldsymbol{T}, \boldsymbol{\sigma}) = \mathrm{h}((T_{n-1}, \sigma_{n-1}), (K, L, X))$. It suffices to show that $(\boldsymbol{T}, \boldsymbol{\sigma})$ satisfies the properties in Lemma 3.6.

First, condition (i) follows from the construction of $\boldsymbol{\sigma}$ and the distributions of both K and σ_{n-1}. Second, conditioning on $\boldsymbol{\sigma}$, which is equivalent to conditioning on both σ_{n-1} and K, we get

$$\begin{aligned}\{p_{\boldsymbol{T}}(v),\ v \in V(\boldsymbol{T}) \setminus \{r(\boldsymbol{T})\}\} &= \{p_{\boldsymbol{T}}(v),\ 1 < \boldsymbol{\sigma}(v) < \boldsymbol{\sigma}(n)\}\\ &\quad \cup \{p_{\boldsymbol{T}}(v),\ \boldsymbol{\sigma}(n) \leqslant \boldsymbol{\sigma}(v) \leqslant n\}\\ &= \{p_{T_{n-1}}(v),\ v \in 1 < \sigma_{n-1}(v) < K\}\\ &\quad \cup \{p_{\boldsymbol{T}}(v),\ (2 \vee K) \leqslant \boldsymbol{\sigma}(v) \leqslant n\},\end{aligned}$$

where the last two sets are conditionally independent given $\boldsymbol{\sigma}$. Now, since (T_{n-1}, σ_{n-1}) is uniformly random in $\mathscr{D}_{n-1}$, the parents $\{p_{T_{n-1}}(v),\ v \in 1 < \sigma_{n-1}(v) < K\}$ are independent, conditionally given σ_{n-1} (and thus also conditionally given $\boldsymbol{\sigma}$). On the other hand, for v with $\boldsymbol{\sigma}(v) \geqslant K$,

$$p_{\boldsymbol{T}}(v) = \begin{cases} n & \text{if } X_{\boldsymbol{\sigma}(v)-1} = 1,\\ p_{T_{n-1}}(v) & \text{if } X_{\boldsymbol{\sigma}(v)} = 0,\\ \boldsymbol{\sigma}^{-1}(L) & \text{if } \boldsymbol{\sigma}(v) = K.\end{cases}$$

Note that $p_{\boldsymbol{T}}(v)$ is determined independently from other vertices, thus $\{p_{\boldsymbol{T}}(v),\ K \leqslant \boldsymbol{\sigma}(v) \leqslant n\}$ are also independent, conditionally given $\boldsymbol{\sigma}$. This implies that condition (ii) is satisfied.

Third, fix $1 \leqslant i < j \leqslant n$ and fix distinct $v, w \in [n]$. We consider three cases, namely $v = n$, $w = n$ and $\{v, w\} \subset [n-1]$. Let

$$\begin{aligned}A_1 &= \{p_{\boldsymbol{T}}(n) = w,\ \boldsymbol{\sigma}(n) = j,\ \boldsymbol{\sigma}(w) = i\},\\ A_2 &= \{p_{\boldsymbol{T}}(v) = n,\ \boldsymbol{\sigma}(v) = j,\ \boldsymbol{\sigma}(n) = i\},\\ A_3 &= \{p_{\boldsymbol{T}}(v) = w,\ \boldsymbol{\sigma}(v) = j,\ \boldsymbol{\sigma}(w) = i\}.\end{aligned}$$

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

It remains to show that the probabilities of A_1, A_2, A_3 are given by (3.1) for all $i, j \in [n]$. The event $p_{\boldsymbol{T}}(n) = w$ implies that $\sigma_{n-1}(w) = L < K$. Therefore A_1 occurs precisely when $K = j$, $L = i$ and $\sigma_{n-1}(w) = i$. Then

$$\mathbb{P}(A_1) = \mathbb{P}(K = j, L = i)\,\mathbb{P}(\sigma_{n-1}(w) = i) = \frac{1}{n(j-1)(n-1)}.$$

Next, $p_{\boldsymbol{T}}(v) = n$ implies that $\sigma_{n-1}(v) \geqslant K$ and thus $\boldsymbol{\sigma}(v) = \sigma_{n-1}(v) + 1$. It then follows that A_2 occurs when $K = i$, $\sigma_{n-1}(v) = j - 1$ and $X_{j-1} = 1$. Therefore

$$\mathbb{P}(A_2) = \mathbb{P}(K = i, X_{j-1} = 1)\,\mathbb{P}(\sigma_{n-1}(v) = j-1) = \frac{1}{n(j-1)(n-1)}.$$

For the last case, since $u, v < n$, it follows that $K \notin \{i, j\}$. For each $k \in [n] \setminus \{i, j\}$, let

$$A_{3,k} = \{p_{\boldsymbol{T}}(v) = w,\ \boldsymbol{\sigma}(v) = j,\ \boldsymbol{\sigma}(w) = i,\ K = k\}.$$

In computing the probabilities $\mathbb{P}(A_{3,k})$ we use that (T_{n-1}, σ_{n-1}) is uniformly random in RD_{n-1}. If $K > j$, then both $\sigma_{n-1}(v) = \boldsymbol{\sigma}(v)$ and $\sigma_{n-1}(w) = \boldsymbol{\sigma}(w)$; in addition, $p_{\boldsymbol{T}}(v) = w$ only if $p_{T_{n-1}}(v) = w$. Therefore, if $k > j$, then

$$\begin{aligned}\mathbb{P}(A_{3,k}) &= \mathbb{P}(K = k)\,\mathbb{P}(p_{T_{n-1}}(v) = w,\ \sigma_{n-1}(v) = j,\ \sigma_{n-1}(w) = i)\\ &= \frac{1}{n(n-1)(n-2)(j-1)}.\end{aligned}$$

Similarly, if $K < j$, then $\sigma_{n-1}(v) = \boldsymbol{\sigma}(v) - 1$, $\sigma_{n-1}(w) = \boldsymbol{\sigma}(w) - \mathbf{1}_{[K<i]}$ and additionally $X_{j-1} = 0$. It then follows that if $k < j$,

$$\begin{aligned}\mathbb{P}(A_{3,k}) &= \mathbb{P}(K = k,\ X_{j-1} = 0)\,\mathbb{P}(p_{T_{n-1}}(v) = w,\ \sigma_{n-1}(v) = j-1,\ \sigma_{n-1}(w) = i - \mathbf{1}_{[K<i]})\\ &= \frac{1}{n} \cdot \frac{j-2}{j-1} \cdot \frac{1}{(n-1)(n-2)(j-2)}.\end{aligned}$$

We have shown that $\mathbb{P}(A_{3,k})$ is uniform for all $k \in [n] \setminus \{i, j\}$, and we get

$$\mathbb{P}(A_3) = \sum_{k \neq i,j} \mathbb{P}(A_{3,k}) = \frac{1}{n(n-1)(j-1)}.$$

Altogether, we have shown that condition (iii) is satisfied and so the proof is complete. □

4. The Poisson approximation

Recall that (T_n, σ_n) is a uniform decorated tree and that T_n has the shape of a recursive tree. In fact Proposition 1.2 implies that the following distributional identity holds, for all $n \in \mathbb{N}$,

$$(\mathrm{d}_{T_n}(\boldsymbol{\sigma}_n^{-1}(i)); i \in [n]) \overset{\text{dist}}{=} (\mathrm{d}_{R_n}(i); i \in [n]).$$

