# A NON-INCREASING TREE GROWTH PROCESS FOR RECURSIVE TREES AND APPLICATIONS 

LAURA ESLAVA


#### Abstract

We introduce a non-increasing tree growth process $\left(\left(T_{n}, \sigma_{n}\right), n \geq 1\right)$, where $T_{n}$ is a rooted labeled tree on $n$ vertices and $\sigma_{n}$ is a permutation of the vertex labels. The construction of ( $T_{n}, \sigma_{n}$ ) from $\left(T_{n-1}, \sigma_{n-1}\right)$ 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 monotonous in the process.

We present two applications. First, while couplings between Kingman's coalescent and random recursive trees where known for any fixed $n$, this new process provides a nonstandard 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.


## 1. Introduction

In a paper of 1970 [21], Na and Rapoport presented the problem of modeling how the structure of networks (as sociograms, communication and acquaintance networks) emerge through time. They considered two cases: 'growing' trees and 'static' trees. The 'growing' model is now know as uniform attachment model and each instance is usually named (random) recursive tree. These are part of a broad class of tree growth models where vertices are sequentially added and connected to a random vertex in the current tree. On the other side, the term 'static' was motivated by the fact that this construction starts with the $n$ vertices the tree is aimed to have and $n-1$ edges are added one by one (without creating cycles). The 'static' model was an early description of what is now referred to as coalescent processes. The seemingly two 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, Kingman's coalescents correspond, for any fixed number of vertices $n$, to recursive trees; see Remark 2.3.

Here we present a non-increasing tree growth process $\left(\left(T_{n}, \sigma_{n}\right), n \geq 1\right)$ where $T_{n}$ is a rooted labeled tree on $n$ vertices and $\sigma_{n}$ is a permutation of the vertex labels. The three key features of this new growth process are:
(1) The shape of $T_{n}$ has the same distribution as that of recursive trees (vertices are labeled uniformly at random),

[^0](2) adding edges according to the permutation $\sigma_{n}$ (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 $\mathcal{D}_{n}$ of decorated trees on $n$ vertices and a random mapping $\mathrm{H}_{n}: \mathcal{D}_{n-1} \rightarrow \mathcal{D}_{n}$ such that $\mathrm{H}_{n}\left(T_{n-1}, \sigma_{n-1}\right) \stackrel{\text { dist }}{=}\left(T_{n}, \sigma_{n}\right)$ for all $n>1$. Our main result, Theorem 1.1, states that recursively applying the mappings $\mathrm{H}_{n}$ to the unique element in $\mathcal{D}_{1}$ gives uniformly random decorated trees on $\mathcal{D}_{n}$; from which 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 them in the last section.

We call Robin-Hood pruning to the random mapping $\mathrm{H}_{n}$ that builds ( $T_{n}, \sigma_{n}$ ) from $\left(T_{n-1}, \sigma_{n-1}\right)$; it 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 connection had been rarely exploited, with exception to [6, 23], where an equivalent construction was used by to study union-find trees and then related 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 allow us to understand how the vertex-degree distributions change when we condition on one such vertex-degrees 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 $[n]=\{1, \ldots, n\}$ and $\mathcal{S}_{n}$ 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 labeled tree $T=(V(T), E(T))$, write $|T|=|V(T)|$ and call $|T|$ the size of $T$. We write $\mathcal{T}_{n}$ 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 $e=u v$ for an edge with tail $u$ and head $v$. For $u \in V(T) \backslash\{r(T)\}$ we write $p_{T}(u)$ for the parent of $u$, that is, the unique vertex $v$ with $u v$ 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 $\mathrm{d}_{T}(v)=\#\left\{u: p_{T}(u)=v\right\}$.

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

Recursive trees on $n$ vertices, which we denote $R_{n}$, are usually constructed as follows. Start with $R_{1}$ as a single node with label 1 . For each $1<j \leq n, R_{j}$ 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 $1<j \leq n$. It is readily seen that $R_{n}$ is a uniformly random tree in $\mathcal{I}_{n}$. It follows that the process $\left(R_{n}, n \geq 1\right)$ is a random increasing tree growth process.


Figure 1. A decorated tree $(T, \sigma) \in \mathcal{D}_{6}$ on the left; the permutation $\sigma$ is depicted with bold numbers next to the vertices in $T$ (so for example $\sigma(1)=5$ and $\sigma(6)=2)$. On the right, the increasing tree $\sigma(T)$.
1.2. The new growth process. In what follows we extend the concept of increasing trees. If $T \in \mathcal{T}_{n}$ and $\sigma \in \mathcal{S}_{n}$ then $\sigma(T)$ is the tree $T^{\prime} \in \mathcal{T}_{n}$ with edges $\{\sigma(u) \sigma(v): u v \in E(T)\}$. In words, $T^{\prime}$ is obtained from $T$ by relabeling the vertices of $T$ according to the permutation $\sigma$; see Figure 1 for an example. We say that $\sigma$ is an stamp history for $T$ if $\sigma(T)$ is increasing. If $\sigma$ is an stamp history for $T$ then we say that the pair $(T, \sigma)$ is a recursively decorated tree or decorated tree, and that vertex $v$ has time stamp $\sigma(v)$. We denote the set of decorated trees of size $n$ by

$$
\mathcal{D}_{n}=\left\{(T, \sigma): T \in \mathcal{T}_{n}, \sigma \text { is an stamp history of } T\right\}
$$

For each $n \geq 2$, the Robin-Hood pruning $\mathrm{H}_{n}: \mathcal{D}_{n-1} \rightarrow \mathcal{D}_{n}$ is a random mapping that can be applied to any decorated tree. The exact definition of $\mathrm{H}_{n}$ will be given in Section 3. Broadly speaking, $\mathrm{H}_{n}(T, \sigma)$ is obtained from $(T, \sigma)$ by pruning some subtrees of $T$ and placing them as subtrees of a new vertex labeled $n$; 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 $\sigma$ such 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 $n$.

The content of our main theorem says that, when the input of $\mathrm{H}_{n}$ is uniformly random in $\mathcal{D}_{n-1}$ the output is uniformly random in $\mathcal{D}_{n}$. For the remainder of the paper, for any $n \geq 1$, the pair $\left(T_{n}, \sigma_{n}\right)$ denotes a uniformly random element in $\mathcal{D}_{n}$. Such result boils down to carefully setting up the distribution of the random parameters involved in the Robin-Hood pruning.

Theorem 1.1. For each $n \geq 2$, the Robin-Hood pruning provides a coupling between $\left(T_{n-1}, \sigma_{n-1}\right)$ and $\left(T_{n}, \sigma_{n}\right)$ such that $\left(T_{n}, \sigma_{n}\right)=\mathrm{H}_{n}\left(T_{n-1}, \sigma_{n-1}\right)$.

Note that $\left|\mathcal{D}_{1}\right|=1$, thus the Robin-Hood pruning can be unambiguously applied to decorated trees starting from $\mathcal{D}_{1}$. Theorem 1.1 implies that the tree growth process $\left(\left(T_{n}, \sigma_{n}\right), n \geq\right.$ 1) given by $\left(T_{n}, \sigma_{n}\right)=\mathrm{H}_{n}\left(\left(T_{n-1}, \sigma_{n-1}\right)\right)$ is composed of uniformly random decorated trees, but it yields a non-increasing growth process on trees. This occurs since 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 proven by a straightforward bijection between $\mathcal{D}_{n}$ and $\mathcal{I}_{n} \times \mathcal{S}_{n}$.

Proposition 1.2. For each $n \in \mathbb{N},\left|\mathcal{D}_{n}\right|=n!(n-1)$ ! and if $\left(T_{n}, \sigma_{n}\right) \in \mathcal{D}_{n}$ is chosen uniformly at random then $\sigma_{n}\left(T_{n}\right) \stackrel{\text { dist }}{=} R_{n}$ is a recursive tree of size $n$ and $\sigma_{n}$ is a uniformly chosen permutation in $\mathcal{S}_{n}$.
Proof. By definition, if $(T, \sigma) \in \mathcal{D}_{n}$, then $\sigma(T) \in \mathcal{I}_{n}$. Let $\varphi: \mathcal{D}_{n} \rightarrow \mathcal{I}_{n} \times \mathcal{S}_{n}$ be defined such that $\varphi(T, \sigma)=(\sigma(T), \sigma)$. For an increasing tree $T$ and $\sigma \in \mathcal{S}_{n}$, let $T^{\prime}=\sigma^{-1}(T)$ then $\varphi\left(T^{\prime}, \sigma\right)=(T, \sigma)$, it is also straightforward that $\varphi$ is injective. Therefore, $\left|\mathcal{D}_{n}\right|=$ $\left|\mathcal{I}_{n}\right| \cdot\left|\mathcal{S}_{n}\right|=n!(n-1)!$. 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 $\left(T_{n}, \sigma_{n}\right)$ 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, for the moment it suffices to say that $\mathbf{C}=\left(F_{n}, \ldots, F_{1}\right)$ denotes a Kingman's coalescent, where the $F_{j}$ are forests.
Proposition 1.3. Let $\left(T_{n}, \sigma_{n}\right)$ be uniformly random in $\mathcal{D}_{n}$ and $\mathbf{C}=\left(F_{n}, \ldots, F_{1}\right)$ be a Kingman's coalescent. Denote $F_{1}=\left\{T_{\mathbf{C}}\right\}$, then $T_{\mathbf{C}} \stackrel{\text { dist }}{=} T_{n}$ and the forests evolution is given by $\sigma_{n}$.

Typically there is no simple coupling of finite $n$-coalescent processes as $n$ varies. The first application of Theorem 1.1 is that the Robin-Hood pruning produces, given a Kingman's coalescent on $n$ vertices, a Kingman's coalescent on $n+1$ vertices.
Corollary 1.4. The tree growth process $\left(\left(T_{n}, \sigma_{n}\right), n \geq 1\right)$, 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. Phase changes occurs 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 infinite 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 tend 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 $0<m \leq n$, let us count the number of high-degree vertices by

$$
Z_{m}^{(n)}=\#\left\{v \in[n]: \mathrm{d}_{R_{n}}(v) \geq m\right\}
$$

and write $\lambda_{n, m}=\mathbb{E}\left[Z_{m}^{(n)}\right]$. The following estimates where implicitly given in [2] and a proof can be found in Appendix A, Proposition 6.3. For each $c \in(0,2)$, there is $\gamma=\gamma(c)$ such that uniformly over $m<c \ln n$,

$$
\begin{equation*}
2^{-m+\log n}\left(1-o\left(n^{-\gamma}\right)\right)=\mathbb{E}\left[Z_{m}^{(n)}\right] \leq 2^{-m+\log n} \tag{1}
\end{equation*}
$$

It thus follows that the phase change occurs when $m=m(n) \approx \log n$. Using a Poisson approximation together with (1) we obtain the following phase change for the counts on high-degree vertices.

