Properties for extreme-valued degrees in recursive trees

Laura Eslava Fernández

Doctor of Philosophy

Department of Mathematics and Statistics

McGill University

Montreal,Quebec

2016-12-15

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

DEDICATION

To Guillermo, for he has never set limits to our lifes.

ACKNOWLEDGEMENTS

My years as a graduate student at McGill University have been an enriching experience; largely because of the support and guidance I received from my supervisor, Louigi Addario-Berry.

As an undergraduate student, I came to McGill having almost no idea about the research and social aspects of academic life. Louigi showed me how to go about doing research, pointing towards a wide range of collaborations and learning opportunities. Equally important to me, he was outspoken about social issues, taking a stance and acting accordingly. Louigi was a role model, in all aspects of life, and I am deeply grateful to him.

Henning Sulzbach and Guillem Perarnau provided generous, detailed feedback at several stages of my career and I enjoyed many helpful discussions about my research. Maia Kaplan, with her never-ending energy and contagious positive view of life, provided me with essential confidence during the final excruciating months of editing the thesis. I will always be indebted to them.

Many thanks to all my friends, graduate and climbing fellows, who turned my time in Montreal into a wonderful experience. By staying in touch, friends and family in Mexico remind me that, in fact, I am not too far away from home.

Special thanks to my parents, Laura and Clemente for having my back through all my life. And to Talina Eslava, Atefeh Mohajeri, Yovanna Chacón and Lena Yudistky who made sure I felt welcome in their lives.

Finally, in so many ways, I would like to thank Guillermo. Our entangled lives have been fulfilled beyond we ever imagined, and I am fortunate to have him on my side.

PREFACE

The line of research we present in this manuscript was triggered during my attendance to the Random Networks Oberwolfach seminar in 2013. It consists of a series of three papers on the extreme-valued degrees of recursive trees which contain original research performed during my PhD studies at McGill University. The first paper of the series was written in collaboration with my supervisor Louigi Addario-Berry.

This dissertation, particularly the chapter on Depth of high-degree vertices has has been greatly benefited from discussions with Henning Sulzbach. Some of the proofs were shortened thanks to his suggestions; he also assisted me with the translation to French of the abstract. Maia Kaplan gave me a wealth of editorial advice which substantially improved the presentation of this work.

ABSTRACT

In this work we are concerned with recursive trees and linear preferential attachment trees. These are random tree growth processes; that is random networks, with no cycles that evolve with time by sequentially connecting new vertices to the existing network.

Despite having similar descriptions, there are several qualitative differences between recursive trees and linear preferential attachment trees; for example, their degree distributions and maximum degrees. Although the vertices with highest degree have been described for linear preferential attachment trees, an study for recursive trees with the same level of detail is missing in the literature.

We obtain a description, in very much detail, of the number and location of vertices with near-maximal degrees in recursive trees. From this, we comment on the qualitatively different behavior compared with linear preferential attachment trees. Additionally, we establish central limit theorems for the number of nodes with large degree, strengthen previous results about the limiting distribution of the maximum degree, and apply our methodology to raise a question about targeted cuttings in recursive trees.

ABRÉGÉ

Dans cet ouvrage, nous nous intéressons aux arbres récursifs et aux arbres d'attachement préférentiel linéaire. Ce sont des processus aléatoires de croissance des arbres; c'est-à-dire des réseaux aléatoires, sans cycles, qui évoluent avec le temps en reliant séquentiellement de nouveaux sommets au réseau existant. En dépit de descriptions similaires, il existe plusieurs différences qualitatives entre les arbres récursifs et les arbres d'attachement préférentiel linéaire; par exemple, leurs suites de degrés et leurs degrés maximaux. Bien que les sommets des degrés les plus élevés ont été décrits pour les arbres d'attachement préférentiel linéaire, une étude aussi approfondie de ceux-ci dans les arbres récursifs est absente dans la littérature.

Nous obtenons une description avec autant de détails du nombre et de l'emplacement des sommets des degrés près du maximal dans les arbres récursifs. De ce point de vue, nous soulignons le comportement qualitativement différent par rapport aux arbres d'attachement préférentiel linéaires. De plus, nous établissons des théorèmes de limite centrale pour le nombre de nœuds avec hauts degrés. Ainsi, on renforce des résultats précédents sur la distribution limite du degré maximal. Enfin, nous appliquons notre méthodologie pour élever à une question sur les attaques ciblées dans les arbres récursifs.

TABLE OF CONTENTS

LIST OF FIGURES

Part I

Background

CHAPTER 1
Introduction

Probability models allow us to have a rough (and sometimes rather precise) idea of phenomena occurring within a large, complex system whose actual evolution escapes our control or understanding; in social sciences, probability can be used to account for our inability to predict human behavior. Around the 1870s, Galton and Watson motivated what would later become the theory of branching processes [55]. Galton and Watson investigated, independent from Bienaymé's work on the same problem [72], why many family names had disappeared in spite of the growing population in England. Their model assumed that all individuals have a random number of offspring and the only parameter studied was the number of individuals at each generation. Naturally, the complexity of probabilistic models has increased over time. Today there is a vast range of research areas that utilizes combinatorial structures, mainly graphs, to describe the interactions between the elements of interest.

In this work, we are concerned with random networks; mainly the properties of exceptional elements (vertices) with many more connections (we call these the degree of a vertex) than usual. In particular we consider tree growth processes; that is, stochastic processes on graphs in which vertices are sequentially added to a graph by applying a probabilistic rule, and without creating cycles.

Our main object of study, recursive trees, was introduced by Na and Rapoport in 1970 [67]. It was a first attempt to understand how, for example, acquaintance networks are built over time. Many alternatives to these models were later studied over the years, in particular with applications in computer science. Almost 30 years later, Barabási and Albert popularized a range of processes now known as preferential attachment models [7].

The line of research during my doctoral studies was triggered by the problem of targeted attacks on random networks. To approach this problem, one needs to understand the role that highly connected elements play within the network. Although the vertices with highest degree have been described for linear PA trees [13, 19], a thorough study for those in recursive trees is missing in the literature.

This dissertation presents a series of three manuscripts describing distinct aspects of vertices with high degree in recursive trees. We specify the number and location of vertices with near-maximal degrees, and obtain central limit theorems for the number of vertices with large degree. In turn, this gives us better insight on the depth distribution of a random vertex and strengthens previous results about the limiting distribution of the maximum degree. Additionally, we explain how these results can be used to partially answer the problem of targeted attacks in recursive trees.

Our approach is based on two alternative constructions of recursive trees. The first has been known from some time; it consists of a modified version of the standard Kingman's coalescent on a finite number of elements. The second is a coupling of the aforementioned coalescent for distinct tree sizes. To the best of our knowledge, the latter introduces a novel idea which points to an unexplored feature of the growth of real-world networks.

1.1 Outline

This thesis comprises three distinct parts. Within the first one, Chapter 2 briefly introduces the tree growth processes we are considering and provides two main frameworks used for their study; namely, the increasing tree model of analytic combinatorics and age-dependent branching processes. Chapter 3 offers a panoramic view of both properties and techniques of recursive trees and linear PA trees; whenever possible, we make connections to related models of random trees and graphs. Chapter 4 precises the difficulties that arise in studying high-degree vertices in recursive trees and we present the successful approach of mapping Kingman's coalescents to recursive trees, together with its key property. Chapter 5 contains a detailed description of the results of this thesis.

There is a natural line of evolution in the main results obtained in each of the manuscripts. Therefore, Part II presents the manuscripts without further comments. Chapter 6 contains our first description on the number of vertices with near-maximal degree in recursive trees using the theory of point processes. Chapter 7, deepens our knowledge about the vertices with near-maximum degree by adding the information of their depths as marks in the point processes studied in Chapter 6. On a different note, Chapter 8 obtains precise convergence rates for the number of vertices with high degree.

Finally, Part III closes this work with an application of our results to the initial problem of targeted attacks and an outline of further avenues of research.

1.2 Notation

For $n \in \mathbb{N}$, we write $[n] = \{1, \dots, n\}$ which usually denotes the set of vertices in a graph with n vertices. We use $\ln n$ to denote natural logarithms and $\log n$ to denote logarithms base 2. For functions f, g we write $f(x) = o(g(x))$ and $f(x) = O(g(x))$ if, respectively, $\lim_{x\to\infty} |f(x)/g(x)| = 0$ and, for some $C > 0$, $\limsup_{x\to\infty} |f(x)/g(x)| \leq C$.

We use $\stackrel{\mathcal{L}}{=}$ and $\xrightarrow{\mathcal{L}}$ to denote distributional equivalence and convergence in distribution, respectively. Bernoulli and Geometric variables with parameter

$p \in [0, 1]$ are defined as follows. We say $X \overset{\mathcal{L}}{=} \text{Bernoulli}(p)$ if $\mathbf{P}(X = 1) = p = 1 - \mathbf{P}(X = 0)$; and $Y \overset{\mathcal{L}}{=} \text{Geo}(p)$ if for all integers $k \geq 0$, $\mathbf{P}(Y = k) = (1-p)p^k$.

Given a rooted labeled tree $t = (V(t), E(t))$, we write $|t| = |V(t)|$ and denote $|t|$ for the size of t. We denote the root of t by $r(t)$. An edge is denoted $e = uv$ if it connects vertices u and v, we say that e is incident to u and v.

Let $\mathcal{T}_n$ the set of rooted trees of size n with vertex set $V(t) = [n]$. In rooted trees, the edges can be naturally directed towards the root. As a convention, an edge $e = uv$ is denoted by $u \to v$ and we say that e is directed from u to v and that v is the parent of u. For a rooted tree t and $v \in V(t)$, the depth $\mathrm{h}_t(v)$ is the distance between v and $r(t)$, while the degree $\deg_t(v)$ (also denoted $\mathrm{d}_t(v)$) is the number of edges directed towards v.

CHAPTER 2
Tree growth processes

Through both continuous and discrete models, tree growth processes have historically represented population models such as genealogies and infection spreading. The book by Bacaër gives an excellent historical account for population models [6].

In the discrete setting, a tree growth process is a sequence $(t_n,\, n \geq 1)$ of rooted labeled trees with $V(t_n) = [n]$ for each n, or equivalently $t_n \in \mathcal{T}_n$ for all $n \geq 1$. Now, we denote the recursive tree process by $(T_n, n \geq 1)$, where T_n is a recursive tree of size n. Recursive trees received their name due to their construction: Start with a vertex as the single element of a tree T_1 and for $n > 1$, obtain T_n by adding to T_{n-1} a vertex labeled n and connecting it to a uniformly chosen vertex $v_n \in [n-1]$, independently for each n; see Figure 2–1.

The evolutionary properties of this definition have been useful in contexts ranging from practical social applications to probability theoretical ones. For example, mathematically proving that offers from pyramid schemes are scams

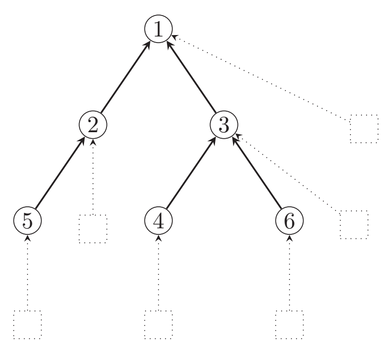


Figure 2–1: A recursive tree t on $n = 6$ vertices. Adding a new vertex to t can be done in $n - 1$ distinct ways, depicted with dotted-line squares.

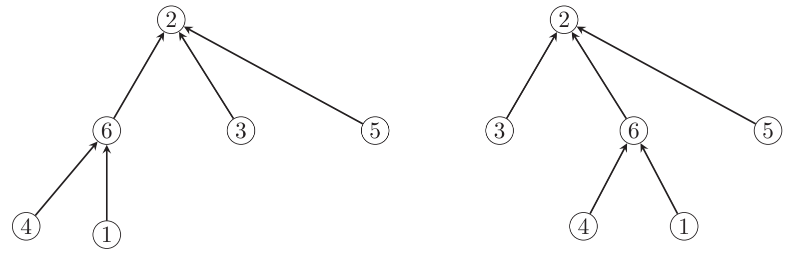


Figure 2–2: Two distinct rooted plane-oriented trees.

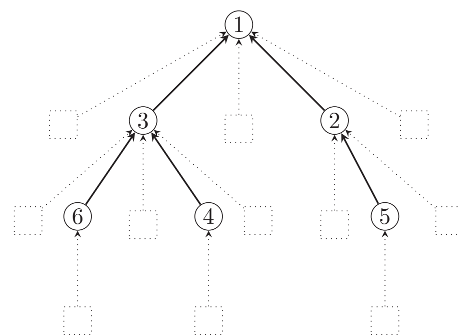


Figure 2–3: A recursive plane-oriented t tree on $n = 6$ vertices. Adding a new vertex to t can be done in $2n - 1$ distinct ways, depicted with dotted-line squares. There are, e.g., two distinct ways to add the edge $7 \to 2$.

[38], analyzing sorting and searching algorithms in computer science, and describing more complex dynamics like the Bolthausen-Sznitman coalescent [41].

Szymański introduced plane-oriented recursive trees in 1985 [79]. Plane-oriented tree are thought of being embedded in the plane (up to homeomorphisms); see Figure 2–2. Therefore, for each vertex in a tree t, the ordering of offspring is relevant. That is, if $\mathrm{d}_t(v) = j$, then the are $j!$ possible ways to draw the edges connecting v with its children in the plane. If a new child is added to v, this can be done in $j + 1$ distinct ways; see Figure 2–3.

Plane-oriented recursive trees (PA_n, $n \geq 1$), the choice of notation will be apparent shortly, are constructed in an analogous fashion as to recursive trees. To construct PA_n from PA_{n-1}, the position of the new vertex n is chosen uniformly at random among all possibilities; which are

$$\sum_{v \in [n-1]} (\deg_{PA_{n-1}}(v) + 1) = 2n - 3.$$

In particular, denoting by w_n to the parent of vertex n, then for $v \in [n-1]$,

$$\mathbf{P}\left(w_n = v \,|\, PA_m\, m \leq n-1\right) = \tfrac{\deg_{PA_{n-1}}(v)+1}{2n-3};$$

that is, the probability distribution of w_n is proportional to $\deg_{PA_{n-1}}(w_n)+1$.

Both recursive trees and plane-oriented recursive trees have similar properties in terms of depth and height. Their qualitative differences arise mainly on the degree distribution, and consequently, maximum degree. As Albert and Barabási would later describe the phenomenon, the attaching probabilities for plane-oriented trees make ‘the rich (vertices) get richer’. This endows their degrees with a power-law distribution, which is frequently observed in real-world networks.

Barabási and Albert pointed out that, compared with the ubiquitous theoretical model of Erdös-Rényi random graphs, there was still missing a growing process for graphs which would yield the qualitative properties of real-world networks [7]. In this paper of 1999, they exhibit an informal construction, which indeed, heuristically resulted on the graphs having a power-law degree distribution. Several rigorous constructions have been proposed; for random graphs, Bollobás, Riordan, Spencer and Tusnády [18] give a description of preferential attachment graphs using the ‘Linearized Chord Diagram’ which allows both loops and multiple edges.

For preferential tree growth models [21], Bubeck, Devroye and Lugosi define an ensemble of processes $(PA^\alpha_n, n \geq 1)$ as follows. Fix $\alpha \geq 0$; start with a vertex as the single element of a tree PA^α_1 and for $n > 1$, obtain PA^α_n by adding to PA^α_{n-1} a vertex labeled n, connecting it to a vertex $p_n \in [n-1]$ with probability proportional to proportional to the degrees in PA^α_{n-1} raised to the power α. The case $\alpha = 0$ yields recursive trees $(T_n,\, n \geq 1)$, also known as uniform preferential trees; while plane-oriented trees $(PA_n,\, n \geq 1)$, when

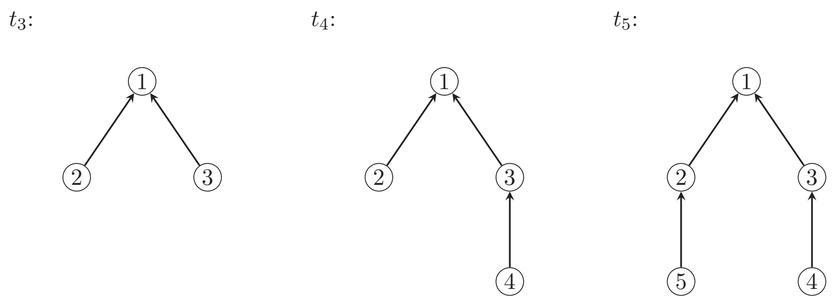


Figure 2–4: Example: t_3, t_4, t_5 of an increasing tree growth process $(t_n,\, 1 \geq 1)$.

devoid of their planar embedding, arise when $\alpha = 1$. For this reason, we slightly abuse notation and refer to plane-oriented trees as linear preferential attachment trees.

As a final example of tree growth processes, we define the ensemble of affine preferential attachment trees which encompasses both recursive trees and plane-oriented trees [75]. For each $b \geq 0$ let $(T_n^b,\, n \geq 1)$ be defined as follows. In T_{n+1}^b we attach vertex $n+1$ to vertex v_n chosen with probability proportional to $b\mathrm{d}_{T_n^b}(v_n) + 1$. Note that $b = 0$ yields recursive trees, while $b = 1$ yields linear PA trees.

In the remainder of this chapter we introduce two wide frameworks which are related to recursive trees. First, we discuss the class of increasing trees, which have a combinatorial flavour and several points of connections with computer science. Second, we define a special ensemble of age-dependent branching processes (Crump-Mode-Jagers processes), which can be regarded as the continuous version of increasing trees.

2.1 Increasing trees

For any $n \geq 1$, the class of increasing trees $\mathcal{I}_n$ contains all rooted labeled trees $t \in \mathcal{T}_n$ for which labels along any path from a leaf to the root are in increasing order. A tree growth process is increasing if t_n is a subtree of t_{n+1} for all n; this implies that $t_n \in \mathcal{I}_n$ for all n; see an instance in Figure 2–4.

Recursive trees are a stochastic example of increasing tree growth processes. Note, that a tree $t \in \mathcal{I}_n$ is uniquely determined by a tree $t' \in \mathcal{I}_{n-1}$ and $p_t(n)$, the parent of n in t. In other words, for an arbitrary $t' \in \mathcal{I}_{n-1}$, there are exactly $n-1$ distinct trees in $\mathcal{I}_n$ for which the deletion of vertex n yields t'. It follows that $|\mathcal{I}_n| = (n-1)!$. Consequently, from the construction of recursive trees $(T_n,\, n \geq 1)$, we can see that T_n has the uniform distribution in $\mathcal{I}_n$.

The fact that a tree is planar or not is not revelant for describing the depth and degree of its vertices. In a slight abuse of notation, we consider the trees in $(PA_n, n \geq 1)$ devoided of their planar embedding and thus $(PA_n, n \geq 1)$ is also a tree growth process on $\mathcal{I}_n$. However, we remark on the distribution of PA_n as a tree embedded in the plane. Let $\mathcal{IO}_n$ be the set of plane-oriented trees on n vertices; in particular, $|\mathcal{IO}_2| = 1$. Since a plane-oriented tree on n vertices can be extended to a plane-oriented tree on $n+1$ vertices in $2n-1$ distinct ways, it follows that

$$|\mathcal{IO}_n| = 1 \cdot 3 \cdot 5 \cdots (2n-3) = (2n-3)!! \tag{2.1}$$

for $n \geq 1$. As a consequence, plane-oriented recursive trees have the uniform distribution on $\mathcal{IO}_n$.

Another example of increasing tree growth processes are binary search trees and, more generally, m-ary search trees. For these, the parent of each newly added vertex is chosen independently among all vertices with degree strictly less than m. There are many other distributions for increasing growth processes; see [10, 20]. A detailed analysis of several varieties of increasing trees, planar or non-planar, can be performed through the analysis of of generating functions [71, 10]; we present this setting in terms of non-planar trees.

First, the degree-weight generating function, which is defined by

$$\varphi(x) = \sum_{k\geq 0} \varphi_k \tfrac{x^k}{k!};$$

with $\varphi_k \geq 0$ for each k. Second, an exponential generation function (EGF),

$$F(z) = \sum_{k\geq 1} f_n \tfrac{z^n}{n!};$$

where $f_n = \sum_{t\in\mathcal{I}_n} w_t$ is the sum of *weights* of trees in $\mathcal{I}_n$. The information to compute the weights is kept in the degree-weight generating function; the parameter φ_k^{-1} may be viewed as the bias towards distinct degrees in the tree (see e.g. [37, Section III.6.2, Example VII.24])

For recursive trees, there is no bias; that is $\varphi_k = 1$ for all $k \geq 0$. Therefore, $w_t = 1$ to each of the trees $t \in \mathcal{I}_n$, giving T_n the uniform distribution in $\mathcal{I}_n$. Hence its exponential generating functions are $\varphi(x) = e^x$ and

$$F_{\mathcal{T}}(z) = \sum_{n\geq 1} \tfrac{(n-1)!z^n}{n!} = -\ln(1-z).$$

Plane-oriented trees [1] have $\varphi_k = k!$ and thus $\varphi(x) = (1-x)^{-1}$. The weight w_t counts in the number of plane-oriented trees in $\mathcal{IO}_n$ that yield $t \in \mathcal{I}_n$ when forgetting the planar embedding. By (2.1), $\sum_{t\in\mathcal{I}_n} w_t = |\mathcal{IO}_n| = (2n-1)!!$ which gives

$$F_{\mathcal{PA}}(z) = 1 - \sqrt{1-2z}.$$

We note that the EGF of a tree growth process with bounded maximum degree m, such as m-ary search trees, boils down to a polynomial function. This is because the weights φ_k are set to zero for all $k > m$. It is important to note that not all increasing tree families obtained through weights $(\varphi_k)_{k\geq 0}$ can

[1] Plane-oriented trees are called heap ordered trees in [71]

be obtained from an increasing tree growth process. For example, uniformly random full-binary trees (vertices are either leaves or have two children); are obtained with the sequence $\varphi_0 = \varphi_2 = 1$ and $\varphi_k = 0$ for $k = 1$ and $k \geq 3$.

2.2 Branching processes

As we mentioned before, Galton-Walton processes analyse only the size of a population as generations come along. However a richer process is obtained if we follow the genealogy of each of the individuals in the process. The Ulam-Harris tree is designed to formalize this approach; *see* [ection 13.6, Chapter VI]Harris63.

As a convention, for any $u = (u_1, \ldots, u_n) \in \mathbb{N}^n$ and $v = (v_1, \ldots, v_m) \in \mathbb{N}^m$ we write $uv = (u_1, \ldots, u_n, v_1, \ldots, v_m) \in \mathbb{N}^{n+m}$; additionally, for any $v \in \mathbb{N}^n$, we write $\emptyset v = v\emptyset = v$. The Ulam-Harris (infinite) tree $\mathbb{T}$ is defined through its vertex set

$$\mathbb{T} = \{\emptyset\} \cup (\cup_{n\geq 1}\mathbb{N}^n)\,;$$

the edge set is implicity given by $\{v \to w \,:\, v = (v_1, \ldots, v_n) \in \mathcal{T},\, w = (v_1 \ldots, v_{n-1})\}$. We say that the *individual* $v = (v_1, \ldots, v_n) \in \mathbb{T}$ is the v_n-th child of $(v_1, \ldots, v_{n-1})$ and that $\emptyset$ is the initial ancestor.

Again, we abuse notation by defining a *proper* tree in terms of its vertex set. A subset $t \subset \mathbb{T}$ is a proper subtree of $\mathbb{T}$ if it satisfies the following two conditions. First, $\emptyset \in t$. Second, $v = (v_1, \ldots, v_n) \in t$, ; that is $(v_1, \ldots, v_{n-1}) \in t$ and $(v_1, \ldots, j) \in t$ for all $j \in [v_n]$. With these conditions we can think of proper subtrees as representing the genealogy of a family. For any individual in a proper subtree t, has its parent and 'older' sibilings are contained in the t as well.

Galton Watson trees can be defined as random proper subtrees of the Ulam-Harris tree, see for example [76]. Branching random walks assign, furthermore, a collection $\{N(v),\, v \in \mathbb{T}\}$ of *positions*. In the case we describe below each position $N(v) \in \mathbb{R}_{\geq 0}$ is interpreted as the birth time of v, and all individuals are assumed to live forever.

We introduce a particular type of age-dependent branching processes (also known as Crump-Mode-Jagers processes, see e.g. [48],[44, Chapter VI]) which serve as a continuous model for increasing tree growth processes. A thorough presentation of the applications of branching processes in the analysis of random trees is given in [28].

Let $\{\xi^v, v \in \mathcal{V}\}$ of independent copies of a point process $\xi = (\xi_i, i \geq 1)$ on $\mathbb{R}_{\geq 0}$ where $0 \leq \xi_1 \leq \xi_2 \ldots$; each ξ^v represents the birth times of all the offspring of vertex v. We remark that in general, the point process ξ^v may have a finite number of elements, say $|\xi^v| = K^v < \infty$. In that case, we use the convention that x_i is infinite for $i > K^v$ and say that the $(K^v + 1)$-th child of v is never born.

Now, the age-dependent process with birth distribution ξ is defined by the birth times $N : \mathbb{T} \to \mathbb{R}_{\geq 0}$ as follows. First, $N(\emptyset) = 0$ and, for each $v = wj \in \mathbb{T}$, set $N(v) = N(w) + \xi^w_j$. For each $s \in \mathbb{R}$, let $\mathcal{T}_\xi(s)$ be defined by

$$\mathcal{T}_\xi(s) = \{v \in \mathbb{T} : N(v) \leq s\}.$$

Next, let $Z(s) = |\mathcal{T}_\xi(s)|$ be the number of individuals alive at time s and let τ_i be the birth time of the i-th individual in the process; that is, $Z(\tau_i) = i$ but $Z(\tau_i^-) = i - 1$. Denote by V_i the individual born at time τ_i.

Now, we present an ensemble of distributions ξ_b for which the process $(\mathcal{T}_{\xi_b}(\tau_n)_{n\geq 1})$ is probabilistically equivalent to the affine PA trees $(T^b_n, n \geq 1)$. For each $b \geq 0$, let $\xi_b = (\xi_i, i \geq 1)$ be a point process whose interarrival times $(\xi_{i+1} - \xi_i, i \geq 0)$ are distributed as independent exponential variables with mean $\lambda(i) = bj + 1$ (we set $\xi_0 = 0$). For example, the point process ξ_0 is the standard Poisson point process as in this case $\lambda(i) = 1$ for all $i \geq 1$.

In what follows, fix $b \geq 0$ and write $\mathcal{T}(s) = \mathcal{T}_{\xi_b}(s)$. The analysis of the total population $|\mathcal{T}(s)|$ is facilitated by the fact that its transition rate depends only on the number of individuals in the process, that is

$$\sum_{v \in \mathcal{T}(s)} (b \cdot \mathrm{d}_{\mathcal{T}(s)}(v) + 1) = |\mathcal{T}(s)| + b \sum_{v \in \mathcal{T}(s)} \mathrm{d}_{\mathcal{T}(s)}(v) = (b+1)|\mathcal{T}(s)| - b.$$

It follows that for $j < i$,

$$\mathbf{P}\left(p_{\mathcal{T}(\tau_i)}(V_i) = V_j | \mathcal{T}(\tau_i^-)\right) = \tfrac{(b \cdot \mathrm{d}_{\mathcal{T}(\tau_i^-)}(V_j)+1)}{(b+1)(i-2)-b};$$

or in other words, the parent of V_i is chosen to be w with probability proportional to $bj + 1$, where $\mathrm{d}_{\mathcal{T}(\tau_i^-)}(w) = j$.

Therefore $(\mathcal{T}(\tau_i),\, i \geq 0)$ is an stochastic tree growth process with the desired attachement probabilities (and we can relabel vertices in increasing order of their birth times). In particular, to obtain the recursive tree process we use $b = 0$; and $b = 1$ for the linear PA process; this observation was first exploited in [74], see Theorem 3.3.1.

An important concept is the Malthusian parameter: it measures, for a wide range of branching processes, the exponential rate at which the population grows with time [68, 15]. For the age-depedendent models with ξ_b, the Malthusian parameter α_b can be explicitly computed [28, Theorem 5.2]. A consequence of this computation is the fact that, almost surely,

$$\lim_{n \to \infty} \frac{\tau_n}{\ln n} = \frac{1}{\alpha_b} = \frac{1}{1+b}. \tag{2.2}$$

CHAPTER 3
Recursive trees: Properties and techniques

There are many ways to construct and to study recursive trees; through branching processes, Pólya urns, random permutations, generating functions, renewal theory and Kingman's coalescent, to name a few. Each perspective allow us to understand distinct characteristics of recursive trees. Some of the approaches can be used to prove results for other models, including linear PA trees but also DAGs (directed acyclic graphs), split trees and increasing trees. In this chapter we briefly sample some of these techniques by reviewing main properties of depth and degrees of recursive trees and linear PA trees.

The Kingman's coalescent approach is not as widely exploited, as the other representations for recursive trees. It underlies the connections found between recursive trees and the data structure known as union-find [27, 75]. The construction of recursive trees via Kingman's coalescent is presented separatedly in Section 4.1. Aside from the work presented in this thesis, [75] seems to be the only research which uses the Kingman's coalescent representation of recursive trees.

3.1 Degree sequences and urn models

The first variables to be studied for random recursive trees were the counting variables of the degree distribution. For $m \in \mathbb{N}$, let

$$Z_m^{(n)} = \#\{v \in [n] : \mathrm{d}_{T_n}(v) = m\}.$$

Na and Rapoport [67] first studied the mean of such variables through a system of difference equations obtained by a first-step analysis. Recall that $v_{n+1} \in [n]$ denotes the parent of $n+1$. Note that at the n-th step of the process, the new vertex arrives having degree zero. Therefore, the total number of leaves in the tree either increases by one or remains the same depending on the degree of v_{n+1} in T_n. More precisely,

$$Z_0^{(n+1)} = Z_0^{(n)} + 1 - \mathbf{1}_{[\mathrm{d}_{T_n}(v_{n+1})=0]};$$

and in general, for $m \geq 1$, we have

$$Z_m^{(n+1)} = Z_m^{(n)} + \mathbf{1}_{[\mathrm{d}_{T_n}(v_{n+1})=m-1]} - \mathbf{1}_{[\mathrm{d}_{T_n}(v_{n+1})=m]}.$$

Therefore, taking expectations we get

$$\begin{aligned}\mathbf{E}\left[Z_0^{(n+1)}\right] &= 1 + \left(1 - \tfrac{1}{n}\right)\mathbf{E}\left[Z_0^{(n)}\right],\\ \mathbf{E}\left[Z_m^{(n+1)}\right] &= \left(1 - \tfrac{1}{n}\right)\mathbf{E}\left[Z_m^{(n)}\right] + \tfrac{1}{n}\mathbf{E}\left[Z_{m-1}^{(n)}\right].\end{aligned}$$

Using that $Z_0^{(1)} = Z_0^{(2)} = 1$, we have, $\mathbf{E}\left[Z_0^{(n)}\right] = n/2$ for all $n \in \mathbb{N}$. Solving the remaining system of equations yields the next proposition.

Proposition 3.1.1 ([67]). *For all $m \in \mathbb{N}$, as $n \to \infty$*

$$\lim_{n\to\infty} \frac{\mathbf{E}\left[Z_m^{(n)}\right]}{n} = 2^{-m-1}.$$

More detailed information about the joint distributions of $(Z_0^{(n)}, Z_1^{(n)}, Z_2^{(n)})$ appeared in increments until the work of Mahmoud and Smythe, where the limit was proven to be asymptotically normal with an explicit covariance matrix [62]. For a detailed account of this history, see their survey [77, Section 3.3]. Mahmoud and Smythe observed that their technique could, in principle, yield the joint limiting distribution of the vector $(Z_m^{(n)}, m \geq 0)$. Their idea was to use the theory of Pólya urns, also known as generalized Friedman urns; see e.g. [61].

Let $k \in \mathbb{N}$ be fixed and consider a generalized Pólya urn with $k+1$ types (colours) of balls. Let $B(n) = (B_0(n), \ldots, B_k(n))$ describe the urn at time n

by setting

$$B_j(n) = \#\{\text{Balls with color } j \text{ at time } n\}.$$

The urn represents the vertices in T_n. The balls are colored according to their degree; having the first k colours for the degrees zero to $k-1$ and the last colour for all the vertices with degree at least k. The dynamic of the urn is defined so that when we select a random vertex v_{n+1} to be the parent of vertex $n+1$, then for $j \le k-1$,

$$\mathbf{P}\left(\mathrm{d}_{T_n}(v_{n+1}) = j\right) = \frac{B_j}{n}$$
$$\mathbf{P}\left(\mathrm{d}_{T_n}(v_{n+1}) \ge k\right) = \frac{B_k}{n}.$$

More precisely, at each step, regardless of the selected ball, we always add one ball with the colour zero; this accounts for the newly added vertex having zero degree. If we take a ball with colour $j < k$, then we take it out and return one of colour $j+1$; that is, if $\mathrm{d}_{T_n}(v_n) = j$ then $\mathrm{d}_{T_{n+1}}(v_n) = j+1$. The case $j = k$ is distinct, we put back the ball coloured k as $\mathrm{d}_{T_{n+1}}(v_n) = \mathrm{d}_{T_n}(v_n) + 1 \ge k$. The last case arises since we are only considering a finite number of colours.

In this way we have,

$$(Z_m^{(n)}, 0 \le m \le k-1) \stackrel{\mathcal{L}}{=} (B_0(n), \ldots, B_{k-1}(n)). \tag{3.1}$$

The limiting distribution of the infinite vector $(Z_m^{(n)}, m \ge 0)$ required new results on generalized Pólya urns, which were established by Janson [49, 50]. Interestingly, Janson's work is based on the branching process perspective of Pólya urns by Athreya and Karlin [5].

Theorem 3.1.2 ([50]). *As $n \to \infty$, $n^{-1} Z_m^{(n)} \to 2^{-m-1}$ a.s., and*

$$n^{-1/2}(Z_m^{(n)} - 2^{-m-1} n) \xrightarrow{\mathcal{L}} Z_m,$$

jointly for all $m \geq 0$, where the Z_m are jointly Gaussian variables with zero means and an explicit covariance matrix is given.

A similar analysis for linear PA trees can be performed. For $m \geq 0$, let

$$Y_m^{(n)} = \#\{v \in [n] : \mathrm{d}_{PA_n}(v) = m\}. \tag{3.2}$$

In this case, the balls in the urn do not directly represent vertices in the tree. Rather, if the urn has $k+1$ types of balls, then for $0 \leq j < k$,

$$B_j(n) = (j+1)Y_j^{(n)};$$

and $B_k(n) = \sum_{i \geq k}(i+1)Y_i^{(n)}$. The dynamics of the urn are now designed so that at each step, each vertex v is represented by either $\mathrm{d}_{PA_n}(v) + 1$ balls of type j if $\mathrm{d}_{PA_n}(v) = j < k$, or $\mathrm{d}_{PA_n}(v) + 1$ balls of type k if $\mathrm{d}_{PA_n}(v) \geq k$. That is, at each step we remove a ball of type j; following we add one ball of colour zero, remove j balls of type j, and replace $j+2$ balls of type $j+1$ or type k according to $j < k$ or $j = k$, respectively.