It follows that the distribution of $(Z_m^{(n)},\ m \geqslant 1)$ and Δ_n does not change if we redefine them as $Z_m^{(n)} = \#\{v \in [n] \colon \mathrm{d}_{T_n}(v) \geqslant m\}$ and $\Delta = \max\{\mathrm{d}_{T_n}(v) \colon v \in [n]\}$. However, the correlations in $(\mathrm{d}_{T_n}(v); v \in [n])$ have a subtle difference in comparison with those in $(\mathrm{d}_{R_n}(i); i \in [n])$. To see this, observe that $(\mathrm{d}_{R_n}(i),\ i \in [n])$ is negative orthant-dependent; for a definition see [8]. This fact can be proved by induction from the two-vertex case $(\mathrm{d}_{R_n}(i), \mathrm{d}_{R_n}(j))$, which, in turn, follows essentially from the negative orthant dependence of multinomial distributions; see *e.g.* [7, Lemma 1]. As a consequence, for all $i, j \in [n]$,

$$\mathbb{P}(\mathrm{d}_{R_n}(i) \geqslant m,\ \mathrm{d}_{R_n}(j) \geqslant m) \leqslant \mathbb{P}(\mathrm{d}_{R_n}(i) \geqslant m)\,\mathbb{P}(\mathrm{d}_{R_n}(j) \geqslant m). \tag{4.1}$$

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

On the other hand, the following proposition gives conditions on m for the degrees in T_n to have a pairwise 'almost' negative correlation.

Proposition 4.1. *For any $c \in (0,2)$ there exists $\alpha = \alpha(c) > 0$ such that, uniformly for $m = m(n) < c \ln n$ and distinct $v, w \in [n]$,*

$$\mathbb{P}(d_{T_n}(v) \geqslant m,\ d_{T_n}(w) \geqslant m) \leqslant \mathbb{P}(d_{T_n}(v) \geqslant m)\,\mathbb{P}(d_{T_n}(w) \geqslant m) + O(2^{-2m-\alpha \log n}). \tag{4.2}$$

Moreover, $\alpha < \frac{1}{4}(1 - c + \sqrt{1 + 2c - c^2}) < 1$.

We make the constraints on α precise as this is crucial to Theorem 1.5. A weaker version of Proposition 4.1, without explicit error bounds, was proved in [2, Proposition 4.2]; a complete proof of Proposition 4.1 appears in Appendix A.

Although we do not claim the bounds in Proposition 4.1 are optimal, it seems that the property in (4.1) is lost when randomizing the vertex labels of R_n to obtain T_n. The bound in (4.2) will be an important input to the Chen–Stein method.

Briefly explained, our application of the Chen–Stein method compares, in total variation distance, the sum $Z_m^{(n)}$ with respect to a Poisson variable with mean $\mathbb{E}[Z_m^{(n)}]$. The strength of the bounds depend on finding suitable couplings between $(\mathbf{1}_{[\mathrm{d}_{T_n}(v) \geqslant m]};\ v \in [n])$ and conditional versions of such variables. More precisely, we use the pruning procedure to obtain T_n and Fact 5.1 describes $(\mathrm{d}_{T_n}(i),\ i \in [n])$ in terms of the independent elements $(\mathrm{d}_{T_{n-1}}(i),\ i \in [n-1])$ and $\mathrm{d}_{T_n}(n)$. This allows us to analyse the conditional law of $(\mathrm{d}_{T_n}(i),\ i \in [n-1])$ given $\{\mathrm{d}_{T_n}(n) \geqslant m\}$.

Before going into further details we lay out the necessary notation. Given probability measures μ and ν, a *coupling* of μ and ν is a pair (X, Y) of random variables (either real- or vector-valued) with $X \sim \mu$ and $Y \sim \nu$. Let $I = (I_a,\ a \in \mathscr{A})$ be a collection of $\{0,1\}$-valued random variables. Let μ be the law of $W = \sum_{a \in \mathscr{A}} I_a$ and for $a \in \mathscr{A}$ let ν_a be the conditional law of W given that $I_a = 1$, so

$$\nu_a(B) = \mathbb{P}(W_a \in B) = \mathbb{P}(W \in B \mid I_a = 1).$$

We use the Chen–Stein method stated below.

Theorem 4.2 ([14, Theorem 3.7])**.** *Let $I = (I_a,\ a \in \mathscr{A})$ be a collection of $\{0,1\}$-valued random variables and let $W = \sum_{a \in \mathscr{A}} I_a$. For each $a \in \mathscr{A}$ fix a coupling (W, W_a) of μ and ν_a. Then, with $\lambda = \mathbb{E}[W]$, we have*

$$\mathrm{d_{TV}}(W, \mathrm{Poi}\,(\lambda)) \leqslant \min\{\lambda^{-1}, 1\} \sum_{a \in \mathscr{A}} \mathbb{E}[I_a]\mathbb{E}[|W - (W_a - 1)|].$$

To apply Theorem 4.2 with bounds as tight as possible, one can exploit properties of the variables I_a or construct couplings of μ and ν with specific properties.

Corollary 4.3. *Let $I = (I_a,\ a \in \mathscr{A})$ be a collection of $\{0,1\}$-valued random variables and let $W = \sum_{a \in \mathscr{A}} I_a$. If the variables $I = (I_a,\ a \in \mathscr{A})$ are exchangeable, then for any fixed $a \in \mathscr{A}$ and coupling (W, W_a) of μ and ν_a, we have*

$$\mathrm{d_{TV}}(W, \mathrm{Poi}\,(\lambda)) \leqslant \mathbb{E}[|W - (W_a - 1)|]. \tag{4.3}$$

If, moreover, $W_a = (J_{ab},\ b \in \mathscr{A})$ and there is a coupling (W, W_a) of μ and ν_a satisfying $J_{ab} \leqslant I_a$ for all $b \in \mathscr{A} \setminus \{a\}$, then

$$\mathrm{d_{TV}}(W, \mathrm{Poi}\,(\lambda)) \leqslant \mathbb{E}[I_a] + \sum_{b \in \mathscr{A} \setminus \{a\}} \mathbb{E}[I_a - J_{ab}]. \tag{4.4}$$

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

Now, for the remainder of the section, fix m and, for all $v \in [n]$, let $I_v = \mathbf{1}_{[\mathrm{d}_{T_n}(v) \geqslant m]}$, so that $Z_m^{(n)} \stackrel{\text{dist}}{=} \sum_{v\in[n]} I_v$. Let $(I, J) = ((I_v,\ v \in [n]), (J_v,\ v \in [n]))$ be a coupling of μ and $\nu = \nu_n$ where μ is the law of $(I_1, \ldots, I_n)$ and $\nu = \nu_n$ is the conditional law of $(I_1, \ldots, I_n)$ given that $I_n = 1$.

If we had orthant negative correlation for $(\mathrm{d}_{T_n}(v),\ v \in [n])$, then it would follow that for all $v \in [n-1]$, $\mathbb{E}[I_n I_v] - \mathbb{E}[I_n]\mathbb{E}[I_v] \leqslant 0$ and so the conditions for (4.4) would be satisfied. Although such a strong property has not been yet established, Proposition 4.1 implies, for each $v \in [n-1]$,

$$\mathbb{E}[I_n I_v] - \mathbb{E}[I_n]\mathbb{E}[I_v] \leqslant O(2^{-2m-\alpha \log n}).$$

This suggests that there are couplings of μ and ν for which, with high probability, $I_v \leqslant J_v$ for all $v \in [n-1]$. The existence of such couplings is delicate as the inequality $I_v \leqslant J_v$ has to hold for all $v \in [n-1]$ simultaneously.

The next proposition is the key ingredient in applying the Chen–Stein method to prove Theorem 1.8. The coupling is based on the Robin Hood pruning and its proof is the content of Section 5.