Theorem 1.5. For each $c \in(1, \log e)$ there exists $c^{\prime} \in(1, c)$ such that if $c^{\prime} \ln n<m<c \ln n$, then

$$
\begin{equation*}
\frac{Z_{m}^{(n)}-\lambda_{n, m}}{\sqrt{\lambda_{n, m}}} \xrightarrow{\text { dist }} N(0,1) \tag{2}
\end{equation*}
$$

If $m \geq \log n$, under sequences $n_{j}$ for which $\lambda_{n_{j}, m} \rightarrow \lambda$ we have that

$$
Z_{m}^{\left(n_{j}\right)} \xrightarrow{\text { dist }} \operatorname{Poi}(\lambda) .
$$

Remark 1.6. Counting the high-degree vertices is equivalent to count fringe trees (of all sizes) with a high-degree root. Therefore, the asymptotic normal distribution of $Z_{m}^{(n)}$, (2), follows from [16, Corollary 1.25]; however, the computation of both the mean and variance for the renormalization of the variables is not seemingly straightforward. Nevertheless, we remark that the associated convergence rates in Theorem 1.7 are strong and novel.

Previous results on the profile of recursive trees consider $X_{m}^{(n)}=\#\left\{v \in[n]: \mathrm{d}_{R_{n}}(v)=m\right\}$, for $m<n$. For finite values of $m$, Janson established the joint limiting distribution of $\left(X_{m}^{(n)}, m \geq 1\right)$ in [17]. Addario-Berry and the author addressed the case $m=m(n) \rightarrow$ $\infty$, providing all the possible limiting distributions of $\left(X_{\lfloor\log n\rfloor+k}^{(n)}, k \in \mathbb{Z}\right)$ 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 $\left(\mathbf{1}_{[]} \mathrm{d}_{T_{n}}(v) \geq m, v \in[n]\right)$ change when conditioning to $\mathrm{d}_{T_{n}}(v) \geq m$. The details of this approach are somewhat delicate, so we defer the discussion to Section 4.

Theorem 1.7. Fix $1<c^{\prime}<c<2$. There are constants $\alpha=\alpha(c) \in(0,1)$ and $\beta=\beta\left(c^{\prime}\right)>0$ such that uniformly for $m=m(n)$ satisfying $c^{\prime} \ln n<m<c \ln n$,

$$
\mathrm{d}_{\mathrm{TV}}\left(Z_{m}^{(n)}, \operatorname{Poi}\left(\lambda_{n, m}\right)\right) \leq O\left(2^{-m+(1-\alpha) \log n}\right)+O\left(n^{-\beta}\right)
$$

Remark 1.8. A detailed but simple track of the conditions on $\alpha$, see Proposition 4.1, shows that there is a non-empty interval $\mathcal{I}=((1-\alpha) \log e, c)$ such that if $c^{\prime} \in \mathcal{I} \cap(1,2)$, then the bounds in Theorem 1.7 are, in fact, tending to zero.

Remark 1.9. The exponent $\alpha$ is determined by almost negative correlation between pairs of vertices in $T_{n}$ (see Proposition 4.1), while the exponent $\beta$ depends on an auxiliary coupling based on the Robin-Hood pruning (see Proposition 4.4). We believe that the constraint on $c^{\prime}>1$ could be relaxed by obtaining uniform bounds on $\mathbb{P}\left(\mathrm{d}_{R_{n}}(i)=m\right)$ rather than $\mathbb{P}\left(\mathrm{d}_{R_{n}}(i) \geq m\right)$.

Finally, consider now the maximum degree $\Delta_{n}$ of a recursive tree $R_{n}$. Note that $Z_{m}^{(n)}>$ 0 if and only if $\Delta_{n} \geq m$. Therefore, having $\mathbb{E}\left[Z_{\log n}^{(n)}\right] \approx 1$ indicates $\Delta_{n} \approx \log n$. In fact, Devroye and Lu showed that $\Delta_{n} / \log n \rightarrow 1$ a.s. [7]. The first tail bounds on $\Delta_{n}$ where obtained for $\mathbb{P}\left(\Delta_{n}<\lfloor\log n\rfloor+i\right)$ with $i \in \mathbb{Z}$ using singularity analysis of generating functions [13]. The relation between recursive trees and Kingman's coalescent provided simpler proofs to such results, extending it also to $i<2 \ln n-\log n$ [2]. The bounds in Theorem 1.7 yield broader, tighter bounds.

Corollary 1.10. There exists $C>0$ such that uniformly over $0<i=i(n)<\log e \ln \ln n-$ C,

$$
\mathbb{P}\left(\Delta_{n}<\lfloor\log n\rfloor-i\right)=\exp \left\{-2^{i+\varepsilon_{n}}\right\}(1+o(1))
$$

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 $\Delta_{n}$. Goh and Schmutz provide an alternative heuristic based on the fact that $\mathrm{d}_{R_{n}}(i)$, with $i \rightarrow \infty$ slowly, is asymptotically normal [13].

Outline. The paper is divided into two parts. First, we discuss more on the connection between recursive trees, Kingman's coalescents and other tree models in Section 2. The precise definition of the Robin-Hood pruning $\mathrm{H}_{n}$ together with the proof of Theorem 1.1 is given in Section 3. Second, 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 on how we use the Chen-Stein method is given 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. And 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 with 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 an more traditional construction using binary search trees (BST) 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-know bijection between BST's 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.
2.1. Recursive trees perspective. Na and Rapoport loosely described this process as the construction of 'static' trees with $n$ vertices [21]:
"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 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]). A formal description follows.

A forest $f$ is a set of trees with pairwise disjoint vertex sets. Denote by $V(f)$ and $E(f)$, respectively, the union of the vertex and edge sets of the trees in $f$. For $n \geq 1$, an $n$-chain is a sequence $C=\left(f_{n}, \ldots, f_{1}\right)$ of elements of $\mathcal{F}_{n}=\{f: V(f)=[n]\}$ such that, for $1<i \leq n$, $f_{i-1}$ 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 $\left(f_{n}, \ldots, f_{1}\right)$ and $1 \leq i \leq n$, list the trees of $f_{i}$ in increasing order of their smallest-labeled vertex as $T_{1}^{(i)}, \ldots, T_{i}^{(i)}$. Independently for each $1<i \leq n$ let $\left\{a_{i}, b_{i}\right\} \subset$ $\{\{a, b\}: 1 \leq a<b \leq i\}$ be uniformly chosen at random; in addition, let $\xi_{i}$ be independent Bernoulli random variables with mean $1 / 2$.

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

In other words, if $\mathbf{C}=\left(F_{n}, \ldots, F_{1}\right)$ 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 labeling that tracks the number of trees left in the forest when a give edge comes along. Fix $C=\left(f_{n}, \ldots, f_{1}\right)$, for each $e \in E\left(T_{C}\right)$, let

$$
\rho_{C}(e)=\max \left\{i \in[n-1]: e \in E\left(f_{i}\right)\right\} .
$$

We next define a vertex labeling $\sigma_{C}: V\left(T_{C}\right) \rightarrow[n]$. Let $\sigma_{C}\left(r\left(T_{C}\right)\right)=1$, and for each $u v \in E\left(t_{C}\right)$, let

$$
\sigma_{C}(u)=\rho_{C}(u v)+1
$$

The following proposition shows that the pair $\left(T_{C}, \sigma_{C}\right) \in \mathcal{D}_{n}$ contains all the information to recover the original $n$-chain $C$; in other words, if $\mathcal{C}_{n}$ denotes the set of $n$-chains, then $\mathcal{D}_{n}$ and $\mathcal{C}_{n}$ are in bijection.

Proposition 2.2. Let $\Upsilon: \mathcal{C}_{n} \rightarrow \mathcal{D}_{n}$ be defined as follows. For an n-chain $C=\left(f_{n}, \ldots, f_{1}\right)$, let $\Upsilon(C)=\left(T_{C}, \sigma_{C}\right)$. Then $\Upsilon$ is a bijection.

[^1]

Figure 2. An example of an $n$-chain with $n=6$. The edge labelling $\rho_{n}$ is presented with numbers in bold.

Proof. First, we show that $\mathcal{C}_{n}$ and $\mathcal{D}_{n}$ have the same cardinality. To count the number of $n$-chains, consider constructing $\left(f_{n}, \ldots, f_{1}\right)$ 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 $\left(f_{n}, \ldots, f_{k}\right)$, there are $k(k-1)$ possible directed edges to add. Therefore, $\left|\mathcal{C}_{n}\right|=n!(n-1)$ !.

Next, let $C=\left(f_{n}, \ldots, f_{1}\right)$ be an $n$-chain. For each $1 \leq i<n$, 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 $\left\{\rho_{C}(e), e \in E\left(T_{C}\right)\right\}$ decrease along all paths in $T_{C}$ towards the root $r\left(T_{C}\right)$. Consequently, the labels $\left\{\sigma_{C}(v), v \in[n]\right\}$ are, indeed, an stamp history of $T_{C}$. It follows that $\Upsilon$ is well defined.

Finally, let $C=\left(f_{n}, \ldots, f_{1}\right), C^{\prime}=\left(f_{n}^{\prime}, \ldots, f_{1}^{\prime}\right)$ be distinct $n$-chains and write $k=\min \{i$ : $\left.f_{i} \neq f_{i}^{\prime}\right\}$. If $k=1$ then $T_{C} \neq T_{C^{\prime}}$ and clearly, $\Upsilon(C) \neq \Upsilon\left(C^{\prime}\right)$. Otherwise $T_{C}=T_{C^{\prime}}$, $f_{k-1}=f_{k-1}^{\prime}$ and the (unique) edges $e \in E\left(f_{k-1}\right) \backslash E\left(f_{k}\right)$ and $e^{\prime} \in E\left(f_{k-1}^{\prime}\right) \backslash E\left(f_{k}^{\prime}\right)$ are distinct. It follows that $e=u v \in f_{k}^{\prime}$ and so $\sigma_{C}(u)=k>\sigma_{C^{\prime}}(u)$. This shows that $\Upsilon$ is injective, and so $\Upsilon$ is a bijection between $\mathcal{C}_{n}$ and $\mathcal{D}_{n}$.

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

Proof of Proposition 1.3. Let $\mathbf{C}=\left(F_{n}, \ldots, F_{1}\right)$ be a Kingman's coalescent. For any fixed $n$-chain $\left(f_{n}, \ldots, f_{1}\right) \in \mathcal{C}_{n}$,

$$
\mathbb{P}\left(\left(F_{n}, \ldots, F_{1}\right)=\left(f_{n}, \ldots, f_{1}\right)\right)=\prod_{k=1}^{n-1} \mathbb{P}\left(F_{k}=f_{k} \mid\left(F_{n}, \ldots, F_{k+1}\right)=\left(f_{n}, \ldots, f_{k+1}\right)\right) .
$$

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 $\left(f_{n}, \ldots, f_{1}\right)$,

$$
\mathbb{P}\left(\left(F_{n}, \ldots, F_{1}\right)=\left(f_{n}, \ldots, f_{1}\right)\right)=[(n-1)!n!]^{-1} .
$$

Recall $F_{1}=\left\{T_{\mathbf{C}}\right\}$. By Proposition 2.2, $\left(T_{\mathbf{C}}, \sigma_{\mathbf{C}}\right) \in \mathcal{D}_{n}$ and it has a uniform distribution, since the bijection preserves the uniform measure of $\mathbf{C}$. Finally, by Proposition 2.2 , it follows that $T_{\mathbf{C}} \stackrel{\text { dist }}{=} T_{n}$. 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 relabeling of vertices, Kingman's coalescent correspond 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 $\mathcal{B}_{n}$ be the set containing all plane, rooted, unlabeled binary trees with $n$ external nodes. Trees in $\mathcal{B}_{n}$ distinguish between left and right subtrees of any given internal vertex. It can be shown that the sizes $\left|\mathcal{B}_{n}\right|=3 \cdot 5 \cdots(2 n-3)$ are given by the Catalan numbers.