Theorem 3.1.3 ([50])**.** *As $n \to \infty$, $n^{-1}Y_m^{(n)} \to 4/(m+1)(m+2)(m+3)$ a.s., and*

$$n^{-1/2}(Y_m^{(n)} - 4n/(m+1)(m+2)(m+3)) \xrightarrow{\mathcal{L}} Y_m,$$

jointly for all $m \geq 0$, where the Y_m are jointly Gaussian variables with zero means and an explicit covariance matrix is given.

One restriction of the Pólya urn results exploited in [50] is that we are required to work with a finite number of variables $(Z_m^{(n)}, 0 \leq m < k)$. This does not represent an obstacle to study the joint distribution of the degree sequence $(Z_m^{(n)}, m \geq 0)$. However, one of the goals of the work in this thesis was to understand the distribution of $Z_m^{(n)}$ with $m = m(n) \to \infty$, which cannot be approached through this method.

3.2 Depth, renewal theory and records

Recall that the depth of a vertex v in a tree t is the distance from v to the root. For recursive trees, and increasing tree growth models in general, the depth $\mathrm{h}_{T_n}(n)$ has been referred to as the insertion depth. This variable is important for trees representing data structures, such as recursive trees and binary search trees.

For recursive trees, $\mathrm{h}_{T_n}(n)$ was first studied in [27, 59]; however, we follow the approach of [30], which provides a general framework to cover a wider class of random trees.

Theorem 3.2.1 ([30])**.** *As $n \to \infty$, $\mathrm{h}_{T_n}(n)/\ln n \to 1$ in probability and*

$$\frac{\mathrm{h}_{T_n}(n) - \ln n}{\sqrt{\ln n}} \xrightarrow{\mathcal{L}} N(0,1).$$

Proof sketch of Theorem 3.2.1. Let us shift the labels in T_n so that $V(T_n) = \{0, 1 \ldots, n-1\}$ and $v_{n+1} \stackrel{\mathcal{L}}{=} \mathrm{Unif}\{0, \ldots, (n-1)\}$. This is convenient as we can then sample the random vertices $(v_n,\, n \geq 1)$ using i.i.d. random variables $X_i \stackrel{\mathcal{L}}{=} \mathrm{Unif}(0,1)$ by setting $v_{n+1} = \lfloor nX_n \rfloor$.

Now, denote the path from n to the root by $(w_0 = n+1, w_1, \ldots w_k = 1)$; in particular, $\mathrm{h}_{T_{n+1}}(n+1) = k$. It follows that for each $j \in [k]$,

$$v_{w_{j-1}} = w_j = \lfloor w_{j-1} X_{w_{j-1}} \rfloor.$$

Furthermore,

$$k = \min\{j \in \mathbb{N} : \lfloor w_{j-1} X_{w_{j-1}} \rfloor = 0\} = \min\{j \in \mathbb{N} : \lfloor \lfloor \lfloor nX_{w_0} \rfloor X_{w_1} \rfloor \cdots X_{w_{j-1}} \rfloor < 1\}.$$

On the other hand, for each $j \in \mathbb{N}$

$$nX_{w_0} \cdots X_{w_{j-1}} - j < \lfloor \lfloor \lfloor nX_{w_0} \rfloor X_{w_1} \rfloor \cdots X_{w_{j-1}} \rfloor \leq nX_{w_0} \cdots X_{w_{j-1}}.$$

or in other words, the aim is to estimate the distribution of $k = k(n)$, such that

$$\ln n - \ln(1+k) < \sum_{j=1}^{k} (-\log X_j) \leq \ln n;$$

Note that $-\log X_j = \mathrm{Exp}(1)$ and so $\mathbf{E}\left[\mathrm{h}_{T_{n+1}}\right] = E(k) \sim \ln n$. The result then follows by an application of the renewal theory, see e.g. [35, Exercise 3.4.7]. □

This approach is used to generalize the result to distinct types of tree growth processes in which the attaching rule gives $v_{n+1} = \lfloor nX_n \rfloor$ with i.i.d. random variables X_i such that $\mathbf{Var}\left[-\ln X_i\right] < \infty$, see [30].

An interesting proof of Theorem 3.2.1 is through records of i.i.d. variables. Let $(U_i,\, i \geq 1)$ be i.i.d. random variables have continuous distribution and for each $k \in \mathbb{N}$, let $\sigma_k : [k] \to [k]$ satisfy $U_{\sigma_k(1)} > U_{\sigma_k(2)} > \cdots U_{\sigma_k(k)}$ (with probability 1, there are no ties between the random variables). By symmetry, σ_k is uniformly random among all permutations of $[k]$.

Alternative proof sketch of Theorem 3.2.1. Let $M_k = \arg\max_{i \in [k]} U_i$, for each $k \in \mathbb{N}$. Then, using σ_k,

$$M_i = \sigma_k^{-1}(1) \stackrel{\mathcal{L}}{=} \mathrm{Unif}\{1, \ldots, i\}.$$

Now, we say that a record occurs, at time k, if $M_k = k$. Devroye observed in [27], that writing $B_i = \mathbf{1}_{[M_i = i]}$ implies that

$$\mathrm{h}_{T_n}(n) \stackrel{\mathcal{L}}{=} \sum_{i=1}^{n-1} B_i. \tag{3.3}$$

The sum in the right of this distributional equivalence exhibits an asymptotically normal distribution; and so (3.3) establishes Theorem 3.2.1. Surprisingly $(B_i,\, i \in [n-1])$ is a vector of independent Bernoulli random variables, which in addition, satisfy the Lindeberg-Feller conditions for the central limit theorem; see e.g. [35, Theorem 3.4.5]. We close the section by providing a proof of this claim.

First, since the permutations σ_i are uniformly random, we have $\mathbf{P}\left(B_i = 1\right) = 1/i$. Second, for any permutation $\pi : [i] \to [i]$ and $j > i$,

$$\mathbf{P}\left(X_j = 1 | \sigma_i = \pi\right) = \frac{\binom{j-1}{i}(j-i-1)!}{j!/i!} = \frac{1}{j};$$

this follows from three facts about the rankings σ_j: there are $j!/i!$ distinct permutations of $[j]$ that preserve the partial rankings of π for the first i variables, $(j-i-1)!$ possible ways to rank the $j-i$ variables $U_{i+1}, \ldots, U_j$ so that $U_j > U_l$ for $i < l < j$, and $\binom{j-1}{i}$ ways to interlace the first i rankings with the last ones so that U_j is a record.

Thus, that $(X_i,\, i \in [n])$ are independent follows from the fact that for all $i < j$,

$$\mathbf{E}\left[X_j = 1 | \sigma_i\right] = \frac{1}{i!} \sum_{\pi:[i]\to[i]} \mathbf{P}\left(X_j = 1 | \sigma_i = \pi\right) = \frac{1}{j}.$$

Finally, the Linderberg conditions are satisfied since the variables properly renormalized to $\hat{B}_{n,i} = (B_i - 1/i)\ln^{-1/2} n$ satisfy,

$$\sum_{i=1}^{n} \mathbf{E}\left[\hat{B}_{n,i}^2\right] = \ln^{-1} n \sum_{i=1}^{n} \mathbf{Var}\left[B_i^2\right] = \ln^{-1} n \sum_{i=1}^{n} \left(\tfrac{1}{i} - \tfrac{1}{i^2}\right) = 1 + o(1).$$

Additionally, $|\hat{B}_{n,i}| \to 0$ uniformly over i, as $n \to \infty$. Therefore for each $\varepsilon > 0$ there exists sufficiently large n such that $\sum_{i\in[n]} \mathbf{E}\left[\hat{B}_{n,i}^2 \mathbf{1}_{[|\hat{B}_{n,i}|>\varepsilon]}\right] = 0$. $\square$

Both arguments above are not suitable for linear PA trees as the attachment probabilities depend on the degree of the vertices and not only on their labels. Instead, Mahmoud uses the analysis of generating functions to obtain the corresponding result for linear PA trees.

Theorem 3.2.2 ([60]). *As* $n \to \infty$, $\mathbf{E}\left[\mathrm{h}_{PA_n}(n)\right] / \ln n \to 1/2$ *and*

$$\frac{\mathrm{h}_{PA_n}(n) - (1/2)\ln n}{\sqrt{(1/2)\ln n}} \xrightarrow{\mathcal{L}} N(0,1).$$

3.3 Height and branching processes

One of the first results about the maximum depth of a recursive tree

$$H_n = \max\{\mathrm{h}_{T_n}(v),\, v \in [n]\}$$

was given by Szymański in [80]. He showed that

$$\mathbf{P}\left((1-\varepsilon)\ln n < H_n < e \ln n\right) \to 1.$$

Devroye and Pittel used branching processes to obtain that, in probability, $e \ln n$ is the right order of H_n [26, 75]. Other proofs using split trees appear in [29, 20]. We present here the formulation of [75].

Consider the random increasing tree T_n^b, $b \geq 0$ (defined before Section 2.1), and denote the height of T_n^b by H_n^b. Recall that $T_n \overset{\mathcal{L}}{=} T_n^0$ and $PA_n \overset{\mathcal{L}}{=} T_n^1$.

Theorem 3.3.1 ([75]). *For each* $b \geq 0$*, let* $\gamma = \gamma(b) > 0$ *be the positive root of* $b\gamma + \ln\gamma + 1 = 0$*. Then, with probability one,*

$$\lim_{n\to\infty} \frac{H_n^b}{\ln n} = ((1+b)\gamma)^{-1}.$$

It follows that $\gamma^{-1} = e$ for recursive trees, while for linear PA trees, $(2\gamma)^{-1} \approx 1.79$.

Proof sketch. This proof uses the age-dependent branching process with the associated tree process $\mathcal{T} = (\mathcal{T}_{\xi_b}(s),\, s \geq 0)$; see Section 2.2. A very well-known parameter is the minimal position B_k of $\mathcal{T}$ at each generation $k \in \mathbb{N}$. Precisely, B_k is the time at which the first member of the k-th generation in $\mathcal{T}$ is born. A law of large numbers for the value B_k/k was studied, under several conditions, by Hammersley, Kingman and Biggins in [14, 56, 43]. In particular, see e.g. [28, Theorem 5.1], it shows that there is $\gamma = \gamma(\xi_b)$ defined as in Theorem 3.3.1 such that, with probability one,

$$\lim_{k\to\infty} \tfrac{B_k}{k} = \gamma. \tag{3.4}$$

On the other hand, note that if the first member of the k-th generation in $\mathcal{T}$ is born before τ_n, then the tree $\mathcal{T}(\tau_n)$ has at least one element at depth k, that is, $\{H_n \geq k\} = \{B_k \leq \tau_n\}$. Also,

$$B_{H_n} \leq \tau_n \leq B_{H_n+1}. \tag{3.5}$$

Now, by (2.2), $\tau_n \to \infty$ as $n \to \infty$; therefore, $H_n \to \infty$ and

$$\lim_{n\to\infty} \tfrac{B_{H_n}}{H_n} = \lim_{n\to\infty} \tfrac{\tau_n}{H_n} = \gamma.$$

Finally, using both (2.2) and (3.4) we get

$$\lim_{n\to\infty} \tfrac{\ln n}{H_n} = \lim_{n\to\infty} \tfrac{\tau_n}{H_n} \cdot \tfrac{\ln n}{\tau_n} = (1+b)\gamma.$$

□

3.4 The maximum degree and analytic combinatorics

In [80], Szymański also gives bounds for the maximum degree of recursive trees $\Delta_n = \max\{\mathrm{d}_{T_n}(v),\, v \in [n]\}$; showing that, as $n \to \infty$,

$$\mathbf{P}\left((1-\varepsilon)\ln n < \Delta_n < (1+\varepsilon)\log n\right) \to 1. \tag{3.6}$$

The lower bound follows simply from the fact that the degree of the root is asymptotically normal and has $\mathbf{E}\left[\mathrm{d}_{T_n}(1)\right] = \ln n + O(1)$. On the other side, if the limit in Theorem 3.1.1 were to hold for $m = m(n) \to \infty$, then having

$$\mathbf{E}\left[Z^{(n)}_{\lfloor \log n+1 \rfloor}\right] \approx 1$$

would suggest that the maximum degree is of order $\log n$. Devroye and Lu [31] showed that this is correct. Their method involves upper bounds on $\mathbf{P}\left(\mathrm{d}_{T_n}(v) \geq m\right)$ which have to be tailored for disticnt ranges of $v \in [n]$.

Theorem 3.4.1 ([31])**.** *As* $n \to \infty$, $\mathbf{E}\left[\Delta_n / \log n\right] \to 1$ *and*

$$\lim_{n\to\infty} \frac{\Delta_n}{\log n} = 1 \ a.s.$$

An analogous limit holds for the maximum degree in uniform random recursive DAGs. In a DAG, vertices are sequentially added and connected to r uniformly chosen vertices in the current tree. Then Theorem 3.4.1 holds for the maximum degree of a DAG by replacing $\log n$ with $\log_{1+1/r} n$ [31].

The convergence in L^1 can be proved using analytic combinatorics, a proof sketch can be found in [32, following Theorem 6.12]. The upper bound is obtained by a simple argument. For each $k \geq 1$, write

$$Z_{\geq k}^{(n)} = \#\{v \in [n] : \mathrm{d}_{T_n} \geq k\}.$$

Since $\{\Delta_n \geq j\} = \{Z_{\geq j}^{(n)} > 0\}$, Markov's inequality yields

$$\mathbf{P}\left(\Delta_n \geq j\right) = \mathbf{P}\left(Z_{\geq j}^{(n)} > 0\right) \leq \mathbf{E}\left[Z_{\geq j}^{(n)}\right].$$

Thus,

$$\mathbf{E}\left[\Delta_n\right] = \sum_{j\geq 0} \mathbf{P}\left(\Delta_n \geq j\right) \leq \log n + \sum_{j>\log n} \mathbf{E}\left[Z_{\geq j}^{(n)}\right] = \log n + O(1).$$

The last equality holds by an explicit uniform bound for the error in approximating $\mathbf{E}\left[Z_{\geq k}^{(n)}\right]$. Precisely, by [32, Lemma 6.14], uniformly for all $k \in \mathbb{N}$,

$$\mathbf{E}\left[Z_{\geq k}^{(n)}\right] = \tfrac{n}{2^k} + O\left(\tfrac{(\ln n)^k}{n(k!)}\right);$$

then, using $k = \lceil \log n \rceil$, and Stirling's formula, we get $\sum_{j\geq k} \frac{(\ln n)^j}{n(j!)} = O(1)$.

Another remarkable achievement of the analysis of singularities for generating functions is the approximation of the limiting distribution of Δ_n by Goh and Schmutz.

Theorem 3.4.2 ([39])**.** *For $d \in \mathbb{N}$ fixed,*

$$\mathbf{P}\left(\Delta_n < \lfloor \log n \rfloor + d\right) = \exp\{-2^{\log n - \lfloor logn \rfloor - d}\} + o(1).$$

We briefly explain the analytic combinatoric setup for the proof of Theorem 3.4.2.

Proof sketch of Theorem 3.4.2. Fix $k \in \mathbb{N}$. Let $\mathcal{T}_{n,k}$ be the class of increasing trees with maximum degree at most k and let $y_{n,k} = |\mathcal{T}_{n,k}|$. The EGF we analyze is

$$Y_k(z) = \sum_{n \geq 1} y_{n,k} \tfrac{z^n}{n!}.$$

A standard technique in analytic combinatorics is noting that, under some conditions on the combinatorial class, differentiation of generating functions corresponds to deleting an element within the the given structure. In this case, consider $t \in \mathcal{T}_{n,k}$, a tree on n vertices and maximum degree k. By deleting the root of the tree t, we are left with a sequence of j increasing trees with degree at most k; moreover, we have the additional restriction that $j \leq k$. Therefore, we get

$$Y_k'(z) = \sum_{n \geq 0} y_{n+1,k} \tfrac{z^n}{n!} = \textstyle\sum_{j=0}^{k} \tfrac{Y_k(z)^m}{m!}. \tag{3.7}$$

On the other hand, using Cauchy's integral formula we have that

$$\mathbf{P}\left(\Delta_n \leq k\right) = \tfrac{y_{n,k}}{(n-1)!} = \tfrac{n}{2\pi i} \oint_{C_k} \tfrac{Y_k(z)}{z^{n+1}} dz;$$

where C_k is a suitable circle inside the radius of convergence of $Y_k(z)$. The rest of the proof follows from (3.7) and relies heavily on asymptotic approximations for the partial sums of the exponential series. $\square$

We conclude this overview on recursive trees with one of the starkest differences with linear PA trees. The maximum degree Δ_n^{PA} of PA_n is of a different order of magnitude.

Theorem 3.4.3 ([66])**.** *There exists a finite random variable D with absolutely continuous distribution such that, as $n \to \infty$,*

$$\lim_{n\to\infty} \frac{\Delta_n^{PA}}{\sqrt{n}} = D \ \ a.s.$$

Moreover, with probability tending to one, there is exactly one vertex attaining the maximum degree and its label is bounded; see [13]. The work in [19] gives a description of high-degree vertices in linear PA trees, that is, Y_m with $m = m(n) \to \infty$.

In these series of manuscripts, we develop techniques to study the behaviour of both near-maximum degree and high-degree vertices for recursive trees. Our findings show that their properties are remarkably different compared with linear PA trees. Most importantly, the labels of near-maximum degree vertices are constantly changing, see Theorem 5.1.2 and the remarks afterward. Chapter 5 provides a complete description of our results.

CHAPTER 4
Object of study and challenges

In this collection of papers we focus on describing both the depth and the number of vertices with high and near-maximal degree in random recursive trees. More precisely, vertices with $\mathrm{d}_{T_n}(v) \geq c \ln n$ for some $c > 0$ and with $\mathrm{d}_{T_n}(v) = \lfloor \log n \rfloor + b$ for some integer $b \in \mathbb{Z}$, respectively.

Our investigation faced two important challenges; first, the distribution of each vertex degree changes quite drastically, for each $j \in [n]$,

$$\mathrm{d}_{T_n}(j) \stackrel{\mathcal{L}}{=} \sum_{i=j}^{n-1} B_i; \tag{4.1}$$

where the variables $B_i \stackrel{\mathcal{L}}{=} \text{Bernoulli}(1/i)$ are independent. It follows that only for j fixed or slowly tending to infinity we have $\mathrm{d}_{T_n}(j)$ being asymptotically normally distributed with mean $\mathbf{E}\left[\mathrm{d}_{T_n}(j)\right] = (1 - o(1)) \ln n$. Moreover, these degrees are correlated, although this correlation is weak between any bounded number of vertices.

Second, the depth of a vertex is determined at its arrival to the network, while its degree depends only on the process afterwards. Through the standard construction of recursive trees (see Section 2.1), it is not clear how conditioning on a given vertex having large or near-maximum degree is changing the global distribution of the rest of the tree, particularly, the depth of such a vertex.

To overcome the difficulties of studying high-degree vertices, we turn to the Kingman's representation of recursive trees, which we describe in the next section. In addtion, the convergence rates obtained in Chapter 8 required the introduction of a distinct tree growth process, which couples Kingman's coalescents on $[n]$, for all finite n. To the best of our knowledge, this is a novel representation of Kingman's coalescent dynamics.

The connection between recursive trees and Kingman's coalescent is central to understanding the close relation between degree and depth of vertices in recursive trees; see Lemma 4.1.5. Given the importance of such point of view in obtaining our results, we extracted from the manuscripts the definition of Kingman's coalescent and the key observation on degree and depth of a given vertex.

4.1 A Kingman's coalescent approach

In this section we give a representation of Kingman's coalescent in terms of labeled forests and connect this with recursive trees. For a general description of Kingman's coalescent, see [9, Chapter 2]; the construction below is based on that given in [1]. Recall that we write $\mathrm{d}_t(v)$ and $\mathrm{h}_t(v)$ for the degree and depth of vertex v in a tree t.

A forest f is a set of trees whose vertex sets are pairwise disjoint. Denote by $V(f)$ and $E(f)$, respectively, the unions of the vertex and edge sets of the trees contained in f. For each $n \geq 1$, we consider the set of forests $\mathcal{F}_n = \{f : V(f) = [n]\}$ with vertex labels $[n]$. An n-chain is a sequence $C = (f_n, \ldots, f_1)$ of elements of $\mathcal{F}_n$ if for $1 < i \leq n$, f_{i-1} is obtained from f_i by adding an edge connecting two of the roots in f_i. In particular, f_n contains n one-vertex trees, and f_1 contains exactly one tree denoted by $t_C \in \mathcal{F}_n$.

For an n-chain $(f_n, \ldots, f_1) \in \mathcal{CF}_n$ and $1 \leq i \leq n$, we always list the trees in $f_i = \{t_1^{(i)}, \ldots, t_i^{(i)}\}$ in increasing order of their smallest-labeled vertex.

Definition 4.1.1. *The following constructs Kingman's n-coalescent as a random n-chain $\mathbf{C} = (F_n, \ldots, F_1)$.*

For each $1 < i \leq n$, choose $\{a_i, b_i\} \subset \{\{a,b\} : 1 \leq a < b \leq i\}$ independently and uniformly at random; also let $(\xi_i,\, i \in [n-1])$ be a sequence of independent Bernoulli(1/2) *random variables.*

For $1 \leq i < n$, F_i is obtained from $F_{i+1} = \{T_1^{(i+1)}, \ldots, T_{i+1}^{(i+1)}\}$ as follows. Add an edge e_i between the roots of $r(T_{a_{i+1}}^{(i+1)})$ and $r(T_{b_{i+1}}^{(i+1)})$; direct e_i towards $r(T_{a_{i+1}}^{(i+1)})$ if $\xi_i = 1$, and towards $r(T_{b_{i+1}}^{(i+1)})$ otherwise. Then F_i contains the new tree and the remaining $i-1$ unaltered trees from F_{i+1}.

For an example of the process see Figure 4–1.

Lemma 4.1.2. *Kingman's n-coalescent $\mathbf{C}$ is uniformly random in $\mathcal{CF}_n$, the set of n-chains.*

Proof. Any $(f_n, \ldots, f_1) \in \mathcal{CF}_n$ is determined by the order in which the edges of t_C are added. For each $2 \leq i < n$, there are $(i+1)i$ possible oriented edges between the roots in f_{i+1} and only one of them is $e \in E(f_i) \setminus E(f_{i+1})$. Thus,

$$\mathbf{P}\left((F_n, \ldots, F_1) = (f_n, \ldots, f_1)\right) = \frac{\prod_{k=1}^{n-1} \mathbf{P}\left(F_k = f_k | F_j = f_j,\, k < j \leq n\right)}{n!(n-1)!}.$$

This expression holds for all $(f_n, \ldots, f_1) \in \mathcal{CF}_n$, so the result follows. □

Let $e_{n-1}, \ldots, e_1$ be the edges of t_C ordered as they were added to the chain C. That is, $e_i \in E(F_i)$ while $e_i \notin E(F_{i+1})$ for all $1 \leq i < n$. Now, write $e_i = v_i w_i$. Let $\sigma_C : V(t_C) \to [n]$ be defined as $\sigma_C(r(t_C)) = 1$ and for each $e_i = v_i w_i \in E(t_C)$,

$$\sigma_C(v_i) = i + 1.$$

This is well defined as all edges are directed towards the root, so $v_i \neq v_j$ for all $i, j \in [n-1]$. Note that for each $1 \leq i < n$, e_i is directed towards the root of the new tree in f_i. Thus, the labels $\{\sigma_C(v),\, v \in [n]\}$ decrease along leaf-to-root paths in t_C. As a consequence, we obtain an increasing tree by relabeling the vertices of t_C using σ_C.

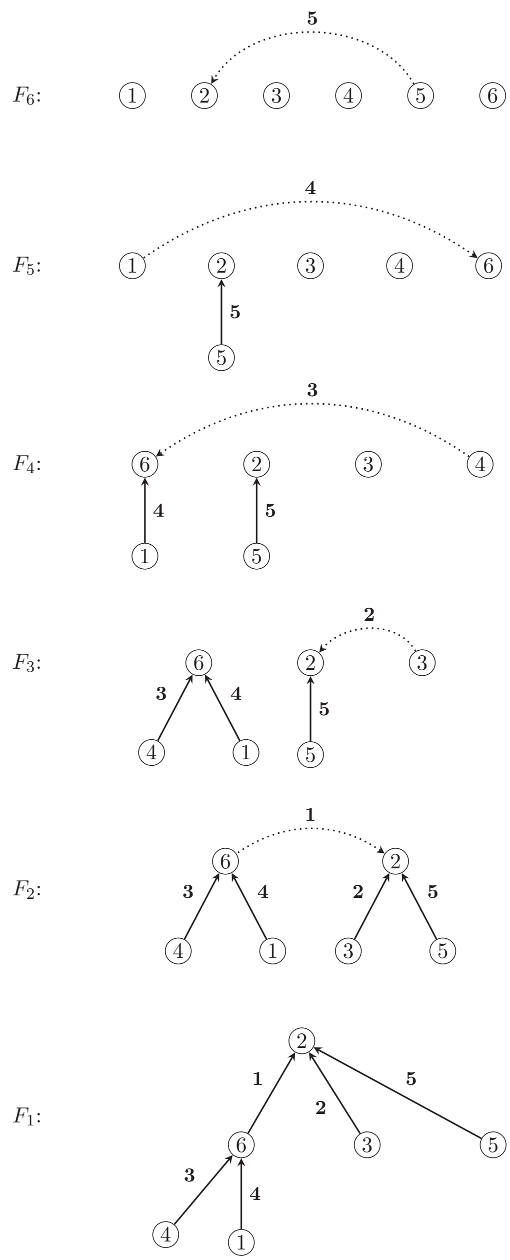


Figure 4–1: An example of Kingman's n-coalescent $\mathbf{C} = (F_n, \ldots, F_1)$ for $n = 6$. For $1 < i \leq n$, we present the edge $E(F_{i-1}) \setminus E(F_i)$ with a dotted line in F_i. Edges are marked with the labels ρ_C; $n - \rho_C(e)$ is the first forest where e is present. In this case, $\xi_6 = \xi_4 = \xi_3 = 1$, $\xi_5 = \xi_2 = 0$ and $\{a_5, b_5\} = \{2, 5\}$, $\{a_4, b_4\} = \{1, 5\}$, $\{a_3, b_3\} = \{1, 4\}$, $\{a_2, b_2\} = \{2, 3\}$, $\{a_1, b_1\} = \{1, 2\}$.

Proposition 4.1.3. *For each $C = (f_n, \dots, f_1) \in \mathcal{CF}_n$, relabel the vertices in t_C with σ_C to obtain $\phi(C) \in \mathcal{I}_n$. Then the law of $\phi(\mathbf{C})$ is that of a recursive tree of size n.*

Proof. From the argument in the proof of Lemma 7.2.2, we have that $|\mathcal{CF}_n| = n!(n-1)!$. Next, we show that ϕ is onto and, additionally, an $n!$-to-1 mapping. Thus ϕ preserves the uniform measure from $\mathcal{CF}_n$ to $\mathcal{I}_n$.

Fix an increasing tree $t \in \mathcal{I}_n$. Every vertex $j > 1$ has outdegree 1 in t, thus we write uniquely define $v_j \in V(t)$ such that $jv_j \in E(t)$. For each $1 < j \leq n$, let $e_{j-1} = jv_j$. Consider an n-chain $C = (f_n, \dots, f_1)$ defined as follows. Let $f_n \in \mathcal{F}_n$ have no edges, and for each $1 \leq i < n$, construct f_i from f_{i+1} by adding the edge e_i. It is easy to see that C satisfies $\sigma_C(i) = i$ for all $i \in [n]$ and t_C. Therefore $\phi(C) = t$, showing that ϕ is onto.

Now, consider $C \in \mathcal{CF}_n$ such that $\phi(C) = t$. For each permutation $\pi : [n] \to [n]$, let C_π be the n-chain obtained from $C = (f_n, \dots, f_1)$ by applying π to each of the labels of $V(f_i)$, $i \in [n]$. The mapping ϕ does not depend of the vertex labels in C, but on the order in which edges are added; therefore, $\phi(C) = \phi(C_\pi)$ for all permutations π. This shows that $|\phi^{-1}(t)| \geq n!$ for any $t \in \mathcal{I}_n$, completing the proof. □

For each n, let $\mathbf{C}$ be a Kingman's n-coalescent and let $T^{(n)} = t_{\mathbf{C}}$ be the unique tree in F_1. Since $\phi(\mathbf{C})$ only relabels vertices in $t_{\mathbf{C}}$, it follows that the shape of the tree is preserved; and so are the degrees and depths of the vertices. That is, as multisets,

$$\{(deg_{T^{(n)}}(v), h_{T^{(n)}}(v))\}_{v \in [n]} = \{(deg_{\phi(\mathbf{C})}(v), h_{\phi(\mathbf{C})}(v))\}_{v \in [n]}.$$

Moreover, for each $t \in \mathcal{I}_n$ the set $\phi^{-1}(t)$ can be indexed by permutations on $[n]$. This directly implies the following key corollary of Proposition 4.1.3.

Corollary 4.1.4. *For all $n \in \mathbb{N}$,*

$$((\mathrm{d}_{T^{(n)}}(i), \mathrm{h}_{T^{(n)}}(i)),\ i \in [n]) = ((\mathrm{d}_{T_n}(\sigma(i)), \mathrm{h}_{T_n}(\sigma(i))),\ i \in [n]);$$

where σ is a uniformly random permutation of $[n]$ and is independent of T_n. Consequently, the following equality in distribution holds jointly for all $i \in \mathbb{Z}$ and $j \in \mathbb{N}$,

$$|\{v \in [n] : \mathrm{d}_{T_n}(v) = i,\ \mathrm{h}_{T_n} = j\}| = |\{v \in [n] : \mathrm{d}_{T^{(n)}}(v) = i,\ \mathrm{h}_{T^{(n)}}(v) = j\}|$$

Proof. For any $n \in \mathbb{N}$, let $\mathcal{P}_n$ be the set of permutations on $[n]$. For any n-chain $C = (f_n, \ldots, f_1)$ let $\varphi(C) = (\phi(C), \sigma_C)$. Then $\varphi : \mathcal{CF}_n \to \mathcal{I}_n \times \mathcal{P}_n$ is a bijection and the result follows. $\square$

4.1.1 The degree and depth relation

Fix $n \in \mathbb{N}$ and consider Kingman's n-coalescent $\mathbf{C} = (F_n, \ldots, F_1)$. For each vertex $v \in [n]$ and $1 \leq i \leq n$, let $T_i(v)$ be the tree in F_i that contains v. We use $\mathrm{d}_{F_i}(v)$ and $\mathrm{h}_{F_i}(v)$ to denote the degree and depth of v in $T_i(v)$. For simplicity, we use $\mathrm{d}_n(v)$ and $\mathrm{h}_n(v)$ for the degree and depth of vertices in $T^{(n)}$.

We next define indicator functions $(s_{i,v},\ 2 \leq i \leq n)$ and the *selection set* $\mathcal{S}_n(v)$ as follows, let $s_{i,v}$ be the indicator that $T_i(v) \in \{T^{(i)}_{a_i}, T^{(i)}_{b_i}\}$; that is, $s_{i,v} = 1$ when $T_i(v) \in F_i$ is chosen to be merged and form a larger tree in F_{i-1}, and otherwise $s_{i,v} = 0$. Now we set

$$\mathcal{S}_n(v) = \{2 \leq i \leq n :\ s_{i,v} = 1\}.$$

The selection set $\mathcal{S}_n(v)$ keeps track of each time i where $T_i(v)$ merges.

The lemma below describes the joint law of the depth and degree of a given vertex.

Lemma 4.1.5. *Fix $v \in [n]$, let G be $\mathrm{Geo}(1/2)$ independent of $\mathcal{S}_n(v)$ and let $D = \min\{G, |\mathcal{S}_n(v)|\}$. Then, $\mathrm{d}_n(v) \stackrel{\mathcal{L}}{=} D$ and for all $k, l \in \mathbb{N}$,*

$$\mathbf{P}\left(\mathrm{d}_n(v) \geq k, \mathrm{h}_n(v) \leq l\right) = 2^{-k}\mathbf{P}\left(\mathrm{Bin}(|\mathcal{S}_n(v)| - k, 1/2) \leq l,\ |\mathcal{S}_n(v)| \geq k\right).$$

Proof. Any vertex starts as the root of a single-vertex tree. If $|\mathcal{S}_n(v)| = m$, then we flip a fair coin m times and set $\mathrm{d}_n(v)$ as the length of the first streak of heads and $\mathrm{h}_n(v)$ as the total number of tails; this proves the distributional identity of $\mathrm{d}_n(v)$.

Moreover, if $\mathrm{d}_n(v) \geq k$, then $|\mathcal{S}_n(v)| \geq k$ and the first k coin flips are determined to be heads, the latter event occurring with probability 2^{-k}. The remaining $|\mathcal{S}_n(v)| - k$ coin flips are independent of the previous tosses. □

One of the challenges in both Chapters 6 and 7 is to understand the correlations between the selection sets $\{\mathcal{S}_1^{(n)}, \ldots, \mathcal{S}_k^{(n)}\}$ for any fixed $k \in \mathbb{N}$.

CHAPTER 5
Summary of results

We recall the use of $\ln n$ to denote natural logarithms and $\log n$ to denote logarithms base 2. In addition, for the remainder of the chapter we write $\varepsilon_n = \log n - \lfloor \log n \rfloor$; this value is used to account for the lattice effect that occurs when looking at degrees around $\log n$, which are required to be integer-valued. We recall that for $m \in \mathbb{N}$, we write

$$Z_m^n = \#\{v \in [n],\, \mathrm{d}_{T_n}(v) = m\},$$

$$Z_{\geq m}^n = \#\{v \in [n],\, \mathrm{d}_{T_n}(v) \geq m\}.$$

Since we will deal with near-maximum degrees, we may simplify the notation by writing $X_d^{(n)} = Z^n_{\lfloor \log n \rfloor + d}$ and $X_{\geq d}^{(n)} = Z^n_{\geq \lfloor \log n \rfloor + d}$ for each $d \in \mathbb{Z}$.

5.1 A Poisson point process for highest-degree vertices

The next theorem describes the asymptotic joint law of the number of vertices with near-maximal degree.

Theorem 5.1.1. *Fix $\varepsilon \in [0,1]$. Let $(n_l)_{l\geq 1}$ be an increasing sequence of integers satisfying $\varepsilon_{n_l} \to \varepsilon$ as $l \to \infty$. Then, as $l \to \infty$*

$$(X_i^{(n_l)},\, i \in \mathbb{Z}) \xrightarrow{\mathrm{d}} (P_i^{\varepsilon},\, i \in \mathbb{Z})$$

jointly for all $i \in \mathbb{Z}$ where the P_i^ε are independent Poisson r.v.'s with mean $2^{-i-1+\varepsilon}$.