Proposition 4.4. *Let $c \in (1, 2)$. There is $\beta = \beta(c) > 0$ such that, for any $m = m(n) > c \ln n$, there exists a coupling $(I, J) = ((I_1, \ldots, I_n), (J_1, \ldots, J_n))$ of μ and ν, in which, for all $v \in [n-1]$,*

$$\mathbb{P}(I_v < J_v) \leqslant O(n^{-1-\beta}).$$

In the next section we assume Proposition 4.4 and complete the proofs of the results on high-degree vertices of R_n.

4.1 Proofs for high-degree vertices

Proof of Theorem 1.8. Fix $1 < c' < c < 2$ and let $c' \ln n < m = m(n) < c \ln n$. We apply the Chen–Stein method to $Z_m^{(n)} \stackrel{\text{dist}}{=} \sum_{v\in[n]} I_v$. First we use the coupling $(I, J) = ((I_1, \ldots, I_n), (J_1, \ldots, J_n))$ of μ and ν given in Proposition 4.4. By (4.3), we have

$$d_{TV}(Z_m^{(n)}, \mathrm{Poi}\,(\mathbb{E}[\lambda_{n,m}])) \leqslant \mathbb{E}[|W - (W_n - 1)|] \leqslant \mathbb{E}[I_n] + \sum_{v\in[n-1]} \mathbb{E}[|I_v - J_v|].$$

It thus remains to show that the terms in the bound above are $O(2^{-m+(1-\alpha)\log n}) + O(n^{-\beta})$, where $\alpha = \alpha(c) \in (0, 1)$ and $\beta = \beta(c') > 0$ are defined as in Propositions 4.1 and 4.4 respectively. For any $v \in [n-1]$,

$$\begin{aligned}\mathbb{E}[I_n]\mathbb{E}[|J_v - I_v|] &= \mathbb{E}[I_n]\mathbb{E}[I_v - J_v] + 2\mathbb{E}[I_n]\mathbb{E}[(J_v - I_v)\mathbf{1}_{[I_v<J_v]}] \\ &= (\mathbb{E}[I_n]\mathbb{E}[I_v] - \mathbb{E}[I_n I_v]) + 2\mathbb{E}[I_n]\,\mathbb{P}(I_v < J_v).\end{aligned}$$

The terms in the last line are bounded by (4.2) and Proposition 4.4, respectively. Since (1.1) gives $\mathbb{E}[I_n] = 2^{-m}(1 + o(1))$, we get

$$\begin{aligned}\sum_{v\neq n} \mathbb{E}[|I_v - J_v|] &= (n-1)\left[\frac{\mathbb{E}[I_n]\mathbb{E}[I_v] - \mathbb{E}[I_n I_v]}{\mathbb{E}[I_n]} + 2\mathbb{P}(I_v < J_v)\right] \\ &= O(2^{-m+(1-\alpha)\log n}) + O(n^{-\beta}).\end{aligned}$$

Finally, (1.1) together with $\alpha < 1$ also gives $\mathbb{E}[I_n] = O(2^{-m+(1-\alpha)\log n})$. □

Proof of Theorem 1.5. Fix $c \in (1, \log e)$ and let $\alpha = \alpha(c)$ be as in Theorem 1.8. Using the upper bound for α in Proposition 4.1 and simple computations yields $(1 - \alpha)\log e < c$. Thus we can choose $c' \in ((1 - \alpha)\log e, c)$. Let $m = m(n)$ be such that $c' \ln n < m < c \ln n$. By the choice of c

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

and c', we have that, as $n \to \infty$, $(1-\alpha)\log n - m < 0$, while (1.1) implies

$$\mathbb{E}[Z_m^{(n)}] = 2^{-m+\log n}(1+o(1)) \to \infty.$$

The result then follows by Theorem 1.8 and the central limit theorem of Poisson variables; see *e.g.* [9, Exercise 3.4.4]. □

Proof of Corollary 1.11. Recall that $\varepsilon_n = \log n - \lfloor \log n \rfloor$. Let $i = i(n)$ satisfy

$$0 < i < \log e \ln\ln n - C,$$

where $C > 0$ is a constant to be determined below, and note that

$$2^{i+\varepsilon_n} \leqslant 2^{i+1} < 2^{-C+1} \ln n.$$

Let $m = \lfloor \log n \rfloor - i$ and $Z \overset{\text{dist}}{=} \text{Poi}\,(\lambda_{m,n})$.

We have that $\{\Delta_n < \lfloor \log n \rfloor - i\}$ if and only if $\{Z_m^{(n)} = 0\}$. Therefore

$$\mathbb{P}(\Delta_n < \lfloor \log n \rfloor - i) = \mathbb{P}(Z_m^{(n)} = 0) \leqslant \mathbb{P}(Z=0) + d_{TV}(Z_m^{(n)}, Z). \tag{4.5}$$

We deal with the two terms on the right-hand side of (4.5) separately. First, using the lower bound on i, there is a constant $c \in (\log e, 2)$ such that, for n large enough, $m - i < c \ln n$. Therefore (1.1) gives $\gamma > 0$ such that $\lambda_{n,m} = 2^{i+\varepsilon_n} + o(n^{-\gamma} \ln n)$. Consequently

$$\mathbb{P}(Z=0) = \exp\{-\lambda_{n,m} x\} = \exp\{-2^{i+\varepsilon_n}\}(1+o(1)).$$

For the second term in (4.5), Theorem 1.8 gives $\alpha, \beta > 0$ such that

$$d_{TV}(Z_m^{(n)}, Z) = O(2^{-m+(1-\alpha)\log n}) + O(n^{-\beta}).$$

It remains to deal with these two error terms. Note that $\exp\{2^{i+\varepsilon_n}\} \leqslant \exp\{2^{-C+1} \ln n\}$. Therefore, if $C > 1 + \log(1/\beta)$, then

$$\exp\{2^{i+\varepsilon_n}\} O(n^{-\beta}) = O(\exp\{(2^{-C+1} - \beta)\ln n\}) \to 0;$$

similarly, for C large enough,

$$\exp\{2^{i+\varepsilon_n}\} O(2^{-m+(1-\alpha)\log n}) = \exp\{2^{i+\varepsilon_n}\} O(2^{i-\alpha \log n}) \to 0.$$

The two limits above imply that $d_{TV}(Z_m^{(n)}, Z) = o(\exp\{-2^{i+\varepsilon_n}\})$, completing the proof. □

5. The coupling for the Chen–Stein method

In this section we define and analyse the auxiliary coupling used in Proposition 4.4. The coupling is based on the following straightforward property of the deterministic pruning.

Fact 5.1. Fix $n \geqslant 2$. For $\mathrm{h}_n((T,\sigma),(k,l,x)) = (T',\sigma')$, we have $\mathrm{d}_{T'}(n) = \sum_{i=k}^{n-1} x_i$, and for $v \in [n-1]$,

$$\mathrm{d}_{T'}(v) = \mathrm{d}_T(v) + \mathbf{1}_{[l=\sigma(v)]} - \sum_{i=k}^{n-1} x_i \mathbf{1}_{[v=p_T(\sigma^{-1}(i))]}.$$

In words, Fact 5.1 specifies when the degree of a vertex $v < n$ changes: either by gaining n as a new child, or by losing children that are rewired towards n. Moreover, the degree of n in T' equals the total number of such rewirings.

The heuristic for the *almost* negative relation obtained in Proposition 4.4 is as follows. Start with (T_{n-1}, σ_{n-1}) and apply the Robin Hood procedure. If the degree of vertex n is large, Fact 5.1

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

implies that a large number of vertices in T_{n-1} were rewired towards n in the new tree; thus, many (parent) vertices decreased their degree by at least one. In short, conditioning on $\deg_{T_n}(n) \geqslant m$ implies that other vertices are (slightly) less likely to satisfy $\deg_{T_n}(v) \geqslant m$.