Binary search trees are the tree representation of the sorting algorithm Quicksort. Simply described, for each $n \geq 1$, 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 $\sigma(2), \ldots, \sigma(n)$ are sequentially added 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 $B \in \mathcal{B}_{n+1}$, 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
all the insertion process. Using this representation, the rotation correspondence maps (one-to-one) 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 side, phylogenetic trees on $n$ species are represented by elements in $\mathcal{B}_{n}$. 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, 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 YuleHarding process does not yield uniform phylogenetic trees (as neither the BST is uniform in $\mathcal{B}_{n}$ ). For further discussion between the two models, see e.g. [4, Section 3].

It has been presumed that Kingman's coalescent is the bottom's 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 through the bijections and $n!$-to- 1 mappings from Propositions 2.2 and 1.2 , there should be a correspondence between BST's with time stamps and Kingman's coalescent.

The construction of Kingman's coalescent as a binary tree in $\mathcal{B}_{n}$ with time stamps is the following. 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 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 set 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 labeled exchangeably. However, considering only the final tree (either in $\mathcal{T}_{n}$ or $\mathcal{B}_{n}$ ) 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 $\left|\mathcal{T}_{n}\right|=n^{n-1}$ and $\left|\mathcal{B}_{n}\right|=3 \cdot 5 \cdots(2 n-3)$ different rooted, labeled trees and phylogenetic trees, respectively.

## 3. The Robin-Hood pruning

The Robin-Hood pruning $\mathrm{H}_{n}: \mathcal{D}_{n-1} \rightarrow \mathcal{D}_{n}$ is a random procedure based on randomizing the parameters of a deterministic mapping $\mathrm{h}_{n}: \mathcal{D}_{n-1} \times \mathcal{P}_{n} \rightarrow \mathcal{D}_{n}$ where the set $\mathcal{P}_{n}$ defines all possible ways to prune a decorated tree on $n-1$ vertices. The distribution on $\mathcal{P}_{n}$ is tailored so that the Robin-Hood pruning, in fact, yields a coupling of $\left(\left(T_{n}, \sigma_{n}\right) ; n \geq 1\right)$.

First we introduce the necessary notation to define $\mathcal{P}_{n}$, the deterministic pruning $\mathrm{h}_{n}$ and verify that, indeed, the mapping $h_{n}$ is well defined. We then continue to define the distribution on $\mathcal{P}_{n}$ 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 underline the difference between deterministic elements $(T, \pi)$ of $\mathcal{D}_{n}$ 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 $(k, l, x) \in \mathcal{P}_{n}$ : the time stamp $k$ of the new vertex, its point of attachment $l$ given by time stamp, and the vertices to be rewired encoded by time stamp in the sequence $x=\left(x_{1}, \ldots, x_{n-1}\right)$. Once a stamp history is given to a tree $T, \mathcal{V}_{n}$ contains the vertices to be pruned and rewired towards the new vertex $n$.

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

$$
\mathcal{P}_{n}=\left\{(k, l, x): 1 \leq l<k \leq n, x \in\{0,1\}^{n-1}\right\} \cup\left\{(1,0, x): x \in\{0,1\}^{n-1}, x_{1}=1\right\}
$$

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

$$
\mathcal{V}_{n}(k, l, x, \sigma)=\mathcal{V}_{n}(k, x, \sigma)=\left\{v \in[n-1]: x_{\sigma(v)}=1, \sigma(v) \geq k\right\}
$$

Remark 3.1. The definition of $\mathcal{P}_{n}$ is such that $\sigma^{-1}(1) \in \mathcal{V}_{n}$ if and only if $k=1$.
The following deterministic pruning is illustrated in Figure 3.
Definition 3.2. Fix $n \geq 2,(T, \sigma) \in \mathcal{D}_{n-1}$ and $(k, l, x) \in \mathcal{P}_{n}$. We define $\left(T^{\prime}, \sigma^{\prime}\right)$ and set

$$
\mathrm{h}_{n}((T, \sigma),(k, l, x))=\left(T^{\prime}, \sigma^{\prime}\right)
$$

as follows. First, let $\mathcal{V}=\mathcal{V}_{n}(k, x, \sigma)$ and construct $T^{\prime}$ from $T$ : For each $v \in \mathcal{V} \backslash\{r(T)\}$, replace the edge $v p_{T}(v)$ with an edge connecting $v$ to a new vertex labeled $n$. Now, if $k=1$ then attach $r(T)$ to $n$; otherwise, attach vertex $n$ to $\sigma^{-1}(l)$. In other words, the edges of $T^{\prime}$ are given by

$$
E\left(T^{\prime}\right)=\left\{\begin{array}{cl}
(E(T) \cup\{v n ; v \in \mathcal{V}\}) \backslash\left\{v p_{T}(v) ; v \in \mathcal{V}\right\} & \text { if } k=1 \\
\left\{n \sigma^{-1}(l)\right\} \cup(E(T) \cup\{v n ; v \in \mathcal{V}\}) \backslash\left\{v p_{T}(v) ; v \in \mathcal{V}\right\} & \text { if } k>1
\end{array}\right.
$$

Second, let $\sigma^{\prime}:[n] \rightarrow[n]$ be defined by $\sigma^{\prime}(n)=k$ and for $v<n$,

$$
\sigma^{\prime}(v)=\sigma(v)+\mathbf{1}_{[\sigma(v) \geq k]}
$$

Lemma 3.3. For any $n \geq 2, \mathrm{~h}_{n}: \mathcal{D}_{n-1} \times \mathcal{P}_{n} \rightarrow \mathcal{D}_{n}$ is well defined. That is, for any $(T, \sigma) \in \mathcal{D}_{n-1}$ and $(k, l, x) \in \mathcal{C}_{n}$,

$$
\mathrm{h}_{n}((T, \sigma),(k, l, x)) \in \mathcal{D}_{n}
$$

Proof. Write $\mathrm{h}_{n}((T, \sigma),(k, l, x))=\left(T^{\prime}, \sigma^{\prime}\right)$. When $k=1$, it is clear that $T^{\prime}$ is a tree. When $k>1$, let $w=\sigma^{-1}(l)$ be the parent of $n$ in $T^{\prime}$ and let $\left(w=v_{1}, \ldots, v_{j}=r(T)\right)$ be the path from $w$ to the root of $T$. Since $\sigma$ is a stamp history of $T, l=\sigma\left(v_{1}\right)>\sigma\left(v_{2}\right)>\cdots>\sigma\left(v_{j}\right)=1$; moreover, $l<k$. It follows that $v_{i} \notin \mathcal{V}(k, l, x, \sigma)$ for all $i \in[j]$ and consequently, no edges in the path from $n$ to the root in $T^{\prime}$ closes a cycle by connecting to $n$.

Now, we show that $\sigma^{\prime}$ is a stamp history for $T^{\prime}$. It is clear that $\sigma^{\prime}$ is a permutation of $[n]$, so it suffices to prove that $\sigma^{\prime}(v)>\sigma^{\prime}\left(p_{T^{\prime}}(v)\right)$, for all $v \in V(T) \backslash\left\{r\left(T^{\prime}\right)\right\}$. First, for vertices $v$ with $p_{T^{\prime}}(v)=n$ we have $\sigma(v) \geq k$ and consequently $\sigma^{\prime}(v)=\sigma(v)+1>k=\sigma^{\prime}(n)$.

Second, consider $v, w<n$ with $p_{T^{\prime}}(v)=w$. It follows that $v w \in E(T)$ and thus $\sigma(v)>$ $\sigma(w)$. Consequently, $\mathbf{1}_{[\sigma(v) \geq k]} \geq \mathbf{1}_{[\sigma(w) \geq k]}$ and so $\sigma^{\prime}(v)>\sigma^{\prime}(w)$. The last case occurs when $k>1$ and $p_{T^{\prime}}(n)=w=\sigma^{-1}(l)$. We then have $\sigma^{\prime}(n)=k>l=\sigma(w)=\sigma^{\prime}(w)$.

Remark 3.4. Whenever $(k, l, x) \in \mathcal{P}_{n}$ has $x_{j}=1$ for some $j \geq k$, setting $\left(T^{\prime}, \sigma^{\prime}\right)=$ $\mathrm{h}_{n}((T, \sigma),(k, l, x))$ and $v=\sigma^{-1}(j)$ yields $n=p_{T^{\prime}}(v) \neq p_{T}(v) \in[n-1]$. This implies that $T \not \subset T^{\prime}$ 。

(A) A tree $(T, \sigma)$ in $\mathcal{D}_{9}$. The permutation $\sigma$ is depicted with bold numbers.

(B) Vertices in gray satisfy $X_{\sigma(v)}=1$ and underlined are time stamps $\sigma(i) \geq k$.

(c) Nodes $i$ with $\sigma(i) \geq k$ and $x_{i}=1$ have been pruned and time stamps have been adjusted.

(D) The resulting tree $\mathrm{h}_{10}((T, \sigma),(k, l, x)) \in \mathcal{D}_{10}$.

Figure 3. An example of the Robin-Hood pruning for $(T, \sigma)$ with $k=$ $6, l=5$ and $x_{1}=x_{2}=x_{7}=x_{8}=1$; all other $x_{i}=0$.
3.2. The random process. The $\mathrm{H}_{n}$-set is a sample of $\mathcal{P}_{n}$ according to the following distribution.

Definition 3.5. Fix $n \geq 1$. Let $K \stackrel{\text { dist }}{=} \operatorname{Unif}(1,2, \ldots, n)$; if $K=1$ let $L=0$, and if $K>1$ let $L=\operatorname{Unif}(1,2, \ldots K-1)$. Independently, let $X=\left(X_{1}, \ldots, X_{n-1}\right)$ where $X_{i} \stackrel{\text { dist }}{=}$ Bernoulli $(1 / i)$ are independent variables. An $\mathrm{H}_{n}$-set is a triple of random variables with the same law as $(K, L, X) \in \mathcal{P}_{n}$.