We present a stronger version of this result. Denote by $v_n^1, \ldots, v_n^n$ the vertices in T_n in decreasing order of degree; use a uniformly random ordering of vertices with the same degree to break ties. Write d_n^i and h_n^i for the degree and depth of v_n^i, respectively. Next, let $\mathcal{P}$ be a Poisson point process on $\mathbb{R}$ with rate function $\lambda(x) = 2^{-x} \ln 2$. Since $\mathbf{E}[\mathcal{P}(x, \infty)] < \infty$, for all $x \in \mathbb{R}$, we can list the elements in $\mathcal{P}$ in decreasing order. That is, for $i \geq 1$ let P_i be the i-th largest point of $\mathcal{P}$, or equivalently, such that $\mathcal{P}[P_i, \infty) = i$ but $\mathcal{P}(P_i, \infty) = i - 1$.

Theorem 5.1.2. *Let N_i be independent standard Gaussian variables, $i \in \mathbb{N}$. For each $\varepsilon \in [0, 1]$ and for any sequence of integers $(n_l,\, l \geq 1)$ for which $\log n_l - \lfloor \log n \rfloor \to \varepsilon$ as $l \to \infty$, then*

$$\left(d_{n_l}^i - \lfloor \log n_l \rfloor, \frac{h_{n_l}^i - (1 - (\log e)/2) \ln n_l}{\sqrt{(1 - (\log e)/4) \ln n_l}},\, i \geq 1 \right) \xrightarrow{\mathcal{L}} ((\lfloor P_i + \varepsilon \rfloor, i \geq 1), (N_i, i \geq 1)) .$$

For recursive trees, there is not only one vertex attaining the maximum degree. Instead, such number converges to a random variable. Let

$$\mathcal{M}_n = \{ v \in [n] : \mathrm{d}_n(v) = \Delta_n \}.$$

For each $\varepsilon \in [0, 1]$, consider the positive integer-valued random variable M_ε whose distribution is given, for each integer $k \geq 1$, by

$$\mathbf{P}(\mathbf{M}_\varepsilon = k) = \sum_{m \in \mathbb{Z}} e^{-2^{-m+\varepsilon}} \frac{2^{-(m+1-\varepsilon)k}}{k!}.$$

Proposition 5.1.3. *For any sequence of integers $(n_l,\, l \geq 1)$ for which $\log n_l - \lfloor \log n_l \rfloor \to \varepsilon$ and $n_l \to \infty$ as $l \to \infty$, then $|\mathcal{M}_{n_l}|$ converges to M_ε in distribution, and*

$$\left(\frac{\mathrm{h}_{T_{n_l}}(v) - (1 - (\log e)/2) \ln n_l}{\sqrt{(1 - (\log e)/4) \ln n_l}},\, v \in \mathcal{M}_{n_l} \right) \xrightarrow{\mathcal{L}} (N_i,\, 1 \leq i \leq M_\varepsilon),$$

where N_i are independent standard Gaussian variables.

Theorem 5.1.3 contrasts the behaviour of highest-degree vertices in linear PA trees; it implies that there is an interesting process of vertices that are involved in streaks of gaining degree that position them in the set $\mathcal{M}_n$ but are overtaken by other vertices when they run out of luck.

Additionally, Theorem 5.1.2 implies the following convergence in distribution.

Proposition 5.1.4. *For any sequence of integers $(n_l,\, l \geq 1)$ for which $\log n_l - \lfloor \log n_l \rfloor \to \varepsilon$ and $n_l \to \infty$ as $l \to \infty$. For each $i \in \mathbb{Z}$,*

$$X_{\geq i}^{(n_l)} \xrightarrow{\mathcal{L}} \mathrm{Poi}(2^{-i+\varepsilon}).$$

5.2 Central limit theorems for high-degree vertices

First, we can formalize the heuristic given in Section 3.4 for the order of Δ_n.

Proposition 5.2.1. *For fixed $c \in (0,2)$, uniformly over $m = m(n) < c \ln n$,*

$$\mathbf{E}\left[Z_m\right] = (1+o(1))2^{-m+1+\log n},$$

$$\mathbf{E}\left[Z_{\geq m}\right] = (1+o(1))2^{-m+\log n}.$$

Write $\lambda_{n,m} = \mathbf{E}\left[Z_{\geq m}^{(n)}\right]$. By restricting the range of $m = m(n)$, we obtain both normal asymptotic behavior and explicit convergence rates for $Z_{\geq m}^{(n)}$.

Theorem 5.2.2. *For each $c' \in (1, \log e]$ there exists $c \in (1, c')$ such that if $c \ln n < m < c' \ln n$, and $\lambda_{n,m} \to \infty$ as $n \to \infty$, then*

$$\frac{Z_{\geq m}^{(n)} - \lambda_{n,m}}{\sqrt{\lambda_{n,m}}} \xrightarrow{\mathcal{L}} N(0,1).$$

Theorem 5.2.3. *Fix $1 < c < c' < 2$. There are constants $\alpha = \alpha(c') \in (0,1)$ and $\beta = \beta(c) > 0$ such that uniformly for $m = m(n)$ satisfying $c \ln n < m < c' \ln n$,*

$$\mathrm{d_{TV}}\left(Z_m^{(n)}, \mathrm{Poi}(\lambda_{n,m})\right) \leq O(2^{-m+(1-\alpha)\log n}) + O(n^{-\beta}).$$

Theorem 5.2.2 follows directly from the central limit theorem for Poisson variables and Theorem 5.2.3; analogous results for linear PA trees are given in [19].

Finally, we are able to track the conditional depths of a finite number of vertices, given that their degrees are lower bounded, possibly by a high-degree value.

Theorem 5.2.4. *Fix $k \in \mathbb{N}$ and let $(u_i,\, i \in [k])$ be k distinct vertices in T_n chosen uniformly at random. For every $(a_1, \ldots, a_k) \in [0,1]^k$ and $(b_1, \ldots, b_k) \in \mathbb{Z}^k$, the conditional law of*

$$\left(\frac{\mathrm{h}_{T_n}(u_i) - (1 - (a_i \log e)/2) \ln n}{\sqrt{(1 - (a_i \log e)/4) \ln n}},\, i \in [k] \right),$$

given that $\mathrm{d}_{T_n}(u_i) \geq \lfloor a_i \log n \rfloor + b_i$ for all $i \in [k]$, converges to the law of k independent standard Gaussian variables.

5.3 Gumbel approximation for maximum degree

Using that for all $d \in \mathbb{Z}$, $\{\Delta_n \geq \lfloor \log n \rfloor + d\} = \{X_{\geq d} > 0\}$. The Poisson approximations of Proposition 5.1.4 and Theorem 5.2.3 yield the next two approximations; strengthening the results of Theorem 3.4.2.

Theorem 5.3.1. *For any $i = i(n)$ with $i + \log n < 2 \ln n$ and $\liminf_{n \to \infty} i(n) > -\infty$,*

$$\mathbf{P}(\Delta_n \geq \lfloor \log n \rfloor + i) = (1 - \exp\{-2^{-i+\varepsilon_n}\})(1 + o(1)).$$

Theorem 5.3.2. *Uniformly over $0 < i = i(n) < \log e \ln \ln n - C$, for some $C > 0$,*

$$\mathbf{P}\left(\Delta_n < \lfloor \log n \rfloor - i\right) = \exp\{-2^{i+\varepsilon_n}\}(1 + o(1)).$$

Part II

Manuscripts

CHAPTER 6
High degrees in random recursive trees

For $n \geq 1$, let T_n be a random recursive tree (RRT) on the vertex set $[n] = \{1, \ldots, n\}$. Let $\deg_{T_n}(v)$ be the degree of vertex v in T_n, that is, the number of children of v in T_n. Devroye and Lu [31] showed that the maximum degree Δ_n of T_n satisfies $\Delta_n/\lfloor \log_2 n \rfloor \to 1$ almost surely; Goh and Schmutz [39] showed distributional convergence of $\Delta_n - \lfloor \log_2 n \rfloor$ along suitable subsequences. In this work we show how a version of Kingman's coalescent can be used to access much finer properties of the degree distribution in T_n.

For any $i \in \mathbb{Z}$, let $X_i^{(n)} = |\{v \in [n] : \deg_{T_n}(v) = \lfloor \log n \rfloor + i\}|$. Also, let $\mathcal{P}$ be a Poisson point process on $\mathbb{R}$ with rate function $\lambda(x) = 2^{-x} \cdot \ln 2$. We show that, up to lattice effects, the vectors $(X_i^{(n)}, i \in \mathbb{Z})$ converge weakly in distribution to $(\mathcal{P}[i, i+1), i \in \mathbb{Z})$. We also prove asymptotic normality of $X_i^{(n)}$ when $i = i(n) \to -\infty$ slowly, and obtain precise asymptotics for $\mathbf{P}(\Delta_n - \log_2 n > i)$ when $i(n) \to \infty$ and $i(n)/\log n$ is not too large. Our results recover and extend the previous distributional convergence results on maximal and near-maximal degrees in random recursive trees.

6.1 Statement of results

The process of random recursive trees $(T_n, n \geq 1)$ is defined as follows. T_1 has a single node with label 1, which its root. The tree T_{n+1} is obtained from T_n by directing an edge from a new vertex $n+1$ to $v \in [n]$; the choice of v is uniformly random and independent for each $n \in \mathbb{N}$. We call T_n a random recursive tree (RRT) of size n.

As a consequence of the construction, vertex-labels in T_n increase along root-to-leaf paths. Rooted labelled trees with such property are called *increasing trees*. It is not difficult to see that, in fact, T_n is uniformly chosen among the set $\mathcal{I}_n$ of increasing trees with vertex set $[n]$.

We write $\deg_{T_n}(v)$ to denote the number of children of v in T_n. The degree distribution of T_n is encoded by the variables $Z_i^{(n)} = |\{v \in [n] : \deg_{T_n}(v) = i\}|$, for $i \geq 0$. In fact, the study of RRT's started with a paper by Na and Rapoport [67] in which they obtained, for any *fixed* $i \geq 0$, the convergence $\mathbb{E}(Z_i^{(n)})/n \to 2^{-i-1}$ as $n \to \infty$; this result was extended to convergence in probability by Meir and Moon in [64]. Mahmoud and Smythe [62] derived the asymptotic joint normality of $Z_i^{(n)}$ for $i \in \{0,1,2\}$; and finally, Janson [50] extended the joint normality to $Z_i^{(n)}$ for $i \geq 0$ and gave explicit formulae for the covariance matrix.

The above results concern typical degrees; the focus in this work is large degree vertices, and in particular the maximum degree in T_n, which we denote $\Delta_n = \max_{v\in[n]} \deg_{T_n}(v)$. For the rest of the paper we write log to denote logarithms with base 2, and ln to denote natural logarithms. For $n \in \mathbb{N}$ let $\varepsilon_n = \log n - \lfloor \log n \rfloor$.

A heuristic to find the order of Δ_n is that, if $\mathbb{E}(Z_i^{(n)}) \approx n2^{-i-1}$ were to hold for all i, as it does when i is fixed, then we would have $\mathbb{E}(Z_{\lfloor \log n \rfloor}^{(n)}) \approx n2^{-\lfloor \log n \rfloor -1} = 2^{-1+\varepsilon_n}$. This heuristic suggests that Δ_n is of order $\log n$. This is indeed the case: Szymanski [80] proved that $\mathbf{E}\left[\Delta_n\right]/\log n \to 1$ as $n \to \infty$, and Devroye and Lu [31] later established that $\Delta_n/\log n \to 1$ a.s.. Finally, Goh and Schmutz [39] showed that $\Delta_n - \lfloor \log n \rfloor$ converges in distribution along suitable subsequences, and identified the possible limiting laws.

Since we focus on maximal degrees, it is useful to let

$$X_i^{(n)} = Z_{i+\lfloor \log n \rfloor}^{(n)} = |\{v \in [n] : \deg_{T_n}(v) = \lfloor \log n \rfloor + i\}|,$$

for $n \in \mathbb{N}$ and $i \geq -\lfloor \log n \rfloor$. The following is a simplified version of one of our main results.

Theorem 6.1.1. *Fix $\varepsilon \in [0,1]$. Let $(n_l)_{l\geq 1}$ be an increasing sequence of integers satisfying $\varepsilon_{n_l} \to \varepsilon$ as $l \to \infty$. Then, as $l \to \infty$*

$$(X_i^{(n_l)},\, i \in \mathbb{Z}) \xrightarrow{\mathrm{d}} (P_i^{\varepsilon},\, i \in \mathbb{Z})$$

jointly for all $i \in \mathbb{Z}$ where the P_i^{ε} are independent Poisson r.v.'s with mean $2^{-i-1+\varepsilon}$.

The random variables $X_i^{(n)}$ do not converge in distribution as $n \to \infty$ without taking subsequences; this is essentially a lattice effect caused by the floor $\lfloor \log n \rfloor$ in the definition of $X_i^{(n)}$.

Theorem 6.1.1 can be stated in terms of weak convergence of point processes (which is equivalent to convergence of finite dimensional distributions (FDD's); see Theorem 11.1.VII in [25]). In fact, we will also prove convergence (along subsequences) of

$$X_{\geq i}^{(n)} = \sum_{k \geq i} X_k^{(n)} = |\{v \in [n] : \deg_{T_n}(v) \geq \lfloor \log n \rfloor + i\}|.$$

This is useful as it yields information about Δ_n which cannot be derived from Theorem 6.1.1. We formulate this result as a statement about convergence of point processes, and now provide the relevant definitions. Let $\mathbb{Z}^* = \mathbb{Z} \cup \{\infty\}$. Endow $\mathbb{Z}^*$ with the metric defined by $\mathrm{d}(i,j) = |2^{-j} - 2^{-i}|$ and $\mathrm{d}(i,\infty) = 2^{-i}$ for $i, j \in \mathbb{Z}$. Let $\mathcal{M}_{\mathbb{Z}^*}^{\#}$ be the space of boundedly finite measures of $\mathbb{Z}^*$.

Let $\mathcal{P}$ be a Poisson point process on $\mathbb{R}$ with rate function $\lambda(x) = 2^{-x} \cdot \ln 2$. For each $\varepsilon \in [0,1]$ let $\mathcal{P}^{\varepsilon}$ be the point process on $\mathbb{Z}^*$ given by

$$\mathcal{P}^{\varepsilon} = \sum_{x \in \mathcal{P}} \delta_{\lfloor x+\varepsilon \rfloor}.$$

Similarly, for all $n \in \mathbb{N}$ let

$$\mathcal{P}^{(n)} = \sum_{v \in [n]} \delta_{\deg_{T_n}(v) - \lfloor \log n \rfloor}.$$

Then, for each $i \in \mathbb{Z}$ we have that

$$\mathcal{P}^{\varepsilon}(\{i\}) := |\{x \in \mathcal{P} : \lfloor x + \varepsilon \rfloor = i\}| = |\{x \in \mathcal{P} : x \in [i - \varepsilon, i + 1 - \varepsilon)\}|$$

has distribution $\mathrm{Poi}(2^{-i-1+\varepsilon})$; also $\mathcal{P}^{(n)}(\{i\}) = X_i^{(n)}$. We abuse notation by writing, e.g., $\mathcal{P}^{(n)}(i) = \mathcal{P}^{(n)}(\{i\})$.

It is clear that $\mathcal{P}^{(n)}$ and $\mathcal{P}^{\varepsilon}$ are elements of $\mathcal{M}^{\#}_{\mathbb{Z}^*}$. The advantage of working on the state space to $\mathbb{Z}^*$ is that intervals $[k, \infty]$ are compact. In particular, the convergence of FDD's of $\mathcal{P}^{(n_l)}$ implies the convergence in distribution of $X_{\geq i}^{(n_l)} = \mathcal{P}^{(n_l)}[i, \infty)$.

Theorem 6.1.2. *Fix $\varepsilon \in [0,1]$. Let $(n_l)_{l \geq 1}$ be an increasing sequence of integers satisfying $\varepsilon_{n_l} \to \varepsilon$ as $l \to \infty$. Then in $\mathcal{M}^{\#}_{\mathbb{Z}^*}$, $\mathcal{P}^{(n_l)}$ converges weakly to $\mathcal{P}^{\varepsilon}$ as $l \to \infty$. Equivalently, for any $i < i' \in \mathbb{Z}$, jointly as $l \to \infty$*

$$(X_i^{(n_l)}, \ldots, X_{i'-1}^{(n_l)}, X_{\geq i}^{(n_l)}) \xrightarrow{\mathrm{d}} (\mathcal{P}^{\varepsilon}(i), \ldots, \mathcal{P}^{\varepsilon}(i'-1), \mathcal{P}^{\varepsilon}[i', \infty)).$$

Note that Theorem 6.1.1 follows from Theorem 6.1.2. We finish this section stating two additional results. The first is an extension of the main theorem from [39], that result being essentially the case $i = O(1)$.

Theorem 6.1.3. *For any $i = i(n)$ with $i + \log n < 2 \ln n$ and $\liminf_{n \to \infty} i(n) > -\infty$,*

$$\mathbf{P}(\Delta_n \geq \lfloor \log n \rfloor + i) = (1 - \exp\{-2^{-i+\varepsilon_n}\})(1 + o(1)).$$

When $i = O(1)$, the assertion of Theorem 6.1.3 is a straight-forward consequence of Theorem 6.1.2. For the case that $i(n) \to \infty$ we use estimates for the first and second moments of $X_{\geq i}^{(n)}$; note that $\{\Delta_n < \lfloor \log n \rfloor + i\} = \{X_{\geq i}^{(n)} = 0\}$.

Finally, we also obtain the asymptotic normality for $X_i^{(n)}$ when i tends to $-\infty$ slowly enough.

Theorem 6.1.4. *If $i = i(n) \to -\infty$ and $i = o(\ln n)$, then as $n \to \infty$*

$$\frac{X_i^{(n)} - 2^{-i-1+\varepsilon_n}}{\sqrt{2^{-i-1+\varepsilon_n}}} \xrightarrow{\mathrm{d}} N(0,1).$$

Remark 6.1.5. *Up to lattice effects, Theorems 6.1.2 and 6.1.4 extend the range of $i = i(n)$ for which the heuristic that $Z_i^{(n)} \approx n2^{-i-1}$ holds.*

A key novelty of our approach is that for each n we use *Kingman's coalescent* to generate a tree $T^{(n)}$ whose vertex degrees $\{\deg_{T^{(n)}}(v)\}_{v\in[n]}$ are exchangeable but otherwise have the same law as degrees in T_n. (See [9], Chapter 2 for a description of Kingman's coalescent, and [1], Section 2.2 for a description of the connection with random recursive trees which we exploit in this paper.) By this we mean that if $\sigma : [n] \to [n]$ is a uniformly random permutation then the following distributional identiy holds:

$$(\deg_{T^{(n)}}(v),\, v \in [n]) \stackrel{\mathrm{d}}{=} (\deg_{T_n}(\sigma(v)),\, v \in [n]). \tag{6.1}$$

We describe the trees $T^{(n)}$, $n \in \mathbb{N}$ in Section 6.3.

An essentially equivalent construction was used by Devroye [26] to study union-find trees. In [75], Pittel related the results of [26] on union-find trees to the height of RRT's. It is worth mentioning that both Kingman's coalescent and the union-find trees can be equivalently represented as binary trees or, as we will see in Section 6.3, as RRT's. Aside from the works [26] and [75], it seems that the use of Kingman's coalescent or of union-find trees to study RRT's is rare. However, it turns out to provide just the right perspective for studying high degree vertices.

6.2 Outline

In this section we sketch the approach used in the paper. The proofs of the theorems relay on the computation of the moments of the FDD's of $\mathcal{P}^{(n)}$; these estimates are given in Proposition 6.2.1. In particular, the proofs of Theorems 6.1.2 and 6.1.4 use the method of moments (e.g., see [52] Section 6.1, and [16] Section 1.5).

Any FDD of $\mathcal{P}^{(n)}$ can be recovered from suitable marginals of the joint distribution of $(X_i^{(n_l)}, \ldots, X_{i'-1}^{(n_l)}, X_{\geq i'}^{(n_l)})$ for some $i < i' \in \mathbb{Z}$. For simplicity, we focus for the moment on collections of variables $X_i^{(n)}, \ldots, X_{i'}^{(n)}$ for $i \leq i'$. For $r \in \mathbb{R}$ and $a \in \mathbb{N}$ write $(r)_a = r(r-1)\cdots(r-a+1)$, also let $(r)_0 = 1$. We will prove that for any non-negative integers $a_i, \ldots, a_{i'}$, as $n \to \infty$, we have

$$\mathbf{E}\left[\prod_{i\leq k\leq i'} (X_k^{(n)})_{a_k}\right] - \prod_{i\leq k\leq i'} \left(2^{-(k+1)+\varepsilon_n}\right)^{a_k} \to 0. \tag{6.2}$$

This immediately yields Theorem 6.1.1.

By the linearity of expectation, proving (6.2) reduces to understanding the probabilities

$$\mathbf{P}\left(\deg_{T_n}(v_k) = \lfloor \log n \rfloor + i_k,\, k \in [K]\right) \tag{6.3}$$

for all $i_1, \ldots i_K \in \mathbb{N}$ and $v_1, \ldots v_K \in [n]$, $K \in \mathbb{N}$; see Section 6.5 for more details.

In the standard model for RRT's described at the beginning, $\deg_{T_n}(v)$ is a sum of Bernoulli variables:

$$\deg_{T_n}(v) = \sum_{v<u\leq n} \mathbf{1}_{\{u\to v\}}.$$

The lack of symmetry of the degrees $\{\deg_{T_n}(v)\}_{v\in[n]}$ complicates the analysis of (6.3). In proving that $\Delta_n/\log n \overset{\text{a.s.}}{\to} 1$, Devroye and Lu [31] used that $\{\deg_{T_n}(v)\}_{v\in[n]}$ are negatively orthant dependent (see [53] for a definition), which in particular means that for all $S \subset [n]$ and $m_1, \ldots, m_n \in \mathbb{N}$

$$\mathbf{P}\left(\deg_{T_n}(v) \geq m_v,\, v \in S\right) \leq \prod_{v \in S} \mathbf{P}\left(\deg_{T_n}(v) \geq m_v\right)$$

and then obtained upper bounds for $\mathbf{P}\left(deg_{T_n}(v) \geq c \ln n\right)$ for each $v \in [n]$.

One approach to studying high degrees in T_n would be to obtain matching lower bounds for $\mathbf{P}\left(\deg_{T_n}(v) \geq m_v,\, v \in S\right)$, with uniform error terms even when m_v is large. Instead, we study trees $T^{(n)}$, mentioned in (6.1), above, for which we can obtain precise asymptotics for the analogous probabilities

$$\mathbf{P}\left(\deg_{T^{(n)}}(v) \geq m_v,\, v \in [K]\right). \tag{6.4}$$

The core of the paper lies in Proposition 6.4.2, which gives precise estimates of (6.4) for $m_1, \ldots, m_K$ in a suitable range. Broadly speaking, $\deg_{T^{(n)}}(v)$ depends on a set of random *selection times* $\mathcal{S}_v$ and the first streak of heads in a sequence of $|\mathcal{S}_v|$ fair coin flips. As mentioned in the previous section, the degrees of $T^{(n)}$ have the same distribution as the degrees in T_n. Consequently, our estimation of (6.4) allows us to obtain the following moments estimate.

Proposition 6.2.1. *For all $c \in (0,2)$ and $K \in \mathbb{N}$ there is $\alpha = \alpha(c,K) > 0$ such that the following holds. Fix any integers i, i' with $0 < i + \log_n < i' + \log_n < c \ln n$. Then for any non-negative integers $a_i, \ldots, a_{i'}$ with $a_i + \ldots + a_{i'} = K$, we have*

$$\mathbf{E}\left[(X^{(n)}_{\geq i'})_{a_{i'}} \prod_{i \leq k < i'} (X^{(n)}_k)_{a_k}\right] = \left(2^{-i'+\varepsilon_n}\right)^{a_{i'}} \prod_{i \leq k < i'} \left(2^{-(k+1)+\varepsilon_n}\right)^{a_k} (1 + o(n^{-\alpha})).$$

Equipped with Proposition 6.2.1, the proofs of the theorems are straightforward. The rest of the paper is organized as follows. In Section 6.3, we explain how to define the trees $T^{(n)}$ using Kingman's coalescent and establish the distributional relation between $T^{(n)}$ and the RRT; see Corollary 6.3.4. In Section 6.4, we define the random sets $(\mathcal{S}_v, v \in T^{(n)})$ and explain their relation with degrees in $T^{(n)}$. The proof of Proposition 6.4.2, which is our estimate of (6.4), is then presented using a decoupling of the events in (6.4) and the concentration of the random variables $|\mathcal{S}_v|$. Finally, the proof of Proposition 6.2.1 is given in Section 6.5 and the proof of Theorems 6.1.2-6.1.4 are in Section 6.6.

6.3 Random Recursive Trees and Kingman's coalescent

In this section we give a representation of Kingman's coalescent in terms of labelled forests, and relate it to RRT's. All trees in the remainder of the paper are rooted, and we write $r(t)$ for the root of tree t. By convention, edges of a tree are directed towards the root of the tree and we write uv to denote an edge directed from u to v. A forest f is a set of trees whose vertex sets are pairwise disjoint. The vertex set of a forest, denoted $V(f)$, is the union of the vertex sets of its trees. Similarly, $E(f)$ denotes the set of edges in the trees of f. For $n \geq 1$, let

$$\mathcal{F}_n = \{f : V(f) = [n]\}$$

be the set of forests with vertex set $[n]$.

A sequence $C = (f_1, \ldots, f_n)$ of elements of $\mathcal{F}_n$ is an n-chain if f_1 is the forest in $\mathcal{F}_n$ with n one-vertex trees and, for $1 \leq i < n$, f_{i+1} is obtained from f_i by adding a directed edge between the roots of some pair of trees in f_i. If $(f_1, \ldots, f_n)$ is an n-chain then for $1 \leq i \leq n$, the forest f_i consists of $n+1-i$ trees, and in this case we list its elements in increasing order of their smallest-labelled vertex as $t_1^{(i)}, \ldots, t_{n+1-i}^{(i)}$.

Definition 6.3.1. *Kingman's n-coalescent is the random n-chain $\mathbf{C} = (F_1, \ldots, F_n)$ built as follows. Independently for each $1 \leq i \leq n-1$ let $\{a_i, b_i\}$ be a random pair uniformly chosen from $\{\{a, b\} : 1 \leq a < b \leq n+1-i\}$ and let ξ_i be independent with* Bernoulli$(1/2)$ *distribution.*

For $1 \leq i < n$, construct F_{i+1} from F_i as follows. If $\xi_i = 1$ then add an edge from $r(T_{b_i}^{(i)})$ to $r(T_{a_i}^{(i)})$ and if $\xi_i = 0$ then add an edge from $r(T_{a_i}^{(i)})$ to

$r(T^{(i)}_{b_i})$. *The forest* F_{i+1} *consists of the new tree and the remaining* $n-1-i$ *unaltered trees from* F_i.

For an example of the process see Figure 6–1.

Lemma 6.3.2. *Let* $\mathcal{CF}_n$ *be the set of* n*-chains of elements in* $\mathcal{F}_n$. *Then* $|\mathcal{CF}_n| = n!(n-1)!$ *and Kingman's* n*-coalescent is a uniformly random element of* $\mathcal{CF}_n$.

Proof. Fix an n-chain $(f_1, \ldots, f_n) \in \mathcal{CF}_n$. Then

$$\mathbf{P}\left((F_1, \ldots, F_n) = (f_1, \ldots, f_n)\right) = \prod_{k=1}^{n-1} \mathbf{P}\left(F_{k+1} = f_{k+1} | F_j = f_j,\, 1 \le j \le k\right).$$

Among the $(n+1-k)(n-k)$ possible oriented edges between roots of f_k, there is exactly one whose addition yields f_{k+1}. It follows that the k-th term in the above product is $((n+1-k)(n-k))^{-1}$, so $\mathbf{P}\left((F_1, \ldots, F_n) = (f_1, \ldots, f_n)\right) = [n!(n-1)!]^{-1}$. The result follows since this expression does not depend on $(f_1, \ldots, f_n) \in \mathcal{CF}_n$. □

Recall that $\mathcal{I}_n$ is the set of increasing trees with vertex set $[n]$. It is not difficult to see that $|\mathcal{I}_n| = (n-1)!$ and that a RRT is a uniformly random element of $\mathcal{I}_n$.

There is a natural mapping ϕ between n-chains and increasing trees. Given an n-chain $C = (f_1, \ldots, f_n)$, write $t^{(n)} := t^{(n)}_1$ for the unique tree in f_n. Let $L^-_C : E(t^{(n)}) \to [n-1]$ be defined as follows. For each $e \in E(t^{(n)})$, let

$$L^-_C(e) = \max\{i \in [n-1] : e \notin E(t^{(i)})\}.$$

We think of L^-_C as a function that keeps track of the *time of addition* of the edges along the n-chain C. Now, we define a vertex labelling $L_C : V(t^{(n)}) \to [n]$ as follows. Let $L_C(r(t^{(n)})) = 1$ and for each $uv \in E(t^{(n)})$, let

$$L_C(u) = n + 1 - L^-_C(uv);$$

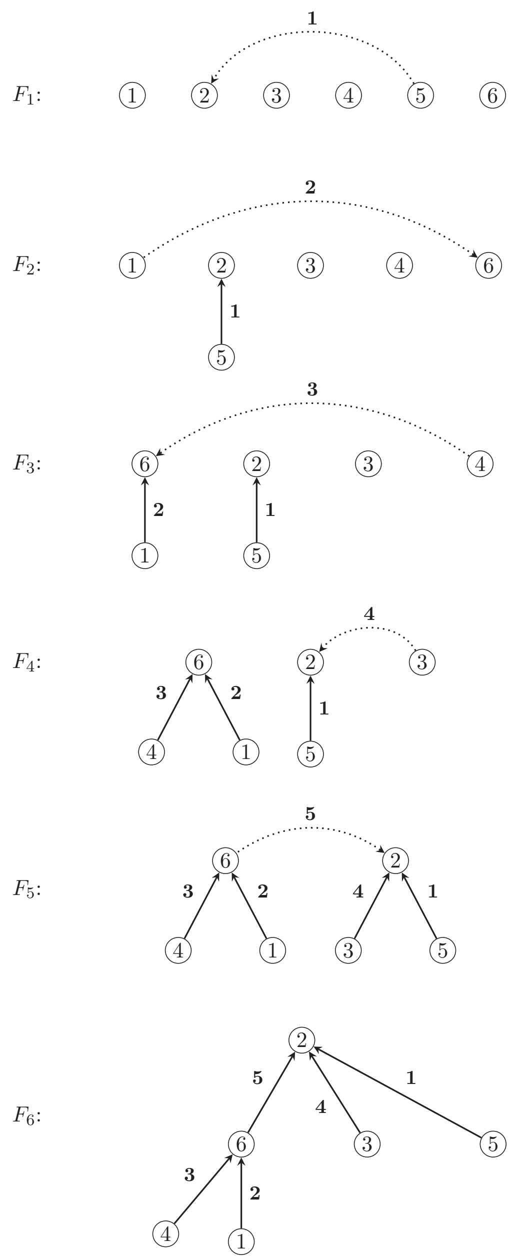
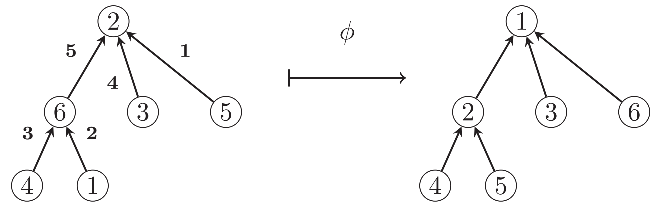


Figure 6–1: An example of Kingman's n-coalescent $\mathbf{C} = (F_1, \ldots, F_n)$ for $n = 6$. For $1 \leq i < n$, F_i has, in dotted line, the edge in $E(F_{i+1}) \setminus E(F_i)$. Edges are marked with their time of addition; this is the function $L^-_{\mathbf{C}}$ defined after Lemma 6.3.2. In this instance, $\xi_1 = \xi_3 = \xi_4 = 1$, $\xi_2 = \xi_5 = 0$ and $\{a_1, b_1\} = \{2, 5\}$, $\{a_2, b_2\} = \{1, 5\}$, $\{a_3, b_3\} = \{1, 4\}$, $\{a_4, b_4\} = \{2, 3\}$, $\{a_5, b_5\} = \{1, 2\}$.



$t^{(n)}$, (L_C^- in bold) $\qquad\qquad$ $\phi(f_1, \ldots, f_n)$

Figure 6–2: On the left a tree $t^{(n)}$; edges are marked with L_C^-, from which the n-chain $C = (f_1, \ldots, f_n)$ can be recovered. On the right, the increasing tree $\phi(f_1, \ldots, f_n)$; it has the shape of $t^{(n)}$ and the vertex labels $\{L_C(v),\, v \in V(t^{(n)})\}$.

then $L_C(u)$ is the number of trees in the forest just before uv is added.

Note that for each $i \in [n-1]$, the new edge in f_{i+1} joins the roots of two trees in f_i and is directed towards the root of the resulting tree. Thus, the labels $\{L_C^-(e),\, e \in E(t^{(n)})\}$ increase along all paths in $t^{(n)}$ towards the root $r(t^{(n)})$ and consequently, the labels $\{L_C(v),\, v \in V(t^{(n)})\}$ increase along root-to-leaf paths in $t^{(n)}$. This shows that relabelling the vertices of $t^{(n)}$ with L_C yields an increasing tree (specifically, an element of $\mathcal{I}_n$). See Figure 6–2 for an example.

Proposition 6.3.3. *Let $\phi : \mathcal{CF} \to \mathcal{I}_n$ be defined as follows. For an n-chain $C = (f_1, \ldots, f_n)$ let $\phi(C)$ be the tree obtained from $t^{(n)}$ by relabelling its vertices with L_C. Then $\phi(\mathbf{C})$, the push-forward of Kingman's n-coalescent by ϕ, has the law of a RRT of size n.*

Proof. First, we prove that ϕ is onto. Fix an increasing tree $t \in \mathcal{I}_n$. For each $j \in V(t) \setminus \{1\}$, let $v_j \in V(t)$ be such that $jv_j \in E(t)$, recall that edges are directed toward the root of t, thus v_j is uniquely defined. For each $1 < j \leq n$, let $e_{n-j+1} = ju_j$.

Now construct an n-chain C as follows. Let f_1 be the forest with n one-vertex trees. For each $1 < i \leq n$ construct f_i from f_{i-1} by adding the edge e_{i-1}. In other words, for each $1 \leq i < n$, $L_C^-(e_i) = i$ and so $L_C(n+1-i) =$

$n+1-L_C^-(e_i) = n+1-i$; also since $r(t) = 1$, we have $L_C(1) = 1$. Consequently, $\phi(C) = t$.