For the remainder of the section, fix $n \in \mathbb{N}$, $c \in (1,2)$ and $m = m(n) > c \ln n$. Let (T_{n-1}, σ_{n-1}) be uniformly random in $\mathscr{D}_{n-1}$, let (K, L, X) be an H_n-set, and let (K', L', X') be distributed as an H_n-set conditioned to satisfy $\sum_{i=K}^{n-1} X_i' \geqslant m$. Now write

$$(T_n, \sigma_n) = \mathrm{h}((T_{n-1}, \sigma_{n-1}), (K, L, X)), \tag{5.1}$$

$$(\boldsymbol{T}, \boldsymbol{\sigma}) = \mathrm{h}((T_{n-1}, \sigma_{n-1}), (K', L', X')). \tag{5.2}$$

To avoid cluttered notation, we omit the dependence on m of the conditional random variables (K', L', X') and $(\boldsymbol{T}, \boldsymbol{\sigma})$. By Fact 5.1 and Theorem 1.1, the distribution of $(\boldsymbol{T}, \boldsymbol{\sigma})$ is a uniform element (T_n, σ_n) conditioned on the event that $\mathrm{d}_{T_n}(n) \geqslant m$. Consequently, if $I_v = \mathbf{1}_{[\mathrm{d}_{T_n}(v) \geqslant m]}$ and $J_v = \mathbf{1}_{[\mathrm{d}_{\boldsymbol{T}}(v) \geqslant m]}$ for all $v \in [n]$, then any coupling between (K, L, X) and (K', L', X') yields a coupling for the measures μ and ν in Proposition 4.4.

Our goal is then to couple (K, L, X) and (K', L', X') in such a way that the negative relation between I_v and J_v fails on a negligible set. More precisely, we construct a coupling so that there is $\beta = \beta(c) > 0$ satisfying

$$\mathbb{P}(I_v < J_v) = \mathbb{P}(\mathrm{d}_{T_{n-1}}(v) < m \leqslant \mathrm{d}_{\boldsymbol{T}}(v)) = O(n^{-1-\beta}). \tag{5.3}$$

Lemmas 5.2–5.4 provide the coupling between (K, L, X) and (K', L', X'), while Proposition 5.5 gives necessary conditions, under the coupling, for $I_v < J_v$ to hold. The proof of Proposition 4.4 then follows from bounding the probability that such necessary conditions occur.

5.1 Construction of the coupling

For any integer $n - m \leqslant k < n$, let $X^k = (X_i^k,\ i \in [n-1])$ be a conditional version of X given that $\sum_{i=k}^{n-1} X_i \geqslant m$. The following observation is quite standard but we include a proof for completeness. For $a = (a_1, \ldots, a_d)$ and $b = (b_1, \ldots, b_d) \in \{0,1\}^d$, $a \leqslant b$ only if $a_i \leqslant b_i$ for all $i \in [d]$. We say that $S \subset \{0,1\}^d$ is monotone if $a \leqslant b$ and $a \in S$ imply $b \in S$.

Lemma 5.2. *For each $k < n$, there exists a coupling of X^k and X such that $X_i \leqslant X_i^k$ for all $i \in [n-1]$.*

Proof. Fix $k < n$. Note that $S_k = \{a \in \{0,1\}^{n-1} \colon a_k + \ldots + a_{n-1} \geqslant m\}$ is a monotone increasing subset of $\{0,1\}^{n-1}$. Harris inequality implies $\mathbb{P}(X \in S \cap S_k) \geqslant \mathbb{P}(X \in S_k)\,\mathbb{P}(X \in S)$, for any monotone subset $S \in \{0,1\}^{n-1}$. Dividing through by $\mathbb{P}(X \in S_k)$ yields $\mathbb{P}(X^k \in S) \geqslant \mathbb{P}(X \in S)$. Therefore X^k stochastically dominates X. The existence of the coupling is then guaranteed by Strassen's theorem [19]. □

Before the next coupling, we gather two observations. First, for fixed (k, l), we have $\mathbb{P}(L = l \mid K = k) = \mathbb{P}(L' = l \mid K' = k)$. To see this, observe that $\mathbb{P}(L' = l \mid K' = k)$ can be rewritten as

$$\frac{\mathbb{P}(L = l,\ K = k,\ \sum_{i=K}^{n-1} X_i \geqslant m)}{\mathbb{P}(K = k,\ \sum_{i=K}^{n-1} X_i \geqslant m)} = \frac{\mathbb{P}(L = l,\ K = k,\ \sum_{i=k}^{n-1} X_i \geqslant m)}{\mathbb{P}(K = k,\ \sum_{i=k}^{n-1} X_i \geqslant m)}.$$

The claim then follows by the independence between X and (K, L). Second, the sequence $p_k = \mathbb{P}(K = k \mid \sum_{i=K}^{n-1} X_i \geqslant m)$ is proportional to $\mathbb{P}(\sum_{i=k}^{n-1} X_i \geqslant m)$ and thus it is decreasing in k. Clearly

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

the latter sequence of probabilities is decreasing in k, while both are proportional with a factor $Z = n\mathbb{P}(\sum_{i=K}^{n-1} X_i \geqslant m)$. To see this, use the independence between X and K to obtain

$$\mathbb{P}\left(K = k \,\middle|\, \sum_{i=K}^{n-1} X_i \geqslant m\right) = \frac{\mathbb{P}(K = k, \sum_{i=k}^{n-1} X_i \geqslant m)}{\mathbb{P}(\sum_{i=K}^{n-1} X_i \geqslant m)} = \frac{1}{Z}\,\mathbb{P}\left(\sum_{i=k}^{n-1} X_i \geqslant m\right).$$

Lemma 5.3. *There exists a coupling of (K, L) and (K', L') such that $K' \leqslant K$ and $L' \leqslant L$.*

Proof. Let $X = (X_1, \ldots, X_{n-1})$ be independent with $X_i \overset{\text{dist}}{=} \text{Bernoulli}\,(1/i)$ and independently, let U_1, U_2 be i.i.d. Unif $(0, 1)$. By a slight abuse of notation we *redefine* the variables (K, L) and (K', L') using the variables U_1, U_2 and argue that the *original* law is preserved.

Let $(K, L) = (\lceil nU_1 \rceil, \lceil (K-1)U_2 \rceil)$ and $(K', L') = (K', \lceil (K'-1)U_2 \rceil)$ with

$$K' = \max\left\{k\colon U_1 > \sum_{j=1}^{k-1} p_j\right\}.$$

It is straightforward that (K, L) and K' have the correct law by construction, while L' has the correct law since $\mathbb{P}(L = l \mid K = k) = \mathbb{P}(L' = l \mid K' = k)$ for each $0 \leqslant l < k \leqslant n$. Moreover, since p_k is decreasing, it follows that $K' = j$ implies

$$U_1 > \sum_{i=1}^{j-1} p_i \geqslant \frac{j-1}{n}.$$

It follows that $K \geqslant j = K'$, and so $L = \lceil (K-1)U_2 \rceil \geqslant \lceil (K'-1)U_2 \rceil = L'$. □

Lemma 5.4. *There exists a coupling of (K, L, X) and (K', L', X') such that $K' \leqslant K$, $L' \leqslant L'$ and $X_i \leqslant X_i'$ for all $i \in [n-1]$.*

Proof. Let U_1, U_2 be i.i.d. Unif $(0, 1)$ and independently, let $X = (X_1, \ldots, X_{n-1})$ be independent with $X_i \overset{\text{dist}}{=} \text{Bernoulli}\,(1/i)$. For each $1 \leqslant k < n$ fix a vector X^k coupled with X according to Lemma 5.2. The dependence structure of $X^1, \ldots, X^{n-1}$ is unimportant to the argument, but for concreteness we may, for example, take them to be conditionally independent given X. On the other hand, it is important to insist that the X^k are independent of (K', L'). Since we will define (K', L') using U_1, U_2, the existence of such joint coupling is straightforward.