We are ready to define the Robin-Hood pruning. For each $n \geq 2$, let $(K, L, X) \in \mathcal{P}_{n}$ be an $\mathrm{H}_{n}$-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 $(T, \sigma)$; however, the distribution of the $\mathrm{H}_{n}$-set is tailored so that $\mathrm{H}_{n}\left(T_{n-1}, \sigma_{n-1}\right)$ preserves the uniform measure on decorated trees. In order to prove Theorem 1.1, we start with a characterization of $\left(T_{n}, \sigma_{n}\right)$.

Lemma 3.6. Let $n \geq 1$ be an integer. A random decorated tree $(\boldsymbol{T}, \boldsymbol{\sigma}) \in \mathcal{D}_{n}$ is uniformly random if and only if the following properties are satisfied.
i) The permutation $\boldsymbol{\sigma}$ is uniformly random on $\mathcal{S}_{n}$.
ii) Conditionally given $\boldsymbol{\sigma}$, the vertices $\left(p_{\boldsymbol{\sigma}(\boldsymbol{T})}\left(\boldsymbol{\sigma}^{-1}(v)\right), v \in V(\boldsymbol{T}) \backslash\{r(\boldsymbol{T})\}\right)$ are independent.
iii) For all vertices $v, w \in[n]$ and indices $i, j \in[n]$,

$$
\begin{equation*}
\mathbb{P}\left(p_{\boldsymbol{T}}(v)=w, \boldsymbol{\sigma}(v)=j, \boldsymbol{\sigma}(w)=i\right)=\frac{1}{n(n-1)(j-1)} \mathbf{1}_{[j>i]} \tag{3}
\end{equation*}
$$

Proof. Let $(\boldsymbol{T}, \boldsymbol{\sigma})=\left(T_{n}, \sigma_{n}\right)$ be uniformly random on $\mathcal{D}_{n}$. Condition $\left.i\right)$ 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

$$
\left(p_{\boldsymbol{\sigma}(\boldsymbol{T})}\left(\boldsymbol{\sigma}^{-1}(v)\right), v \in V(\boldsymbol{T}) \backslash\{r(\boldsymbol{T})\}\right) \stackrel{\text { dist }}{=}\left\{p_{R_{n}}(j), 1<j \leq n\right\}
$$

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

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

Now consider a random decorated tree $(\boldsymbol{T}, \boldsymbol{\sigma}) \in \mathcal{D}_{n}$ satisfying conditions $i$-iii). Fix a decorated tree $(T, \pi) \in \mathcal{D}_{n}$, and for $v \in V(T) \backslash\{r(T)\}$, let $w_{v}=p_{T}(v)$. Condition $\left.i i\right)$ on the conditional independence of parents gives, for $v \neq r(T)$,

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

Using that $\pi$ is an stamp history for $T$, so $\pi(v)>\pi\left(w_{v}\right)$, and that $\boldsymbol{\sigma}$ is uniformly random, it follows from (3) that

$$
\begin{equation*}
\mathbb{P}\left(p_{\boldsymbol{T}}(v)=w_{v} \mid \boldsymbol{\sigma}=\pi\right)=\frac{1}{\pi(v)-1} \tag{4}
\end{equation*}
$$

Any increasing tree $T^{\prime} \in \mathcal{I}_{n}$ is determined by the set of parents $\left\{p_{T^{\prime}}(v), 1<v \leq n\right\}$. Using that $\pi(T) \in \mathcal{I}_{n}$ and the conditional independence from condition $\left.i i\right)$ we get

$$
\begin{aligned}
\mathbb{P}(\boldsymbol{\sigma}(\boldsymbol{T})=\pi(T) \mid \boldsymbol{\sigma}=\pi) & =\mathbb{P}\left(p_{\boldsymbol{\sigma}(\boldsymbol{T})}(j)=p_{\pi(T)}(j), 1<j \leq n \mid \boldsymbol{\sigma}=\pi\right) \\
& =\mathbb{P}\left(p_{\boldsymbol{T}}(v)=p_{T}(v), v \in V(t) \backslash\{r(T)\} \mid \boldsymbol{\sigma}=\pi\right) \\
& =\prod_{v \in V(T) \backslash r(T)} \mathbb{P}\left(p_{\boldsymbol{T}}(v)=p_{T}(v) \mid \boldsymbol{\sigma}=\pi\right) \\
& =[(n-1)!]^{-1} ;
\end{aligned}
$$

the last equality holds by (4) and the fact that $\{\pi(v), v \in V(T) \backslash\{r(T)\}\}=\{2, \ldots, n\}$. Finally, using the equation above and 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 $(T, \pi)$, so $(\boldsymbol{T}, \boldsymbol{\sigma})$ is uniformly random in $\mathcal{D}_{n}$.

We are now ready to prove Theorem 1.1.
Proof of Theorem 1.1. Let $\left(T_{n-1}, \sigma_{n-1}\right) \in \mathcal{D}_{n-1}$ be a uniformly random decorated tree. Let $(K, L, X)$ be an $\mathrm{H}_{n}$-set and let $(\boldsymbol{T}, \boldsymbol{\sigma})=\mathrm{h}\left(\left(T_{n-1}, \sigma_{n-1}\right),(K, L, X)\right)$. 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 $\sigma_{n-1}$. Second, once conditioning on $\boldsymbol{\sigma}$, which is equivalent to conditioning on both $\sigma_{n-1}$ and $K$, we get

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

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

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

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

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

$$
\begin{aligned}
& A_{1}=\left\{p_{\boldsymbol{T}}(n)=w, \boldsymbol{\sigma}(n)=j, \boldsymbol{\sigma}(w)=i\right\}, \\
& A_{2}=\left\{p_{\boldsymbol{T}}(v)=n, \boldsymbol{\sigma}(v)=j, \boldsymbol{\sigma}(n)=i\right\}, \\
& A_{3}=\left\{p_{\boldsymbol{T}}(v)=w, \boldsymbol{\sigma}(v)=j, \boldsymbol{\sigma}(w)=i\right\} .
\end{aligned}
$$

It remains to show that the probabilities of $A_{1}, A_{2}, A_{3}$ are given by (3) 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,

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

Next, $p_{\boldsymbol{T}}(v)=n$ implies that $\sigma_{n-1}(v) \geq K$ 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,

$$
\mathbb{P}\left(A_{2}\right)=\mathbb{P}\left(K=i, X_{j-1}=1\right) \mathbb{P}\left(\sigma_{n-1}(v)=j-1\right)=\frac{1}{n(j-1)(n-1)}
$$

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

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

In computing the probabilities $\mathbb{P}\left(A_{3, k}\right)$ we use that $\left(T_{n-1}, \sigma_{n-1}\right)$ is uniformly random in $R D_{n-1}$. 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

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

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 $k<j$,

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

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

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

Altogether, we have shown that condition $i i i$ ) is satisfied and so the proof is complete.

## 4. The Poisson approximation

Recall that $\left(T_{n}, \sigma_{n}\right)$ 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 $n \in \mathbb{N}$,

$$
\left(\mathrm{d}_{T_{n}}\left(\boldsymbol{\sigma}_{n}^{-1}(i)\right) ; i \in[n]\right) \stackrel{\text { dist }}{=}\left(\mathrm{d}_{R_{n}}(i) ; i \in[n]\right)
$$

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

$$
\begin{equation*}
\mathbb{P}\left(\mathrm{d}_{R_{n}}(i) \geq m, \mathrm{~d}_{R_{n}}(j) \geq m\right) \leq \mathbb{P}\left(\mathrm{d}_{R_{n}}(i) \geq m\right) \mathbb{P}\left(\mathrm{d}_{R_{n}}(j) \geq m\right) \tag{5}
\end{equation*}
$$

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 $v, w \in[n]$,
(6) $\mathbb{P}\left(\mathrm{d}_{T_{n}}(v) \geq m, \mathrm{~d}_{T_{n}}(w) \geq m\right) \leq \mathbb{P}\left(\mathrm{d}_{T_{n}}(v) \geq m\right) \mathbb{P}\left(\mathrm{d}_{T_{n}}(w) \geq m\right)+O\left(2^{-2 m-\alpha \log n}\right)$.

Moreover, $\alpha<\frac{1}{4}\left(1-c+\sqrt{1+2 c-c^{2}}\right)<1$.
We make precise the constraints on $\alpha$ 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 (5) is lost when randomizing the vertex labels of $R_{n}$ to obtain $T_{n}$. The bound in (6) 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 $\mathbb{E}\left[Z_{m}^{(n)}\right]$. The strength of the bounds depend on finding suitable couplings between $\left(\mathbf{1}_{\left[\mathrm{d}_{T_{n}}(v) \geq m\right]} ; v \in[n]\right)$ and conditional versions of such variables. More precisely, we use the pruning procedure to obtain $T_{n}$ and Fact 5.1 describes $\left(\mathrm{d}_{T_{n}}(i), i \in[n]\right)$ in terms of the independent elements $\left(\mathrm{d}_{T_{n-1}}(i), i \in[n-1]\right)$ and $\mathrm{d}_{T_{n}}(n)$. This allows us to analyze the conditional law of $\left(\mathrm{d}_{T_{n}}(i), i \in[n-1]\right)$ given $\left\{\mathrm{d}_{T_{n}}(n) \geq m\right\}$ holds.

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

$$
\nu_{a}(B)=\mathbb{P}\left(W_{a} \in B\right)=\mathbb{P}\left(W \in B \mid I_{a}=1\right)
$$

We use the Chen-Stein method stated below.
Theorem 4.2 ([14, Theorem 3.7]). Let $I=\left(I_{a}, a \in \mathcal{A}\right)$ be a collection of $\{0,1\}$-valued random variables and let $W=\sum_{a \in \mathcal{A}} I_{a}$. For each $a \in \mathcal{A}$ fix a coupling $\left(W, W_{a}\right)$ of $\mu$ and $\nu_{a}$. Then with $\lambda=\mathbb{E}[W]$, we have

$$
\mathrm{d}_{\mathrm{TV}}(W, \operatorname{Poi}(\lambda)) \leq \min \left\{\lambda^{-1}, 1\right\} \sum_{a \in \mathcal{A}} \mathbb{E}\left[I_{a}\right] \mathbb{E}\left[\left|W-\left(W_{a}-1\right)\right|\right]
$$

To apply Theorem 4.2 with as tight as possible bounds, one can exploit properties of the variables $I_{a}$ or construct couplings of $\mu$ and $\nu$ with specific properties.