We claim that $|\phi^{-1}(t)| \geq n!$ for any $t \in \mathcal{I}_n$. To see this, consider an n-chain C and a permutation $\sigma : [n] \to [n]$. Let C_σ be the n-chain obtained from C by permuting the vertices in each forest of C by σ. Since $L_C(v)$ depends only on the time of addition of its outgoing edge (if any), it follows that $\phi(C) = \phi(C_\sigma)$ for all permutations σ. By Lemma 6.3.2, this shows that ϕ is $n!$-to-1 and that $\phi(\mathbf{C})$ is a uniform element in $\mathcal{I}_n$. $\square$

Since $\phi(\mathbf{C})$ preserves the shape of $T^{(n)}$ and only relabels its vertices, the degrees in $T^{(n)}$ and $\phi(\mathbf{C})$ are equal as multisets: $\{deg_{T^{(n)}}(v)\}_{v\in[n]} = \{deg_{\phi(\mathbf{C})}(v)\}_{v\in[n]}$. This immediately gives the following key corollary of Proposition 6.3.3, on which the rest of the paper relies.

Corollary 6.3.4. *For all $n \in \mathbb{N}$, we have the following equality in distribution holds jointly for all $i \in \mathbb{Z}$,*

$$X_i^{(n)} \stackrel{\mathrm{d}}{=} |\{v \in [n] : \deg_{T^{(n)}}(v) = \lfloor \log n \rfloor + i\}|.$$

We now proceed to the study of the joint distribution of the vertex degrees in $T^{(n)}$.

6.4 Degree distribution: Selection sets and coin flips

By construction, the vertex degrees $\{\deg_{T^{(n)}}(v)\}_{v\in[n]}$ are exchangeable. Our next goal is to explain how to approximate (6.4); that is, for any fixed $k \in \mathbb{N}$ and integers $m_1, \ldots, m_k < 2\ln n$, to obtain estimates for $\mathbf{P}\left(\deg_{T^{(n)}}(v) \geq m_v,\, v \in [k]\right)$.

The key to analyse the degrees in $T^{(n)}$ is to understand how the degrees of a vertex $v \in [n]$ change in Kingman's coalescent $\mathbf{C} = (F_1, \ldots, F_n)$. For any vertex v and $1 \leq i \leq n$, denote $\deg_{F_i}(v)$ the number of children of v in F_i. Also, we will simply write $\deg(v) = \deg_{F_n}(v) = \deg_{T^{(n)}}(v)$. For each $1 \leq i < n$, if $\xi_i = 1$ we say that ξ_i *favours* the vertices of $T_{a_i}^{(i)}$, and otherwise

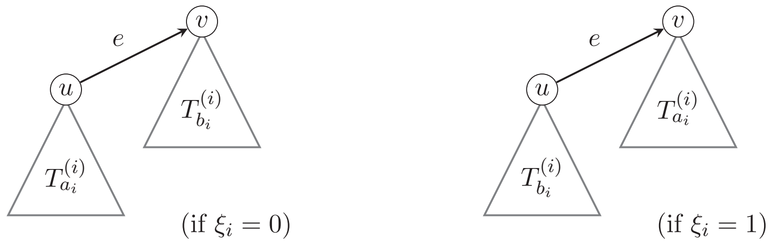


Figure 6–3: If v is a root in $T_{a_i}^{(i)} \cup T_{b_i}^{(i)}$ and ξ_i favours v, then v increases its degree and remains a root in F_{i+1}.

that it favours the vertices of $T_{b_i}^{(i)}$. For $v \in [n]$, let

$$\mathcal{S}_v = \{i \in [n-1] : v \in T_{a_i}^{(i)} \cup T_{b_i}^{(i)}\}.$$

For any vertex v, and $1 \le i < n$, $\deg_{F_{i+1}}(v)$ increases by one only if v is a root in F_i, $i \in \mathcal{S}_v$ and ξ_i favours v; see Figure 6–3. Conversely, let $p_v = \min\{i \in \mathcal{S}_v, \xi_i \text{ does not favour } v\}$, then the first F_{i+1} in which v is not a root is exactly $i = p_v$. In this case, in F_{p_v+1} there is an outgoing edge from v, and v is not a root of any subsequent forests. As a consequence, $\deg_{F_j}(v) = \deg_{F_{p_v}}(v)$ for $p_v < j \le n$.

Fact 6.4.1. *For $v \in [n]$, $\deg(v) = \deg_{F_{p_v}}(v) = |\mathcal{S}_v \cap [p_v - 1]|$.*

In other words, $\deg(v)$ depends only on its first streak of favourable random variables ξ_i with $i \in \mathcal{S}_v$. More precisely, given $|\mathcal{S}_v|$, the degree $\deg(v)$ is distributed as $\min\{|\mathcal{S}_v|, G\}$, where G is a Geometric(1/2) r.v. independent of $\mathcal{S}_v$.

Thus, it is relevant to observe that $|\mathcal{S}_v|$ is distributed as an sum of independent (though not identically distributed) Bernoulli random variables and so it is concentrated around its mean $\mathbf{E}[|\mathcal{S}_v|] = 2\ln n + O(1)$; a more precise statement can be found in Proposition 6.4.5 below. Since $|\mathcal{S}_v| \to \infty$ in probability as $n \to \infty$, it follows easily that $\deg(v)$ is asymptotically geometric for any fixed node v. More strongly, the following proposition shows that for any fixed k, the random variables $\{\deg_{T^{(n)}}(v)\}_{v\in[k]}$ asymptotically behave like independent Geometric random variables, even if they are conditioned to be quite large.

Proposition 6.4.2. *Fix $c \in (0,2)$ and $k \in \mathbb{N}$. There exists $\alpha = \alpha(c,k) > 0$ such that uniformly over positive integers $m_1,\dots,m_k < c\ln n$,*

$$\mathbf{P}\left(\deg_{T^{(n)}}(v) \geq m_v,\, v \in [k]\right) = 2^{-\sum_v m_v}(1+o(n^{-\alpha})).$$

We now explain how the events in the proposition above can be decoupled into a product of two probabilities, one of them corresponding to tail bounds for the random variables $|\mathcal{S}_v|$. We start with an upper bound for Proposition 6.4.2.

Lemma 6.4.3. *For any $k \in \mathbb{N}$ and positive integers $m_1,\dots,m_k < n$,*

$$\mathbf{P}\left(\deg(v) \geq m_v,\, v \in [k]\right) \leq 2^{-\sum_v m_v}\,\mathbf{P}\left(|\mathcal{S}_v| \geq m_v,\, v \in [k]\right).$$

Equality holds for $k = 1$.

Proof. For each $v \in [k]$ list $\mathcal{S}_v$ in increasing order as $(i_{v,j},\, 1 \leq j \leq |\mathcal{S}_v|)$. Let $\mathcal{A}$ be the set of sequences $A = (A_1,\dots,A_k)$ satisfying $A_v \subset [n-1]$ and $|A_v| = m_v$ for all $v \in [k]$. For every $A \in \mathcal{A}$, let D_A be the event that $|\mathcal{S}_v| \geq m_v$ and $\{i_{v,1},\dots,i_{v,m_v}\} = A_v$, for all $v \in [k]$. By Fact 6.4.1, if $\deg(v) \geq m_v$ then necessarily $|\mathcal{S}_v| \geq m_v$ so

$$\{\deg(v) \geq m_v,\, v \in [k]\} \cap D_A = \{\xi_{i_{v,j}} \text{ favours } v \text{ for all } j \in [m_v],\, v \in [k]\} \cap D_A.$$

Now, ξ_i are i.i.d Bernoulli(1/2) r.v.'s. Thus, if D_A has positive probability then

$$\mathbf{P}\left(\xi_{i_{v,j}} \text{ favours } v \text{ for all } j \in [m_v],\, v \in [k] | D_A\right) = \begin{cases} 2^{-\sum_v m_v} & \text{if } |A_u \cap A_v| = 0,\ \forall u \neq v \in [k] \\ 0 & \text{o.w.} \end{cases}$$

The second case follows from the fact that if $i \in \mathcal{S}_u \cap \mathcal{S}_v$ for some $u \neq v$, then ξ_i cannot favour both u and v. The events $(D_A,\, A \in \mathcal{A})$ are pairwise disjoint, and if $\deg(v) \geq m_v$ for all $v \in [k]$ then one of the events D_A must occur. It follows that

$$\begin{aligned}
\mathbf{P}\left(\deg(v) \geq m_v,\, v \in [k]\right) &= \sum_{A \in \mathcal{A}} \mathbf{P}\left(D_A,\, \deg(v) \geq m_v,\, v \in [k]\right) \\
&\leq \sum_{A \in \mathcal{A}} 2^{-\sum_v m_v} \mathbf{P}\left(D_A\right) \\
&= 2^{-\sum_v m_v} \mathbf{P}\left(|S_v| \geq m_v,\, v \in [k]\right).
\end{aligned}$$

Finally, the second line holds with equality when $k = 1$. □

For the lower bound we restrict to events D_A where the sets A_v are already disjoint. To do so, we consider instead the vertex degrees in F_I for some $I < n$. For $k \geq 2$ let

$$\tau_k = \min\{i \in [n-1] : \{a_i, b_i\} \subset [k]\}.$$

Since $F_i \subset F_j$ for all $i \leq j \in [n]$ we have that for any $I < n$

$$\begin{aligned}
\mathbf{P}\left(\deg(v) \geq m_v,\, v \in [k]\right) &\geq \mathbf{P}\left(\deg_{F_{I+1}}(v) \geq m_v,\, v \in [k]\right) \\
&\geq \mathbf{P}\left(I < \tau_k,\, \deg_{F_{I+1}}(v) \geq m_v,\, v \in [k]\right). \qquad (6.5)
\end{aligned}$$

Recall that trees in F_i are listed in increasing order of their least elements; this implies that indices of the trees of vertices $1, \ldots, k$ do not change until two trees indexed by $a, b \leq k$ are merged. Therefore, for all $v \in [k]$, $v \in T_v^{(i)}$ for $i \leq \tau_k$. This implies the sets $\{\mathcal{S}_v \cap [\tau_k - 1],\, v \in [k]\}$ are pairwise disjoint. These observations allow us to obtain a lower bound analogous to Lemma 6.4.3.

Lemma 6.4.4. *For any positive integers $k \geq 2$ and $m_1, \ldots, m_k, I < n$,*

$$\mathbf{P}\left(\deg(v) \geq m_v,\, v \in [k]\right) \geq 2^{-\sum_v m_v} \mathbf{P}\left(I < \tau_k, |S_v \cap [I]| \geq m_v, v \in [k]\right).$$

Proof. By (6.5), it suffices to bound $\mathbf{P}\Big(I < \tau_k,\, \deg_{F_{I+1}}(v) \geq m_v,\, v \in [k]\Big)$.

Let $\mathcal{A}^*$ be the set of sequences $A = (A_1, \ldots, A_k)$ of pairwise disjoint subsets of $[I]$ satisfying $|A_v| = m_v$ for all $v \in [k]$. For each $A \in \mathcal{A}^*$, let D_A be the event that for all $v \in [k]$, $\{i_{v,j}, \ldots, i_{v,m_v}\} = A_v$ (and so $|\mathcal{S}_v \cap [I]| \geq m_v$).

As in the proof of Lemma 6.4.3, we have that

$$\{\deg_{F_{I+1}}(v) \geq m_v,\, v \in [k]\} \cap D_A = \{\xi_{i_{v,j}} \text{ favours } v \text{ for all } j \in [m_v],\, v \in [k]\} \cap D_A.$$

In this case, the sets A_v are pairwise disjoint. If $\mathbf{P}(D_A) > 0$ then

$$\mathbf{P}\left(\xi_{i_{v,j}} \text{ favours } v \text{ for all } j \in [m_v],\, v \in [k] | D_A\right) = 2^{-\sum_v m_v}.$$

Recall that $I < \tau_k$ if and only if the sets $\{S_v \cap [I],\, v \in [k]\}$ are pairwise disjoint; that is, if one of the events D_A occur. We then have

$$\begin{aligned}
\mathbf{P}\Big(I < \tau_k,\, \deg_{F_{I+1}}(v) \geq m_v,\, v \in [k]\Big) &= \sum_{A \in \mathcal{A}^*} \mathbf{P}\Big(D_A,\, \deg_{F_{I+1}}(v) \geq m_v,\, v \in [k]\Big) \\
&= \sum_{A \in \mathcal{A}^*} 2^{-\sum_v m_v} \mathbf{P}(D_A) \\
&= 2^{-\sum_v m_v} \mathbf{P}\left(I < \tau_k,\, |S_v \cap [I]| \geq m_v,\, v \in [k]\right).
\end{aligned}$$

□

To use Lemma 6.4.4 we need tail bounds for $|\mathcal{S}_v \cap [I]|$ for some suitable $I < n$; these are provided by the following proposition.

Proposition 6.4.5. *Fix $\varepsilon \in (0,1)$ and $c \in (0, 2(1-\varepsilon))$. Then there exists $\beta = \beta(c, \varepsilon) > 0$ such that for any vertex v,*

$$\mathbf{P}\left(|\mathcal{S}_v \cap [n - \lceil n^\varepsilon \rceil]| < c \ln n\right) = o(n^{-\beta}).$$

Proof. Fix $\varepsilon \in (0,1)$ and $c \in (0, 2(1-\varepsilon))$. Let $\{B_i,\, i \in \mathbb{N}\}$ be a collection of independent Bernoulli r.v.'s, with $\mathbf{E}[B_i] = \frac{2}{i}$. Recall the definition of $\mathcal{S}_v$ at the beginning of the section.

For any fixed vertex $v \in [n]$, and each $i \in [n-1]$, the probability of the event $\{v \in T^{(i)}_{a_i} \cup T^{(i)}_{b_i}\}$ is $2/(n-i+1)$; this is because, in the forest F_i, there are $n-i+1$ trees and the trees $T^{(i)}_{a_i}$, $T^{(i)}_{b_i}$ are chosen uniformly at random among them. Since each of these events are independent we have $|\mathcal{S}_v| \stackrel{\mathrm{d}}{=} \sum_{i=2}^{n} B_i$. Moreover, writing $W_{n,\varepsilon} = \sum_{i=n-\lceil n^\varepsilon \rceil}^{n} B_i$, we also have

$$W_{n,\varepsilon} \stackrel{\mathrm{d}}{=} |\mathcal{S}_v \cap [n - \lceil n^\varepsilon \rceil]|.$$

We now apply Bernstein's inequality (see, e.g., [52], Theorem 2.8) to obtain that for any $t > 0$,

$$\mathbf{P}\left(W_{n,\varepsilon} \le \mathbf{E}\left[W_{n,\varepsilon}\right] - t\right) \le \exp\left\{-\frac{t^2}{2\mathbf{E}\left[W_{n,\varepsilon}\right]}\right\}.$$

We take $t = \mathbf{E}\left[W_{n,\varepsilon}\right] - c\ln n$. Since

$$\mathbf{E}\left[W_{n,\varepsilon}\right] = \sum_{i=n-\lceil n^\varepsilon \rceil}^{n} \frac{2}{i} = 2(1-\varepsilon)\ln n + O(1),$$

setting $\delta = 2(1-\varepsilon) - c > 0$ we have $t = \delta \ln n + O(1)$, so

$$\mathbf{P}\left(|\mathcal{S}_v \cap [n - \lceil n^\varepsilon \rceil]| < c\ln n\right) = \mathbf{P}\left(W_{n,\varepsilon} \le \mathbf{E}\left[W_{n,\varepsilon}\right] - t\right) = O(1) \cdot n^{-\delta^2/(4(1-\varepsilon))}.$$

Choosing $0 < \beta < \delta^2/4(1-\varepsilon)$, the result follows. □

The following lemma is the last ingredient for Proposition 6.4.2.

Lemma 6.4.6. *Fix an integer $k \ge 2$ and let $\varepsilon \in (0,1)$. Then, for n large enough,*

$$\mathbf{P}\left(\tau_k \le n - \lceil n^\varepsilon \rceil\right) \le \frac{2k^2}{\lceil n^\varepsilon \rceil - 1}.$$

Proof. By the definition of τ_k, if $\tau_k > n - \lceil n^\varepsilon \rceil$ then $\{a_i, b_i\} \not\subset [k]$ for all $1 \le i \le n - \lceil n^\varepsilon \rceil$. The events that $\{a_i, b_i\} \not\subset [k]$ are independent for distinct i

and $\mathbf{P}(\{a_i, b_i\} \subset [k]) = \frac{k(k-1)}{(n+1-i)(n-i)}$, so we have that

$$\mathbf{P}(\tau_k > n - \lceil n^\varepsilon \rceil) = \prod_{i=1}^{n-\lceil n^\varepsilon \rceil} \left(1 - \frac{k(k-1)}{(n+1-i)(n-i)}\right) \geq 1 - \sum_{i=1}^{n-\lceil n^\varepsilon \rceil} \frac{2k^2}{(n-i)^2}$$

The last inequality holds for n large enough. Since $\sum_{j=m}^{\infty} j^{-2} \leq \int_{m-1}^{\infty} x^{-2} dx = (m-1)^{-1}$, we get

$$\mathbf{P}(\tau_k \leq n - \lceil n^\varepsilon \rceil) \leq \sum_{i=1}^{n-\lceil n^\varepsilon \rceil} \frac{2k^2}{(n-i)^2} \leq \sum_{j=\lceil n^\varepsilon \rceil}^{\infty} \frac{2k^2}{j^2} = \frac{2k^2}{\lceil n^\varepsilon \rceil - 1}. \qquad \square$$

We finish this section with the proof of Proposition 6.4.2.

Proof of Proposition 6.4.2. Fix $c \in (0, 2)$, $k \in \mathbb{N}$ and let $m_1, \ldots, m_k < c \ln n$ be positive integers. Let $\varepsilon = (2-c)/4$ so that Proposition 6.4.5 holds for some $\beta(c) = \beta(c, \varepsilon) > 0$. For $k = 1$, the result follows from the equality in Lemma 6.4.3 and Proposition 6.4.5 since

$$\mathbf{P}(|\mathcal{S}_1| < m_1) \leq \mathbf{P}(|\mathcal{S}_1 \cap [n - \lceil n^\varepsilon \rceil]| < c \ln n) = o(n^{-\beta}).$$

For $k \geq 2$, the upper bound is likewise established immediately by Lemma 6.4.3. For the lower bound, letting $I = n - \lceil n^\varepsilon \rceil$, by Lemma 6.4.6 and Proposition 6.4.5 we have

$$\mathbf{P}(I < \tau_k, |\mathcal{S}_v \cap [I]| \geq m_v, v \in [k]) \geq 1 - \mathbf{P}(I \geq \tau_k) - \sum_{v \in [k]} \mathbf{P}(|\mathcal{S}_v \cap [I]| < m_v) \geq 1 - o(n^{-\alpha}),$$

where $\alpha < \min\{\beta, \varepsilon\}$. By Lemma 6.4.4, it follows that

$$\mathbf{P}(\deg(v) \geq m_v,\ v \in [k]) = 2^{-\sum_v m_v}(1 + o(n^{-\alpha})),$$

as required. $\square$

6.5 Proof of Proposition 6.2.1

By Corollary 6.3.4 we can study vertex degrees in $T^{(n)}$ and derive conclusions about the variables $X_i^{(n)}, X_{\geq i}^{(n)}$, $i \in \mathbb{Z}$. Recall that we write $\deg(v) = \deg_{T^{(n)}}(v)$, for $v \in [n]$.

Lemma 6.5.1. *For any $k \in \mathbb{N}$ and integers $m_1, \ldots, m_k$,*

$$\mathbf{P}\left(\deg(u) = m_u,\, u \in [k]\right) = \sum_{j=0}^{k} \sum_{\substack{S \subset [k] \\ |S|=j}} (-1)^j \mathbf{P}\left(\deg(u) \geq m_u + \mathbf{1}_{[u \in S]},\, u \in [k]\right).$$

Furthermore, for $k' \in \mathbb{N}$ and integers $m_{k+1}, \ldots m_{k+k'}$,

$$\begin{aligned} &\mathbf{P}\left(\deg(u) = m_u,\, \deg(v) \geq m_v,\, 1 \leq u \leq k < v \leq k+k'\right) \\ =&\sum_{j=0}^{k} \sum_{\substack{S \subset [k] \\ |S|=j}} (-1)^j \mathbf{P}\left(\deg(v) \geq m_v + \mathbf{1}_{[v \in S]},\, v \in [k+k']\right). \end{aligned}$$

Proof. The second equation follows by intersecting the event $\{\deg(v) \geq m_v,\, k < v \leq k+k'\}$ along all probabilities in the first equation. The first is straightforwardly proved using the inclusion-exclusion principle. $\square$

We are now ready to prove Proposition 6.2.1.

Proof of Proposition 6.2.1. Let $c \in (0,2)$ and $K \in \mathbb{N}$. Let $i < i'$ be integers such that $0 < i + \log_n < i' + \log_n < c \ln n$ and let a_j, $i \leq j \leq i'$ be non-negative integers with $a_i + \cdots a_{i'} = K$. We are interested in the factorial moments $\mathbf{E}\left[(X_{\geq i'}^{(n)})_{a_{i'}} \prod_{i \leq k < i'} (X_k^{(n)})_{a_k}\right]$.

For $i \leq k \leq i'$, for each v with $\sum_{l=i}^{k-1} a_l < v \leq \sum_{l=i}^{k} a_l$ let $m_v = \lfloor \log n \rfloor + k$. Let $K' = K - a_{i'}$, by Corollary 6.3.4 and the exchangeability of the vertex degrees of $T^{(n)}$,

$$\mathbf{E}\left[(X^{(n)}_{\geq i'})_{a_{i'}} \prod_{i\leq k<i'} (X^{(n)}_k)_{a_k}\right] = (n)_K \mathbf{P}\left(\deg(u) = m_u,\ \deg(v) \geq m_v,\ 1 \leq u \leq K' < v \leq K\right)$$
$$= (n)_K \sum_{l=0}^{K'} \sum_{\substack{S\subset[K'] \\ |S|=l}} (-1)^l \mathbf{P}\left(\deg(v) \geq m_v + \mathbf{1}_{[v\in S]},\ v\in[K]\right),$$

the last equality by Lemma 6.5.1. At this point we can apply Proposition 6.4.2 to each of the terms. Since $m_v \leq c\ln n$ for $v \in [K]$, there is $\alpha' = \alpha'(c,K) > 0$ such that

$$\sum_{l=0}^{K'} \sum_{\substack{S\subset[K'] \\ |S|=l}} (-1)^l \mathbf{P}\left(\deg(v) \geq m_v + \mathbf{1}_{[v\in S]},\ v\in[K]\right)$$
$$= \sum_{l=0}^{K'} \sum_{\substack{S\subset[K'] \\ |S|=l}} (-1)^l 2^{-l-\sum_v m_v}(1+o(n^{-\alpha'}))$$
$$= 2^{-\sum_v m_v}(1+o(n^{-\alpha'})) \sum_{l=0}^{K'} \sum_{\substack{S\subset[K'] \\ |S|=l}} (-1)^l 2^{-l}$$
$$= 2^{-K'-\sum_v m_v}(1+o(n^{-\alpha'})).$$

Using that $(n)_K = n^K(1+o(n^{-1}))$, we get

$$\mathbf{E}\left[(X^{(n)}_{\geq i'})_{a_{i'}} \prod_{i\leq k<i'} (X^{(n)}_k)_{a_k}\right] = 2^{K\log n - K' - \sum_{v=1}^{K} m_v}(1+o(n^{-\alpha}));$$

where $\alpha = \min\{\alpha', 1\}$. Finally, to complete the proof, note that

$$K\log n - K' - \sum_{v=1}^{K} m_v = \sum_{v=K'+1}^{K} (\log n - m_v) + \sum_{v=1}^{K'} (\log n - 1 - m_v)$$
$$= (-i' + \varepsilon_n)a_{i'} + \sum_{k=i}^{i'-1} (-k-1+\varepsilon_n)a_k.$$

□

6.6 Proofs of the main theorems

Proof of Theorem 6.1.2. By Theorem 11.1.VII of [25], weak convergence in $\mathcal{M}^{\#}_{\mathbb{Z}^*}$ is equivalent to convergence of FDD's, that is, convergence of every finite family of bounded continuity sets; see Definition 11.1.IV of [25]. For any point process ξ on $\mathbb{Z}$ and any $i \in \mathbb{Z}$, we have that $\mathbb{Z} \cap [i, \infty)$ is a bounded stochastic continuity set for the underlying measure of ξ in $\mathcal{M}^{\#}_{\mathbb{Z}^*}$. Thus, any FDD of ξ can be recovered from suitable marginals of the joint distribution of $(\xi(i), \ldots, \xi(i-1'), \xi[i, \infty))$ for some $i < i' \in \mathbb{Z}$.

Let $\varepsilon \in [0,1]$ and $(n_l)_{l \geq 1}$ be an increasing sequence with $\varepsilon_{n_l} \to \varepsilon$. The goal then is to prove that, for any integers $i < i'$, the joint distribution of

$$X_i^{(n_l)}, \ldots, X_{i'-1}^{(n_l)}, X_{\geq i'}^{(n_l)}$$

converges to the joint distribution of

$$\mathcal{P}^{\varepsilon}(i), \ldots, \mathcal{P}^{\varepsilon}(i'-1), \mathcal{P}^{\varepsilon}[i', \infty),$$

that is, to the law of independent Poisson r.v.'s with parameters $2^{-i-1+\varepsilon}, \ldots, 2^{-i'-2+\varepsilon}, 2^{-i'+\varepsilon}$.

We compute the limit of the factorial moments of $X_i^{(n_l)}, \ldots, X_{i'-1}^{(n_l)}, X_{\geq i'}^{(n_l)}$. For any non-negative integers $a_i, \ldots, a_{i'}$, by Proposition 6.2.1,

$$\begin{aligned}\mathbf{E}\left[(X_{\geq i'}^{(n)})_{a_{i'}} \prod_{i \leq k < i'} (X_k^{(n)})_{a_k}\right] &= \left(2^{-i'+\varepsilon_n}\right)^{a_{i'}} \prod_{i \leq k < i'} \left(2^{-(k+1)+\varepsilon_n}\right)^{a_k} (1 + o(n^{-\alpha})) \\ &\to \left(2^{-i'+\varepsilon}\right)^{a_{i'}} \prod_{i \leq k < i'} \left(2^{-(k+1)+\varepsilon}\right)^{a_k},\end{aligned}$$

as $n_l \to \infty$. The limit correspond to the factorial moment

$$\mathbf{E}\left[(\mathcal{P}^{\varepsilon}[i', \infty))_{a_i'} \prod_{i \leq k < i'} (\mathcal{P}^{\varepsilon}(k))_{a_k}\right].$$

The result follows (by, e.g. Theorem 6.10 of [52]). $\square$

Proof of Theorem 6.1.3. Since $\{\Delta_n \geq \lfloor \log n \rfloor + i\} = \{X_{\geq i}^{(n)} > 0\}$, we need only to estimate $\mathbf{P}\left(X_{\geq i}^{(n)} > 0\right)$. If $i = O(1)$, then $\exp\{-2^{-i+\varepsilon_n}\} = O(1)$ and so it suffices to prove that

$$\mathbf{P}\left(X_{\geq i}^{(n)} = 0\right) - \exp\{-2^{-i+\varepsilon_n}\}) \to 0,$$

as $n \to \infty$. This follows from Theorem 6.1.2 and the subsubsequence principle. Suppose that there exists $\delta > 0$ and a subsequence n_k for which $|\mathbf{P}\left(X_{\geq i}^{(n_k)} = 0\right) - \exp\{-2^{-i+\varepsilon_{n_k}}\}| > \delta$. Since $\{\varepsilon_{n_k}\}_{k\geq 1}$ is a bounded set there is a subsubsequence n_{k_l} such that $\varepsilon_{n_{k_l}} \to \varepsilon$ for some $\varepsilon \in [0, 1]$. By Theorem 6.1.2, $\mathbf{P}\left(X_{\geq i}^{(n_{k_l})} = 0\right) \to \exp\{-2^{-i+\varepsilon}\}$; this contradicts our assumption on the subsequence n_k.

Now consider the case $i \to \infty$ with $i + \log_n < 2 \ln n$. By a standard inclusion-exclusion argument (see, e.g., [16] Corollary 1.11),

$$\mathbf{P}\left(X_{\geq i}^{(n)} = 0\right) = \sum_{r=0}^{n} (-1)^r \frac{\mathbf{E}\left[(X_{\geq i}^{(n)})_r\right]}{r!}, \tag{6.6}$$

and this sum has the so called *alternating inequalities* property; this means that partial sums alternatively serve as upper and lower bounds for $\mathbf{P}\left(X_{\geq i}^{(n)} = 0\right)$. Consequently [1] ,

$$\mathbf{E}\left[X_{\geq i}^{(n)}\right] - \frac{1}{2}\mathbf{E}\left[(X_{\geq i}^{(n)})_2\right] \leq \mathbf{P}\left(X_{\geq i}^{(n)} > 0\right) \leq \mathbf{E}\left[X_{\geq i}^{(n)}\right]. \tag{6.7}$$

Using Proposition 6.2.1 and the fact that $i \to \infty$, we have that $\mathbf{E}\left[X_{\geq i}^{(n)}\right] = 2^{-i+\varepsilon_n}(1 + o(1))$ and

$$\mathbf{E}\left[X_{\geq i}^{(n)}\right] - \frac{1}{2}\mathbf{E}\left[(X_{\geq i}^{(n)})_2\right] = 2^{-i+\varepsilon_n}(1 + o(1)) = (1 - \exp\{-2^{-i+\varepsilon_n}\})(1 + o(1)).$$

[1] A similar lower bound for $\mathbf{P}\left(X_{\geq i}^{(n)} > 0\right)$ could be obtained from Paley-Zigmund's inequality.

The result follows. $\square$

Proof of Theorem 6.1.4. We again use the method of moments. By Theorem 1.24 of [16], it suffices to prove that, as $n \to \infty$

$$\mathbf{E}\left[(X_i^{(n)})_a\right] - (2^{-i-1+\varepsilon_n})^a = o(2^{-(i+1-\varepsilon_n)b}), \tag{6.8}$$

for all fixed $1 \leq a \leq b$. Since $i = o(\ln n)$, we have that $2^{-i-1+\varepsilon_n} = n^{o(1)}$. On the other hand, by Proposition 6.2.1 there is $\alpha > 0$ such that

$$\mathbf{E}\left[(X_i^{(n)})_a\right] - (2^{-i-1+\varepsilon_n})^a = o(n^{-\alpha}2^{-(i+\varepsilon_n)a}) = n^{-\alpha+o(1)} = o(n^{o(1)}).$$

Therefore, condition (6.8) is satisfied and the proof is complete. $\square$

CHAPTER 7
Depth of vertices with high degree in random recursive trees

Let T_n be a random recursive tree with n nodes. List vertices of T_n in decreasing order of degree as $v^1, \dots, v^n$, and write d^i and h^i for the degree of v^i and the distance of v^i from the root, respectively. We prove that, as $n \to \infty$ along suitable subsequences,

$$\left(d^i - \lfloor \log_2 n \rfloor, \frac{h^i - \mu \ln n}{\sqrt{\sigma^2 \ln n}}\right) \to ((P_i, i \geq 1), (N_i, i \geq 1)),$$

where $\mu = 1 - (\log_2 e)/2$, $\sigma^2 = 1 - (\log_2 e)/4$, $(P_i, i \geq 1)$ is a Poisson point process on $\mathbb{Z}$ and $(N_i, i \geq 1)$ is a vector of independent standard Gaussians. We additionally establish joint normality for the depths of uniformly random vertices in T_n, which extends results from [27, 59]. The joint holds even if the random vertices are conditioned to have large degree, provided the normalization is adjusted accordingly.

Our results are based on a simple relationship between random recursive trees and Kingman's n-coalescent; a utility that seems to have been largely overlooked.

7.1 Introduction

Random recursive trees have been heavily studied since their introduction in 1970 [67], and are closely related to binary search trees, preferential attachment trees and increasing trees in general, see e.g. [10, 32]. In the current work we obtain strong information about the joint law of degrees and depths of maximum and near-maximum degrees and contrast our results to similar results established for linear preferential attachment trees, see [13, 66]. We first recall basic notation and the standard construction of both random recursive trees (RRTs) and linear preferential attachment trees. We use ln to denote natural logarithms and log to denote logarithms with base 2.

For $n \geq 1$, let T_n be a random recursive tree with vertex set $[n] = \{1, \ldots, n\}$. The standard construction of RRTs, which couples the elements of $(T_n,\, n \geq 1)$, is the following: Let T_1 be a single vertex labeled 1, which is the root. For $n \in \mathbb{N}$, the tree T_{n+1} is obtained from T_n by adding an edge from a new vertex $n+1$ to a vertex $v_n \in [n]$; the choice of v_n is uniformly random, and is independent for each $n \in \mathbb{N}$. For $v \in [n]$, the depth $\mathrm{h}_{T_n}(v)$ is the distance from v to the root in T_n. We write $\mathrm{d}_{T_n}(v)$ for the number of children of v in T_n and call this the degree of v in T_n. A particular characteristic of RRTs, as contrasted with other increasing trees e.g. m-ary trees, is that for each $v \in \mathbb{N}$, almost surely $\mathrm{d}_{T_n}(v) \to \infty$ as $n \to \infty$. Let $\Delta_n = \max_{v \in T_n} \mathrm{d}_{T_n}(v)$ be the maximum degree in T_n and let $\mathcal{M}_n$ be the set of vertices in T_n attaining Δ_n.

Linear preferential attachment trees are also constructed recursively, except that the parent v_n of vertex $n+1$ is chosen with probability proportional to the degree of v_n in the current tree. More precisely, for $\alpha > 0$, the linear preferential attachment process $(T_{\alpha,n},\, n \geq 1)$ is defined as follows. Let $T_{\alpha,1}$ be a single vertex labeled 1. For $n \in \mathbb{N}$ let $T_{\alpha,n+1}$ be the tree obtained from $T_{\alpha,n}$ by adding an edge from a new vertex $n+1$ to a vertex $v_n \in [n]$. In this case, the $\mathbf{P}(v_n = v)$ is proportional to $\alpha \mathrm{d}_{T_{\alpha,n}}(v) + 1$. Note that, in this context, RRTs correspond to the case $\alpha = 0$.

For the linear preferential attachment models, it has been proven that the renormalized maximum degree $n^{-1/(2+1/\alpha)}\Delta_{\alpha,n}$ converges a.s. and in L_p to a positive, finite random variable with absolutely continuous distribution, [66].

Furthermore, the label of the vertex attaining the maximum degree is finite a.s. [13].

For random recursive trees, the picture is quite different. Naturally, if $i < j$ then $\mathrm{d}_{T_n}(i)$ stochastically dominates $\mathrm{d}_{T_n}(j)$. However, it is unlikely that the root of T_n will attain the maximum degree in T_n. By construction, $\mathrm{d}_{T_n}(i)$ is distributed as $\sum_{j=i+1}^{n} B_i$ where the summands are independent and B_j is distributed as Bernoulli$(1/j)$. It follows easily that $\mathrm{d}_{T_n}(1) = \ln n(1 + o_p(1))$. However, it is known that the maximum degree satisfies $\Delta_n / \log n \to 1$ a.s. as $n \to \infty$ [31].

It is also known that the limiting distribution of $\Delta_n - \log n$ is, up to lattice effects, a Gumbel distribution [3, 39]. The latter can be explained since the Gumbel distribution arises as the limiting distribution of the maximum of independent random variables under rather general hypotheses on the laws of such variables. The degrees of vertices in T_n are correlated and are not identically distributed, but between pairs of vertices in T_n the correlation is weak and the Gumbel limit still occurs. This was first shown by Goh and Schmutz [39] using singularity analysis of generating functions. Our approach to RRTs provides a probabilistic explanation of this phenomenon; see [3] for more details.