Again, by a slight abuse of notation we *redefine* the variables and argue that the *original* law is preserved. Define (K, L),(K', L') as in Lemma 5.3 and let $X' = X^{K'}$. Clearly (K, L, X) is an H_n-set. It remains to show that (K', L', X') has the conditional distribution of (K, L, X) given that $\sum_{i=K}^{n-1} X_i \geqslant m$. For any $(k, l, x) \in \mathscr{P}_n$, the probability

$$\mathbb{P}\left((K, L, X) = (k, l, x) \,\middle|\, \sum_{i=K}^{n-1} x_i \geqslant m\right)$$

can be rewritten as

$$\frac{\mathbb{P}(K = k,\ L = l,\ X = x,\ \sum_{i=k}^{n-1} x_i \geqslant m)}{\mathbb{P}(\sum_{i=K}^{n-1} x_i \geqslant m)} = \frac{\mathbb{P}(K = k,\ L = l)\,\mathbb{P}(X = x,\ \sum_{i=k}^{n-1} x_i \geqslant m)}{\mathbb{P}(\sum_{i=K}^{n-1} x_i \geqslant m)}.$$

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

Adding two factors of $\mathbb{P}(\sum_{i=k}^{n-1} x_i \geqslant m)$ and using the independence between (K, L) and X, we can factorize these probabilities as

$$\frac{\mathbb{P}(K=k,\ L=l,\ \sum_{i=k}^{n-1} x_i \geqslant m)}{\mathbb{P}(\sum_{i=K}^{n-1} x_i \geqslant m)} \cdot \frac{\mathbb{P}(X=x,\ \sum_{i=k}^{n-1} x_i \geqslant m)}{\mathbb{P}(\sum_{i=k}^{n-1} x_i \geqslant m)}.$$

These probabilities correspond, respectively, to the distributions of (K', L') and X^k, which are independent. Therefore

$$\begin{aligned}\mathbb{P}\left((K,L,X)=(k,l,x) \,\middle|\, \sum_{i=K}^{n-1} x_i \geqslant m\right) &= \mathbb{P}((K',L')=(k,l))\,\mathbb{P}(X^k=x)\\ &= \mathbb{P}((K',L',X')=(k,l,x))\end{aligned}$$

as desired. Finally, the variables (K, L, X) and (K', L', X') satisfy the desired inequalities by Lemmas 5.2 and 5.3. □

5.2 Analysis of the coupling

The proof of Proposition 4.4 boils down to understanding necessary conditions for $\mathrm{d}_{T_n}(v) < m \leqslant \mathrm{d}_{\boldsymbol{T}}(v)$ to hold under the coupling of Lemma 5.4.

Proposition 5.5. *Consider (K, L, X) and (K', L', X') defined in Lemma 5.4 and their corresponding decorated trees (T_n, σ_n), $(\boldsymbol{T}, \boldsymbol{\sigma})$ defined in (5.1) and (5.2). For any $v \in [n-1]$,*

$$\{d_{T_n}(v) < m \leqslant d_{\boldsymbol{T}}(v)\} \subset \{L' = \sigma_{n-1}(v)\} \cap \{d_{T_{n-1}}(v) \geqslant m-1\}.$$

Proof. From the properties of the coupling in Lemma 5.4,

$$\sum_{i=K}^{n-1} X_i\, \mathbf{1}_{[v=p_{T_{n-1}}(\sigma_{n-1}^{-1}(i))]} \leqslant \sum_{i=K'}^{n-1} X'_i\, \mathbf{1}_{[v=p_{T_{n-1}}(\sigma_{n-1}^{-1}(i))]}. \tag{5.4}$$

Consequently, using Fact 5.1 we have $\mathrm{d}_{\boldsymbol{T}}(v) - \mathrm{d}_{T_n}(v) \leqslant \mathbf{1}_{[L'=\sigma_{n-1}(v)]}$. On the other hand, if $\{\mathrm{d}_{T_n}(v) < m \leqslant \mathrm{d}_{\boldsymbol{T}}(v)\}$ holds, then it follows that $\mathrm{d}_{\boldsymbol{T}}(v) - \mathrm{d}_{T_n}(v) > 0$ and so it is necessary that $\{L' = \sigma_{n-1}(v)\}$ holds. Finally, $\{m \leqslant \mathrm{d}_{\boldsymbol{T}}(v)\}$ implies that

$$m \leqslant \mathrm{d}_{\boldsymbol{T}}(v) = \mathrm{d}_{T_{n-1}}(v) + \mathbf{1}_{[L'=\sigma_{n-1}(v)]} - \sum_{i=K'}^{n-1} X'_i\, \mathbf{1}_{[v=p_{T_{n-1}}(\sigma_{n-1}^{-1}(i))]} \leqslant \mathrm{d}_{T_{n-1}}(v) + 1,$$

or equivalently that $\{\mathrm{d}_{T_{n-1}}(v) \geqslant m-1\}$. □

We can also argue, more specifically, that

$$\{\mathrm{d}_{T_n}(v) < m \leqslant \mathrm{d}_{\boldsymbol{T}}(v)\} \subset \{L' = \sigma_{n-1}(v)\} \cap \{\mathrm{d}_{T_{n-1}}(v) = m-1\}.$$

However, the approach we choose allows us to use uniform bounds for all $v \in [n-1]$. We will frame the events $\{\mathrm{d}_{T_{n-1}}(v) \geqslant m-1\}$ from the perspective of recursive trees where the degree distributions are distinct for each vertex. Recall the following version of Bernstein inequalities (see *e.g.* [18, Theorem 2.8, (2.5)]). For a sum S of $\{0,1\}$-valued variables and $\varepsilon > 0$,

$$\mathbb{P}(S > (1+\varepsilon)\mathbb{E}[S]) \leqslant \exp\left\{-\frac{3\varepsilon^2}{2(3+\varepsilon)}\mathbb{E}[S]\right\}.$$

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

By the construction of R_n we have $\mathrm{d}_{R_n}(i) \overset{\text{dist}}{=} \sum_{k=i}^{n} B_k \leqslant \sum_{k=1}^{n} B_k$, where $(B_k, k \geqslant 1)$ are independent Bernoulli variables with mean $1/k$. Therefore

$$\mathbb{P}(\mathrm{d}_{R_n}(i) > m) \leqslant \mathbb{P}\left(\sum_{k=1}^{n} B_k > c \ln n\right).$$

Using the fact that $\mathbb{E}[\sum_{k=1}^{n} B_k] = \ln n + O(1) < c \ln n$, we can apply Bernstein's inequality with $\varepsilon = c - 1 + o(1)$ and set $\beta = 3\varepsilon^2/(2(3+\varepsilon))$. It follows that there is $\beta = \beta(c) > 0$ such that, uniformly over $m > c \ln n$, and $i \in [n]$,

$$\mathbb{P}(\mathrm{d}_{R_n}(i) > m) = O(n^{-\beta}). \tag{5.5}$$

Proof of Proposition 4.4. Fix $c \in (1, 2)$. Let $m = m(n) > c \ln n$ and $\beta = \beta(c) > 0$ be as in (5.5). Let T_n and $\boldsymbol{T}$ be as defined in (5.1) and (5.2) with $((K, L, X), (K', L', X'))$ as in Lemma 5.4. Set $I_v = \mathbf{1}_{[\mathrm{d}_{T_n}(v) \geqslant m]}$ and $J_v = \mathbf{1}_{[\mathrm{d}_{\boldsymbol{T}}(v) \geqslant m]}$ for all $v \in [n]$, so that $(I, J) = ((I_1, \ldots, I_n), (J_1, \ldots J_n))$ is a coupling of the measures μ and ν.