Corollary 4.3. Let $I=\left(I_{a}, a \in \mathcal{A}\right)$ be a collection of $\{0,1\}$-valued random variables and let $W=\sum_{a \in \mathcal{A}} I_{a}$. If the variables $I=\left(I_{a}, a \in \mathcal{A}\right)$ are exchangeable, then for any fixed $a \in \mathcal{A}$ and coupling $\left(W, W_{a}\right)$ of $\mu$ and $\nu_{a}$, we have

$$
\begin{equation*}
\mathrm{d}_{\mathrm{TV}}(W, \operatorname{Poi}(\lambda)) \leq \mathbb{E}\left[\left|W-\left(W_{a}-1\right)\right|\right] \tag{7}
\end{equation*}
$$

If, moreover, $W_{a}=\left(J_{a b}, b \in \mathcal{A}\right)$ and there is a coupling $\left(W, W_{a}\right)$ of $\mu$ and $\nu_{a}$ satisfying $J_{a b} \leq I_{a}$ for all $b \in \mathcal{A} \backslash\{a\}$, then

$$
\begin{equation*}
\mathrm{d}_{\mathrm{TV}}(W, \operatorname{Poi}(\lambda)) \leq \mathbb{E}\left[I_{a}\right]+\sum_{b \in \mathcal{A} \backslash\{a\}} \mathbb{E}\left[I_{a}-J_{a b}\right] \tag{8}
\end{equation*}
$$

Now, for the remainder of the section, fix $m$ and let, for all $v \in[n], I_{v}=\mathbf{1}_{\left[\mathrm{d}_{T_{n}}(v) \geq m\right]}$, so that $Z_{m}^{(n)} \stackrel{\text { dist }}{=} \sum_{v \in[n]} I_{v}$. Let $(I, J)=\left(\left(I_{v}, v \in[n]\right),\left(J_{v}, v \in[n]\right)\right.$ be a coupling of $\mu$ and $\nu=\nu_{n}$ where $\mu$ is the law of $\left(I_{1}, \ldots, I_{n}\right)$ and $\nu=\nu_{n}$ is the conditional law of $\left(I_{1}, \ldots, I_{n}\right)$ given that $I_{n}=1$.

If we would have orthant negative correlation for $\left(\mathrm{d}_{T_{n}}(v), v \in[n]\right)$ then it would follow that for all $v \in[n-1], \mathbb{E}\left[I_{n} I_{v}\right]-\mathbb{E}\left[I_{n}\right] \mathbb{E}\left[I_{v}\right] \leq 0$ and so the conditions for (8) would be satisfied. Although such strong property has not been yet established, Proposition 4.1 implies for each $v \in[n-1]$,

$$
\mathbb{E}\left[I_{n} I_{v}\right]-\mathbb{E}\left[I_{n}\right] \mathbb{E}\left[I_{v}\right] \leq O\left(2^{-2 m-\alpha \log n}\right)
$$

This suggests that there are couplings of $\mu$ and $\nu$ for which, with high probability, $I_{v} \leq J_{v}$ for all $v \in[n-1]$. The existence for such couplings is delicate as the inequality $I_{v} \leq J_{v}$ 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 1.7. 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 $(I, J)=\left(\left(I_{1}, \ldots, I_{n}\right),\left(J_{1}, \ldots, J_{n}\right)\right)$ of $\mu$ and $\nu$, in which for all $v \in[n-1]$,

$$
\mathbb{P}\left(I_{v}<J_{v}\right) \leq O\left(n^{-1-\beta}\right)
$$

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 1.7. Fix $1<c^{\prime}<c<2$ and let $c^{\prime} \ln n<m=m(n)<c \ln n$. We apply the Chen-Stein method to $Z_{m}^{(n)} \stackrel{\text { dist }}{=} \sum_{v \in[n]} I_{v}$. First, we use the coupling $(I, J)=$ $\left(\left(I_{1}, \ldots, I_{n}\right),\left(J_{1}, \ldots, J_{n}\right)\right)$ of $\mu$ and $\nu$ given in Proposition 4.4. By (7), we have

$$
\mathrm{d}_{\mathrm{TV}}\left(Z_{m}^{(n)}, \operatorname{Poi}\left(\mathbb{E}\left[\lambda_{n, m}\right]\right)\right) \leq \mathbb{E}\left[\left|W-\left(W_{n}-1\right)\right|\right] \leq \mathbb{E}\left[I_{n}\right]+\sum_{v \in[n-1]} \mathbb{E}\left[\left|I_{v}-J_{v}\right|\right]
$$

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

$$
\begin{aligned}
\mathbb{E}\left[I_{n}\right] \mathbb{E}\left[\left|J_{v}-I_{v}\right|\right] & =\mathbb{E}\left[I_{n}\right] \mathbb{E}\left[I_{v}-J_{v}\right]+2 \mathbb{E}\left[I_{n}\right] \mathbb{E}\left[\left(J_{v}-I_{v}\right) \mathbf{1}_{\left[I_{v}<J_{v}\right]}\right] \\
& =\left(\mathbb{E}\left[I_{n}\right] \mathbb{E}\left[I_{v}\right]-\mathbb{E}\left[I_{n} I_{v}\right]\right)+2 \mathbb{E}\left[I_{n}\right] \mathbb{P}\left(I_{v}<J_{v}\right)
\end{aligned}
$$

The terms in the last line are bounded by (6) and Proposition 4.4, respectively. Since (1) gives $\mathbb{E}\left[I_{n}\right]=2^{-m}(1+o(1))$ we get

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

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

Proof of Theorem 1.5. Fix $c \in(1, \log e)$ and let $\alpha=\alpha(c)$ be as in Theorem 1.7. Using the upper bound for $\alpha$ in Proposition 4.1 and simple computations yield $(1-\alpha) \log e<c$. Thus, we can chose $c^{\prime} \in((1-\alpha) \log e, c)$. Let $m=m(n)$ be such that $c^{\prime} \ln n<m<c \ln n$. By the choice of $c$ and $c^{\prime}$, we have that, as $n \rightarrow \infty,(1-\alpha) \log n-m<0$; while (1) implies

$$
\mathbb{E}\left[Z_{m}^{(n)}\right]=2^{-m+\log n}(1+o(1)) \rightarrow \infty
$$

The result then follows by Theorem 1.7 and the central limit theorem of Poisson variables, see e.g. [9, Exercise 3.4.4].
Proof of Theorem 1.10. 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 $2^{i+\varepsilon_{n}} \leq$ $2^{i+1}<2^{-C+1} \ln n$. Let $m=\lfloor\log n\rfloor-i$ and $Z \stackrel{\text { dist }}{=} \operatorname{Poi}\left(\lambda_{m, n}\right)$.

We have that $\left\{\Delta_{n}<\lfloor\log n\rfloor-i\right\}$ if and only if $\left\{Z_{m}^{(n)}=0\right\}$. Therefore,

$$
\begin{equation*}
\mathbb{P}\left(\Delta_{n}<\lfloor\log n\rfloor-i\right)=\mathbb{P}\left(Z_{m}^{(n)}=0\right) \leq \mathbb{P}(Z=0)+\mathrm{d}_{\mathrm{TV}}\left(Z_{m}^{(n)}, Z\right) \tag{9}
\end{equation*}
$$

We deal with the two terms on the right-hand side of (9) 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) gives $\gamma>0$ such that $\lambda_{n, m}=2^{i+\varepsilon_{n}}+o\left(n^{-\gamma} \ln n\right)$. Consequently,

$$
\mathbb{P}(Z=0)=\exp \left\{-\lambda_{n, m}\right\}=\exp \left\{-2^{i+\varepsilon_{n}}\right\}(1+o(1))
$$

For the second term in (9), Theorem 1.7 gives $\alpha, \beta>0$ such that

$$
\mathrm{d}_{\mathrm{TV}}\left(Z_{m}^{(n)}, Z\right)=O\left(2^{-m+(1-\alpha) \log n}\right)+O\left(n^{-\beta}\right)
$$

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

$$
\exp \left\{2^{i+\varepsilon_{n}}\right\} O\left(n^{-\beta}\right)=O\left(\exp \left\{\left(2^{-C+1}-\beta\right) \ln n\right\}\right) \rightarrow 0
$$

similarly, for $C$ large enough,

$$
\exp \left\{2^{i+\varepsilon_{n}}\right\} O\left(2^{-m+(1-\alpha) \log n}\right)=\exp \left\{2^{i+\varepsilon_{n}}\right\} O\left(2^{i-\alpha \log n}\right) \rightarrow 0
$$

The two limits above imply that $\mathrm{d}_{\mathrm{TV}}\left(Z_{m}^{(n)}, Z\right)=o\left(\exp \left\{-2^{i+\varepsilon_{n}}\right\}\right)$, completing the proof.

## 5. The coupling for the Chen-Stein Method

In this section we define and analyze 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 $n \geq 2$. For $\mathrm{h}_{n}((T, \sigma),(k, l, x))=\left(T^{\prime}, \sigma^{\prime}\right)$, we have $\mathrm{d}_{T^{\prime}}(n)=\sum_{i=k}^{n-1} x_{i}$, and for $v \in[n-1]$,

$$
\mathrm{d}_{T^{\prime}}(v)=\mathrm{d}_{T}(v)+\mathbf{1}_{[l=\sigma(v)]}-\sum_{i=k}^{n-1} x_{i} \mathbf{1}_{\left[v=p_{T}\left(\sigma^{-1}(i)\right)\right]}
$$

In words, Fact 5.1 specifies when the degree of a vertex $v<n$ changes: either for having $n$ as a new child or for losing children that are rewired towards $n$. Clearly, the degree of $n$ equals the total number of such rewirings.

The heuristic for the almost negative relation obtained in Proposition 4.4 is the following. Start with $\left(T_{n-1}, \sigma_{n-1}\right)$ and apply the Robin-Hood procedure. If the degree of vertex $n$ is large, Fact 5.1 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 $\operatorname{deg}_{T_{n}}(n) \geq m$ implies that other vertices are (slightly) less likely to satisfy $\operatorname{deg}_{T_{n}}(v) \geq m$.

For the remainder of the section, fix $n \in \mathbb{N}, c \in(1,2)$ and $m=m(n)>c \ln n$. Let ( $T_{n-1}, \sigma_{n-1}$ ) be uniformly random in $\mathcal{D}_{n-1},(K, L, X)$ be an $\mathrm{H}_{n}$-set, and ( $K^{\prime}, L^{\prime}, X^{\prime}$ ) be distributed as an $\mathrm{H}_{n}$-set conditioned to satisfy $\sum_{i=K}^{n-1} X_{i}^{\prime} \geq m$. Now, write

$$
\begin{align*}
\left(T_{n}, \sigma_{n}\right) & =\mathrm{h}\left(\left(T_{n-1}, \sigma_{n-1}\right),(K, L, X)\right),  \tag{10}\\
(\boldsymbol{T}, \boldsymbol{\sigma}) & =\mathrm{h}\left(\left(T_{n-1}, \sigma_{n-1}\right),\left(K^{\prime}, L^{\prime}, X^{\prime}\right)\right) . \tag{11}
\end{align*}
$$

To avoid cluttery notation, we omit the dependency on $m$ of the conditional random variables ( $K^{\prime}, L^{\prime}, X^{\prime}$ ) and $(\boldsymbol{T}, \boldsymbol{\sigma})$. By Fact 5.1 and Theorem 1.1, $(\boldsymbol{T}, \boldsymbol{\sigma})$ is a conditional version of $\left(T_{n}, \sigma_{n}\right)$ given that $\mathrm{d}_{T_{n}}(n) \geq m$. Consequently, if $I_{v}=\mathbf{1}_{\left[\mathrm{d}_{T_{n}}(v) \geq m\right]}$ and $J_{v}=\mathbf{1}_{\left[\mathrm{d}_{\boldsymbol{T}}(v) \geq m\right]}$ for all $v \in[n]$, then any coupling between $(K, L, X)$ and $\left(K^{\prime}, L^{\prime}, \bar{X}^{\prime}\right)$ yields a coupling for the measures $\mu$ and $\nu$ in Proposition 4.4.