In [3], Addario-Berry and the author describe the number of high-degree vertices in T_n via the sequence $(\mathrm{d}_{T_n}(v) - \lfloor \log n \rfloor, v \in [n])$. They show that, along suitable subsequences, this sequence converges in distribution to a Poisson point process $\mathcal{N}$ in $\mathbb{Z}$ with $\mathbf{E}\left[|\mathcal{N} \cap [j, \infty)|\right] = \Theta(2^{-j})$ for all $j \in \mathbb{Z}$.

7.1.1 Statement of results

This work provides a detailed description of the degrees and depths of high-degree vertices in T_n. In particular we show that the number of vertices attaining the maximum degree is random and their depths are independent and asymptotically normal. Write $\mu = 1 - (\log e)/2$ and $\sigma^2 = 1 - (\log e)/4$.

Theorem 7.1.1. *For each $\varepsilon \in [0,1]$, there exists a positive integer-valued random variable M_ε such that, for any increasing sequence of integers $(n_l,\, l \geq 1)$ for which $\log n_l - \lfloor \log n_l \rfloor \to \varepsilon$ as $l \to \infty$, then $|\mathcal{M}_{n_l}|$ converges to M_ε in distribution, and*

$$\left(\frac{\mathrm{h}_{T_{n_l}}(v) - \mu \ln n_l}{\sqrt{\sigma^2 \ln n_l}},\, v \in \mathcal{M}_{n_l} \right) \xrightarrow{\mathcal{L}} (N_i,\, 1 \leq i \leq M_\varepsilon),$$

where N_i are independent standard Gaussian variables.

We remark that Theorem 7.1.1 implies that maximum-degree vertices of RRTs are constantly changing along the process $(T_n,\, n \geq 1)$.

Our main result gives a more general description of the depths of all vertices in T_n, indexed in decreasing order of their degrees. List vertices of T_n in decreasing order of degree as $v^1, \ldots, v^n$; here we break ties between vertices with the same degree by ordering them uniformly at random. Write d^i and h^i for the degree and depth of v^i, respectively. Let $\mathcal{P}$ be a Poisson point process in $\mathbb{R}$ with intensity $\lambda(x) = 2^{-x} \ln 2$. Then for $i \geq 1$, let P_i be the i-th largest point of $\mathcal{P}$ so $|\mathcal{P} \cap [P_i, \infty)| = i$ and $|\mathcal{P} \cap (P_i, \infty)| = i - 1$. This ordering is well defined as $|\mathcal{P} \cap [0, \infty)| < \infty$ almost surely.

Theorem 7.1.2. *Let N_i be independent standard Gaussian variables, $i \in \mathbb{N}$. For each $\varepsilon \in [0,1]$ and for any increasing sequence of integers $(n_l,\, l \geq 1)$ for which $\log n_l - \lfloor \log n_l \rfloor \to \varepsilon$ as $l \to \infty$, then*

$$\left(d^i - \lfloor \log n_l \rfloor, \frac{h^i - \mu \ln n_l}{\sqrt{\sigma^2 \ln n_l}} \right) \xrightarrow{\mathcal{L}} \left((\lfloor P_i + \varepsilon \rfloor, i \geq 1), (N_i, i \geq 1) \right).$$

The condition on the subsequence n_l in Theorems 7.1.1 and 7.1.2 is due to a lattice effect on the law of $(\lfloor P_i + \varepsilon \rfloor, i \geq 1)$ caused by the fact that degrees are integer-valued.

Our last result provides information about vertices with degree near $a\Delta_n$ for fixed $a \in [0,1]$. For $a \in [0,1]$, let $\mu_a = 1 - (a \log e)/2$ and $\sigma_a^2 = 1 - (a \log e)/4$; note that $\mu = \mu_1$ and $\sigma = \sigma_1$.

Theorem 7.1.3. *Fix $k \in \mathbb{N}$ and let $(u_i,\, i \in [k])$ be k distinct vertices in T_n chosen uniformly at random. For every $(a_1, \ldots, a_k) \in [0,1]^k$ and $(b_1, \ldots, b_k) \in \mathbb{Z}^k$, the conditional law of*

$$\left(\frac{\mathrm{h}_{T_n}(u_i) - \mu_{a_i} \ln n}{\sqrt{\sigma_{a_i}^2 \ln n}},\, i \in [k] \right),$$

given that $\mathrm{d}_{T_n}(u_i) \geq \lfloor a_i \log n \rfloor + b_i$ for all $i \in [k]$, converges to the law of k independent standard Gaussian variables.

Note that the case $b_i = a_i = 0$ for all $i \in [k]$ of Theorem 7.1.3 involves no conditioning, and thus yields the joint distribution for the depths of k uniformly random vertices in T_n. This extends the results of the papers [27, 59] where the case for $k = 1$, $a_1 = b_1 = 0$ of Theorem 7.1.3 is established. These results were obtained in the context of analyzing the *insertion depth*, $\mathrm{h}_{T_n}(n)$ of RRTs, important for the analysis of data structures in computer science.

Theorem 7.1.1 is a quite straightforward consequence of Theorem 7.1.2, whose proof relies essentially on Theorem 7.1.3. The proof of Theorem 7.1.3 exploits the relation between degrees and depths of vertices in a different random tree $T^{(n)}$ whose shape has the same law as that of T_n. This alternative tree $T^{(n)}$ is constructed through Kingman's coalescent, as described in Section 7.2.1. A binary tree representation of Kingman's coalescent had been previously used to study a data structure known as union-find trees, [27]. Pittel mentions the connection between the results of [27] and the height of RRTs in [75]. However, although the connection between Kingman's coalescent and random recursive trees had been observed, prior to our previous work with

Addario-Berry [3], its utility in studying vertex degrees seems to have gone unremarked.

7.1.2 The point process in Theorem 7.1.2

In this section we briefly explain how we use the method of moments (e.g., see [52] Section 6.1) to obtain the limiting distribution of a sequence of (marked) point processes. In particular, we present an alternative characterization of the processes involved in Theorem 7.1.2. Although this change of perspective requires the introduction of further notation, the problem of establishing Theorem 7.1.2 becomes, in fact, more tractable.

We start by considering the unmarked processes $(d^i - \lfloor \log n \rfloor,\ i \in [n])$ and $\mathcal{P} = (P_i,\ i \geq 1)$. Define, for each $n \in \mathbb{N}$, $\varepsilon_n = \log n - \lfloor \log n \rfloor$. We consider a fixed $\varepsilon \in [0,1]$ and increasing sequence n_l such that $\varepsilon_{n_l} \to \varepsilon$ as $l \to \infty$.

For $j \in \mathbb{Z}$, we define the following counting measures of the sequence $\mathcal{P}^\varepsilon = (\lfloor P_i + \varepsilon \rfloor, i \geq 1)$;

$$X_j = \#\{i \geq 1 : \lfloor P_i + \varepsilon \rfloor = j\},$$

$$X_{\geq j} = \#\{i \geq 1 : \lfloor P_i + \varepsilon \rfloor \geq j\}.$$

Note that $X_j \stackrel{\mathcal{L}}{=} \mathrm{Poi}(2^{-j+\varepsilon-1})$ and $X_{\geq j} \stackrel{\mathcal{L}}{=} \mathrm{Poi}(2^{-j+\varepsilon})$; in particular, the number of points of $\mathcal{P}^e ps$ on any interval $[j, \infty]$ is finite almost surely. Therefore, $\mathcal{P}^e ps$ is characterized by the collection of joint distributions $(X_{j'}, \ldots, X_{j-1}, X_{\geq j})$ for any integers $j' < j$; see e.g. Section 3.1 of [24] and Section 9.2 of [25]. Similarly, the collection of the joint distribution of the variables

$$X_j^{(n)} = \#\{v \in [n] : \mathrm{d}_{T_n}(v) = \lfloor \log n \rfloor + j\},$$

$$X_{\geq j}^{(n)} = \#\{v \in [n] : \mathrm{d}_{T_n}(v) \geq \lfloor \log n \rfloor + j\}$$

characterizes the law of the sequence $(\mathrm{d}^i - \lfloor \log n \rfloor, i \geq 1)$. Finally, to prove that $(\mathrm{d}^i - \lfloor \log n_l \rfloor, i \geq 1) \xrightarrow{\mathcal{L}} \mathcal{P}^\varepsilon$ it suffices to show that, for all $j' < j \in \mathbb{Z}$

$$(X_{j'}^{(n_l)}, \ldots, X_{j-1}^{(n_l)}, X_{\geq j}^{(n_l)}) \xrightarrow{\mathcal{L}} (X_{j'}, \ldots, X_{j-1}, X_{\geq j}). \tag{7.1}$$

Next, for any $r \in \mathbb{R}$ and $a \in \mathbb{N}$, let $(r)_a = r(r-1)\cdots(r-a+1)$ and set $(r)_0 = 1$. Recall that, if $X \overset{\mathcal{L}}{=} \mathrm{Poi}(\lambda)$, $\mathbf{E}[(X)_a] = \lambda^a$ for all integers $a \geq 0$. Now, using the method of moments, the following estimates imply (7.1).

Proposition 7.1.4 (Proposition 2.1, [3]). *For all $c \in (0,2)$ and $A \in \mathbb{N}$ there is $\beta = \beta(c, A) > 0$ such that the following holds. If $j' = j'(n)$ and $j = j(n)$ are integer-valued functions with $0 < j' + \log n < j + \log n < c \ln n$, then uniformly over non-negative integers $a_{j'}, \ldots, a_j$ with $a_{j'} + \ldots + a_j = A$, we have*

$$\mathbf{E}\left[(X_{\geq j}^{(n)})_{a_j} \prod_{j' \leq k < j} (X_k^{(n)})_{a_k}\right] = \left(2^{-j+\varepsilon_n}\right)^{a_j} \prod_{j' \leq k < j} \left(2^{-(k+1)+\varepsilon_n}\right)^{a_k} (1 + o(n^{-\beta})).$$

Marked point processes are, in fact, point processes in a larger space; thus, the same approach can be used when we add the information of the depths $((\mathrm{h}^i - \mu \ln n)/\sqrt{\sigma^2 \ln n}, i \geq 1)$ and the marks $(N_i,\ i \geq 1)$. Let us define subsets of $\mathbb{Z} \times \mathcal{B}(\mathbb{R})$ that will help us define the FDDs of our marked point processes; see Figure 7–1 for an example. It suffices to consider the set

$$\mathcal{B}_I = \{(-\infty, b],\ (a, b],\ (a, \infty);\ -\infty < a < b < \infty\}.$$

Definition 7.1.5. *Fix positive integers $K' < K$. If the pairs $(j_k, B_k) \in \mathbb{Z} \times \mathcal{B}_I$, $k \in [K]$, satisfy*

1. *$j_1 \leq j_2 \leq \cdots \leq j_{K'} < j = j_{K'+1} = \cdots = j_K$ and*
2. *for all $1 \leq k < l \leq K$, if $j_k = j_l$ then $B_k \cap B_l = \emptyset$;*

*then we say that $((j_k, B_k),\ k \in [K])$ is a (K', K)-*canonical FDD *sequence.*

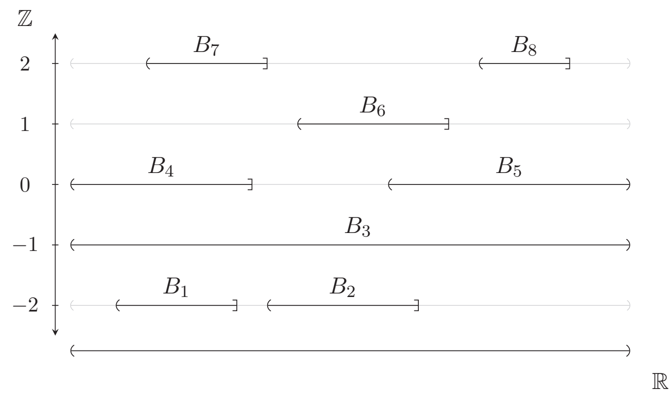


Figure 7–1: An example of a (K', K)-canonical set. In this case, $K' = 6$, $K = 8$ and $j_1 = -2$, $j_8 = 2$.

Also, for $(j, B) \in \mathbb{Z} \times \mathcal{B}_I$, let

$$X_j(B) = \#\{i \geq 1 : \lfloor P_i + \varepsilon \rfloor = j,\ N_i \in B\}$$
$$X_j^{(n)}(B) = \# \left\{ v \in [n] : \mathrm{d}_{T_n}(v) = \lfloor \log n \rfloor + j,\ \frac{\mathrm{h}_{T_n}(v) - \mu_1 \ln n}{\sqrt{\sigma_1^2 \ln n}} \in B \right\};$$

and let $X_{\geq j}(B)$, $X_{\geq j}^{(n)}(B)$ be defined accordingly.

Now the convergence in distribution of point processes is equivalent to the convergence of its finite dimensional distributions (FDD); see [25, Theorem 11.1.VII]. In our case, this leads to the following lemma.

Lemma 7.1.6. *The following are equivalent.*

a) As $l \to \infty$,

$$\left(d^i - \lfloor \log n_l \rfloor, \frac{h^i - \mu \ln n_l}{\sqrt{\sigma^2 \ln n_l}} \right) \xrightarrow{\mathcal{L}} ((\lfloor P_i + \varepsilon \rfloor, i \geq 1), (N_i, i \geq 1)),$$

b) For every (K', K)*-canonical FDD sequence* $((j_k, B_k),\ 1 \leq k \leq K)$ *as* $l \to \infty$,

$$\begin{aligned} &(X_{j_1}^{(n_l)}(B_1), \ldots, X_{j_{K'}}^{(n_l)}(B_{K'}), X_{\geq j_{K'+1}}^{(n_l)}(B_{K'+1}), \ldots, X_{\geq j_K}^{(n_l)}(B_K)) \\ &\xrightarrow{\mathcal{L}} (X_{j_1}(B_1), \ldots, X_{j_{K'}}(B_{K'}), X_{\geq j_{K'+1}}(B_{K'+1}), \ldots, X_{\geq j_K}(B_K)). \end{aligned}$$

Let Φ denote the measure of a standard Gaussian variable; that is $\Phi(A) = \int_A e^{-x^2/2} dx/\sqrt{2\pi}$ for any $A \subset \mathbb{R}$.

Fact 7.1.7. *For all $j \in \mathbb{Z}$ and $B \subset \mathbb{R}$, $X_j(B) \stackrel{\mathcal{L}}{=} \mathrm{Poi}(2^{-j+\varepsilon-1}\Phi(B))$ and $X_{\geq j}(B) \stackrel{\mathcal{L}}{=} \mathrm{Poi}(2^{-j+\varepsilon}\Phi(B))$; additionally, for any (K', K)-canonical FDD sequence, the variables*

$$(X_{j_1}(B_1), \ldots, X_{j_{K'}}(B_K), X_{\geq j_{K'+1}}(B_{K'+1}), \ldots, X_{\geq j_K}(B_K))$$

are independent.

Below is the more general form of moment estimation which is required to obtain Theorem 7.1.2.

Proposition 7.1.8. *Fix $c \in (0, 2)$ and $M \in \mathbb{N}$. Let $j = j(n)$ and $j' = j(n)$ are integer-valued functions with $0 \leq j'(n) + \log n < j(n) + \log n < c \ln n$, and let $K \in \mathbb{N}$ and non-negative integers $(a_k,\, k \in [K])$ $K \in \mathbb{N}$ be such that $\sum_{k\in[K]} a_k = M$. Then uniformly over (K', K)-canonical sequences $((j_k, B_k),\, 1 \leq k \leq K)$ with $j' = j_1$ and $j = j_K$, we have*

$$\begin{aligned}
&\mathbf{E}\left[\prod_{k=1}^{K'} \left(X_{j_k}^{(n)}(B_k)\right)_{a_k} \prod_{k=K'+1}^{K} \left(X_{\geq j}^{(n)}(B_k)\right)_{a_k}\right] \\
=&\prod_{k=1}^{K'} \left(2^{-j_k+\varepsilon_n-1}\Phi(B_k)\right)^{a_k} \prod_{k=K'+1}^{K} \left(2^{-j'+\varepsilon_n}\Phi(B_k)\right)^{a_k} (1 + o(1)).
\end{aligned}$$

Proof of Theorem 7.1.2 (assuming Proposition 7.1.8). Fix $\varepsilon \in [0, 1]$ and let n_l be an increasing sequence with $\varepsilon_{n_l} \to \varepsilon$ as $l \to \infty$. Let $K' < K$ and $((j_k, B_k),\, 1 \leq k \leq K)$ be a fixed (K', K)-canonical FDD sequence. Set $c = 3/2$, which implies for n large enough that $0 \leq j_1 + \log n < j_K + \log n < c \ln n$. Thus, Proposition 7.1.8 implies, by the method of moments, that the first condition in Lemma 7.1.6 is satisfied for each (K, K')-canonical FDD sequence. This completes the proof of Theorem 7.1.2. $\square$

We briefly explain a key ingredient to proving Proposition 7.1.8. Note that each $X_j^{(n)}(B)$ and $X_{\geq j}^{(n)}(B)$ is a sum of indicator variables. Therefore, the expectations of their factorial moments are reduced to a sum of probabilities as follows: for each $S \subset [n]$, collection $B_j \subset \mathcal{B}_I$ and sequence $m_j < 2 \ln n$,

$$\begin{aligned}&\mathbf{P}\left(\deg_{T_n}(v_j) \geq m_j,\, \mathrm{h}_{T_n}(v_j) \in B_j,\, v_j \in S\right)\\ =&\mathbf{P}\left(\deg_{T_n}(v_j) \geq m_j,\, v_j \in S\right) \mathbf{P}\left(\mathrm{h}_{T_n}(v_j) \in B_j,\, v_j \in S \,|\, \deg_{T_n}(v_j) \geq m_j,\, v_j \in S\right).\end{aligned} \tag{7.2}$$

The first factor in (7.2) has been analyzed in [3]; the result we need from that paper is restated below as Proposition 7.5.4. The second factor in (7.2) is bounded in Theorem 7.2.5.

We now turn to describing Kingman's coalescent.

7.2 A Kingman's coalescent approach

The connection between RRTs and Kingman's coalescent is central to understanding the close relation between degree and depth of vertices reflected in Theorem 7.1.3, which is key to the proofs of Theorems 7.1.1 and 7.1.2. Therefore, we briefly sketch the role that Theorem 7.1.3 plays in the proof of Theorem 7.1.2. In the next section we define the tree $T^{(n)}$, after which we discuss the contents of the remainder of the paper.

7.2.1 Kingman's coalescent process

In this section we give a representation of Kingman's coalescent in terms of labeled forests and connect this with RRTs. For a general description of Kingman's coalescent, see [9, Chapter 2]; the construction below is based on that given in [1]. For the remainder of the paper, all trees are rooted and we use $r(t)$ to denote the root of tree t. We write $V(t)$ and $E(t)$ for the sets of vertices and edges of t, respectively. By convention, we assume that edges of a tree t are directed towards $r(t)$ and an edge directed from u to v is denoted by uv. If t has n vertices, we say that t has size n; we also write $\mathrm{d}_t(v)$ and $\mathrm{h}_t(v)$ for the degree and depth of vertex v in t.

A rooted labeled tree t is *increasing* if its labels are increasing along root-to-leaf paths. Let us write $\mathcal{I}_n = \{t : t \text{ is increasing}, V(t) = [n]\}$ to denote the set of increasing trees on $[n]$. It is not difficult to see that T_n is a uniformly random element of $\mathcal{I}_n$ and that $|\mathcal{I}_n| = (n-1)!$.

A forest f is a set of trees whose vertex sets are pairwise disjoint. Denote by $V(f)$ and $E(f)$, respectively, the union of the vertex and edge sets of the trees contained in f. For each $n \geq 1$, we consider the set of forests $\mathcal{F}_n = \{f : V(f) = [n]\}$ with vertex labels $[n]$. An n-chain is a sequence $C = (f_n, \ldots, f_1)$ of elements of $\mathcal{F}_n$ if for $1 < i \leq n$, f_{i-1} is obtained from f_i by adding an edge connecting two of the roots in f_i. In particular, f_n contains n one-vertex trees, and f_1 contains exactly one tree denoted by $t_C \in \mathcal{F}_n$.

For an n-chain $(f_n, \ldots, f_1) \in \mathcal{CF}_n$ and $1 \leq i \leq n$, we write $f_i = \{t_1^{(i)}, \ldots, t_i^{(i)}\}$, ordering of the trees is in increasing order of their smallest-labeled vertex.

Definition 7.2.1. *The following constructs Kingman's n-coalescent as a random n-chain $\mathbf{C} = (F_n, \ldots, F_1)$.*

For each $1 < i \leq n$, choose $\{a_i, b_i\} \subset \{\{a, b\} : 1 \leq a < b \leq i\}$ independently and uniformly at random; also let $(\xi_i,\, i \in [n-1])$ be a sequence of independent Bernoulli$(1/2)$ *random variables.*

For $1 \leq i < n$, F_i is obtained from F_{i+1} as follows. Add an edge e_i between the roots of $r(T_{a_{i+1}}^{(i+1)})$ and $r(T_{b_{i+1}}^{(i+1)})$; direct e_i towards $r(T_{a_{i+1}}^{(i+1)})$ if $\xi_i = 1$, and towards $r(T_{b_{i+1}}^{(i+1)})$ otherwise. Then F_i contains the new tree and the remaining $i-1$ unaltered trees from F_{i+1}.

For an example of the process see Figure 7–2.

Lemma 7.2.2. *Kingman's n-coalescent $\mathbf{C}$ is uniformly random in $\mathcal{CF}_n$, the set of n-chains.*

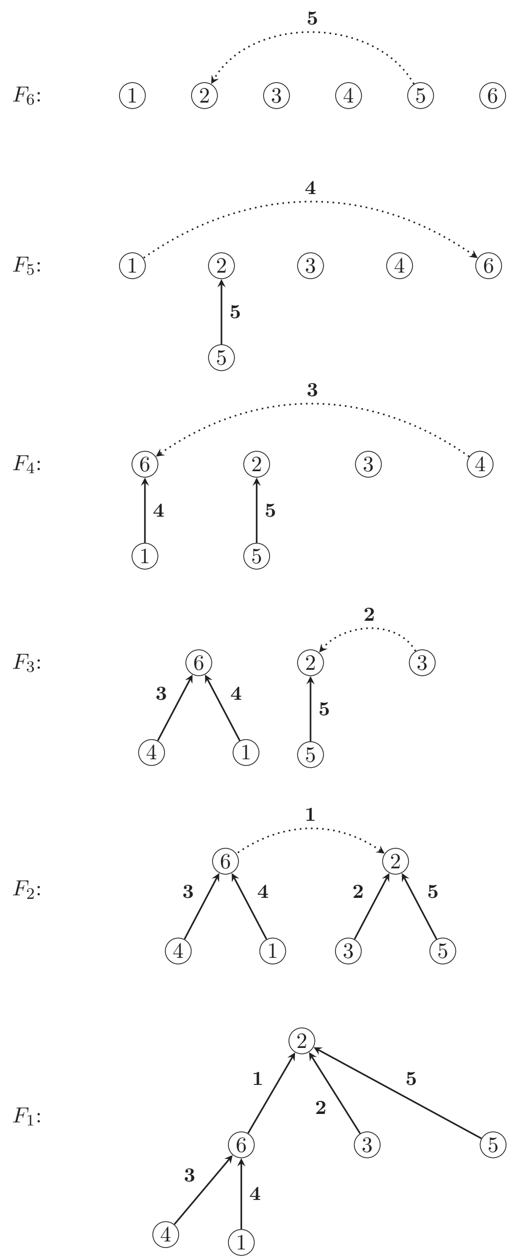


Figure 7–2: An example of Kingman's n-coalescent $\mathbf{C} = (F_n, \ldots, F_1)$ for $n = 6$. For $1 < i \leq n$, we present the edge $E(F_{i-1}) \setminus E(F_i)$ with a dotted line in F_i. Edges are marked with the labels ρ_C; $n - \rho_C(e)$ is the first forest where e is present. In this case, $\xi_6 = \xi_4 = \xi_3 = 1$, $\xi_5 = \xi_2 = 0$ and $\{a_5, b_5\} = \{2, 5\}$, $\{a_4, b_4\} = \{1, 5\}$, $\{a_3, b_3\} = \{1, 4\}$, $\{a_2, b_2\} = \{2, 3\}$, $\{a_1, b_1\} = \{1, 2\}$.

Proof. Any $(f_n, \ldots, f_1) \in \mathcal{CF}_n$ is determined by the order in which the edges of t_C are added. For each $2 \le i < n$, there are $(i+1)i$ possible oriented edges between the roots in f_{i+1} and only one of them is $e \in E(f_i) \setminus E(f_{i+1})$. Thus,

$$\mathbf{P}\left((F_n, \ldots, F_1) = (f_n, \ldots, f_1)\right) = \prod_{k=1}^{n-1} \mathbf{P}\left(F_k = f_k | F_j = f_j,\, k < j \le n\right) \left[n!(n-1)!\right]^{-1}.$$

This expression holds for all $(f_n, \ldots, f_1) \in \mathcal{CF}_n$, so the result follows. □

Let $e_{n-1}, \ldots, e_1$ be the edges of t_C ordered as they were added to the chain C. That is, $e_i \in E(F_i)$ while $e_i \notin E(F_{i+1})$ for all $1 \le i < n$. Now, write $e_i = v_i w_i$. Let $\sigma_C : V(t_C) \to [n]$ be defined as $\sigma_C(r(t_C)) = 1$ and for each $e_i = v_i w_i \in E(t_C)$,

$$\sigma_C(v_i) = i + 1.$$

This is well defined as all edges are directed towards the root, so $v_i \ne v_j$ for all $i, j \in [n-1]$. Note that for each $1 \le i < n$, e_i is directed towards the root of the new tree in f_i. Thus, the labels $\{\sigma_C(v),\, v \in [n]\}$ decrease along leaf-to-root paths in t_C. As a consequence, we obtain an increasing tree by relabeling the vertices of t_C using σ_C.

Proposition 7.2.3. *For each $C = (f_n, \ldots, f_1) \in \mathcal{CF}_n$, relabel the vertices in t_C with σ_C to obtain $\phi(C) \in \mathcal{I}_n$. Then the law of $\phi(\mathbf{C})$ is that of a RRT of size n.*

Proof. From the argument in the proof of Lemma 7.2.2, we have that $|\mathcal{CF}_n| = n!(n-1)!$. Next, we show that ϕ is onto and, additionally, an $n!$-to-1 mapping. Thus ϕ preserves the uniform measure from $\mathcal{CF}_n$ to $\mathcal{I}_n$.

Fix an increasing tree $t \in \mathcal{I}_n$. Every vertex $j > 1$ has outdegree 1 in t, thus we write uniquely define $v_j \in V(t)$ such that $jv_j \in E(t)$. For each $1 < j \le n$, let $e_{j-1} = jv_j$. Consider an n-chain $C = (f_n, \ldots, f_1)$ defined as follows. Let $f_n \in \mathcal{F}_n$ have no edges, and for each $1 \le i < n$, construct f_i from f_{i+1} by adding the edge e_i. It is easy to see that C satisfies $\sigma_C(i) = i$ for all $i \in [n]$ and t_C. Therefore $\phi(C) = t$, showing that ϕ is onto.

Now, consider $C \in \mathcal{CF}_n$ such that $\phi(C) = t$. For each permutation $\pi : [n] \to [n]$, let C_π be the n-chain obtained from $C = (f_n, \ldots, f_1)$ by applying π to each of the labels of $V(f_i)$, $i \in [n]$. The mapping ϕ does not depend of the vertex labels in C, but on the order in which edges are added; therefore, $\phi(C) = \phi(C_\pi)$ for all permutations π. This shows that $|\phi^{-1}(t)| \geq n!$ for any $t \in \mathcal{I}_n$, completing the proof. $\square$

For each n, let $\mathbf{C}$ be a Kingman's n-coalescent and let $T^{(n)} = t_{\mathbf{C}}$. Since $\phi(\mathbf{C})$ only relabels vertices in T_C, it follows that the shape of the tree is preserved; and so are the degrees and depths of the vertices. That is, as multisets,

$$\{(deg_{T^{(n)}}(v), h_{T^{(n)}}(v))\}_{v\in[n]} = \{(deg_{\phi(\mathbf{C})}(v), h_{\phi(\mathbf{C})}(v))\}_{v\in[n]}.$$

Moreover, for each $t \in \mathcal{I}_n$ the set $\phi^{-1}(t)$ can be indexed by permutations on $[n]$. This directly implies the following key corollary of Proposition 7.2.3.

Corollary 7.2.4. *For all $n \in \mathbb{N}$,*

$$((\mathrm{d}_{T^{(n)}}(i), \mathrm{h}_{T^{(n)}}(i)),\ i \in [n]) = ((\mathrm{d}_{T_n}(\sigma(i)), \mathrm{h}_{T_n}(\sigma(i))),\ i \in [n]);$$

where σ is a uniformly random permutation of $[n]$ and is independent of T_n. Consequently, the following equality in distribution holds jointly for all $i \in \mathbb{Z}$ and $j \in \mathbb{N}$,

$$|\{v \in [n] : \mathrm{d}_{T_n}(v) = i,\ \mathrm{h}_{T_n} = j\}| = |\{v \in [n] : \mathrm{d}_{T^{(n)}}(v) = i,\ \mathrm{h}_{T^{(n)}}(v) = j\}|$$

Proof. For any $n \in \mathbb{N}$, let $\mathcal{P}_n$ be the set of permutations on $[n]$. For any n-chain $C = (f_n, \ldots, f_1)$ let $\varphi(C) = (\phi(C), \sigma_C)$. Then $\varphi : \mathcal{CF}_n \to \mathcal{I}_n \times \mathcal{P}_n$ is a bijection and the result follows. $\square$

7.2.2 Conditional depths of high-degree vertices

In this section we provide a heuristic for the approach we use to study the conditional distributions involved in Theorem 7.2.5 below, which is equivalent to Theorem 7.1.3, and also outline the remainder of the paper.

Fix $n \in \mathbb{N}$ and consider Kingman's n-coalescent $\mathbf{C} = (F_n, \ldots, F_1)$. For each vertex $v \in [n]$ and $1 \leq i \leq n$, let $T_i(v)$ be the tree in F_i that contains v. We use $\mathrm{d}_{F_i}(v)$ and $\mathrm{h}_{F_i}(v)$ to denote the degree and depth of v in $T_i(v)$. Recall that $T^{(n)} = t_{\mathbf{C}}$ is the unique tree in F_1; for simplicity, we use $\mathrm{d}_n(v)$ and $\mathrm{h}_n(v)$ for the degree and depth of vertices in $T^{(n)}$.

Theorem 7.2.5. *Fix $k \in \mathbb{N}$. For any $(a_1, \ldots, a_k) \in [0,1]^k$ and $(b_1, \ldots, b_k) \in \mathbb{Z}^k$, the conditional law of*

$$\left(\frac{\mathrm{h}_n(i) - \mu_{a_i} \ln n}{\sqrt{\sigma_{a_i}^2 \ln n}}, \, i \in [k] \right),$$

given that $\mathrm{d}_n(i) \geq \lfloor a_i \log n \rfloor + b_i$, $i \in [k]$, converges to the law of k independent standard Gaussian variables.

Remark 7.2.6 (Proof of Theorem 7.1.3)**.** *By Corollary 7.2.4, Theorem 7.1.3 follows from Theorem 7.2.5.*

In this section we give a heuristic for Theorem 7.2.5, when $k = 1$. First, we analyze the case $m_1 = m_1(a_1, b_1, n) = \lfloor a_1 \log n \rfloor + b_1 \leq 0$, in which $\{\mathrm{d}_n(1) \geq m_1\}$ occurs; and second $m_1 > 0$. Finally, we discuss the obstacles in treating several vertices, that is, when $k \geq 2$.

We next define indicator functions $(s_{i,v}, \, 2 \leq i \leq n)$ and the *selection set* $\mathcal{S}_n(v)$ as follows, let $s_{i,v}$ be the indicator that $T_i(v) \in \{T_{a_i}^{(i)}, T_{b_i}^{(i)}\}$; that is, $s_{i,v} = 1$ when $T_i(v) \in F_i$ is chosen to be merged and form a larger tree in F_{i-1}, and otherwise $s_{i,v} = 0$. Now we set

$$\mathcal{S}_n(v) = \{2 \leq i \leq n : \, s_{i,v} = 1\}.$$

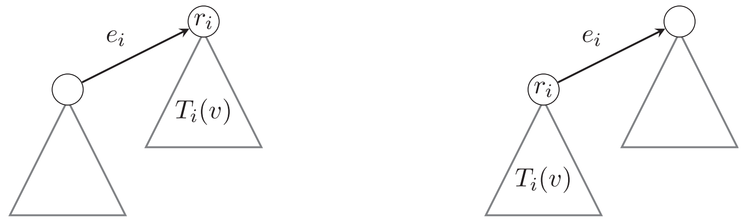


Figure 7–3: For $1 < i \leq n$ let $r_i = r(T_i(v))$ and suppose $i \in \mathcal{S}_n(v)$. If e_i is directed towards r_i, then the degree of r_i increases by one in F_{i-1}. If e_i is directed outwards r_i, then the depth of each $u \in T_i(v)$ increases by one in F_{i-1}.

The selection set $\mathcal{S}_n(v)$ keeps track of each time i where $T_i(v)$ merges. The choice of trees to be merged at each step is both independent and uniform. Thus, for fixed $v \in [n]$, the variables $(s_{i,v},\, 2 \leq i \leq n)$ are independent Bernoulli random variables, with $\mathbf{E}[s_{i,v}] = 2/i$. This implies that $\mathbf{E}[|\mathcal{S}_n(v)|] = \sum_{i=1}^{n} \frac{2}{i} = 2\ln n + O(1)$ and $\mathbf{Var}[|\mathcal{S}_n(v)|] = \sum_{i=1}^{n} \left(\frac{2}{i} - \frac{4}{i^2}\right) = 2\ln n + O(1)$. It is straightforward to see that the Lindenberg conditions are satisfied by $|\mathcal{S}_n(v)|$ and thus, the following holds for any vertex $v \in \mathbb{N}$,

$$\frac{|\mathcal{S}_n(v)| - 2\ln n}{\sqrt{2\ln n}} \xrightarrow{\mathcal{L}} N; \tag{7.3}$$

as $n \to \infty$ and where N is a standard Gaussian variable. Moreover, Bernstein's inequalities (see, e.g. [52, Theorem 2.8 and (2.9)]) yield that, for any $\delta > 0$,

$$\mathbf{P}\left(||\mathcal{S}_n(v)| - \mathbf{E}[|\mathcal{S}_n(v)|]| > \delta\mathbf{E}[|\mathcal{S}_n(v)|]\right) = o(1). \tag{7.4}$$

Now, consider the indicator random variables $(\kappa_{i,v},\, 2 \leq i \leq n)$ where $\kappa_{i,v} = 1$ precisely when $s_{i,v} = 1$ and the edge added to F_i is directed outwards of $r(T_i(v))$. The latter condition depends only on ξ_i and thus $\mathbf{E}[\kappa_{i,v}] = 1/i$. Recall that e_i is the edge added to F_{i+1} to obtain F_i. If e_i is directed towards $r(T_{i+1}(v))$, the degree of $r(T_{i+1}(v))$ increases by one in F_i. Otherwise, e_i is directed outwards $r(T_{i+1}(v))$ and all vertices in $T_{i+1}(v)$ increase their depth by one in F_i. Therefore $\mathrm{h}_{F_j}(v) = \sum_{i=j+1}^{n} \kappa_{i,v}$, and in particular

$$\mathrm{h}_n(v) = \sum_{i=2}^{n} \kappa_{i,v}.$$

Similarly to (7.3), it follows that $(h_n(v) - \ln n)/\sqrt{\ln n}$ converges in distribution to a standard Gaussian variable; this already solves the case when $m_1 \leq 0$. However, such arguments cannot be directly applied to the case when $m_1 > 0$ or when $k \geq 2$. We next describe a slightly different proof that $(\mathrm{h}_n(v) - \ln n)/\sqrt{\ln n}$ is asymptotically normal, which we later extend to cover the general case of Theorem 7.2.5.