Our goal is to bound $\mathbb{P}(I_v < J_v) = \mathbb{P}(\mathrm{d}_{T_n}(v) < m \leqslant \mathrm{d}_{\boldsymbol{T}}(v))$. First, by Proposition 5.5,

$$\mathbb{P}(\mathrm{d}_{T_n}(v) < m \leqslant \mathrm{d}_{\boldsymbol{T}}(v)) \leqslant \sum_{j=1}^{n-1} \mathbb{P}(L' = j,\ \sigma_{n-1}(v) = j,\ \mathrm{d}_{T_{n-1}}(v) \geqslant m - 1).$$

Next we obtain uniform bounds for the terms on the right-hand side. Recall that σ_{n-1} is a uniformly random permutation independent of L' and that $\sigma_{n-1}(T_{n-1}) \overset{\text{dist}}{=} R_{n-1}$. These facts together with (5.5) give, for each $j \in [n-1]$,

$$\begin{aligned}
\mathbb{P}(L' = j,\ \sigma_{n-1}(v) = j,\ \mathrm{d}_{T_{n-1}}(v) \geqslant m - 1) &= \frac{\mathbb{P}(L' = j)}{n-1} \mathbb{P}(\mathrm{d}_{T_{n_1}}(v) \geqslant m - 1 \mid \sigma_{n-1}(v) = j) \\
&= \frac{\mathbb{P}(L' = j)}{n-1} \mathbb{P}(\mathrm{d}_{R_{n-1}}(j) \geqslant m - 1) \\
&\leqslant \mathbb{P}(L' = j) O(n^{-1-\beta}).
\end{aligned}$$

Combining these bounds, we obtain for any $v \in [n-1]$

$$\begin{aligned}
\mathbb{P}(I_v < J_v) &\leqslant \sum_{j=1}^{n-1} \mathbb{P}(L' = j,\ \sigma_{n-1}(v) = j,\ \mathrm{d}_{T_{n-1}}(v) \geqslant m - 1) \\
&= O(n^{-1-\beta}) \sum_{j=1}^{n-1} \mathbb{P}(L' = j) = O(n^{-1-\beta}).
\end{aligned}$$

□

6. Conclusions and further research

The Robin Hood pruning yields an interesting process $((T_n, \sigma_n),\ n \geqslant 1)$. By Theorem 1.1 and Proposition 1.2, $\sigma_n(T_n) \overset{\text{dist}}{=} R_n$ for all $n \geqslant 1$; that is, T_n has the shape of a recursive tree. The novelty of this process is that the Robin Hood pruning is a fairly complex dynamic of trees which has potential connections to mathematical models of social and economic networks and raises challenging theoretical questions.

First, only asymptotically about half the time T_n is obtained from T_{n-1} by simply attaching n to a uniformly random vertex. To see this, recall that

$$\mathrm{d}_{T_n}(n) \overset{\text{dist}}{=} \min\{\mathrm{Geo}\,(1/2), |\mathscr{S}|\},$$

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

where $|\mathscr{S}| \to \infty$ (see Fact A.1 and Lemma A.2). It follows that with probability tending to 1/2 the newly added vertex will be a leaf. Second, for all $n \geqslant 1$, $\mathrm{d}_{R_n}(n) = 0$ a.s., while Fact 5.1 and the distribution of the H-set yield

$$\mathbb{E}[\mathrm{d}_{T_n}(n)] = \mathbb{E}\left[\mathbb{E}\left[\sum_{i=k}^{n-1} X_i \mid M = k\right]\right] = \sum_{k=1}^{n}\sum_{i=k}^{n-1}\frac{1}{n\cdot i} = \sum_{i=1}^{n-1}\sum_{k=1}^{i}\frac{1}{n\cdot i} = 1 - \frac{1}{n}.$$

Third, from time to time, a large proportion of edges will be rewired towards the newly added vertex, drastically reshaping the structure of the tree. For example, for any $a \in [0, 1)$,

$$\mathbb{E}[\mathrm{d}_{T_n}(n) \mid M \leqslant n^a] \geqslant \mathbb{E}\left[\sum_{i=n^a}^{n} X_i\right] = (1-a)\ln n.$$

As for applications, in the context of random networks, the Robin Hood pruning has an interpretation in terms of 'trends'; for example, a new vertex brings in a new idea to the network which may drastically rewire the interests or connections of established individuals in the network. The stamp history σ_n gives a ranking between the elements of T_n that determines the susceptibility of changing parents in the tree. Preferential attachment models are considered better models for real-world networks. It would be interesting to devise a similar pruning procedure that, acting on preferential attachment trees, preserves their scale-free degree distribution.

In the context of biology, Kingman's coalescent is usually represented with increasing binary trees, keeping individuals as external nodes and *adding* an internal node for each merge between two lineages. The representation using n-chains breaks the symmetry between the pairs of trees merging at each step. Thus it is not clear how the Robin Hood pruning process would have a significant interpretation in terms of the genealogical information.

Regardless of the perspective we use to motivate the process $((T_n, \sigma_n),\ n \geqslant 1)$, there are many interesting theoretical questions that would be worth pursuing. Here we list just a few.

(1) Understand the process describing how the parent and descendants of a given vertex change with time.
   – Describe how the size of the subtree rooted at a fixed node j evolves.
   – How does the maximum size of such subtree grow?
(2) Understand the maximum degree dynamics in both $(R_n,\ n \geqslant 1)$ and $(T_n,\ n \geqslant 1)$.
   – How often do vertices attaining the maximum degree change?
   – Are these dynamics the same for both processes?
(3) Determine whether there is a coupling for which the sequence $(\mathbf{1}_{[\mathrm{d}_{T_n}(v)]}, v \in [n])$ is negative related (*i.e.* that conditions for (4.4) are satisfied), or similarly, whether the sequence is negative orthant-dependent.

Acknowledgements

I would like to thank Louigi Addario-Berry and Henning Sulzbach for some very helpful discussions, and to the anonymous referees who provided insight into how to improve the presentation of the results and additional references. This research was supported by FQRNT via a PBEEE scholarship with number 169888 and by NSF grant DMS-1703516.

References

[1] Addario-Berry, L. (2015) Partition functions of discrete coalescents: from Cayley's formula to Frieze's $\zeta(3)$ limit theorem. In *XI Symposium on Probability and Stochastic Processes*, Vol. 68 of Progress in Probability, Birkhäuser.
[2] Addario-Berry, L. and Eslava, L. (2018) High degree of random recursive trees. *Random Struct. Algorithms* **52** 560–575.