Our goal is then to couple $(K, L, X)$ and $\left(K^{\prime}, L^{\prime}, X^{\prime}\right)$ 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

$$
\begin{equation*}
\mathbb{P}\left(I_{v}<J_{v}\right)=\mathbb{P}\left(\mathrm{d}_{T_{n-1}}(v)<m \leq \mathrm{d}_{\boldsymbol{T}}(v)\right)=O\left(n^{-1-\beta}\right) . \tag{12}
\end{equation*}
$$

Lemmas 5.2-5.4 provide the coupling between $(K, L, X)$ and ( $K^{\prime}, L^{\prime}, X^{\prime}$ ), 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 $n-m \leq k<n$, let $X^{k}=\left(X_{i}^{k}, i \in\right.$ $[n-1]$ ) be a conditional version of $X$ given that $\sum_{i=k}^{n-1} X_{i} \geq m$. The following observation is quite standard but we include a proof for completeness. For $a=\left(a_{1}, \ldots, a_{d}\right)$ and $b=$ $\left(b_{1}, \ldots, b_{d}\right) \in\{0,1\}^{d}, a \leq b$ only if $a_{i} \leq b_{i}$ for all $i \in[d]$. We say that $S \subset\{0,1\}^{d}$ is monotone if $a \leq b$ 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 $X_{i} \leq X_{i}^{k}$ for all $i \in[n-1]$.
Proof. Fix $k<n$. Note that $S_{k}=\left\{a \in\{0,1\}^{n-1}: a_{k}+\ldots+a_{n-1} \geq m\right\}$ is a monotone subset of $\{0,1\}^{n-1}$. Harris inequality implies $\mathbb{P}\left(X \in S \cap S_{k}\right) \geq \mathbb{P}\left(X \in S_{k}\right) \mathbb{P}(X \in S)$, for any monotone subset $S \in\{0,1\}^{n-1}$. Dividing through by $\mathbb{P}\left(X \in S_{k}\right)$ yields $\mathbb{P}\left(X^{k} \in S\right) \geq$ $\mathbb{P}(X \in S)$. 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 $\mathbb{P}(L=l \mid K=k)=\mathbb{P}\left(L^{\prime}=l \mid K^{\prime}=k\right)$. To see this, observe that $\mathbb{P}\left(L^{\prime}=l \mid K^{\prime}=k\right)$ can be rewritten as

$$
\frac{\mathbb{P}\left(L=l, K=k, \sum_{i=K}^{n-1} X_{i} \geq m\right)}{\mathbb{P}\left(K=k, \sum_{i=K}^{n-1} X_{i} \geq m\right)}=\frac{\mathbb{P}\left(L=l, K=k, \sum_{i=k}^{n-1} X_{i} \geq m\right)}{\mathbb{P}\left(K=k, \sum_{i=k}^{n-1} X_{i} \geq m\right)} ;
$$

the claim then follows by the independence between $X$ and ( $K, L$ ). Second, the sequence $p_{k}=\mathbb{P}\left(K=k \mid \sum_{i=K}^{n-1} X_{i} \geq m\right)$ is proportional to $\mathbb{P}\left(\sum_{i=k}^{n-1} X_{i} \geq m\right)$ and, thus, it is decreasing in $k$. Clearly, the latter sequence of probabilities is decreasing in $k$, while both
are proportional with a factor $Z=n \mathbb{P}\left(\sum_{i=K}^{n-1} X_{i} \geq m\right)$. To see this, use the independence between $X$ and $K$ to obtain

$$
\mathbb{P}\left(K=k \mid \sum_{i=K}^{n-1} X_{i} \geq m\right)=\frac{\mathbb{P}\left(K=k, \sum_{i=k}^{n-1} X_{i} \geq m\right)}{\mathbb{P}\left(\sum_{i=K}^{n-1} X_{i} \geq m\right)}=\frac{1}{Z} \mathbb{P}\left(\sum_{i=k}^{n-1} X_{i} \geq m\right)
$$

Lemma 5.3. There exists a coupling of $(K, L)$ and $\left(K^{\prime}, L^{\prime}\right)$ such that $K^{\prime} \leq K$ and $L^{\prime} \leq L$.
Proof. Let $X=\left(X_{1}, \ldots, X_{n-1}\right)$ be independent with $X_{i} \stackrel{\text { dist }}{=}$ 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 $\left(K^{\prime}, L^{\prime}\right)$ using the variables $U_{1}, U_{2}$ and argue that the original law is preserved.

Let $(K, L)=\left(\left\lceil n U_{1}\right\rceil,\left\lceil(K-1) U_{2}\right\rceil\right)$ and $\left(K^{\prime}, L^{\prime}\right)=\left(K^{\prime},\left\lceil\left(K^{\prime}-1\right) U_{2}\right\rceil\right)$ with

$$
K^{\prime}=\max \left\{k: U_{1}>\sum_{j=1}^{k-1} p_{j}\right\}
$$

It is straightforward that $(K, L)$ and $K^{\prime}$ have the correct law by construction, while $L^{\prime}$ has the correct law since $\mathbb{P}(L=l \mid K=k)=\mathbb{P}\left(L^{\prime}=l \mid K^{\prime}=k\right)$ for each $0 \leq l<k \leq n$. Moreover, since $p_{k}$ is decreasing, it follows that $K^{\prime}=j$ implies $U_{1}>\sum_{i=1}^{j-1} p_{i} \geq \frac{j-1}{n}$. It follows that $K \geq j=K^{\prime}$, and so $L=\left\lceil(K-1) U_{2}\right\rceil \geq\left\lceil\left(K^{\prime}-1\right) U_{2}\right\rceil=L^{\prime}$.

Lemma 5.4. There exists a coupling of $(K, L, X)$ and $\left(K^{\prime}, L^{\prime}, X^{\prime}\right)$ such that $K^{\prime} \leq K$, $L^{\prime} \leq L^{\prime}$ and $X_{i} \leq X_{i}^{\prime}$ for all $i \in[n-1]$.

Proof. Let $U_{1}, U_{2}$ be i.i.d. Unif $(0,1)$ and independently, let $X=\left(X_{1}, \ldots, X_{n-1}\right)$ be independent with $X_{i} \stackrel{\text { dist }}{=}$ Bernoulli $(1 / i)$. For each $1 \leq k<n$ 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, e.g., take them to be conditionally independent given $X$. On the other hand, it is important to insist that the $X^{k}$ are independent of $\left(K^{\prime}, L^{\prime}\right)$. Since we will define $\left(K^{\prime}, L^{\prime}\right)$ 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 $(K, L),\left(K^{\prime}, L^{\prime}\right)$ as in Lemma 5.3 and let $X^{\prime}=X^{K^{\prime}}$. Clearly, $(K, L, X)$ is an $\mathrm{H}_{n}$-set. It remains to show that $\left(K^{\prime}, L^{\prime}, X^{\prime}\right)$ has the conditional distribution of $(K, L, X)$ given that $\sum_{i=K}^{n-1} X_{i} \geq m$. For any $(k, l, x) \in \mathcal{P}_{n}$, the probability $\mathbb{P}\left((K, L, X)=(k, l, x) \mid \sum_{i=K}^{n-1} x_{i} \geq m\right)$ can be rewritten as

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

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

$$
\frac{\mathbb{P}\left(K=k, L=l, \sum_{i=k}^{n-1} x_{i} \geq m\right)}{\mathbb{P}\left(\sum_{i=K}^{n-1} x_{i} \geq m\right)} \cdot \frac{\mathbb{P}\left(X=x, \sum_{i=k}^{n-1} x_{i} \geq m\right)}{\mathbb{P}\left(\sum_{i=k}^{n-1} x_{i} \geq m\right)}
$$

These probabilities correspond, respectively, to the distributions of $\left(K^{\prime}, L^{\prime}\right)$ and $X^{k}$, which are independent. Therefore,

$$
\begin{aligned}
\mathbb{P}\left((K, L, X)=(k, l, x) \mid \sum_{i=K}^{n-1} x_{i} \geq m\right) & =\mathbb{P}\left(\left(K^{\prime}, L^{\prime}\right)=(k, l)\right) \mathbb{P}\left(X^{k}=x\right) \\
& =\mathbb{P}\left(\left(K^{\prime}, L^{\prime}, X^{\prime}\right)=(k, l, x)\right)
\end{aligned}
$$

as desired. Finally, the variables $(K, L, X)$ and $\left(K^{\prime}, L^{\prime}, X^{\prime}\right)$ 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 $\mathrm{d}_{T_{n}}(v)<m \leq \mathrm{d}_{\boldsymbol{T}}(v)$ to hold under the coupling of Lemma 5.4.

Proposition 5.5. Consider $(K, L, X)$ and $\left(K^{\prime}, L^{\prime}, X^{\prime}\right)$ defined in Lemma 5.4 and their corresponding decorated trees $\left(T_{n}, \sigma_{n}\right),(\boldsymbol{T}, \boldsymbol{\sigma})$ defined in (10) and (11). For any $v \in[n-1]$,

$$
\left\{\mathrm{d}_{T_{n}}(v)<m \leq \mathrm{d}_{\boldsymbol{T}}(v)\right\} \subset\left\{L^{\prime}=\sigma_{n-1}(v)\right\} \cap\left\{\mathrm{d}_{T_{n-1}}(v) \geq m-1\right\}
$$

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

$$
\begin{equation*}
\sum_{i=K}^{n-1} X_{i} \mathbf{1}_{\left[v=p_{T_{n-1}}\left(\sigma_{n-1}^{-1}(i)\right)\right]} \leq \sum_{i=K^{\prime}}^{n-1} X_{i}^{\prime} \mathbf{1}_{\left[v=p_{T_{n-1}}\left(\sigma_{n-1}^{-1}(i)\right)\right]} \tag{13}
\end{equation*}
$$