The direction of the edge e_i is determined by a Bernoulli$(1/2)$, independent of the choice of trees to be merged. Thus, we have the following distributional equality,

$$\mathrm{h}_n(1) \stackrel{\mathcal{L}}{=} \mathrm{Bin}(|\mathcal{S}_n(1)|, 1/2). \tag{7.5}$$

Now, from (7.3), it follows that there exist random variables $X_n \xrightarrow{\mathcal{L}} N$ such that

$$S_n = |\mathcal{S}_n(1)| = 2\ln n + X_n\sqrt{2\ln n}.$$

Similarly, the central limit theorem allows us to write $\mathrm{Bin}(2m, 1/2) = m + \frac{Y_m}{2}\sqrt{2m}$ with $Y_m \xrightarrow{\mathcal{L}} N'$, N' a standard Gaussian variable. We then have

$$\begin{aligned}\mathrm{Bin}(S_n, 1/2) = \frac{S_n}{2} + \frac{Y_{S_n/2}}{2}\sqrt{S_n} &= \frac{2\ln n + X_n\sqrt{2\ln n}}{2} + \frac{Y_{S_n/2}}{2}\sqrt{S_n} \\ &\approx \ln n + \frac{X_n + Y_{\ln n}}{\sqrt{2}}\sqrt{\ln n};\end{aligned}$$

in the last approximation, we neglect the variations of S_n around $2\ln n$. The Binomial variable is determined by the coin flips ξ_i which are independent of

$\mathcal{S}_n(v)$. Thus their (limiting) fluctuations, N and N', should behave independently. It now follows that

$$\frac{\mathrm{h}_n(1) - \ln n}{\sqrt{\ln n}} \approx \frac{1}{\sqrt{2}}(X_n + Y_{\ln n}) \approx \frac{1}{\sqrt{2}}(N + N') \tag{7.6}$$

where the latter expression has a standard Gaussian distribution. This gives a heuristic of the limiting distribution of $\mathrm{h}_n(1)$ without any conditioning.

To prepare for the proof of Theorem 7.2.5, we next state a lemma describing the joint law of the depth and degree of a given vertex.

Lemma 7.2.7. *Fix $v \in [n]$, let G be $\mathrm{Geo}(1/2)$ independent of $\mathcal{S}_n(v)$ and let $D = \min\{G, |\mathcal{S}_n(v)|\}$. Then, $\mathrm{d}_n(v) \stackrel{\mathcal{L}}{=} D$ and for all $k, l \in \mathbb{N}$,*

$$\mathbf{P}\left(\mathrm{d}_n(v) \geq k, \mathrm{h}_n(v) \leq l\right) = 2^{-k}\mathbf{P}\left(\mathrm{Bin}(|\mathcal{S}_n(v)| - k, 1/2) \leq l,\ |\mathcal{S}_n(v)| \geq k\right).$$

Proof. Any vertex starts as the root of a single-vertex tree. If $|\mathcal{S}_n(v)| = m$, then we flip a fair coin m times and set $\mathrm{d}_n(v)$ as the length of the first streak of heads and $\mathrm{h}_n(v)$ as the total number of tails; this proves the distributional identity of $\mathrm{d}_n(v)$.

Moreover, if $\mathrm{d}_n(v) \geq k$, then $|\mathcal{S}_n(v)| \geq k$ and the first k coin flips are determined to be heads, the latter event occurring with probability 2^{-k}. The remaining $|\mathcal{S}_n(v)| - k$ coin flips are independent of the previous tosses. $\square$

Using Lemma 7.2.7, we have for all $k \geq m_1 = \lfloor a_1 \ln n \rfloor + b_1$,

$$\mathbf{P}\left(|\mathcal{S}_n(v)| \geq k \,|\, \mathrm{d}_n(v) \geq m_1\right) = \frac{\mathbf{P}\left(|\mathcal{S}_n(v)| \geq k\right)}{\mathbf{P}\left(|\mathcal{S}_n(v)| \geq m_1\right)} = (1 + o(1))\mathbf{P}\left(|\mathcal{S}_n(v)| \geq k\right);$$

the last equality by use the bounds in (7.4) and the fact that for any $a_1 \in [0, 1]$ and $b_1 \in \mathbb{Z}$, we have $m_1 < (3/2)\ln n$ for n large enough.

Thus, conditioning on the event $\{\mathrm{d}_n(1) \geq m_1\}$ does not have a real impact on the distribution of $|\mathcal{S}_n(1)|$. Therefore, $\mathrm{h}_n(1)$ depends essentially on $(2 - a_1)\ln n$ fair coin flips. In other words, the conditional law of $h_n(1)$, given that

$\mathrm{d}_n(1) \geq m_1$ satisfies

$$\mathrm{h}_n(1) \approx \mathrm{Bin}(S_n - a\lfloor \log n \rfloor - m, 1/2) \approx (1 - (a_1 \log e)/2) \ln n + \frac{X_n + Y_{\ln n}}{\sqrt{2}} \sqrt{\ln n}. \tag{7.7}$$

This suggests that, using a suitable choice of renormalizing constants, the conditional law of $\mathrm{h}_n(1)$ given that $\mathrm{d}_n(1) \geq m_1$ has an asymptotic normal distribution.

To conclude the proof outline for Theorem 7.2.5, we briefly explain how the depths of distinct vertices are correlated. For $k \geq 2$, the joint distribution of $(\mathrm{h}_n(v),\, v \in [k])$ does not depend only on the sizes of the selection sets $(\mathcal{S}_n(v),\, v \in [k])$, but also on their overlaps (i.e. on the sets $\mathcal{S}_n(v) \cap \mathcal{S}_n(w)$, for $v, w \in [k]$).

For distinct vertices v, w, let $\lambda_{v,w} = \max\{2 \leq l \leq n :\, l \in \mathcal{S}_n(v) \cap \mathcal{S}_n(w)\}$. Then, $\lambda_{v,w}$ is the first time that both the trees containing v and w are merged together; moreover, the merging of v, w coincide for the rest of the process. In terms of theirs depths, this implies that $\kappa_{\lambda_{v,w},v} = 1 - \kappa_{\lambda_{v,w},w}$, i.e. exactly one of v or w increases its depth at step $\lambda_{v,w}$, and also $\kappa_{i,v} = \kappa_{i,w}$, for all $i < \lambda_{v,w}$.

We proceed to outline the contents of the remainder of the paper. In the next section, Section 7.3, we make rigorous the heuristics in (7.6) and (7.7). To do so, we express the cumulative distribution function of $\mathrm{h}_n(v)$ as the expected value of a function of $|\mathcal{S}_n|$.

In Section 7.4, we address the correlations between $(\mathrm{h}_n(v),\, v \in [k])$. We work with the coalescent process stopped at the moment where there are $\ln^2 n$ remaining trees, $F_{\ln^2 n}$. Using $F_{\ln^2 n}$ we define, for each $v \in [n]$, the *truncated selection sets*

$$\mathcal{S}_{n,1}(v) = \mathcal{S}_n(v) \setminus [\ln^2 n]$$

and a partial depth $\mathrm{h}_{n,1}(v) = \mathrm{h}_{F_{\ln^2 n}}(v)$. In Section 7.5 we show that if

$$h_n(v) = \mathrm{h}_{n,1}(v) + \mathrm{h}_{n,2}(v),$$

then $\mathrm{h}_{n,2}$ is negligible for the asymptotic distribution of $\mathrm{h}_n(v)$; this holds even if we condition on a finite set of vertices (that includes v) to have large degree. Stopping the process, instead, at $F_{\ln\ln n}$ would facilitate the analysis of $\mathrm{h}_{n,1}(v) - \mathrm{h}_{F_{\ln\ln n}}(v)$, but estimates on $(S_n(v)\setminus[\ln\ln n],\ v \in [k])$ would become much more delicate.

In Section 7.6 we study the joint limiting distribution of $(\mathrm{h}_{n,1}(v); v \in [k])$ and complete the proof of Theorem 7.2.5. Finally, Sections 7.7 and 7.8 contain the proofs of Proposition 7.1.8 and Theorem 7.1.1, respectively.

7.3 Proof of Theorem 7.2.5, case $k = 1$

In this section we fix $a \in [0, 1]$, $b \in \mathbb{Z}$ and write $m = m(a, b, n) = \lfloor a \log n \rfloor + b$. We establish the conditional limiting distribution of $(\mathrm{h}_n(1) - \mu_a \ln n)/\sqrt{\sigma_a^2 \ln n}$ given that $\mathrm{d}_n(1) \geq m$. Our approach consists on averaging over the size of the selection set $\mathcal{S}_n(1)$, and applying the following straightforward lemma for the renormalized version of $|\mathcal{S}_n(1)|$.

Lemma 7.3.1. *Let $f : \mathbb{R} \to \mathbb{R}$ be a uniformly continuous bounded function, $g : \mathbb{R} \to \mathbb{R}$ a continuous function and $(g_n : \mathbb{R} \to \mathbb{R},\ n \in \mathbb{N})$ be a sequence of functions uniformly converging to g over any compact set of $\mathbb{R}$. Let X_n be a sequence of random variables which converges in distribution to X, then*

$$\lim_{n\to\infty} \mathbf{E}\left[f(g_n(X_n))\right] = \mathbf{E}\left[f(g(X))\right].$$

We describe here a straightforward computation which arises in our proofs.

Lemma 7.3.2. *Let N be a standard Gaussian variable. Then, for every $x \in \mathbb{R}$ and $b > 0$,*

$$\mathbf{E}\left[\Phi\left(\frac{\sqrt{1+b^2}x - N}{b}\right)\right] = \Phi(x).$$

Proof. Let N' be a standard Gaussian variable, independent of N. The result follows since $N'b + N$ has a Gaussian distribution with variance $1 + b^2$ and so

$$\mathbf{E}\left[\Phi\left(\frac{\sqrt{1+b^2}x - N}{b}\right)\right] = \mathbf{P}\left(\frac{N'b + N}{\sqrt{1+b^2}} \leq x\right) = \Phi(x).$$

□

We first consider the case $m \leq 0$; in other words, the limiting distribution of $\mathrm{h}_n(1)$ without conditioning on its degree. For any fixed $x \in \mathbb{R}$, let $G_{n,x} : \mathbb{N} \to [0,1]$ be defined as

$$G_{n,x}(t) = \mathbf{P}\left(\mathrm{Bin}(t, 1/2) < x\sqrt{\ln n} + \ln n\right). \tag{7.8}$$

The motivation behind this definition is that, conditioning on $|\mathcal{S}_n(1)|$ and using (7.5), we have

$$\mathbf{P}\left(\mathrm{h}_n(1) < x\sqrt{\ln n} + \ln n\right) = \mathbf{E}\left[G_{n,x}(|\mathcal{S}_n(1)|)\right]. \tag{7.9}$$

The following result describes $G_{n,x}(|\mathcal{S}_n(1)|)$ as a function in terms of $\hat{S}_n = \frac{|\mathcal{S}_n(1)| - 2\ln n}{\sqrt{2\ln n}}$ and exploits the Gaussian limit of binomial variables $\mathrm{Bin}(m, p)$ as $m \to \infty$.

Lemma 7.3.3. *Let N be a standard Gaussian variable. For any $x \in \mathbb{R}$ fixed,*

$$\lim_{n\to\infty} \mathbf{E}\left[G_{n,x}(|\mathcal{S}_n(1)|)\right] = \mathbf{E}\left[\Phi(\sqrt{2}x - N)\right].$$

Proof. For each $n \in \mathbb{N}$, let $g_{n,x} : \mathbb{R} \to \mathbb{R}$ be defined as

$$g_{n,x}(r) = (\sqrt{2}x - r)\left(1 + \frac{r}{\sqrt{2\ln n}}\right)^{-1/2},$$

for $r > -\sqrt{2\ln n}$, and zero otherwise. Note that $g_{n,x}$ converges to $g_x(r) = \sqrt{2}x - r$, uniformly over bounded intervals as $n \to \infty$; this is easily proven and we omit the details. Next, we rewrite $\mathbf{E}\left[G_{n,x}(|\mathcal{S}_n(1)|)\right]$ as function of $\hat{S}_n$; this is to exploit the fact that $\hat{S}_n$ converges in distribution to a standard

Gaussian variable by (7.3). We show that

$$\lim_{n\to\infty} \mathbf{E}\left[G_{n,x}(|\mathcal{S}_n(1)|)\right] = \lim_{n\to\infty} \mathbf{E}\left[\Phi(g_{n,x}(\hat{S}_n))\right] = \mathbf{E}\left[\Phi\left(\sqrt{2}x - N\right)\right], \quad (7.10)$$

where N is a standard Gaussian variable. The last equality follows by Lemma 7.3.1 as the necessary conditions are satisfied: Φ is uniformly continuous and bounded, $g_{n,x}$ converges uniformly over bounded intervals, and $\hat{S}_n$ converges in distribution.

It remains to prove the first equality in (7.10). Note that

$$G_{n,x}(t) = \mathbf{P}\left(\frac{2\text{Bin}(t,1/2) - t}{\sqrt{t}} \le \frac{2x\sqrt{\ln n} + 2\ln n - t}{\sqrt{t}}\right);$$

additionally, letting $t = 2\ln n + r\sqrt{2\ln n}$ we have both $r > -\sqrt{2\ln n}$ and

$$\frac{2x\sqrt{\ln n} + 2\ln n - t}{\sqrt{t}} = \frac{\left(\sqrt{2}x - r\right)\sqrt{2\ln n}}{\sqrt{2\ln n - r\sqrt{2\ln n}}} = \left(\sqrt{2}x - r\right)\left(\frac{2\ln n - r\sqrt{2\ln n}}{2\ln n}\right)^{-1/2}.$$

For $t \ge 1$, let

$$\mathcal{E}(t) = G_{n,x}(t) - \Phi\left(g_{n,x}\left(\frac{t - 2\ln n}{\sqrt{2\ln n}}\right)\right).$$

By the Berry-Essen theorem for Gaussian approximation, see e.g. [35, Theorem 3.4.9], we have that $|\mathcal{E}(t)| \le Ct^{-1/2}$ for all $t \ge 1$. Therefore, using the tail bound in (7.4) for $|\mathcal{S}_n(1)|$, we have as $n \to \infty$,

$$\mathbf{E}\left[\mathcal{E}(|\mathcal{S}_n(1)|)\right] \le \mathbf{E}\left[|\mathcal{E}(|\mathcal{S}_n(1)|)|\right] \le \mathbf{P}\left(|\mathcal{S}_n(1)| \le \ln n\right) + C(\ln n)^{-1/2} \to 0.$$

This completes the proof as (7.10) follows from

$$\lim_{n\to\infty} \mathbf{E}\left[G_{n,x}(|\mathcal{S}_n(1)|)\right] = \lim_{n\to\infty} \mathbf{E}\left[\Phi(g_{n,x}(\hat{S}_n))\right] + \lim_{n\to\infty} \mathbf{E}\left[\mathcal{E}(|\mathcal{S}_n(1)|)\right],$$

where both limits in the right-hand side exist and the last one vanishes. $\square$

Despite Lemma 7.3.4 below being an stronger statement than Lemma 7.3.3, we decided to present the detailed proof of Lemma 7.3.3 as the computations are easier to follow. In particular, Lemmas 7.3.2 and 7.3.3 together imply that for any $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbf{P}\left(\mathrm{h}_n(1) < x\sqrt{\ln n} + \ln n\right) = \Phi(x), \tag{7.11}$$

which formalizes the heuristic in (7.6) and already yields a particular case of Theorem 7.2.5.

We now proceed to deal with the case $k = 1$ and a non-trivial conditioning in Theorem 7.2.5. For any $d, l \in \mathbb{N}$ let $\widetilde{G}_{d,l} : \mathbb{N} \to [0, 1]$ be defined as

$$\widetilde{G}_{d,l}(t) = \mathbf{P}\left(\mathrm{Bin}(t - d, 1/2) < l\right) \mathbf{1}_{[t\geq d]}. \tag{7.12}$$

By Lemma 7.2.7 we get

$$\begin{aligned}\mathbf{P}\left(\mathrm{h}_n(1) \leq l,\, \mathrm{d}_n(1) \geq d\right) &= 2^{-d} \sum_{t\geq d} \mathbf{P}\left(\mathrm{h}_n(1) \leq l \,|\, |\mathcal{S}_n(1)| = t\right) \mathbf{P}\left(|\mathcal{S}_n(1)| = t\right) \\ &= 2^{-d}\mathbf{E}\left[\widetilde{G}_{d,l}(|\mathcal{S}_n(1)|)\right].\end{aligned} \tag{7.13}$$

Recall the next definitions given for Theorem 7.1.3; for $a \in [0, 1]$, let $\mu_a = 1 - (a \log e)/2$ and $\sigma_a^2 = 1 - (a \log e)/4$.

Lemma 7.3.4. *Fix $a \in [0, 1]$, $b \in \mathbb{Z}$ and let $x \in \mathbb{R}$. Write $m = m(a, b, n) = \lfloor a \log n\rfloor + b$ and $l = l(a, x, n) = x\sqrt{\sigma_a^2 \ln n} + \mu_a \ln n$. If $m \geq 0$ then*

$$\mathbf{E}\left[\widetilde{G}_{m,l}(|\mathcal{S}_n(1)|)\right] = \mathbf{E}\left[\Phi\left(\frac{\sqrt{2\sigma_a^2}x - N}{\sqrt{\mu_a}}\right)\right].$$

Proof. The proof uses Lemma 7.3.1 and follows the same approach as in Lemma 7.3.3. We also use the renormalization $\hat{S}_n$. Fix a, b, x and set m, l as given in the statement. For the rest of the proof, write $\mu = \mu_a$ and $\sigma = \sigma_a$. We show that

$$\lim_{n\to\infty} \mathbf{E}\left[\widetilde{G}_{m,l}(|\mathcal{S}_n(1)|)\right] = \lim_{n\to\infty} \mathbf{E}\left[\Phi(\tilde{g}_{n,a,x}(\hat{S}_n))\right], \tag{7.14}$$

where $\tilde{g}_{n,a,x} : \mathbb{R} \to \mathbb{R}$ are functions, defined below, such that $\tilde{g}_{n,a,x}(r)$ converges to $\tilde{g}_{a,x}(r) = \frac{\sqrt{2}\sigma x - t}{\sqrt{\mu}}$, uniformly over bounded sets, as $n \to \infty$. Once (7.14) is established, the result follows by Lemma 7.3.1. To do so, we are required to bound the error of approximating $\widetilde{G}_{m,l}(|\mathcal{S}_n(1)|)$ with $\Phi(\tilde{g}_{n,a,x}(\hat{S}_n))$.

Now, write $\varepsilon = \varepsilon(a,n) = a\log n - \lfloor a \log n \rfloor$; then $m = \lfloor a \log n \rfloor + b = 2(1-\mu)\ln n + b - \varepsilon$. A direct calculation shows that

$$\tilde{g}_{n,a,x}(r) = \frac{\sqrt{2}\sigma x - r}{\sqrt{\mu}} \left(1 + \frac{r\sqrt{2\ln n} - b + \varepsilon}{2\mu \ln n}\right)^{-1/2} + \left(\frac{2\mu\ln n + r\sqrt{2\ln n} - b + \varepsilon}{(b-\varepsilon)^2}\right)^{-1/2}$$

if $r \geq \frac{-2\mu \ln n + b - \varepsilon}{\sqrt{2\ln n}}$, and zero otherwise. The uniform convergence of $\tilde{g}_{n,a,x}$ is straightforward, but we omit the details. For $t \geq 1$, let

$$\mathcal{E}(t-m) = \widetilde{G}_{m,l}(t) - \Phi\left(\tilde{g}_{n,a,x}\left(\frac{t - 2\ln n}{\sqrt{2\ln n}}\right)\right).$$

By the Berry-Essen theorem, see e.g. [35, Theorem 3.4.9], we have that $|\mathcal{E}(t)| \leq Ct^{-1/2}$. Finally, for n large enough, $m < (3/2)\ln n$ and so, having $\mathcal{S}_n(1) > (7/4)\ln n$ implies $|\mathcal{S}_n(1)| - m > (1/4)\ln n$. By (7.4) we get,

$$\mathbf{E}\left[\mathcal{E}(|\mathcal{S}_n(1)| - m)\right] \leq \mathbf{E}\left[|\mathcal{E}(|\mathcal{S}_n(1)| - m)|\right] \leq \mathbf{P}\left(|\mathcal{S}_n(1)| \leq (2 - 1/4)\ln n\right) + 2C(\ln n)^{-1/2} = o(1).$$

This completes the proof as

$$\lim_{n\to\infty} \mathbf{E}\left[\widetilde{G}_{m,l}(|\mathcal{S}_n(1)|)\right] = \lim_{n\to\infty} \mathbf{E}\left[\Phi(\tilde{g}_{n,a,x}(\hat{S}_n))\right] = \lim_{n\to\infty} \mathbf{E}\left[\mathcal{E}(|\mathcal{S}_n(1)| - m)\right],$$

and the last limit vanishes. □

Proof of Theorem 7.2.5, case $k = 1$. Fix $a \in [0,1]$, $b \in \mathbb{Z}$ and $x \in \mathbb{R}$. Let $m = m(a,b,n) = \lfloor a \log n \rfloor + b$ and $l = l(a,x,n) = x\sqrt{\sigma_a^2 \ln n} + \mu_a \ln n$. Our goal is to show that

$$\lim_{n\to\infty} \mathbf{P}\left(\mathrm{h}_n(1) < l \,|\, \mathrm{d}_n(1) \geq m\right) = \Phi(x).$$

If $m \leq 0$ then $a = 0$. The result then follows by (7.11) since $\mu_a = \sigma_a = 1$, and so

$$\mathbf{P}\left(\mathrm{h}_n(1) < l \,|\, \mathrm{d}_n(1) \geq m\right) = \mathbf{P}\left(\mathrm{h}_n(1) < x\sqrt{\ln n} + \ln n\right).$$

Consider now the case $m > 0$. Note that $m = \lfloor a \log n \rfloor + b \leq \frac{3}{2} \ln n$ for n large enough. Therefore, by Lemma 7.2.7 and (7.4), we have

$$\lim_{n\to\infty} 2^m \mathbf{P}\left(\mathrm{d}_n(1) \geq m\right) = \lim_{n\to\infty} \mathbf{P}\left(|\mathcal{S}_n(1)| \geq m\right) = 1. \tag{7.15}$$

Using the equations (7.13), (7.15), and Lemma 7.3.4 we get that for any $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbf{P}\left(\mathrm{h}_n(1) < l \,|\, \mathrm{d}_n(1) \geq m\right) = \lim_{n\to\infty} 2^{-m} \frac{\mathbf{E}\left[\widetilde{G}_{m,l}(|\mathcal{S}_n(1)|)\right]}{\mathbf{P}\left(|\mathcal{S}_n(1)| \geq m\right)} = \mathbf{E}\left[\Phi\left(\frac{\sqrt{2}\sigma_a x - N}{\sqrt{\mu_a}}\right)\right].$$

We use the fact that $2\sigma_a^2 = 1 + \mu_a$ to apply Lemma 7.3.2 to the last term above. This yields the desired result. $\square$

In the following section we lay down the necessary approximations to obtain a generalization of (7.13) to several vertices.

7.4 Truncated selection sets

In this section we fix $k \geq 2$ and consider the depths of vertices in $F_{\ln^2 n}$. Recall from Section 7.2.2 that the truncated selection sets are defined by $\mathcal{S}_{n,1}(v) = \mathcal{S}_n(v) \setminus [\ln^2 n]$ for $v \in [n]$. Let $\Omega_1 = \mathcal{P}(\{\log^2 n + 1, \ldots, n\})$. For the remainder of the paper we use, e.g. $\bar{\mathcal{S}}_{n,1} \in \Omega_1^k$ to denote the vector $(\mathcal{S}_{n,1}(i),\, i \in [k])$.

Our main objective is showing that $(\mathcal{S}_{n,1}(i),\, i \in [k])$ behave, asymptotically, as if they were independent sets; see Proposition 7.4.6. Also, the conditional law of the depths $(\mathrm{h}_{n,1}(i),\, i \in [k])$ given the truncated selection sets $(\mathcal{S}_{n,1}(i),\, i \in [k])$ can be approximated by the law of k independent Binomial variables. This holds even if we condition on the final degrees $(\mathrm{d}_n(i),\, i \in [k])$;

see Proposition 7.4.2 and Remark 7.4.3. These properties are crucial to establishing Theorem 7.2.5 in full generality.

The choice of halting the process at $F_{\ln^2 n}$, and not e.g. $F_{\ln\ln n}$, implies that we must also provide the limiting distribution of $(\mathrm{h}_n(i) - \mathrm{h}_{n,1}(i))/\sqrt{\ln n}$ (a priori $\mathrm{h}_n(i) - \mathrm{h}_{n,1}(i) \leq \ln^2 n$). Nevertheless, we use $F_{\ln^2 n}$ since it allows to use simple arguments in the estimates of Proposition 7.4.6 below.

Note that, $\mathbf{E}\left[|\mathcal{S}_{n,1}(v)|\right] = 2\ln n - 2\ln\ln n + o(1) = 2\ln n(1+o(1))$ and thus, similar to (7.4), we get concentration of $|\mathcal{S}_{n,1}(v)|$ around $2\ln n$ and a normal asymptotic limit.

Fact 7.4.1. *For any $v \in [n]$ and $\varepsilon > 0$, $\mathbf{P}\left(||\mathcal{S}_{n,1}(v)| - 2\ln n| > \varepsilon \ln n\right) = o(1)$ and*

$$\frac{|\mathcal{S}_{n,1}(v) - 2\ln n|}{\sqrt{2\ln n}} \xrightarrow{\mathcal{L}} N;$$

where N is a standard Gaussian variable.

The following proposition is used to obtain independent limiting distributions for the depths of k vertices in the final tree $T^{(n)}$.

Proposition 7.4.2. *Fix $\bar{m}, \bar{l} \in \mathbb{N}^k$. For all $\bar{J} \in \Omega_1^k$ such that $\{J_i,\, i \in [k]\}$ are pairwise disjoint, we have*

$$\mathbf{P}\left(\mathrm{h}_{n,1}(i) \leq l_i,\, i \in [k] \,|\, \bar{\mathcal{S}}_{n,1} = \bar{J}\right) = \prod_{i=1}^{k} \mathbf{P}\left(\mathrm{Bin}(|J_i|, 1/2) \leq l_i\right),$$

$$\mathbf{P}\left(\mathrm{d}_{F_{\ln^2 n}}(i) \geq m_i,\, \mathrm{h}_{n,1}(i) \leq l_i,\, i \in [k] \,|\, \bar{\mathcal{S}}_{n,1} = \bar{J}\right) = 2^{-\sum m_i} \prod_{i=1}^{k} \mathbf{P}\left(\mathrm{Bin}(|J_i| - m_i, 1/2) \leq l_i\right) \mathbf{1}_{[|J_i| \geq m_i]}.$$

Proof. Fix $\bar{m}, \bar{l} \in \mathbb{N}^k$ and $\bar{J} \in \Omega_1^k$ as given in the statement. Once the sets $(\mathcal{S}_{n,1}(i),\, i \in [k])$ are fixed, the depth $\mathrm{h}_{n,1}(i)$ of $i \in [k]$ in $F_{\ln^2 n}$ is determined by the variables $(\xi_j,\, j \in \mathcal{S}_{n,1}(i))$. Consequently, given that $\bar{\mathcal{S}}_{n,1} = \bar{J}$ the conditional law of the degrees and depths of vertices in $F_{\ln^2 n}$ depend on disjoint sets of independent variables. Therefore, we can decouple the event $\{\mathrm{d}_{F_{\ln^2 n}}(i) \geq m_i,\, \mathrm{h}_{n,1}(i) \leq l_i\}$.

The first equality in the statement corresponds to the case when $m_i = 0$ for all $i \in [k]$. Now, for the second equality we first note that the product of indicator functions follows since $\mathrm{d}_{F_{\ln^2 n}}(i) \geq d_i$ for all $i \in [k]$ occurs only if $|\mathcal{S}_{n,1}(i)| \geq m_i$ for each $i \in [k]$. Then, for each $i \in [k]$ we flip $|\mathcal{S}_{n,1}(i)|$ independent fair coins. The first m_i coins must be heads and this occurs with probability 2^{-m_i}. The number of tails in the remaining coin flips determine the depth $\mathrm{h}_{n,1}(i)$; this is distributed as $\mathrm{Bin}(|\mathcal{S}_{n,1}(i)| - m_i, 1/2)$. $\square$

Remark 7.4.3. *Furthermore, if $\bar{J} \in \Omega_1^k$ is such that $|J_i| \geq m_i$ for all $i \in [k]$, then*

$$\{\mathrm{d}_n(i) \geq m_i,\, i \in [k],\, \bar{\mathcal{S}}_{n,1} = \bar{J}\} = \{\mathrm{d}_{F_{\ln^2 n}}(i) \geq m_i, i \in [k], \bar{\mathcal{S}}_{n,1} = \bar{J}\}.$$

Now, with high-probability, vertices in $[k]$ still belong to distinct trees in $F_{\ln^2 n}$ which implies that the truncated selection sets $(\mathcal{S}_{n,1}(i),\, i \in [k])$ are disjoint. To see this, let us define

$$\tau_k = \max\{2 \leq j \leq n : s_{j,v} = s_{j,w} = 1 \text{ for some distinct } v, w \in [k]\}.$$

Recall that the trees in F_j are ordered in increasing order of their least element. By definition of τ_k, $|\{a_j, b_j\} \cap [k]| \leq 1$ for $j > \tau_k$. Thus, at no point $j \geq \tau$ are $T_j(v)$ and $T_j(w)$ merged, for distinct $v, w \in [k]$. In other words, $T_j(i) = i$ for all $i \in [k]$, $j \geq \tau_k$. Therefore,

$$\{\tau \leq \ln^2 n\} = \{(\mathcal{S}_{n,1}(i),\, i \in [k]) \text{ are pairwise disjoint}\}. \tag{7.16}$$

Fact 7.4.4. *Fix an integer $k \geq 2$. For n large enough,*

$$\mathbf{P}\left(\tau_k > \ln^2 n\right) \leq 2k^2 \ln^{-2} n.$$

Proof. By definition, $T_j(i) = T_i^{(j)}$ for all $i \in [k]$ and $j \geq \tau_k$. Therefore,

$$\mathbf{P}(\tau_k \leq l) = \prod_{j=l+1}^{n} \mathbf{P}(|\{a_j, b_j\} \cap [k]| < 2) = \prod_{j=l+1}^{n} \left(1 - \frac{k(k-1)}{j(j-1)}\right) \geq \prod_{j=l}^{\infty} \left(1 - \frac{k^2}{j^2}\right).$$

The second equality is since the pairs $(\{a_j, b_j\},\ 2 \leq j \leq n)$ are chosen independently and uniformly at random. For the next approximation we use that $1 - x > e^{-2x}$ for $x > 0$ sufficiently small and that $e^{-x} > 1 - x$ for all $x \in \mathbb{R}$. Then, letting $l = \ln^2 n$ and n large enough, we have

$$\prod_{j=\ln^2 n}^{\infty} \left(1 - \frac{k^2}{j^2}\right) > 1 - \sum_{j=\ln^2 n}^{\infty} \frac{2k^2}{j^2} > 1 - 2k^2 \int_{\ln^2 n}^{\infty} x^{-2} dx = 1 - 2k^2 \ln^{-2} n.$$

□

Finally, we consider the following family of sets as representing the bulk of the probability measure induced by k truncated sets. We add the parameter $\delta > 0$ to cover the distinct possible values of $\bar{a} \in [0,1]^k$ in Theorem 7.2.5. For $\delta \in (0,2)$ let

$$\mathcal{B}_{n,k,\delta} = \{\bar{J} \in \Omega_1^k : (J_1, \ldots, J_k) \text{ are pairwise disjoint and } ||J_i| - 2\ln n| \leq \delta \ln n,\ i \in [k]\}. \tag{7.17}$$

Lemma 7.4.5. *Fix an integer $k \geq 2$ and $\delta \in (0,2)$. Then*

$$\mathbf{P}\left(\bar{\mathcal{S}}_{n,1} \in \mathcal{B}_{n,k,\delta}\right) = 1 + o(1).$$

Proof. This follows directly from (7.16) and Facts 7.4.1 and 7.4.4;

$$\mathbf{P}\left(\bar{\mathcal{S}}_{n,1} \notin \mathcal{B}_{n,k,\delta}\right) \leq \mathbf{P}\left(\tau_k \geq \ln^2 n\right) + k\mathbf{P}\left(||\mathcal{S}_{n,1}(i)| - 2\ln n| < \delta \ln n\right) = o(1).$$

□

Let $(\mathcal{R}_n(i),\ i \in [k])$ be k independent copies of $\mathcal{S}_{n,1}(1)$. We use sets $\mathcal{B}_{n,k,\delta}$ to make explicit the claim that $(\mathcal{S}_{n,1}(i),\ i \in [k])$ are asymptotically independent; this occurs uniformly on such $\mathcal{B}_{n,k,\delta}$.

Proposition 7.4.6. *Fix an integer $k \geq 2$ and $\delta \in (0,2)$. Uniformly for $\bar{J} \in \mathcal{B}_{n,k,\delta}$,*

$$\mathbf{P}\left(\bar{\mathcal{S}}_{n,1} = \bar{J}\right) = (1+o(1))\mathbf{P}\left(\bar{\mathcal{R}}_n = \bar{J}\right).$$

The remainder of the section is devoted to proving Proposition 7.4.6, and to do so we fix $\delta \in (0,2)$ and $\bar{J} \in \mathcal{B}_{n,k,\delta}$. The notation we define below does not reflect the dependency on $\bar{J}$. We use the index m with $\ln^2 n < m \leq n$ unless otherwise specified. Recall that $\mathcal{S}_{n,1}(i) = \{m : s_{m,i} = 1\}$. Similarly, for all m and $i \in [k]$, let $r_{m,i}$ be the random indicator of $m \in \mathcal{R}_n(i)$ and let $j_{m,i} = \mathbf{1}_{[m \in J_i]}$. Also, let $\sigma_m = \sum_{i \in [k]} j_{m,i}$ and note that from the choice of $\bar{J}$ we have that $\sigma_m \leq 1$ for all m.