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

[3] Anderson, C. W. (1970) Extreme value theory for a class of discrete distributions with applications to some stochastic processes. *J. Appl. Probab.* **7** 99–113.
[4] Blum, M., François, O and Janson, S. (2006) The mean, the variance and limiting distributions of two statistics sensitive to phylogenetic tree balance. *Ann. Appl. Probab.* **16** 2195–2114.
[5] Chang, H. and Fuchs, M. (2010) Limit theorems for patterns in phylogenetic trees. *J. Math. Biol.* **60** 481–512.
[6] Devroye, L. (1987) Branching processes in the analysis of the heights of trees. *Acta Inform.* **24** 277–298.
[7] Devroye, L. and Lu, J. (1995) The strong convergence of maximal degrees in uniform random recursive trees and dags. *Random Struct. Algorithms* **7** 1–14.
[8] Dubhashi, D. and Ranjan, D. (1998) Balls and bins: a study in negative dependence. *Random Struct. Algorithms* **13** 99–124.
[9] Durrett, R. (1996) *Probability: Theory and Examples*, second edition, Duxbury.
[10] Eslava, L. (2016) Depth of vertices with high degree in random recursive trees. arXiv:1611.07466
[11] Flajolet, P. and Sedgewick, R. (2009) *Analytic Combinatorics*, Cambridge University Press.
[12] Fuchs, M. (2008) Subtree sizes in recursive trees and binary search trees: Berry–Esseen bounds and Poisson approximations. *Combin. Probab. Comput.* **17** 661–680.
[13] Goh, W. and Schmutz, E. (2002) Limit distribution for the maximum degree of a random recursive tree. *J. Comput. Appl. Math.* **142** 61–82.
[14] Goldschmidt, C. (2000) The Chen–Stein method for convergence of distributions. https://www.stats.ox.ac.uk/~goldschm/chen-stein.ps.gz
[15] Gumbel, E. J. (1935) Les valeurs extrêmes des distributions statistiques. *Ann. Inst. H. Poincaré* **5** 115–158.
[16] Holmgren, C. and Janson, S. (2015) Limit laws for functions of fringe trees for binary search trees and random recursive trees. *Electron. J. Probab.* **20** 4.
[17] Janson, S. (2005) Asymptotic degree distribution in random recursive trees. *Random Struct. Algorithms* **26** 69–83.
[18] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Random Graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience.
[19] Lindvall, T. (1999) On Strassen's theorem on stochastic domination. *Electron. Comm. Probab.* **4** 51–59.
[20] Luczak, M. and Winkler, P. (2004) Building uniformly random subtrees. *Random Struct. Algorithms* **24** 420–443.
[21] Na, H. S. and Rapoport, A. (1970) Distribution of nodes of a tree by degree. *Math. Biosci.* **6** 313–329.
[22] Pitman, J. (1999) Coalescent random forests. *J. Combin. Theory Ser. A* **85** 165–193.
[23] Pittel, B. (1994) Note on the heights of random recursive trees and random m-ary search trees. *Random Struct. Algorithms* **5** 337–347.
[24] Standish, T. A. (1980) *Data Structure Techniques*, Addison-Wesley Longman.

Appendix A. Proof of Proposition 4.1

We use the representation of Kingman's coalescent that consists of a chain $\mathbf{C} = (F_n, \dots, F_1)$ and write $T^{(n)}$ for the unique tree contained in F_1. By Proposition 1.3 we can work with the tree $T^{(n)}$. The proof mimics that of [2, Proposition 4.2], but requires a little more care as we wish to obtain explicit error bounds.

For each $v, j \in [n]$, let $T_j(v)$ denote the tree in F_j that contains vertex v. For each $v \in [n]$, the *selection set* of v is defined as

$$\mathscr{S}_n(v) = \{2 \leqslant j \leqslant n\colon T_j(v) \in \{T^{(j)}_{a_j}, T^{(j)}_{b_j}\}\};$$

this set keeps record of the times when the tree containing v merges. Finally, for each $2 \leqslant j \leqslant n$, we say that ξ_j is *favourable* for vertices in $T^{(j)}_{a_j}$ (resp. vertices in $T^{(j)}_{b_j}$) if $\xi_j = 1$ (resp. $\xi_j = 0$).

The key property of Kingman's coalescent is as follows. For each $j \in \mathscr{S}_n(v)$, if ξ_j favours v, then $r(T_j(v))$ increases its degree by one in the process; otherwise $r(T_j(v))$ attaches to the root of the other merging tree and the degree of $r(T_j(v))$ remains unchanged for the rest of the process. Since all vertices start the process as roots, $\mathrm{d}_{T^{(n)}}(v)$ is equal to the length of the first streak of favourable times for v. Moreover, $(\xi_j,\ j \in [n-1])$ are independent and distributed as Bernoulli $(1/2)$. Therefore we have the following distributional equivalence.

Fact A.1. Let D be a random variable with distribution Geo $(1/2)$ independent of $\mathscr{S}_n(v)$. Then

$$\mathrm{d}_{T^{(n)}}(v) \overset{\text{dist}}{=} \min\{D, |\mathscr{S}_n(v)|\}.$$

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

This fact, together with the next lemma, allows us to get estimates for the tails of $\mathrm{d}_{T^{(n)}}(v)$.

Lemma A.2. *Let $c \in (0,2)$ and $0 < \varepsilon \leqslant 1 - c/2$. Writing $a = 1 - \varepsilon - c/2$, we have*

$$\mathbb{P}(|\mathscr{S}_n(v) \setminus [n^a]| > c \ln n) \leqslant O(1) n^{-\varepsilon^2/(\varepsilon + c/2)}.$$

Proof. First, there are $j(j-1)$ distinct pairs of trees in F_j, exactly $j-1$ of such pairs containing $T_j(v)$; thus $\mathbb{P}(j \in \mathscr{S}_n(v)) = 2/j$. Since the merging trees are chosen independently at each time, for any $a \in [0,1)$ we have

$$|\mathscr{S}_n(v) \setminus [n^a]| \stackrel{\text{dist}}{=} \sum_{j=n^a+1}^{n} B_j,$$

where the variables $B_1, \ldots, B_n$ are independent Bernoulli variables with $\mathbb{E}[B_i] = 2/i$, respectively. The desired bound is then a straightforward application of Bernstein's inequalities (see *e.g.* [18, Theorem 2.8, (2.6)]). For a sum S of $\{0,1\}$-valued variables, we have $\mathbb{P}(S \leqslant \mathbb{E}[S] - t) \leqslant \exp\{-t^2/2\mathbb{E}[S]\}$. In this case $S = \sum_{i=n^a}^{n} B_i$ and

$$\mathbb{E}[S] = \sum_{i=n^a}^{n} \frac{2}{i} = 2(1-a)\ln n + O(1) = (c + 2\varepsilon)\ln n + O(1).$$

The result follows by setting $t = 2\varepsilon \ln n + O(1)$. □

Proposition A.3. *If $c \in (0,2)$ and $m < c \ln n$, then for $\varepsilon = (2-c)^2/4$,*

$$2^{-m}(1 - o(n^{-\varepsilon})) \leqslant \mathbb{P}(d_{T^{(n)}}(1) \geqslant m) \leqslant 2^{-m}.$$

Proof of Proposition A.3. It follows from Lemma A.1 that

$$\mathbb{P}(\mathrm{d}_{T^{(n)}}(v) \geqslant m) = \mathbb{P}(D \geqslant m)\mathbb{P}(|\mathscr{S}_n(v)| \geqslant m).$$

The upper bound on $\mathbb{P}(\mathrm{d}_{T^{(n)}}(1) \geqslant m)$ is then trivial, while the lower bound follows by Lemma A.2 using $\varepsilon = 1 - c/2$ and that $\mathscr{S}_n(v) = \mathscr{S}_n(v) \setminus [1]$. □