Consequently, using Fact 5.1 we have that $\mathrm{d}_{\boldsymbol{T}}(v)-\mathrm{d}_{T_{n}}(v) \leq \mathbf{1}_{\left[L^{\prime}=\sigma_{n-1}(v)\right]}$. On the other hand, if $\left\{\mathrm{d}_{T_{n}}(v)<m \leq \mathrm{d}_{\boldsymbol{T}}(v)\right\}$ holds, then it follows that $\mathrm{d}_{\boldsymbol{T}}(v)-\mathrm{d}_{T_{n}}(v)>0$ and so it is necessary that $\left\{L^{\prime}=\sigma_{n-1}(v)\right\}$ holds. Finally, $\left\{m \leq \mathrm{d}_{\boldsymbol{T}}(v)\right\}$ implies that

$$
m \leq \mathrm{d}_{\boldsymbol{T}}(v)=\mathrm{d}_{T_{n-1}}(v)+\mathbf{1}_{\left[L^{\prime}=\sigma_{n-1}(v)\right]}-\sum_{i=K^{\prime}}^{n-1} X_{i}^{\prime} \mathbf{1}_{\left[v=p_{T_{n-1}}\left(\sigma_{n-1}^{-1}(i)\right)\right]} \leq \mathrm{d}_{T_{n-1}}(v)+1
$$

or equivalently, that $\left\{\mathrm{d}_{T_{n-1}}(v) \geq m-1\right\}$.
We can also argue, more specifically, that

$$
\left\{\mathrm{d}_{T_{n}}(v)<m \leq \mathrm{d}_{\boldsymbol{T}}(v)\right\} \subset\left\{L^{\prime}=\sigma_{n-1}(v)\right\} \cap\left\{\mathrm{d}_{T_{n-1}}(v)=m-1\right\}
$$

however, the approach we chose allow us to use uniform bounds for all $v \in[n-1]$. We will frame the events $\left\{\mathrm{d}_{T_{n-1}}(v) \geq m-1\right\}$ 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 $\varepsilon>0, \mathbb{P}(S>(1+\varepsilon) \mathbb{E}[S]) \leq \exp \left\{-\frac{3 \varepsilon^{2}}{2(3+\varepsilon)} \mathbb{E}[S]\right\}$. By the construction of $R_{n}$ we have that $\mathrm{d}_{R_{n}}(i) \stackrel{\text { dist }}{=} \sum_{k=i}^{n} B_{k} \leq \sum_{k=1}^{n} B_{k}$ where $\left(B_{k}, k \geq 1\right)$ are independent Bernoulli variables with mean $1 / k$. Therefore,

$$
\mathbb{P}\left(\mathrm{d}_{R_{n}}(i)>m\right) \leq \mathbb{P}\left(\sum_{k=1}^{n} B_{k}>c \ln n\right)
$$

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

$$
\begin{equation*}
\mathbb{P}\left(\mathrm{d}_{R_{n}}(i)>m\right)=O\left(n^{-\beta}\right) \tag{14}
\end{equation*}
$$

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 (14). Let $T_{n}$ and $\boldsymbol{T}$ be as defined in (10) and (11) with $\left((K, L, X),\left(K^{\prime}, L^{\prime}, X^{\prime}\right)\right)$ as in Lemma 5.4. Set $I_{v}=\mathbf{1}_{\left[\mathrm{d}_{T_{n}}(v) \geq m\right]}$ and $J_{v}=\mathbf{1}_{\left[\mathrm{d}_{\boldsymbol{T}}(v) \geq m\right]}$ for all $v \in[n]$, so that $(I, J)=\left(\left(I_{1}, \ldots, I_{n}\right),\left(J_{1}, \ldots J_{n}\right)\right)$ is a coupling of the measures $\mu$ and $\nu$.

Our goal is to bound $\mathbb{P}\left(I_{v}<J_{v}\right)=\mathbb{P}\left(\mathrm{d}_{T_{n}}(v)<m \leq \mathrm{d}_{\boldsymbol{T}}(v)\right)$. First, by Proposition 5.5,

$$
\mathbb{P}\left(\mathrm{d}_{T_{n}}(v)<m \leq \mathrm{d}_{\boldsymbol{T}}(v)\right) \leq \sum_{j=1}^{n-1} \mathbb{P}\left(L^{\prime}=j, \sigma_{n-1}(v)=j, \mathrm{~d}_{T_{n-1}}(v) \geq m-1\right)
$$

Next we obtain uniform bounds for the terms on the right-hand side. Recall that $\sigma_{n-1}$ is a uniformly random permutation independent of $L^{\prime}$ and that $\sigma_{n-1}\left(T_{n-1}\right) \stackrel{\text { dist }}{=} R_{n-1}$. These facts, together with (14) gives, for each $j \in[n-1]$,

$$
\begin{aligned}
\mathbb{P}\left(L^{\prime}=j, \sigma_{n-1}(v)=j, \mathrm{~d}_{T_{n-1}}(v) \geq m-1\right) & =\frac{\mathbb{P}\left(L^{\prime}=j\right)}{n-1} \mathbb{P}\left(\mathrm{~d}_{T_{n_{1}}}(v) \geq m-1 \mid \sigma_{n-1}(v)=j\right) \\
& =\frac{\mathbb{P}\left(L^{\prime}=j\right)}{n-1} \mathbb{P}\left(\mathrm{~d}_{R_{n-1}}(j) \geq m-1\right) \\
& \leq \mathbb{P}\left(L^{\prime}=j\right) O\left(n^{-1-\beta}\right)
\end{aligned}
$$

Plugging together these bounds, we get for any $v \in[n-1]$,

$$
\begin{aligned}
\mathbb{P}\left(I_{v}<J_{v}\right) & \leq \sum_{j=1}^{n-1} \mathbb{P}\left(L^{\prime}=j, \sigma_{n-1}(v)=j, \mathrm{~d}_{T_{n-1}}(v) \geq m-1\right) \\
& =O\left(n^{-1-\beta}\right) \sum_{j=1}^{n-1} \mathbb{P}\left(L^{\prime}=j\right)=O\left(n^{-1-\beta}\right)
\end{aligned}
$$

## 6. Conclusions and further research

The Robin-Hood pruning yields an interesting process $\left(\left(T_{n}, \sigma_{n}\right), n \geq 1\right)$. By Theorem 1.1 and Proposition 1.2, $\sigma_{n}\left(T_{n}\right) \stackrel{\text { dist }}{=} R_{n}$ for all $n \geq 1$; 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 attach$\operatorname{ing} n$ to a uniformly random vertex. To see this, recall that $\mathrm{d}_{T_{n}}(n) \stackrel{\text { dist }}{=} \min \{\mathrm{Geo}(1 / 2),|\mathcal{S}|\}$ where $|\mathcal{S}| \rightarrow \infty$ (see Fact 6.1 and Lemma 6.2 ). It follows that with probability tending to $1 / 2$ the newly added vertex will be a leaf. Second, for all $n \geq 1, \mathrm{~d}_{R_{n}}(n)=0$ a.s., while Fact 5.1 and the distribution of the H -set yield

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

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 $a \in[0,1)$,

$$
\mathbb{E}\left[\mathrm{d}_{T_{n}}(n) \mid M \leq n^{a}\right] \geq \mathbb{E}\left[\sum_{i=n^{a}}^{n} X_{i}\right]=(1-a) \ln n
$$

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 $\sigma_{n}$ 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 $\left(\left(T_{n}, \sigma_{n}\right), n \geq 1\right)$, there are many interesting theoretical questions that would be worth pursuing. To name 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 maximum size of such subtree grow?
(2) Understand the maximum degree dynamics in both $\left(R_{n}, n \geq 1\right)$ and $\left(T_{n}, n \geq 1\right)$.
- How often does vertices attaining the maximum degree change?
- Are this dynamics the same for both processes?
(3) Determine whether there is a coupling for which the sequence $\left(\mathbf{1}_{\left[\mathrm{d}_{T_{n}}(v)\right]}, v \in[n]\right)$ is negative related (i.e. that conditions for (8) 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 on how to improve the presentation of the results and additional references. This research was supported by FQRNT through PBEEE scholarship with number 169888.

## Appendix A: Proof of Proposition 4.1

We use the representation of Kingman's coalescent that consists of a chain $\mathbf{C}=\left(F_{n}, \ldots, F_{1}\right)$ 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

$$
\mathcal{S}_{n}(v)=\left\{2 \leq j \leq n: T_{j}(v) \in\left\{T_{a_{j}}^{(j)}, T_{b_{j}}^{(j)}\right\}\right\}
$$

this set keeps record of the times when the tree containing $v$ merges. Finally, for each $2 \leq j \leq n$, we say that $\xi_{j}$ is favorable for vertices in $T_{a_{j}}^{(j)}$ (resp. vertices in $T_{b_{j}}^{(j)}$ ) if $\xi_{j}=1$ (resp. $\xi_{j}=0$ ).

The key property of Kingman's coalescent is the following. For each $j \in \mathcal{S}_{n}(v)$, if $\xi_{j}$ favors $v$, then $r\left(T_{j}(v)\right)$ increases its degree by one in the process; otherwise $r\left(T_{j}(v)\right)$ attaches to the root of the other merging tree and the degree of $r\left(T_{j}(v)\right)$ 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 favorable times for $v$. Moreover, $\left(\xi_{j}, j \in[n-1]\right)$ are independent and distributed as Bernoulli ( $1 / 2$ ). Therefore we have the following distributional equivalence.

Fact 6.1. Let $D$ be a random variable with distribution $\mathrm{Geo}(1 / 2)$ independent of $\mathcal{S}_{n}(v)$, then

$$
\mathrm{d}_{T^{(n)}}(v) \stackrel{\text { dist }}{=} \min \left\{D,\left|\mathcal{S}_{n}(v)\right|\right\} .
$$

This fact, together with the next lemma, allow us to get estimates for the tails of $\mathrm{d}_{T^{(n)}}(v)$.
Lemma 6.2. If $c \in(0,2)$ and $0<\varepsilon \leq 1-c / 2$. Writing $a=1-\varepsilon-c / 2$, we have

$$
\mathbb{P}\left(\left|\mathcal{S}_{n}(v) \backslash\left[n^{a}\right]\right|>c \ln n\right) \leq O(1) n^{-\varepsilon^{2} /(\varepsilon+c / 2)}
$$

Proof. First, there are $j(j-1)$ distinct pair of trees in $F_{j}$, exactly $j-1$ of such pairs contains $T_{j}(v)$; thus $\mathbb{P}\left(j \in \mathcal{S}_{n}(v)\right)=2 / j$. Since the merging trees are chosen independently at each time, we have that for any $a \in[0,1)$ we have

$$
\left|\mathcal{S}_{n}(v) \backslash\left[n^{a}\right]\right| \stackrel{\text { dist }}{=} \sum_{j=n^{a}+1}^{n} B_{j}
$$

where the variables $B_{1}, \ldots B_{n}$ are independent Bernoulli variables with $\mathbb{E}\left[B_{i}\right]=2 / i$, respectively. The desired bound is then a straightforward application of Bernstein's inequalities (see, e.g. [18], Theorem 2.8 and (2.6)). For a sum $S$ of $\{0,1\}$-valued variables, we have $\mathbb{P}(S \leq \mathbb{E}[S]-t) \leq \exp \left\{-t^{2} / 2 \mathbb{E}[S]\right\}$. 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 6.3. If $c \in(0,2)$ and $m<c \ln n$, then for $\varepsilon=(2-c)^{2} / 4$,

$$
2^{-m}\left(1-o\left(n^{-\varepsilon}\right)\right) \leq \mathbb{P}\left(\mathrm{d}_{T^{(n)}}(1) \geq m\right) \leq 2^{-m}
$$

Proof of Proposition 6.3. It follows from Lemma 6.1 that

$$
\mathbb{P}\left(\mathrm{d}_{T^{(n)}}(v) \geq m\right)=\mathbb{P}(D \geq m) \mathbb{P}\left(\left|\mathcal{S}_{n}(v)\right| \geq m\right)
$$

The upper bound on $\mathbb{P}\left(\mathrm{d}_{T^{(n)}}(1) \geq m\right)$ is then trivial, while the lower bound follows by Lemma 6.2 using $\varepsilon=1-c / 2$ and that $\mathcal{S}_{n}(v)=\mathcal{S}_{n}(v) \backslash[1]$.