Claim 7.4.7. *For each m, let $A_m = \{s_{m,i} = j_{m,i},\, i \in [k]\}$. Then*

$$\mathbf{P}\left(A_m \,|\, A_l,\, m < l \leq n\right) = \begin{cases} \frac{(m-k)(m-k-1)}{m(m-1)} & \text{if } \sigma_m = 0, \\ \frac{2(m-k)}{m(m-1)} & \text{if } \sigma_m = 1. \end{cases} \tag{7.18}$$

and furthermore,

$$\mathbf{P}\left(\mathcal{S}_{n,1}(i) = J_i,\, i \in [k]\right) = \prod_m \mathbf{P}\left(A_m \,|\, A_l,\, m < l \leq n\right).$$

Proof. The second equality follows since $\{\mathcal{S}_{n,1}(i) = J_i;\, i \in [k]\} = \{\cap_m A_m\}$. We proceed to prove (7.18) by induction on $n-m$. For $m = n$, the formula is trivial. For $m < n$ note that the condition $\{A_l,\, m < l \leq n\}$ implies that $\sigma_l \leq 1$ for all $m < l \leq n$. That is, there has been no merges of distinct trees $T_l(v), T_l(w)$ for $v, w \in [k]$. In particular, $T_m(i) = T_i^{(m)}$ for all $i \in [k]$. If $\sigma_m = 0$, then none of these trees are selected to be merged in the next step, and this occurs with probability $\frac{(m-k)(m-k-1)}{m(m-1)}$. If $\sigma_m = 1$, then there is exactly one vertex $i \in [k]$ which is selected and the other tree is selected among $(m-k)$ trees. □

Similarly, we have the following estimates for $(\mathcal{R}_n(i),\, i \in [k])$.

Claim 7.4.8. *For each m, let $A'_m = \{r_{m,i} = j_{m,i},\ i \in [k]\}$. Then*

$$\mathbf{P}\left(A'_m\right) = \begin{cases} \left(1 - \frac{2}{m}\right)^k & \text{if } \sigma_m = 0, \\ \frac{2}{m}\left(1 - \frac{2}{m}\right)^{k-1} & \text{if } \sigma_m = 1. \end{cases} \tag{7.19}$$

and furthermore,

$$\mathbf{P}\left(\mathcal{R}_n(i) = J_i,\ i \in [k]\right) = \prod_m \mathbf{P}\left(A_m\right).$$

Proof. It is clear that $\{\mathcal{R}_n(i) = J_i;\ i \in [k]\} = \{\cap_m A'_m\}$. Observe that the events A'_m are independent. Also, (7.19) follows immediately from the distribution of $(r_{m,i}, i \in [k])$ and the fact that these variables are independent. □

Proposition 7.4.6 is obtained by comparing the two products in the claims above. The following claim relates each of the terms in (7.18) and (7.19). Let

$$p_{m,0} = \frac{(m-k)(m-k-1)}{m(m-1)}, \qquad q_{m,0} = \left(1 - \frac{2}{m}\right)^k,$$
$$p_{m,1} = \frac{2(m-k)}{m(m-1)}, \qquad q_{m,1} = \frac{2}{m}\left(1 - \frac{2}{m}\right)^{k-1}.$$

Claim 7.4.9. *There exists a constant $c = c(k) > 0$ such that for m large enough, we have*

$$q_{m,0} \ > \ p_{m,0} \ > \ q_{m,0}\left(1 - \frac{c}{m^2}\right),$$
$$q_{m,1} \ < \ p_{m,1} \ < \ q_{m,1}\left(1 + \frac{c}{m}\right).$$

Proof. First we prove the bounds on $p_{m,0}$. Note that $p_{m,0} = 1 - \frac{2k}{m} + \frac{k(k-1)}{m(m-1)}$ and so

$$0 < q_{m,0} - p_{m,0} = -\frac{k(k-1)}{m(m-1)} + \frac{2k(k-1)}{m^2} + O(m^{-3}) = O(m^{-2}). \tag{7.20}$$

The upper bound on $p_{m,0}$ follows from the first inequality in (7.20). For the lower bound, use that $q_{m,0} \to 1$ as $m \to \infty$ then

$$\frac{q_{m,0} - p_{m,0}}{q_{m,0}} = \frac{q_{m,0} - p_{m,0}}{1 + o(1)} = O(m^{-2}).$$

The bounds on $p_{m,1} = \frac{2}{m} - \frac{2(k-1)}{m(m-1)}$ are obtained similarly. We use that $mq_{m,1} \to 2$ as $m \to \infty$ and

$$0 < p_{m,1} - q_{m,1} = -\frac{2(k-1)}{m(m-1)} + \frac{4(k-1)}{m^2} + O(m^{-3}) = O(m^{-2}).$$

□

Proof of Proposition 7.4.6. Fix $\delta \in (0, 2)$ and $k \geq 2$. The bounds we give below do not depend on the choice of $\bar{J} \in \mathcal{B}_{n,k,\delta}$ and so the bounds obtained are uniform in $\mathcal{B}_{n,k,\delta}$. By Claims 7.4.7 and 7.4.8, it suffices to prove that

$$\prod_m \mathbf{P}\left(A_m \,|\, A_l,\, m < l \leq n\right) = (1 + o(1)) \prod_m \mathbf{P}\left(A'_m\right).$$

The lower bounds in Claim 7.4.9 give, for m large enough,

$$\begin{aligned}
\mathbf{P}\left(A_m\,,\, \ln^2 n < m \leq n\right) &= \prod_{m=\ln^2 n}^{n} \mathbf{P}\left(A_m \,|\, A_l,\, m < l \leq n\right) \\
&\geq \prod_{m=\ln^2 n}^{n} \mathbf{P}\left(A'_m\right)\left(1 - \frac{c}{m^2}\right) \\
&= \mathbf{P}\left(A'_m\,,\, \ln^2 n < m \leq n\right) \prod_{m=\ln^2 n}^{n} \left(1 - \frac{c}{m^2}\right) \\
&\geq \mathbf{P}\left(A'_m\,,\, \ln^2 n < m \leq n\right)\left(1 - \frac{2c}{\ln^2 n}\right).
\end{aligned}$$

The last equality follows in the same manner as the bound for $\mathbf{P}\left(\tau_k \leq \ln^2 n\right)$ obtained in Fact 7.4.5. Now, using the upper bounds in Claim 7.4.9 we have, for m large enough,

$$\begin{aligned}
\mathbf{P}\left(A_m\,,\, \ln^2 n < m \le n\right) &= \prod_{m=\ln^2 n}^{n} \mathbf{P}\left(A_m \mid A_l,\, m < l \le n\right) \\
&\le \prod_{m=\ln^2 n}^{n} \mathbf{P}\left(A'_m\right)\left(1 + \frac{c}{m}\mathbf{1}_{[\sigma_m = 1]}\right) \\
&= \mathbf{P}\left(A'_m\,,\, \ln^2 n < m \le n\right) \prod_{m:\sigma_m=1} \left(1 + \frac{c}{m}\right) \\
&\le \mathbf{P}\left(A'_m\,,\, \ln^2 n < m \le n\right) \left(1 + \frac{2(2+\delta)ck}{\ln n}\right).
\end{aligned}$$

In the last inequality we use that $\sum_m \sigma_m \le (2+\delta)k \ln n$ by the second condition on $\mathcal{B}_{n,k,\delta}$. Thus,

$$\prod_{\sigma_m=1} \left(1 + \frac{c}{m}\right) \le \exp\left(\frac{(2+\delta)ck\ln n}{\ln^2 n}\right) < 1 + \frac{(2+\delta)2ck}{\ln n}.$$

In the first inequality we use that $m \ge \ln^2 n$ and $1 + x \le e^x$ for all $x \in \mathbb{R}$; for the second inequality, we use that $e^x < 1 + 2x$ for x sufficiently small. □

7.5 Negligible depth increase

In this section we fix $k \ge 2$ and prove that the main contribution to $(\mathrm{h}_n(i),\, i \in [k])$ is already found in $F_{\ln^2 n}$. Recall that $\mathrm{h}_{n,1}(i) = \mathrm{h}_{F_{\ln^2 n}}(i)$ and $h_{n,2}(i) = h_n(i) - h_{n,1}(i)$, for $i \in [n]$. The key observation in this section is that the coalescence after $F_{\ln^2 n}$ can be compared with an independent $\ln^2 n$-coalescent.

Fact 7.5.1. *For any $v \in [n]$, $h_{n,2}(v)$ is stochastically dominated by $|\mathcal{S}_{\ln^2 n}(v)|$.*

Proof. In an n-coalescent $\mathbf{C} = (F_n, \ldots, F_1)$ we have that $h_{F_i}(v) = \sum_{j=i+1}^{n} h_{j,v}$ with $h_{j,v} \le s_{j,v}$. Thus,

$$\mathrm{h}_{n,2} = \mathrm{h}_n(v) - \mathrm{h}_{n,1}(v) = \sum_{j=2}^{\ln^2 n} h_{j,v} \le \sum_{j=2}^{\ln^2 n} s_{j,v}.$$

The result then follows since for any $m \le n$, we have that $|\mathcal{S}_m(v)| \stackrel{\mathcal{L}}{=} \sum_{j=2}^m s_{j,v}$.

$\square$

Lemma 7.5.2. *For any vertex $i \in [n]$, we have $\frac{\mathrm{h}_{n,2}(i)}{\sqrt{\ln n}} \to 0$, in probability as $n \to \infty$.*

Proof. By Fact 7.5.1, it suffices to prove that for every $\varepsilon > 0$,

$$\mathbf{P}\left(|\mathcal{S}_{\ln^2 n}(i)| > \varepsilon\sqrt{\ln n}\right) = o(1). \tag{7.21}$$

Write $m = \ln^2 n$ and note that $\mathbf{E}\left[|\mathcal{S}_m(i)|\right] = 2\ln\ln n + O(1)$ and so $\delta = \frac{\varepsilon\sqrt{\ln n} - \mathbf{E}[|\mathcal{S}_m(i)|]}{\mathbf{E}[|\mathcal{S}_m(i)|]} > 0$ for n large enough. Therefore, Lemma 7.4.1 yields

$$\mathbf{P}\left(|\mathcal{S}_{\ln^2 n}(i)| > \varepsilon\sqrt{\ln n}\right) \le \mathbf{P}\left(\left||\mathcal{S}_m(i)| - \mathbf{E}\left[|\mathcal{S}_m(i)|\right]\right| > \delta\mathbf{E}\left[|\mathcal{S}_m(i)|\right]\right) = o(1).$$

$\square$

In fact, for $i \in [k]$, $h_{n,2}(i)$ is also negligible when we condition on the vertices in $[k]$ to have large degree. Let $\Omega = \mathcal{P}([n])$ be the power set of $[n]$ and fix $\bar{m} \in \mathbb{N}^k$. Let

$$\mathcal{A}_{\bar{m}} = \{\bar{J} \in \Omega^k : \mathbf{P}\left(\bar{\mathcal{S}}_n = \bar{J},\, \mathrm{d}_n(i) \ge m_i,\, i \in [k]\right) > 0\},$$
$$\mathcal{L}_{\bar{m}} = \{\bar{J} \in \Omega^k : |J_i \setminus [\ln^2 n]| \ge m_i,\, i \in [k]\}.$$

Lemma 7.5.3. *Fix $\bar{m} \in \mathbb{N}^k$. For any $s \in \mathbb{N}$ and $i \in [k]$, if $\bar{J} \in \mathcal{A}_{\bar{m}}$ we have*

$$\mathbf{P}\left(\mathrm{h}_{n,2}(i) \ge s,\, \mathrm{d}_n(j) \ge m_j,\, j \in [k] \,|\, \bar{\mathcal{S}}_n = \bar{J}\right) \le 2^{-\sum_j m_j};$$

if $\bar{J} \in \mathcal{A}_{\bar{m}} \cap \mathcal{L}_{\bar{m}}$, then

$$\mathbf{P}\left(\mathrm{h}_{n,2}(i) \ge s,\, \mathrm{d}_n(j) \ge m_j,\, j \in [k] \,|\, \bar{\mathcal{S}}_n = \bar{J}\right) \le 2^{-\sum_j m_j}\mathbf{P}\left(\mathcal{S}_{\ln^2 n}(i) \ge s\right).$$

Proof. Recall that the degree of a vertex $i \in [k]$ is determined by the first streak of selection times $j \in \mathcal{S}_n(i)$ where $h_{j,i} = 0$. If $\bar{J} \in \mathcal{A}_{\bar{m}}$, then the event

$\bar{\mathcal{S}}_n = \bar{J}$ has the property that the set of the first m_i selection times in $\mathcal{S}_n(i)$ are pairwise disjoint for all $i \in [k]$; otherwise $\mathbf{P}\left(\bar{\mathcal{S}}_n = \bar{J},\, \mathrm{d}_n(i) \geq m_i,\, i \in [k]\right) = 0$. It then follows that

$$\mathbf{P}\left(\mathrm{d}_n(j) \geq m_j,\, j \in [k] \,|\, \bar{\mathcal{S}}_n = \bar{J}\right) = 2^{-\sum_j m_j},$$

which yields the first inequality. For the second inequality it remains to prove that for $\bar{J} \in \mathcal{A}_{\bar{m}} \cap \mathcal{L}_{\bar{m}}$,

$$\mathbf{P}\left(\mathrm{h}_{n,2}(i) \geq s \,\middle|\, \mathrm{d}_n(j) \geq m_j,\, j \in [k],\, \bar{\mathcal{S}}_n = \bar{J}\right) \leq \mathbf{P}\left(\mathcal{S}_{\ln^2 n}(i) \geq s\right).$$

In this case, the event $\{\mathrm{d}_n(j) \geq m_j,\, j \in [k]\}$ is already determined by the forest $F_{\ln^2 n}$. Consequently, the remaining selection times $\mathcal{S}_n(i) \cap [\ln^2 n]$, which determine $\mathrm{h}_{n,2}(i)$, are independent of the conditioning event and so the argument in Fact 7.5.1 can be applied. $\square$

The next lemma uses a result from [3], whose proof can be derived from this work but we omit its proof for brevity.

Proposition 7.5.4 (Proposition 4.2 in [3])**.** *Fix $c \in (0, 2)$ and $k \in \mathbb{N}$. There exists $\beta = \beta(c, k) > 0$ such that uniformly over positive integers $m_1, \ldots, m_k < c \ln n$,*

$$\mathbf{P}\left(\mathrm{d}_n(j) \geq m_j,\, j \in [k]\right) = 2^{-\sum_j m_j}(1 + o(n^{-\beta})).$$

Lemma 7.5.5. *Fix $c \in (0, 2)$. If $m_j = m_j(n) < c \ln n$ for all $j \in [k]$, then for any $i \in [k]$,*

$$\mathbf{P}\left(\mathrm{h}_{n,2}(i) \geq \varepsilon\sqrt{\ln n},\, \mathrm{d}_n(j) \geq m_j,\, j \in [k]\right) \to 0.$$

Proof. Let $\bar{m}$ satisfy the conditions of the statement. By Lemma 7.5.3, we have for any $i \in [k]$ and $s \in \mathbb{N}$,

$$\begin{aligned}\mathbf{P}\left(\mathrm{h}_{n,2}(i) \geq s,\, \mathrm{d}_n(j) \geq m_j,\, j \in [k]\right) &= \sum_{\bar{J} \in \mathcal{A}_{\bar{m}}} \mathbf{P}\left(\bar{\mathcal{S}}_n = \bar{J},\, \mathrm{h}_{n,2}(i) \geq s,\, \mathrm{d}_n(j) \geq m_j,\, j \in [k]\right) \\ &\leq 2^{-d}[\mathbf{P}\left(|\mathcal{S}_{\ln^2 n}(i)| \geq s\right) + \mathbf{P}\left(\bar{\mathcal{S}}_n \notin \mathcal{L}_{\bar{m}}\right)].\end{aligned}$$

If $s = \varepsilon\sqrt{\ln n}$, the first probability in the last line vanishes as $n \to \infty$ by (7.21). Also, by (7.4) we get

$$\mathbf{P}\left(\bar{\mathcal{S}}_n \notin \mathcal{L}_{\bar{m}}\right) \leq k\mathbf{P}\left(|\mathcal{S}_n(1)| < c\ln n\right) = o(1).$$

Therefore

$$\mathbf{P}\left(\mathrm{h}_{n,2}(i) \geq \varepsilon\sqrt{\ln n} \,\middle|\, \mathrm{d}_n(j) \geq m_j,\, j \in [k]\right) = \frac{2^{-\sum_j m_j} o(1)}{\mathbf{P}\left(\mathrm{d}_n(j) \geq m_j,\, j \in [k]\right)};$$

and the proof is completed since $\mathbf{P}\left(\mathrm{d}_n(j) \geq m_j,\, j \in [k]\right) = 2^{-\sum_j m_j}(1 + o(1))$ by Proposition 7.5.4. $\square$

7.6 Proof of Theorem 7.2.5, case $k \geq 2$

Fix an integer $k \geq 2$. We would like to express $\mathbf{P}\left(\mathrm{h}_n(i) < x_i\sqrt{\ln n} + \ln n,\, i \in [k]\right)$ as a product of expectations of the form in (7.9) since this would yield the independence of the limiting variables. However we have seen previously that this is not possible largely due to the correlations between the selection sets of vertices $i \in [k]$. Instead we consider the depths $(h_{n,1}(i),\, i \in [k])$ and exploit the fact that $(\mathcal{S}_{n,1}(v),\, v \in [k])$ are asymptotically independent. Given a measure μ, we write $\mathbf{E}_\mu$ for expectations with respect to μ.

Lemma 7.6.1. *For each $n \in \mathbb{N}$, let μ_n and ν_n be probability measures in a space Ω_n. Let $B_n \subset \Omega_n$ be such that, uniformly for each $\omega \in B_n$, $\mu_n(\omega) = (1 - o(1))\nu_n(\omega)$; and $\mu_n(B_n) = 1 + o(1) = \nu_n(B_n)$.*

If $f_n, g_n \in \Omega_n \to \mathbb{R}$ are bounded and $f_n(\omega) = g_n(\omega)$ for all $\omega \in B_n$; then

$$\mathbf{E}_{\mu_n}[f_n] = (1+o(1))\mathbf{E}_{\nu_n}[g_n] + o(1).$$

Proof. In the subspace B_n we can interchange f_n and g_n; and since the approximation $\mu_n(\omega) = (1+o(1))\nu_n(\omega)$ is uniform over $\omega \in B_n$ we have that

$$\mathbf{E}_{\mu_n}\left[f_n\mathbf{1}_{[B_n]}\right] = (1+o(1))\mathbf{E}_{\nu_n}\left[f_n\mathbf{1}_{[B_n]}\right] = (1+o(1))\mathbf{E}_{\nu_n}\left[g_n\mathbf{1}_{[B_n]}\right].$$

The result follows by noting that

$$\mathbf{E}_{\mu_n}\left[f_n\mathbf{1}_{[\Omega_n\setminus B_n]}\right] - (1+o(1))\mathbf{E}_{\mu_n}\left[g_n\mathbf{1}_{[\Omega_n\setminus B_n]}\right] = o(1);$$

which is a straightforward consequence of f_n and g_n being bounded and that the measure of $\Omega_n \setminus B_n$ vanishes for both measures as $n \to \infty$. □

Similar to Section 7.3 above, we will start with the unconditional case; that is, the limiting distribution of $(\mathrm{h}_{n,1}(i),\, i \in [k])$.

Proposition 7.6.2. *Fix an integer $k \geq 2$. For any $\bar{x} \in \mathbb{R}^k$,*

$$\lim_{n\to\infty} \mathbf{P}\left(\mathrm{h}_{n,1}(i) \leq x_i\sqrt{\ln n} + \ln n,\, i \in [k]\right) = \prod_{i=1}^{k} \Phi(x_i).$$

Proof. Recall the definition of $G_{n,x}$ in (7.8). We claim that for any $\bar{x} \in \mathbb{R}^k$,

$$\mathbf{P}\left(\mathrm{h}_{n,1}(i) \leq x_i\sqrt{\ln n} + \ln n,\, i \in [k]\right) = (1+o(1))\prod_{i=1}^{k}\mathbf{E}\left[G_{n,x_i}(|\mathcal{S}_{n,1}(i)|)\right] + o(1). \tag{7.22}$$

To see this, let μ_n denote the law of $(\mathcal{S}_{n,1}(i),\, i \in [k])$ and ν_n denote the law of $(\mathcal{R}_n(i),\, i \in [k])$; recall that the latter are k independent copies of $\mathcal{S}_{n,1}(1)$.

Let $\mathcal{B}_{n,k,1/2}$ be as defined in (7.17) and set

$$f_n(\bar{J}) = \mathbf{P}\left(\mathrm{h}_{n,1}(i) \le x_i\sqrt{\ln n} + \ln n,\ i \in [k] \,|\, \bar{\mathcal{S}}_{n,1} = \bar{J}\right),$$

$$g_n(\bar{J}) = \prod_{i=1}^{k} G_{n,x_i}(|J_i|).$$

From the first equation in Proposition 7.4.2, it follows that $f_n(\bar{J}) = g_n(\bar{J})$ for all $\bar{J} \in \mathcal{B}_{n,k,1/2}$. Therefore, the conditions on Lemma 7.6.1 for μ_n, ν_n and $\mathcal{B}_{n,k,1/2}$ are satisfied by Lemma 7.4.5 and Proposition 7.4.6, establishing (7.22).

Finally, it suffices to verify that, for all $i \in [k]$,

$$\lim_{n\to\infty} \mathbf{E}\left[G_{n,x_i}(|\mathcal{S}_{n,1}(i)|)\right] = \mathbf{E}\left[\Phi(\sqrt{2}x_i - N))\right] = \Phi(x_i);$$

where N is a standard Gaussian variable. The proof of this follows with the same argument as that for Lemma 7.3.3 with the main difference being that, instead of using $|\mathcal{S}_n(i)|$, we use $|\mathcal{S}_{n,1}(i)|$. By Lemma 7.4.1, the renormalization $\frac{|\mathcal{S}_{n,1}(i)|-2\ln n}{\sqrt{2\ln n}}$ also converges to a standard Gaussian distribution. □

We now proceed to treat the case with nontrivial conditioning.

Proposition 7.6.3. *Fix an integer $k \ge 2$ and vectors $\bar{a} \in [0,1]^k$, $\bar{b} \in \mathbb{Z}^k$ and $\bar{x} \in \mathbb{R}^k$; we have*

$$\lim_{n\to\infty} \mathbf{P}\left(\mathrm{h}_{n,1}(i) \le l_i,\ i \in [k] \,|\, \mathrm{d}_n(i) \ge m_i,\ i \in [k]\right) = \prod_{i=1}^{k} \Phi(x_i),$$

where $m_i = m_i(a_i, b_i, n) = \lfloor a_i \log n \rfloor + b_i$ and $l_i = l_i(a_i, x_i, n) = x_i\sqrt{\sigma_{a_i}^2 \ln n} + \mu_{a_i} \ln n$.

Proof. Recall the definition of $\widetilde{G}_{m,l}$ in (7.12). In what follows, we assume, without lose of generality, that $m_i \ge 0$ (if $m_i < 0$ then $\mathrm{d}_n(i) \ge m_i$ a.s., so we set $m_i = 0$). Now, we first show that

$$\mathbf{P}\left(\mathrm{h}_{n,1}(i) \le l_i,\, \mathrm{d}_n(i) \ge m_i,\, i \in [k]\right) = (1+o(1))2^{-\sum_i m_i} \prod_{i=1}^{k} \mathbf{E}\left[\widetilde{G}_{m_i,l_i}(|\mathcal{S}_{n,1}(i)|)\right] + o(1). \tag{7.23}$$

To see this, let μ_n denote the law of $(\mathcal{S}_{n,1}(i),\, i \in [k])$ and ν_n denote the law of $(\mathcal{R}_n(i),\, i \in [k])$; recall that the latter are k independent copies of $\mathcal{S}_{n,1}(1)$. Also, write

$$f_n(\bar{J}) = \mathbf{P}\left(\mathrm{h}_{n,1}(i) \le l_i,\, \mathrm{d}_n(i) \ge m_i,\, i \in [k] \,|\, \bar{\mathcal{S}}_{n,1} = \bar{J}\right)$$
$$g_n(\bar{J}) = 2^{-\sum_i m_i} \prod_{i=1}^{k} \widetilde{G}_{m_i,l_i}(|J_i|).$$

Let $\alpha = \max\{a_i : i \in [k]\}$ and set $0 < \delta < 2 - \alpha$. Note that δ is chosen so that, for n large enough, $f_n(\bar{J}) = g_n(\bar{J})$ for all $\bar{J} \in \mathcal{B}_{n,k,\delta}$; this follows from Remark 7.4.3 and the second equation in Proposition 7.4.2. Lemma 7.4.5 and Proposition 7.4.6 yield the remaining conditions on μ_n, ν_n and $\mathcal{B}_{n,k,\delta}$, which applying Lemma 7.6.1 gives (7.23).

Next, let N be a variable with standard Gaussian distribution. For each $i \in [k]$,

$$\lim_{n\to\infty} \mathbf{E}\left[\widetilde{G}_{m_i,l_i}(|\mathcal{S}_{n,1}(i)|)\right] = \mathbf{E}\left[\Phi\left(\frac{\sqrt{1+\mu_{a_i}}x_i - N}{\sqrt{\mu_{a_i}}}\right)\right] = \Phi(x_i). \tag{7.24}$$

The last equality follows by Lemma 7.3.2 and the proof of the first equality follows similar to Lemma 7.3.3 when replacing the variables $|\mathcal{S}_n(i)|$ to $|\mathcal{S}_{n,1}(i)|$, which have the same limiting distribution.

Finally, it follows from (7.23) and (7.24) that

$$\lim_{n\to\infty} \mathbf{P}\left(\mathrm{h}_{n,1}(i) \le l_i,\, \mathrm{d}_n(i) \ge m_i,\, i \in [k]\right) = 2^{-\sum_i m_i} \prod_{i=1}^{k} \Phi(x_i);$$

The result now follows by Proposition 7.5.4, since $\mathbf{P}(\mathrm{d}_n(i) \geq m_i,\, i \in [k])$ converges to $2^{-\sum_i m_i}$. ☐

Proof of Theorem 7.2.5, case $k \geq 2$. Let $\bar{a} \in [0,1]^k$ and $\bar{m} \in \mathbb{Z}^k$ be fixed and set $m_i = \lfloor a_i \log n \rfloor + b_i$. If $m_i \leq 0$ for all $i \in [k]$, the result follows from Proposition 7.6.2 and Lemma 7.5.2. Otherwise, the result follows from Proposition 7.6.3 and Lemma 7.5.5. ☐

7.7 Proof of Proposition 7.1.8

The next lemma appeared in [3]; we include its short proof for completeness.

Lemma 7.7.1. *For any $k' \in \mathbb{N}$ and integers $(m_i,\, i \in [k'])$,*

$$\mathbf{P}(\mathrm{d}_n(i) = m_i,\, i \in [k']) = \sum_{l=0}^{k'} \sum_{\substack{S \subset [k'] \\ |S|=l}} (-1)^l \mathbf{P}\left(\mathrm{d}(i) \geq m_i + \mathbf{1}_{[i \in S]},\, i \in [k']\right).$$

Furthermore, for fixed $c \in (0,2)$; if $m_i < c \ln n$ for $i \in [k]$ and $k' < k \in \mathbb{N}$, then

$$\begin{aligned} &\mathbf{P}(\mathrm{d}_n(i) = m_i, \mathrm{d}_n(j) \geq m_j,\, 1 \leq i \leq k' < j \leq k) \\ =&\sum_{l=0}^{k'} \sum_{\substack{S \subset [k'] \\ |S|=l}} (-1)^l \mathbf{P}\left(\mathrm{d}_n(i) \geq m_i + \mathbf{1}_{[i \in S]},\, i \in [k]\right) \\ =&2^{-k' - \sum_i m_i}(1 + o(1)). \end{aligned}$$

Proof. The first part is proven directly proved using the inclusion-exclusion principle. The second equation follows by intersecting the event $\{\mathrm{d}_n(j) \geq m_j,\, k' < j \leq k\}$ along all probabilities in the first equation; then applying Proposition 7.5.4 to each term:

$$\sum_{l=0}^{k'} \sum_{\substack{S \subset [k'] \\ |S|=l}} (-1)^l \mathbf{P}\left(\mathrm{d}_n(i) \geq m_i + \mathbf{1}_{[i \in S]},\, i \in [k]\right) = (1+o(1)) 2^{-\sum_i m_i} \sum_{l=0}^{k'} \sum_{\substack{S \subset [k'] \\ |S|=l}} (-1)^l 2^{-l}.$$

☐

Corollary 7.7.2. *Let $k' < k \in \mathbb{N}$ and fix $(a_i, A_i) \in \mathbb{Z} \times \mathcal{B}_I$ for $1 \le i \le k$. Write $m_i = \lfloor \log n \rfloor + a_i$ and*

$$\mathcal{D}_{\bar{m}} = \{\mathrm{d}_n(i) = m_i,\, 1 \le i \le k'\} \cup \{\mathrm{d}_n(i) \ge m_i,\, k' < i \le k\},$$
$$\mathcal{H}_{\bar{A}} = \left\{ \frac{\mathrm{h}_n(i) - \mu \ln n}{\sqrt{\sigma^2 \ln n}} \in A_i,\, i \in [k] \right\}.$$

Then

$$\mathbf{P}\left(\mathcal{D}_{\bar{m}},\, \mathcal{H}_{\bar{A}}\right) = \left(2^{-k' - \sum_i d_i}\right) \prod_{i=1}^{k} \Phi(A_i)(1 + o(1)).$$

Proof. We start by intersecting the event $\mathcal{H}_{\bar{A}}$ along all probabilities in the second expression of Lemma 7.7.1; then we use the approximation by independent Gaussian variables given in Theorem 7.2.5. This gives

$$\begin{aligned}
\mathbf{P}\left(\mathcal{D}_{\bar{m}},\, \mathcal{H}_{\bar{A}}\right) =& \sum_{l=0}^{k'} \sum_{\substack{S \subset [k'] \\ |S| = l}} (-1)^l \mathbf{P}\left(\mathcal{H}_{\bar{A}},\, \mathrm{d}_n(i) \ge m_i + \mathbf{1}_{[i \in S]},\, i \in [k]\right) \\
=& \sum_{l=0}^{k'} \sum_{\substack{S \subset [k'] \\ |S| = l}} (-1)^l \mathbf{P}\left(\mathcal{H}_{\bar{A}} \,\middle|\, \mathrm{d}_n(i) \ge m_i + \mathbf{1}_{[i \in S]},\, i \in [k]\right) \mathbf{P}\left(\mathrm{d}_n(i) \ge m_i + \mathbf{1}_{[i \in S]},\, i \in [k]\right) \\
=& (1 + o(1) \prod_{i=1}^{k} \Phi(A_i) \sum_{l=0}^{k'} \sum_{\substack{S \subset [k'] \\ |S| = l}} (-1)^l \mathbf{P}\left(\mathrm{d}_n(i) \ge m_i + \mathbf{1}_{[i \in S]},\, i \in [k]\right) \\
=& (1 + o(1)) \left(2^{-k' - \sum_i m_i}\right) \prod_{i=1}^{k} \Phi(A_i).
\end{aligned}$$

□

Proof of Proposition 7.1.8. Fix $c \in (0, 2)$ and $M \in \mathbb{N}$. Let $j = j(n)$ and $j' = j(n)$ be integer-valued functions with $0 \le j'(n) + \log n < j(n) + \log n < c \ln n$; let $K' < K \in \mathbb{N}$ and $(a_k,\, k \in [K])$ be non-negative integers such that $\sum_{k \in [K]} a_k = M$ and set $M' = \sum_{k \in [K']} a_k$. Consider an arbitrary (K', K)-canonical sequence $((j_k, B_k),\, 1 \le k \le K)$ with $j' = j_1$ and $j = j_K$.

We define $m_i \in \mathbb{N}$ and $A_i \subset \mathbb{R}$ as follows. For each $k \in [K]$, if $\sum_{l=1}^{k-1} a_l < i \leq \sum_{l=1}^{k}$ then set $m_i = \lfloor \log n \rfloor + j_k$ and let $A_i = B_{j_k}$. In this case, consider the sets

$$
\begin{aligned}
\mathcal{D}_{\bar{m}} &= \{\mathrm{d}_n(i) = m_i,\ 1 \leq i \leq M'\} \cup \{\mathrm{d}_n(i) \geq m_i,\ M' < i \leq M\}, \\
\mathcal{H}_{\bar{A}} &= \left\{ \frac{\mathrm{h}_n(i) - \mu \ln n}{\sqrt{\sigma^2 \ln n}} \in A_i,\ i \in [M] \right\}.
\end{aligned}
$$

By Corollary 7.2.4 and the exchangeability of the vertex degrees of $T^{(n)}$,

$$
\begin{aligned}
&\mathbf{E}\left[\prod_{k=1}^{K'} \left(X_{j_k}^{(n)}(B_k)\right)_{a_k} \prod_{k=K'+1}^{K} \left(X_{\geq j}^{(n)}(B_k)\right)_{a_k}\right] \\
&= (n)_M \mathbf{P}\left(\mathcal{D}_{\bar{m}},\ \mathcal{H}_{\bar{A}}\right) \\
&= (1 + o(1)) \left(2^{M \log n - M' - \sum_i m_i}\right) \prod_{i=1}^{M} \Phi(A_i),
\end{aligned}
$$

the last equality holding by Corollary 7.7.2 and since $(n)_M = n^M(1 + o(n^{-1}))$. Finally, note that

$$
M \log n - M' - \sum_{i=1}^{M} m_i = \sum_{k=1}^{K'} (-j_k - 1 + \varepsilon_n) a_k + \sum_{k=K'+1}^{K} (-j' + \varepsilon_n) a_k,
$$

and so

$$
\left(2^{A \log n - A' - \sum_i m_i}\right) = \prod_{k=1}^{K'} \left(2^{-j_k - 1 + \varepsilon_n}\right)^{a_k} \prod_{k=K'+1}^{K} \left(2^{-j' - \varepsilon_n}\right)^{a_k}.
$$

Similarly, $\prod_{i=1}^{M} \Phi(A_i) = \prod_{k=1}^{K} \Phi(B_k)^{a_k}$; which completes the proof. $\square$

7.8 Proof of Theorem 7.1.1

Recall that $\mathcal{M}_n$ is the set of vertices in t_n attaining the maximum degree. In light of Theorem 7.1.2, to prove Theorem 7.1.1 it suffices to prove the convergence of $|\mathcal{M}_n|$ over suitable subsequences.