Now consider two distinct vertices $v, w \in [n]$. For $m \in \mathbb{N}$, let $\mathscr{G}_m \in \{2, \ldots, n\}^2$ contain all pairs of selection sets that enable vertices v and w to have degree at least m; that is, $(A,B) \in \mathscr{G}_m$ only if

$$\mathbb{P}(\mathrm{d}_{T^{(n)}}(v) \geqslant m,\ \mathrm{d}_{T^{(n)}}(w) \geqslant m, (\mathscr{S}_n(v), \mathscr{S}_n(w)) = (A,B)) > 0.$$

Since the ξ_j are independent of the selection times, we have

$$\mathbb{P}(\mathrm{d}_{T^{(n)}}(v) \geqslant m,\ \mathrm{d}_{T^{(n)}}(w) \geqslant m) \geqslant 2^{-2m}\,\mathbb{P}((\mathscr{S}_n(v), \mathscr{S}_n(w)) \in \mathscr{G}_m). \tag{A.1}$$

To estimate $\mathbb{P}((\mathscr{S}_n(v), \mathscr{S}_n(w)) \in \mathscr{G}_m)$ we need more details on the dynamics of the model. We start with a simple tail bound for the following random variable; let

$$\tau = \max\{j\colon j \in \mathscr{S}_n(v) \cap \mathscr{S}_n(w)\}.$$

Lemma A.4. *For $a \in (0,1)$, $\mathbb{P}(\tau > n^a) \leqslant 4n^{-a}$.*

Proof. Vertices in $T^{(n)}$ are exchangeable, so we can take $v = 1$, $w = 2$; these vertices belong to distinct trees in F_j for all $j \geqslant \tau$. Additionally, by the ordering convention of trees in F_j, it follows that $T_j(1) = 1$ and $T_j(2) = 2$ for all $j \geqslant \tau$.

We claim that, for all $2 < k \leqslant n$,

$$\mathbb{P}(\tau \leqslant k) = \prod_{j=k+1}^{n} \left(1 - \frac{2}{j(j-1)}\right).$$

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press

This follows by induction on $n-k$. Clearly $\tau = n$ only if $\{a_n, b_n\} = \{1,2\}$, which occurs with probability $2/(n(n-1))$. Thus $\mathbb{P}(\tau \leqslant n-1)$ satisfies the equation above. For $k < n$, we have

$$\frac{\mathbb{P}(\tau \leqslant k)}{\mathbb{P}(\tau \leqslant k+1)} = \mathbb{P}(\tau \leqslant k \mid \tau \leqslant k+1) = \mathbb{P}(\{a_{k+1}, b_{k+1}\} \neq \{1,2\}) = 1 - \frac{2}{(k+1)k}.$$

Next, for k large enough,

$$\prod_{j=k+1}^{n}\left(1 - \frac{2}{j(j-1)}\right) \geqslant \prod_{j=k}^{n-1}\left(1 - \frac{2}{j^2}\right) > 1 - \sum_{j=k}^{\infty}\frac{2}{j^2} > 1 - 4\int_k^{\infty} x^{-2}dx = 1 - 4/k.$$

The second inequality uses that $1 - x > e^{-2x}$ for $x > 0$ sufficiently small, followed by the fact that $e^{-\sum 2x_j} > 1 - \sum 2x_j$. The result follows with $k = n^a$. □

Lemma A.5. *If $c \in (0,2)$ and $m < c \ln n$, then for any $\gamma < \frac{1}{4}(1 - c + \sqrt{1+2c-c^2})$,*

$$\mathbb{P}((\mathscr{S}_n(v), \mathscr{S}_n(w)) \in \mathscr{G}_m) \geqslant 1 - o(n^{-\gamma}).$$

Proof. For each $\varepsilon \in (0, 1 - c/2]$ write $a = a(\varepsilon) = 1 - \varepsilon - c/2$. Then

$$\mathbb{P}((\mathscr{S}_n(v), \mathscr{S}_n(w)) \notin \mathscr{G}_m) \leqslant \mathbb{P}(\tau > n^a) + 2\mathbb{P}(|\mathscr{S}_n(v) \setminus [n^a]| < c \ln n). \tag{A.2}$$

Before establishing (A.2), we note that the terms on the right-hand side of (A.2) are bounded by Lemmas A.4 and A.2, respectively. Since such bounds depend on the choice of ε, we can use

$$\gamma < \max_{0<\varepsilon\leqslant 1-c/2}\left\{\min\left(1 - \varepsilon - \frac{c}{2}, \frac{\varepsilon^2}{\varepsilon + c/2}\right)\right\} = \frac{1}{4}\left(1 - c + \sqrt{1+2c-c^2}\right).$$

The last equality is valid since the functions to be minimized are decreasing and increasing, respectively, on the $(0,1)$ interval. It then follows that the maximum is attained when $0 < \varepsilon < 1 - c/2$ satisfies $1 - \varepsilon - c/2 = \varepsilon^2/(\varepsilon + c/2)$.

We now proceed to establish equation (A.2). At step τ, exactly one of v and w is favoured by ξ_τ. Thus at least one of v or w gets its degree fixed for the remainder of the process. Therefore

$$\{(\mathscr{S}_n(v), \mathscr{S}_n(w)) \in \mathscr{G}_m\} \subset \{|\mathscr{S}_n(v) \setminus [\tau]| \geqslant m\} \cup \{|\mathscr{S}_n(w) \setminus [\tau]| \geqslant m\}.$$

By intersecting with the event $\tau > n^a$, and the exchangeability of vertices in $T^{(n)}$, we get

$$\begin{aligned}\mathbb{P}((\mathscr{S}_n(v), \mathscr{S}_n(w)) \notin \mathscr{G}_m) &\leqslant \mathbb{P}(\tau > n^a) + 2\mathbb{P}((\mathscr{S}_n(v), \mathscr{S}_n(w)) \notin \mathscr{G}_m, \tau \leqslant n^a)\\ &\leqslant \mathbb{P}(\tau > n^a) + 2\mathbb{P}(|\mathscr{S}_n(v) \setminus [\tau]| < m, \tau \leqslant n^a)\\ &\leqslant \mathbb{P}(\tau > n^a) + 2\mathbb{P}(|\mathscr{S}_n(v) \setminus [n^a]| < m, \tau \leqslant n^a),\end{aligned}$$

from which (A.2) follows. □

Proof of Proposition 4.1. Fix $c \in (0,2)$, $m = m(n) < c \ln n$ and let I_v, J_v be defined as in Proposition 4.1. By Proposition 1.3, it follows that $\mathbb{E}[I_v] = \mathbb{P}(\mathrm{d}_{T^{(n)}}(v) \geqslant m)$ and

$$\begin{aligned}\mathbb{E}[I_v]\mathbb{E}[J_v] = \mathbb{E}[I_v I_n] &= \mathbb{P}(\mathrm{d}_{T^{(n)}}(v) \geqslant m,\ \mathrm{d}_{T^{(n)}}(n) \geqslant m)\\ &= 2^{-2m}\mathbb{P}((\mathscr{S}_n(v), \mathscr{S}_n(n)) \in \mathscr{G}_m),\end{aligned}$$

the last equality by (A.1). Lemmas A.5 and A.3 then give that, for $\alpha < \frac{1}{4}(1 - c + \sqrt{1+2c-c^2})$,

$$\mathbb{E}[I_v]\mathbb{E}[I_n] - \mathbb{E}[I_v]\mathbb{E}[J_{vn}] \leqslant 2^{-2m} - 2^{-2m}(1 + o(n^{-\alpha})) = 2^{-2m}o(n^{-\alpha}).$$ □

Cite this article: Eslava L (2021). A non-increasing tree growth process for recursive trees and applications. *Combinatorics, Probability and Computing* **30**, 79–104. https://doi.org/10.1017/S0963548320000073

https://doi.org/10.1017/S0963548320000073 Published online by Cambridge University Press