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

$$
\mathbb{P}\left(\mathrm{d}_{T^{(n)}}(v) \geq m, \mathrm{~d}_{T^{(n)}}(w) \geq m,\left(\mathcal{S}_{n}(v), \mathcal{S}_{n}(w)\right)=(A, B)\right)>0
$$

Since the $\xi_{j}$ are independent of the selection times, we have that

$$
\begin{equation*}
\mathbb{P}\left(\mathrm{d}_{T^{(n)}}(v) \geq m, \mathrm{~d}_{T^{(n)}}(w) \geq m\right) \geq 2^{-2 m} \mathbb{P}\left(\left(\mathcal{S}_{n}(v), \mathcal{S}_{n}(w)\right) \in \mathcal{G}_{m}\right) \tag{15}
\end{equation*}
$$

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

$$
\tau=\max \left\{j: j \in \mathcal{S}_{n}(v) \cap \mathcal{S}_{n}(w)\right\}
$$

Lemma 6.4. For $a \in(0,1), \mathbb{P}\left(\tau>n^{a}\right) \leq 4 n^{-a}$.

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 $j \geq \tau$. Additionally, by the ordering convention of trees in $F_{j}$, it follows that $T_{j}(1)=1$ and $T_{j}(2)=2$ for all $j \geq \tau$.

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

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

This follows by induction on $n-k$. Clearly, $\tau=n$ only if $\left\{a_{n}, b_{n}\right\}=\{1,2\}$ which occurs with probability $\frac{2}{n(n-1)}$, thus $\mathbb{P}(\tau \leq n-1)$ satisfies the equation above. For $k<n$, we have

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

Next, for $k$ larger enough,

$$
\prod_{j=k+1}^{n}\left(1-\frac{2}{j(j-1)}\right) \geq \prod_{j=k}^{n-1}\left(1-\frac{2}{j^{2}}\right)>1-\sum_{j=k}^{\infty} \frac{2}{j^{2}}>1-4 \int_{k}^{\infty} x^{-2} d x=1-4 / k
$$

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

Lemma 6.5. If $c \in(0,2)$ and $m<c \ln n$, then for any $\gamma<\frac{1}{4}\left(1-c+\sqrt{1+2 c-c^{2}}\right)$,

$$
\mathbb{P}\left(\left(\mathcal{S}_{n}(v), \mathcal{S}_{n}(w)\right) \in \mathcal{G}_{m}\right) \geq 1-o\left(n^{-\gamma}\right)
$$

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

$$
\begin{equation*}
\mathbb{P}\left(\left(\mathcal{S}_{n}(v), \mathcal{S}_{n}(w)\right) \notin \mathcal{G}_{m}\right) \leq \mathbb{P}\left(\tau>n^{a}\right)+2 \mathbb{P}\left(\left|\mathcal{S}_{n}(v) \backslash\left[n^{a}\right]\right|<c \ln n\right) \tag{16}
\end{equation*}
$$

Before, establishing (16), we note that the terms in the right-hand side of (16) are bounded by Lemmas 6.4 and 6.2, respectively. Since such bounds depend on the choice of $\varepsilon$, we can use

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

The last equality 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 $1-\varepsilon-c / 2=\varepsilon^{2} /\left(\varepsilon+\frac{c}{2}\right)$.

We now proceed to establish equation (16). At step $\tau$, exactly one of $v$ and $w$ is favored by $\xi_{\tau}$. Thus, at least one of $v$ or $w$ gets its degree fixed for the remainder of the process. Therefore,

$$
\left\{\left(\mathcal{S}_{n}(v), \mathcal{S}_{n}(w)\right) \in \mathcal{G}_{m}\right\} \subset\left\{\left|\mathcal{S}_{n}(v) \backslash[\tau]\right| \geq m\right\} \cup\left\{\left|\mathcal{S}_{n}(w) \backslash[\tau]\right| \geq m\right\}
$$

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

$$
\begin{aligned}
\mathbb{P}\left(\left(\mathcal{S}_{n}(v), \mathcal{S}_{n}(w)\right) \notin \mathcal{G}_{m}\right) & \leq \mathbb{P}\left(\tau>n^{a}\right)+2 \mathbb{P}\left(\left(\mathcal{S}_{n}(v), \mathcal{S}_{n}(w)\right) \notin \mathcal{G}_{m}, \tau \leq n^{a}\right) \\
& \leq \mathbb{P}\left(\tau>n^{a}\right)+2 \mathbb{P}\left(\left|\mathcal{S}_{n}(v) \backslash[\tau]\right|<m, \tau \leq n^{a}\right) \\
& \leq \mathbb{P}\left(\tau>n^{a}\right)+2 \mathbb{P}\left(\left|\mathcal{S}_{n}(v) \backslash\left[n^{a}\right]\right|<m, \tau \leq n^{a}\right)
\end{aligned}
$$

from which (16) 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, if follows that $\mathbb{E}\left[I_{v}\right]=\mathbb{P}\left(\mathrm{d}_{T^{(n)}}(v) \geq m\right)$ and

$$
\begin{aligned}
\mathbb{E}\left[I_{v}\right] \mathbb{E}\left[J_{v}\right]=\mathbb{E}\left[I_{v} I_{n}\right] & =\mathbb{P}\left(\mathrm{d}_{T^{(n)}}(v) \geq m, \mathrm{~d}_{T^{(n)}}(n) \geq m\right) \\
& =2^{-2 m} \mathbb{P}\left(\left(\mathcal{S}_{n}(v), \mathcal{S}_{n}(n)\right) \in \mathcal{G}_{m}\right)
\end{aligned}
$$

the last equality by (15). Lemmas 6.5 and 6.3 then gives that for $\alpha<\frac{1}{4}\left(1-c+\sqrt{1+2 c-c^{2}}\right)$,

$$
\mathbb{E}\left[I_{v}\right] \mathbb{E}\left[I_{n}\right]-\mathbb{E}\left[I_{v}\right] \mathbb{E}\left[J_{v n}\right] \leq 2^{-2 m}-2^{-2 m}\left(1+o\left(n^{-\alpha}\right)\right)=2^{-2 m} o\left(n^{-\alpha}\right)
$$

## References

[1] L. Addario-Berry. Partition functions of discrete coalescents: from Cayley's formula to frieze's $\xi(3)$ limit theorem. In XI Symposium on Probability and Stochastic Processes. Progress in Probability, volume 68. Birkhauser, Basel.
[2] Louigi Addario-Berry and Laura Eslava. High degree of random recursive trees. Random Structures Algorithms, 52:560-575, 2018.
[3] C. W. Anderson. Extreme value theory for a class of discrete distributions with applications to some stochastic processes. J. Appl. Probability, 7:99-113, 1970.
[4] Blum M., Franois O and Janson, S. The mean, the variance and limiting distributions of two statistics sensitive to phylogenetic tree balance. Ann. Appl. Probability, 16:2195-2114, 2006.
[5] Huilan Chang and Michael Fuchs. Limit theorems for patterns in phylogenetic trees. J. Math. Biol., 60(4):481-512, 2010.
[6] L. Devroye. Branching processes in the analysis of the heights of trees. Acta Inform., 24(3):277-298, 1987.
[7] Luc Devroye and Jiang Lu. The strong convergence of maximal degrees in uniform random recursive trees and dags. Random Structures Algorithms, 7(1):1-14, 1995.
[8] Devdatt Dubhashi and Desh Ranjan. Balls and bins: a study in negative dependence. Random Structures Algorithms, 13(2):99-124, 1998.
[9] Richard Durrett. Probability: theory and examples. Duxbury Press, Belmont, CA, second edition, 1996.
[10] Laura Eslava. Depth of vertices with high degree in random recursive trees. arXiv:1611.07466, Submitted (2016).
[11] Philippe Flajolet and Robert Sedgewick. Analytic combinatorics. Cambridge University Press, Cambridge, 2009.
[12] Michael Fuchs. Subtree sizes in recursive trees and binary search trees: Berry-Esseen bounds and Poisson approximations. Combin. Probab. Comput., 17(5):661-680, 2008.
[13] William Goh and Eric Schmutz. Limit distribution for the maximum degree of a random recursive tree. J. Comput. Appl. Math., 142(1):61-82, 2002. Probabilistic methods in combinatorics and combinatorial optimization.
[14] Christina Goldschmidt. The chen-stein method for convergence of distributions. https://www.stats.ox.ac.uk/ goldschm/chen-stein.ps.gz, 2000.
[15] E.J. Gumbel. Les valeurs extrêmes des distributions statistiques. Ann. Inst. H. Poincaré, 5(2):115-158, 1935.
[16] Cecilia Holmgren and Svante Janson. Limit laws for functions of fringe trees for binary search trees and random recursive trees. Electron. J. Probab., 20:51 pp., 2015.
[17] Svante Janson. Asymptotic degree distribution in random recursive trees. Random Structures Algorithms, 26(1-2):69-83, 2005.
[18] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. Random graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.
[19] Torgny Lindvall. On Strassen's theorem on stochastic domination. Electron. Comm. Probab., 4:51-59 (electronic), 1999.
[20] Malwina Luczak and Peter Winkler. Building uniformly random subtrees. Random Structures Algorithms, 24(4):420-443, 2004.
[21] Hwa Sung Na and Anatol Rapoport. Distribution of nodes of a tree by degree. Math. Biosci., 6:313-329, 1970.
[22] Jim Pitman. Coalescent random forests. J. Comb. Theory A., 85:165-193, 1999.
[23] Boris Pittel. Note on the heights of random recursive trees and random $m$-ary search trees. Random Structures Algorithms, 5(2):337-347, 1994.
[24] Thomas A. Standish. Data Structure Techniques. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1980.

School of Mathematics, Georgia Institute of Technology, 686 Cherry Street NW, Atlanta, Georgia 30332-0160, USA
E-mail address: laura.eslava@math.gatech.edu
URL: http://www.people.math.gatech.edu/~leslava3/


[^0]:    Date: June 22nd 2018.
    2010 Mathematics Subject Classification. 60C05, 05C80.
    Key words and phrases. Tree growth processes, Kingman's coalescent, random recursive trees, coupling, Chen-Stein method, extreme values.

[^1]:    ${ }^{1}$ Unfortunately, it was incorrectly presumed in [21] that 'static' trees build uniformly random unrooted labeled trees.