Proposition 7.8.1. *Let $\varepsilon \in [0,1]$. If n_l is an increasing sequence such that $\varepsilon_{n_l} \to \varepsilon$ as $l \to \infty$, then $\mathbf{M}_{n_l}$ converges in distribution to $\mathbf{M}_\varepsilon$, where $\mathbf{M}_\varepsilon$ is defined by*

$$\mathbf{P}\left(\mathbf{M}_\varepsilon = k\right) = \sum_{m\in\mathbb{Z}} e^{-2^{-m+\varepsilon}} \frac{2^{-(m+1-\varepsilon)k}}{k!}$$

for each integer $k \geq 1$.

Proof. The formula for $\mathbf{P}\left(\mathbf{M}_\varepsilon = k\right)$ may be seen using the following heuristic. Each of the terms in the sum represent the limit of the probability that; given that the maximum degree in T_n equals $\lfloor \log n \rfloor + m$, there exist precisely k vertices attaining such degree.

We first verify that $\sum_{k\geq 1} \mathbf{P}\left(\mathbf{M}_\varepsilon = k\right) = 1$; this follows from a telescopic analysis of the sum.

$$\begin{aligned}
\sum_{k\geq 1} \mathbf{P}\left(\mathbf{M}_\varepsilon = k\right) &= \lim_{M\to\infty} \sum_{m=-M}^{M} \sum_{k\geq 1} e^{-2^{-m+\varepsilon}} \frac{2^{-(m+1-\varepsilon)k}}{k!} \\
&= \lim_{M\to\infty} \sum_{m=-M}^{M} e^{-2^{-m+\varepsilon}} \left(e^{2^{-(m+1-\varepsilon)}} - 1\right) \\
&= \lim_{M\to\infty} \sum_{m=-M}^{M} \left(e^{-2^{-(m+1-\varepsilon)}} - e^{-2^{-m+\varepsilon}}\right) \\
&= \lim_{M\to\infty} \left(e^{-2^{-(M+1-\varepsilon)}} - e^{-2^{M+\varepsilon}}\right) = 1
\end{aligned}$$

We now proceed to the proof of the theorem; we abuse notation by writing, e.g. $X_j = X_j(\mathbb{R})$. Consider $\varepsilon \in [0,1]$ fixed and n_l an increasing sequence for which $\varepsilon_{n_l} \to \varepsilon$ as $l \to \infty$. We assume $\varepsilon = 0$ for simplicity of the formulas below. Fixing $k, M \geq 1$ we have

$$\mathbf{P}\left(\mathbf{M}_{n_l} = k\right) \leq \mathbf{P}\left(X^{(n_l)}_{\geq -M} = 0\right) + \sum_{j=-M}^{M-1} \mathbf{P}\left(X^{(n_l)}_j = k, X^{(n_l)}_{\geq j+1} = 0\right) + \mathbf{P}\left(X^{(n_l)}_{\geq M} > 0\right).$$

By Lemma 7.1.6 and Theorem 7.1.2 we have that for each $m \in \mathbb{N}$,

$$(X^{(n_l)}_{-m}, \dots X^{(n_l)}_{m-1}, \dots, X^{(n_l)}_{\geq m}) \xrightarrow{\mathcal{L}} (X_{-m}, \dots, X_{m-1}, X_{\geq m})$$

and that the limit is a vector of independent vector of Poisson variables. In particular,

$$\mathbf{P}\left(X_j = k, X_{\geq j+1} = 0\right) = \mathbf{P}\left(X_j = k\right)\mathbf{P}\left(X_{\geq j+1} = 0\right) = \frac{e^{-2^{-j}} 2^{-(j+1)k}}{k!};$$

also $X_{\geq -M}^{(n_l)} = X_{\geq M}^{(n_l)} + \sum_{j=-M}^{M-1} X_j^{(n_l)}$, and $X_{\geq -M}^{(n_l)} \xrightarrow{\mathcal{L}} X_{\geq -M} \stackrel{\mathcal{L}}{=} \text{Poi}(2^M)$. Thus, in the limit

$$\begin{aligned}
\limsup_{n_l \to \infty} \mathbf{P}\left(\mathbf{M}_{n_l} = k\right) &\leq \mathbf{P}\left(X_{\geq -M} = 0\right) + \sum_{j=-M}^{M-1} \mathbf{P}\left(X_j = k, X_{\geq j+1} = 0\right) + \mathbf{P}\left(X_{\geq M} > 0\right). \\
&= e^{-2^M} + \frac{1}{k!} \sum_{j=-M}^{M} \left(e^{-2^{-j}} 2^{-(j+1)k}\right) + \left(1 - e^{-2^{-M}}\right);
\end{aligned}$$

This holds for arbitrary $M \in \mathbb{N}$, hence

$$\begin{aligned}
\limsup_{n_l \to \infty} \mathbf{P}\left(\mathbf{M}_{n_l} = k\right) &\leq \liminf_{M} \left\{ e^{-2^{-(-M)}} + \frac{1}{k!} \sum_{j=-M}^{M-1} \left(e^{-2^{-j}} 2^{-(j+1)k}\right) + \left(1 - e^{-2^{-M}}\right) \right\} \\
&= \frac{1}{k!} \sum_{j \in \mathbb{Z}} e^{-2^{-j}} 2^{-(j+1)k}.
\end{aligned}$$

Similarly,

$$\liminf_{n_l \to \infty} \mathbf{P}\left(\mathbf{M}_{n_l} = k\right) \geq \limsup_{M} \left\{ \frac{1}{k!} \sum_{j=-M}^{M-1} \left(e^{-2^{-j}} 2^{-(j+1)k}\right) \right\} = \frac{1}{k!} \sum_{j \in \mathbb{Z}} e^{-2^{-j}} 2^{-(j+1)k}.$$

□

CHAPTER 8
Extremal values in recursive trees via a new tree growth process

We give convergence rates on the number of vertices with degree at least $c \ln n$, $c \in (1, 2)$, in random recursive trees on n vertices. This allows us to extend the range for which the distribution of the number of vertices of a given degree is well understood.

Conceptually, the key innovation of our work lies in a new tree growth process $((T_n, \boldsymbol{\sigma}_n),\ n \geq 1)$ where T_n is a rooted labeled tree on n vertices and $\boldsymbol{\sigma}_n$ is a permutation of the vertex labels. The shape of T_n has the same law as that of a random recursive tree. Interesting on its own right, this process obtains T_n from T_{n-1} by a procedure we call Robin-Hood pruning, which attaches a vertex labeled n to T_{n-1} and rewires some of the edges in T_{n-1} towards the newly added vertex. Additionally, $((T_n, \boldsymbol{\sigma}_n),\ n \geq 1)$ can be understood as a new coupling of all finite Kingman's coalescents.

8.1 Introduction

In a paper of 1970 [67], 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. First, a class of *'growing'* trees whose construction corresponds to the standard construction of random recursive trees (RRTs). These are constructed by sequentially adding new vertices, which are attached to a uniformly random vertex in the previous tree. Second, a class of *'static'* trees formed via a coalescent process beginning with n isolated nodes. They described the construction of a 'static' tree with n vertices as follows.

"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."

The term 'static' was motivated by the fact that this construction starts with the n vertices the tree is aimed to have at the end of the process. This is a description, in fact, of a discrete multiplicative coalescent which is linked to Kruskal's algorithm for the minimum weighted spanning tree problem [1] [1] . A growing process of such coalescent was not foreseen; however, it is now known that, for some coalescent procedures (e.g. additive and Kingman's), the resulting tree can also be constructed by a growth process [1, 58, 73]. In particular, Kingman's coalescents correspond to RRTs; see e.g. [1], or Propositions 8.1.1 and 8.5.3 below.

The key conceptual contribution of this work is what we call the Robin-Hood pruning procedure. This is a random construction which, given a Kingman's coalescent on n vertices, produces a Kingman's coalescent on $n+1$ vertices. The benefits of such construction are twofolded.

First, Kingman's coalescent had already been exploited by Addario-Berry and the author to describe near-maximum degrees in RRTs, [3, 36]. With the new procedure, we are able to extract finer information about extreme degree

[1] Unfortunately, it was incorrectly presumed in [67] to build uniformly random unrooted labeled trees.

values in RRTs. Second, growth procedures naturally couple families of trees as the size varies. However, typically there is no simple coupling of finite n-coalescent processes as n varies. The introduction of the Robin-Hood pruning provides, to the best of our knowledge, a novel tree growth procedure which is interesting on its own; thereby, opening a wide range of further avenues of research.

The Robin-Hood pruning is best described through an auxiliary tree structure that relates to both Kingman's coalescent and RRTs. We proceed to its description, then we present the results obtained in this work.

8.1.1 Notation

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

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 $e = uv$ for an edge with tail u and head v. For $u \in V(t) \setminus \{r(t)\}$ we write $p_t(u)$ for the parent of u, that is, the unique vertex v with uv in $E(t)$. Finally, write $\mathrm{d}_t(v)$ for the number of edges directed toward v in t, and call $\mathrm{d}_t(v)$ the degree of v. Note that $\mathrm{d}_t(v) = \#\{u : p_t(u) = v\}$.

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 $v \in [n] \setminus \{r(t)\}$ (in particular, $r(t) = 1$). We write $\mathcal{I}_n \subset \mathcal{T}_n$ for the set of increasing trees of size n. It is easy to see that $|\mathcal{I}_n| = (n-1)!$ for all n. Next, a tree growth process is a sequence $(t_n,\, n \geq 1)$ of trees with $t_n \in \mathcal{T}_n$ for each n. The process is increasing if t_n is a subtree of t_{n+1} for all n; this implies that $t_n \in \mathcal{I}_n$ for all n.

Tree growth processes select a tree from $\mathcal{T}_n$ for each $n \geq 1$ usually with the characteristic that new vertices attach to some vertex in the previous tree, giving rise to increasing trees.

8.1.2 Recursively decorated trees

We begin with the standard construction of a RRT of size $n \geq 1$, which we denote by R_n. 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 $1 < j \leq n$. The process $(R_n,\, n \geq 1)$ is a random increasing tree growth process. Moreover, it is readily seen that R_n is a random increasing tree uniformly chosen from $\mathcal{I}_n$.

Recursively decorated trees extend the concept of increasing trees. If $t \in \mathcal{T}_n$ and $\sigma \in \mathcal{S}_n$ then $\sigma(t)$ is the tree $t' \in \mathcal{T}_n$ with edges $\{\sigma(u)\sigma(v) : uv \in E(t)\}$. We say σ is an *addition history* for t if $\sigma(t)$ is increasing. If σ is an addition history for t then we say that the pair (t, σ) is a *recursively decorated tree* or *decorated tree*, and that vertex v has addition time $\sigma(v)$, for all $v \in V(t)$. Write

$$\mathcal{RD}_n = \{(t, \sigma) : t \in \mathcal{T}_n,\, \sigma \text{ is an addition history of } t\},$$

for the set of recursively decorated trees of size n. See Figure 8–1 for an example.

For each $n \geq 1$ let $\mathrm{RT}_n = (T_n, \boldsymbol{\sigma}_n)$ be a uniformly chosen decorated tree in $\mathcal{RD}_n$. We remark now that $\boldsymbol{\sigma}_n$ encodes the evolution of Kingman's coalescent on n vertices; the details on such correspondance are given in Section 8.5. The next, straightforward proposition shows that the shape of T_n has the same law as that of R_n.

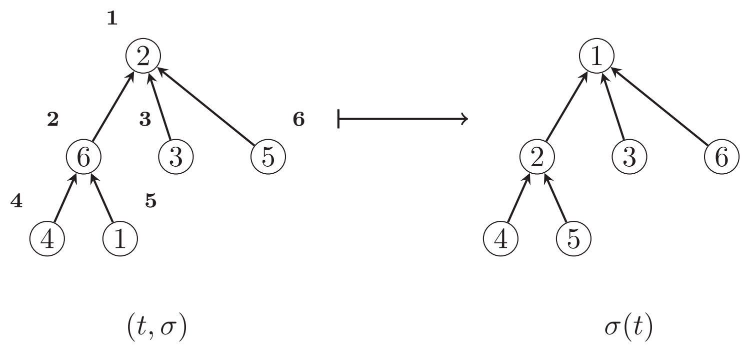


Figure 8–1: A decorated tree $(t, \sigma) \in \mathcal{RD}_6$ on the left; the permutation σ is depicted with bold numbers next to the vertices in t (so for example $\sigma(1) = 5$ and $\sigma(6) = 2$). On the right, the increasing tree $\sigma(t)$.

Proposition 8.1.1. *For each $n \in \mathbb{N}$, $|\mathcal{RD}_n| = n!(n-1)!$ and if $\mathrm{RT}_n = (T_n, \boldsymbol{\sigma}_n) \in \mathcal{RD}_n$ is chosen uniformly at random then $\boldsymbol{\sigma}_n(T_n)$ is a random recursive tree of size n and $\boldsymbol{\sigma}_n$ is a uniformly chosen permutation in $\mathcal{S}_n$.*

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

Corollary 8.1.2. *For all $n \in \mathbb{N}$, the following distributional identity holds.*

$$(\mathrm{d}_{\mathrm{RT}_n}(\boldsymbol{\sigma}_n^{-1}(v));\, v \in [n]) \stackrel{\mathcal{L}}{=} (\mathrm{d}_{R_n}(v);\, v \in [n]).$$

8.1.3 Statement of results

The Robin-Hood pruning $\mathrm{RH}_n : \mathcal{RD}_{n-1} \to \mathcal{RD}_n$ is a random procedure that allows us to construct RT_n from RT_{n-1} while preserving most of the edges in RT_{n-1}.

Broadly speaking, $\mathrm{RH}_n(t, \sigma)$ is obtained from (t, σ) 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 addition history in $\mathrm{RH}_n(t,\sigma)$ is adjusted from σ such that vertex n has a uniformly random *addition time*. Heuristically, the random procedure follows a 'steal from the old to give to the new' scheme; that is, once the addition time of n has been determined, vertices which had been added earlier (according to σ) have larger probability of being reattached to vertex n.

We will write $\mathrm{RH} = \mathrm{RH}_n$ when the size of the input is clear from the context. The exact definition of RH_n will be given in the next section along with the proof of the following result.

Theorem 8.1.3. *For each $n \geq 1$ let $\mathrm{RT}_n = (T_n, \boldsymbol{\sigma}_n)$ be a uniformly random element in $\mathcal{RD}_n$. The Robin-Hood pruning provides a coupling for $((T_n, \boldsymbol{\sigma}_n),\ n \geq 1)$ by setting $(T_n, \boldsymbol{\sigma}_n) = \mathrm{RH}(T_{n-1}, \boldsymbol{\sigma}_{n-1})$ for each $n \geq 2$.*

As we will establish in Proposition 8.5.3, RT_n is a representation of Kingman's coalescent on $[n]$. Therefore, we have the following corollary.

Corollary 8.1.4. *The construction of $((T_n, \boldsymbol{\sigma}_n),\ n \geq 1)$ in Theorem 8.1.3 gives an explicit coupling of all finite Kingman's coalescents.*

A remarkable property of the coupling in Theorem 8.1.3 is that, it yields a tree growth process where all the trees are distributed as RRTs, however the process itself is not increasing. To the best of our knowledge, this is a novel evolution of random networks. Potential applications and open problems are discussed in Section 8.6.

Turning to extreme values in the degree sequence of RRTs, consider the following variables. For integers $0 < m \leq n$, let

$$X_m^{(n)} = \#\{v \in [n] : \mathrm{d}_{R_n}(v) = m\};$$

Janson established the joint limiting distribution of $(X_m^{(n)},\ m \geq 1)$ in [50]; for previous results on the degree distribution of RRTs see the survey [77]. In terms of capturing the degree sequence around the maximum degree range, Addario-Berry and the author provide all the possible limiting distributions of $(X^{(n)}_{\lfloor \log n\rfloor+k}, k \in \mathbb{Z})$ in [3].

In this work we are concerned with high-degree vertices in a broader sense; that is, we consider the variables

$$Z^{(n)}_m = \#\{v \in [n] : \mathrm{d}_{R_n}(v) \geq m\},$$

with $m \sim c \ln n$ and provide results on convergence rates toward their limiting distributions. Throughout the paper, we write $\lambda_{n,m} = \mathbf{E}\left[Z^{(n)}_m\right]$ and record the following estimate.

Lemma 8.1.5 (Lemma 4.3 in [3])**.** *First, $\lambda_{n,m} \leq 2^{-m+\log n}$, and for each $c \in (0, 2)$, there is $\gamma(c)$ such that, uniformly over $m < c \ln n$,*

$$\lambda_{n,m} = n\mathbf{P}\left(\mathrm{d}_{\mathrm{RT}_n}(1) \geq m\right) = 2^{-m+\log n}(1 + o(n^{-\gamma})).$$

Our first result on high-degree vertices is obtained by applying the Chen-Stein method.

Theorem 8.1.6. *Fix $1 < c < c' < 2$. There are constants $\alpha = \alpha(c') \in (0, 1)$ and $\beta = \beta(c) > 0$ such that uniformly for $m = m(n)$ satisfying $c \ln n < m < c' \ln n$,*

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

The exponent $-m + (1 - \alpha) \log n$ in Theorem 8.1.6 is negative when $(1 - \alpha) \log e < c$. A detailed but simple track of the conditions on α, see Proposition 8.3.2, shows that there is a non-empty interval $I_{c'} = ((1-\alpha) \log e, c')$ such that if $c \in I_{c'} \cap (1, 2)$, then the bounds in Theorem 8.1.6 are, in fact, tending to zero. Moreover, by Lemma 8.1.5, if $c < \log e$, then $\lambda_{n,c \ln n} \to \infty$ as $n \to \infty$. This fact, together with Theorem 8.1.6 yields the next corollary.

Corollary 8.1.7. *For each $c' \in (1, \log e)$ there exists $c \in (1, c')$ such that if $c \ln n < m < c' \ln n$, then*

$$\frac{Z_m^{(n)} - \lambda_{n,m}}{\sqrt{\lambda_{n,m}}} \xrightarrow{\mathcal{L}} N(0,1).$$

Corollary 8.1.7 extends the range of $m = m(n)$ for which a central limit theorem exists for $Z_m^{(n)}$; previous results were given for m constant [50] and for $m = \log n - d$ with $d = d(n)$ slowly tending to infinity [3].

We opted to state Theorem 8.1.6 and Corollary 8.1.7 separately to clarify where the bounds on m are limiting the convergence rates. In Section 8.3, we explain how previous results in [3] determine the exponent α, while the exponent β depends on an auxiliary coupling based on the Robin-Hood pruning. The details of such coupling are given in Section 8.4; for the moment we remark that, by Corollary 8.1.2, for all $m \leq n$,

$$Z_m^{(n)} \stackrel{\mathcal{L}}{=} \#\{v \in [n] : \mathrm{d}_{\mathrm{RT}_n}(v) \geq m\}; \tag{8.1}$$

when there is no ambiguity, we write RT_n to refer only to its tree coordinate. The pruning procedure provides a key description of $(\mathrm{d}_{\mathrm{RT}_n}(i),\, i \in [n])$ in terms of both $(\mathrm{d}_{\mathrm{RT}_{n-1}}(i),\, i \in [n-1])$ and $\mathrm{d}_{\mathrm{RT}_n}(n)$, which is independent of the former vector. This allows us to analyze the conditional law of $(\mathrm{d}_{\mathrm{RT}_n}(i),\, i \in [n-1])$ given that $\mathrm{d}_{\mathrm{RT}_n}(n) \geq m$.

Our last result concerns the maximum degree Δ_n of R_n, which by Corollary 8.1.2 satisfies $\Delta_n \stackrel{\mathcal{L}}{=} \max\{\mathrm{d}_{\mathrm{RT}_n}(v) :\, v \in [n]\}$.

Theorem 8.1.8. *There exists $C > 0$ such that uniformly over $0 < i = i(n) < \log e \ln \ln n - C$,*

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

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

The first bounds of this type, for $\mathbf{P}(\Delta_n < \lfloor \log n \rfloor + j)$ with $j \in \mathbb{Z}$, were given in [39]. An extension to $j < 2\ln n - \log n$ was obtained in [3]. Goh and Schmutz provide a heuristic of how the Gumbel, or double-exponential, distribution arises for Δ_n [39]. They do so by looking at the limiting distribution of $\mathrm{d}_{R_n}(i)$ with $i \to \infty$ slowly. Below we present a distinct heuristic in terms of the degree distribution of RT_n.

The maximum of i.i.d. random variables is, under rather general conditions, distributed in the limit as the Gumbel distribution [42]. Lattice distributions are excluded from this regime and, instead, Anderson gives analogous conditions under which the Gumbel distribution serves as an approximation for their maximum [4]. In the case of $(\deg_{\mathrm{RT}_n}(v),\, v \in [n])$, their limiting distributions are geometric, a distribution which satisfies the conditions given in [4]. Although, the degrees of a tree are not independent, their correlations are weak and the Gumbel-type approximation still arises for the distribution of Δ_n.

Outline

The remainder of the paper is organized as follows. The proof of Theorem 8.1.3 is given in the next section, and its connection to Kignman's coalescent in Section 8.5. In Section 8.3, we explain how we apply the Chen-Stein method to $Z_m^{(n)}$ by constructing an auxiliary coupling; the proofs of Theorems 8.1.6 and 8.1.8 and Corollary 8.1.7 are provided in Section 8.3 under the assumption of the existence of such coupling. The auxiliary coupling is based on the Robin-Hood pruning and is defined in Section 8.4. To close this work, we briefly discuss further avenues of research in Section 8.6.

8.2 The Robin-Hood pruning

For each $n \geq 2$, the Robin-Hood pruning RH_n is a random procedure that takes a decorated tree $(t, \sigma) \in \mathcal{RD}_{n-1}$ and outputs a decorated tree

$\mathrm{RH}_n(t,\sigma) \in \mathcal{RD}_n$. We first define a deterministic pruning, which is illustrated in Figure 8–2.

8.2.1 A deterministic process

For $d \geq 1$, we write $x = (x_1, \ldots, x_d) \in \{0,1\}^d$. Let $n > 1$ and set

$$\mathcal{C}_n = \{(k,l,x) : 1 \leq l < k \leq n,\, x \in \{0,1\}^{n-1}\} \cup \{(1,0,x) : x \in \{0,1\}^{n-1},\, x_1 = 1\};$$

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

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

Note that the definition of $\mathcal{C}_n$ is such that $\sigma^{-1}(1) \in \mathcal{V}_n$ if and only if $k = 1$. The set $\mathcal{V}_n$ corresponds to the vertices to be pruned and rewired in the following deterministic *pruning*.

Definition 8.2.1. *For $n \geq 2$, $(t,\sigma) \in \mathcal{RD}_{n-1}$ and $(k,l,x) \in \mathcal{C}_n$ define $(t',\sigma') \in \mathcal{T}_n \times \mathcal{S}_n$ as follows.*

First, t' is obtained from t as follows. Let $\mathcal{V} = \mathcal{V}_n(k,x,\sigma)$. For each $v \in \mathcal{V} \setminus \{r(t)\}$, replace the edge $vp_t(v)$ 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' are given by

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

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

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

We write $\mathrm{rh}_n((t,\sigma),(k,l,x)) = (t',\sigma')$.

Lemma 8.2.2. *Fix $n \geq 2$. For any $(t, \sigma) \in \mathcal{RD}_{n-1}$ and $(k, l, x) \in \mathcal{C}_n$,*

$$\mathrm{rh}_n((t, \sigma), (k, l, x)) \in \mathcal{RD}_n.$$

Proof. Write $\mathrm{rh}_n((t, \sigma), (k, l, x)) = (t'\sigma')$. When $k = 1$, it is clear that t' is a tree. When $k > 1$, let $w = \sigma^{-1}(l)$ be the parent of n in t' and let $(w = v_1, \ldots, v_j = r(t))$ be the path from w to the root of t. Since σ is an addition history of t,

$$l = \sigma(v_1) > \sigma(v_2) > \cdots > \sigma(v_j) = 1;$$

moreover, $l < k$. It follows that $v_i \notin \mathcal{V}(k, l, x, \sigma)$ for $i \in [j]$ and consequently, no edges in the path from n to the root in t' closes a cycle by connecting to n.

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

$$\sigma'(v) = \sigma(v) + 1 > k = \sigma'(n).$$

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

$$\sigma'(n) = k > l = \sigma(w) = \sigma'(w).$$

□

We note here a property of this pruning procedure that will be useful in the proof of Proposition 8.1.6; or more precisely, Proposition 8.4.3.

Fact 8.2.3. *Fix $n \geq 2$. For any $(t, \sigma) \in \mathcal{RD}_{n-1}$ and $(k, l, x) \in \mathcal{C}_n$, write*

$$(t', \sigma') = \mathrm{rh}_n((t, \sigma), (k, l, x)) \in \mathcal{RD}_n.$$

Then $\mathrm{d}_{t'}(n) = \sum_{i=k}^{n-1} x_i$, *and for* $v \in [n-1]$,

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

8.2.2 The random process

The Robin-Hood pruning is defined, for each $(t, \sigma) \in \mathcal{RD}_n$, by

$$\mathrm{RH}_n(t, \sigma) = \mathrm{rh}_n((t, \sigma), (K, L, X));$$

where the element $(K, L, X) \in \mathcal{C}_n$ is an RH_n set of random variables defined as follows.

Definition 8.2.4. *Fix* $n \geq 1$. *Let* $K \overset{\mathcal{L}}{=} \mathrm{Unif}(1, 2, \ldots, n)$; *if* $K = 1$ *let* $L = 0$, *and if* $K > 1$ *let* $L = \mathrm{Unif}(1, 2, \ldots K-1)$. *Independently, let* $X_i = \mathrm{Bernoulli}(1/i)$ *be independent variables for* $i \in [n-1]$ *and write* $X = (X_1, \ldots, X_{n-1})$. *An* RH_n*-set is a triple of random variables with the same law as* (K, L, X).

The law of $\mathrm{RH}_n(t, \sigma)$ depends on the initial input (t, σ); however, the distribution of the RH_n-set of variables is defined so that $\mathrm{RH}_n(\mathrm{RT}_{n-1})$ preserves the uniform measure in decorated trees. To verify this claim, we require the following characterization of RT_n.

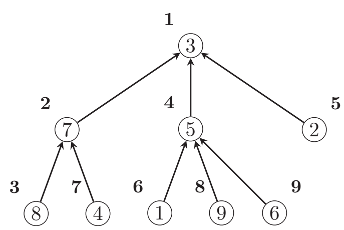
Lemma 8.2.5. *Let* $n \geq 1$ *be an integer. A random decorated tree* $(T, \boldsymbol{\sigma}) \in \mathcal{RD}_n$ *is uniformly random if and only if the following properties are satisfied.*

i) The permutation $\boldsymbol{\sigma}$ *is uniformly random on* $\mathcal{S}_n$.

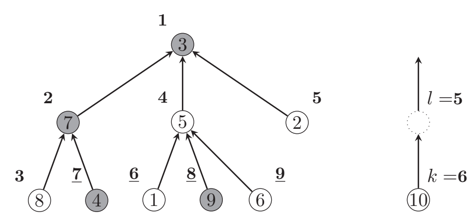
ii) The vertices

$$(p_T(v),\, v \in V(T) \setminus \{r(T)\}) = (p_{\boldsymbol{\sigma}(T)}(\boldsymbol{\sigma}^{-1}(v)),\, v \in V(T) \setminus \{r(T)\})$$

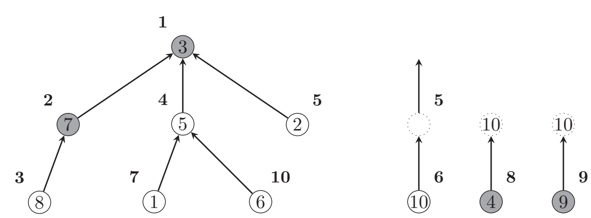
are, conditionally given $\boldsymbol{\sigma}$, *independent.*



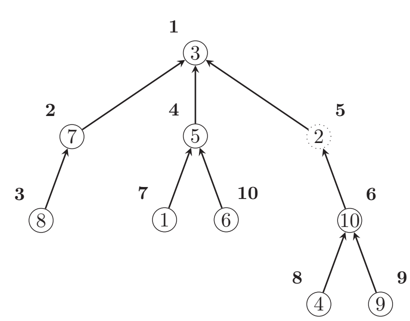
(a) A tree (t, σ) in $\mathcal{RD}_9$. The permutation σ is depicted with bold numbers.



(b) In this case, $k = 6, l = 5$ and $x_1 = x_2 = x_7 = x_8 = 1$; all other $x_i = 0$. Vertices in gray satisfy $X_{\sigma(v)} = 1$ and underlined are addition times $\sigma(i) \geq k$.



(c) Nodes i with $\sigma(i) \geq k$ and $x_i = 1$ have been pruned and addition times have been adjusted.



(d) The resulting tree $\mathrm{rh}_{10}((t, \sigma), (k, l, x)) \in \mathcal{RD}_{10}$.

Figure 8–2: An example of the Robin-Hood pruning for $n = 10$.

iii) For all vertices $v, w \in [n]$ and indices $i, j \in [n]$,

$$\mathbf{P}\left(p_T(v) = w,\, \boldsymbol{\sigma}(v) = j,\, \boldsymbol{\sigma}(w) = i\right) = \frac{1}{n(n-1)(j-1)}\mathbf{1}_{[j>i]}. \tag{8.2}$$

Proof. Let $(T, \boldsymbol{\sigma})$ be uniformly random on $\mathcal{RD}_n$. Then by Proposition 8.1.1, $\boldsymbol{\sigma}$ is a uniformly random permutation and $\boldsymbol{\sigma}(T)$ has the law of R_n. Therefore, as multisets,

$$\{p_T(v),\, v \in V(T) \setminus \{r(T)\}\} \stackrel{\mathcal{L}}{=} \{p_{\boldsymbol{\sigma}(T)}(v),\, 1 < v \leq n\};$$

and parents in RRTs are chosen independently for each of the vertices. The third condition follows immediately: For all $v, w, i, j \in [n]$, we obtain

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

Now consider a random decorated tree $(T, \boldsymbol{\sigma}) \in \mathcal{RD}_n$ satisfying conditions *i)*-*iii)*. Fix a decorated tree $(t, \pi) \in \mathcal{RD}_n$, and for $v \in V(t) \setminus \{r(t)\}$, let $w_v = p_t(v)$, then

$$\begin{aligned}
\mathbf{P}\left(p_T(v) = w_v | \boldsymbol{\sigma} = \pi\right) =& \mathbf{P}\left(p_T(v) = w_v \,|\, \boldsymbol{\sigma}(v) = \pi(v),\, \boldsymbol{\sigma}(w_v) = \pi(w_v)\right) \\
=& \frac{1}{\pi(w_v) - 1}.
\end{aligned} \tag{8.3}$$

The first equality holds by condition *ii)* and the second by both *i)* and *iii)* since $\pi(v) > \pi(w_v)$.

Now, by definition, $\pi(t) \in \mathcal{I}_n$. The increasing tree t' is determined by the set of parents $\{p_{t'}(v),\, 1 < v \leq n\}$. Therefore,

$$\begin{aligned}
\mathbf{P}\left(\boldsymbol{\sigma}(T) = \pi(t) \,|\, \boldsymbol{\sigma} = \pi\right) = \quad & \mathbf{P}\left(p_{\boldsymbol{\sigma}(T)}(v) = p_{\pi(t)}(v),\, 1 < v \leq n \,|\, \boldsymbol{\sigma} = \pi\right) \\
= \quad & \mathbf{P}\left(p_T(v) = p_t(v),\, v \in V(t) \setminus \{r(t)\} \,|\, \boldsymbol{\sigma} = \pi\right) \\
= \quad & \prod_{v \in V(t) \setminus r(T)} \mathbf{P}\left(p_T(v) = p_t(v) \,|\, \boldsymbol{\sigma} = \pi\right) \\
= \quad & [(n-1)!]^{-1}.
\end{aligned}$$

Condition $ii)$ gives the third equality; the last equality holds by (8.3) since

$$\{\pi(w_v),\, v \in V(t) \setminus \{r(t)\}\} = \{2, \ldots, n\}.$$

Finally condition $i)$ and the computations above show that, regardless of the choice of $(t, \pi) \in \mathcal{RD}_n$, we have

$$\begin{aligned}
\mathbf{P}\left((T, \boldsymbol{\sigma}) = (t, \pi)\right) &= \mathbf{P}\left(\boldsymbol{\sigma}(T) = \pi(t) \,|\, \boldsymbol{\sigma} = \pi\right) \mathbf{P}\left(\boldsymbol{\sigma} = \pi\right) \\
&= \frac{1}{n!} \mathbf{P}\left(\boldsymbol{\sigma}(T) = \pi(t) \,|\, \boldsymbol{\sigma} = \pi\right) = [n!(n-1)!]^{-1}.
\end{aligned}$$

□

We are now ready to prove Theorem 8.1.3.

Proof of Theorem 8.1.3. Let $(T, \boldsymbol{\sigma}) \in \mathcal{RD}_{n-1}$ be a uniformly random decorated tree. Let (K, L, X) be an RH_n set and let $(T', \boldsymbol{\pi}) = \mathrm{rh}((T, \boldsymbol{\sigma}), (K, L, X))$. It suffices to show that $(T', \boldsymbol{\pi})$ satisfies the properties in Lemma 8.2.5.

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

$$\begin{aligned}\{p_{T'}(v),\, v\in V(T')\setminus\{r(T')\}\} =&\{p_{T'}(v),\, 1<\boldsymbol{\pi}(v)<\boldsymbol{\pi}(n)\}\\ &\cup\{p_{T'}(v),\, \boldsymbol{\pi}(n)\le\boldsymbol{\pi}(v)\le n\}\\ =&\{p_T(v),\, v\in 1<\boldsymbol{\sigma}(v)<K\}\\ &\cup\{p_{T'}(v),\, (2\vee K)\le\boldsymbol{\pi}(v)\le n\},\end{aligned}$$

where the last two sets are conditionally independent given $\boldsymbol{\pi}$. Now, since $(T,\boldsymbol{\sigma})$ is uniformly random in $\mathcal{RD}_{n-1}$, the parents $\{p_T(v),\, v\in 1<\boldsymbol{\sigma}(v)<K\}$ are independent, conditionally given $\boldsymbol{\sigma}$ (and thus, also conditionally given $\boldsymbol{\pi}$). On the other hand, for v with $\boldsymbol{\pi}(v)\ge K$,

$$p_{T'}(v)=\begin{cases} n & \text{if } X_{\boldsymbol{\pi}(v)-1}=1,\\ p_T(v) & \text{if } X_{\boldsymbol{\pi}(v)}=0,\\ \boldsymbol{\pi}^{-1}(L) & \text{if } \boldsymbol{\pi}(v)=K.\end{cases}$$

Note that $p_{T'}(v)$ is determined independently from other vertices, thus $\{p_{T'}(v),\, K\le\boldsymbol{\pi}(v)\le n\}$ are also independent, conditionally given $\boldsymbol{\pi}$. This implies that condition $ii)$ is satisfied.

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

$$\begin{aligned}A_1&=\{p_{T'}(n)=w,\, \boldsymbol{\pi}(n)=j,\, \boldsymbol{\pi}(w)=i\},\\ A_2&=\{p_{T'}(v)=n,\, \boldsymbol{\pi}(v)=j,\, \boldsymbol{\pi}(n)=i\},\\ A_3&=\{p_{T'}(v)=w,\, \boldsymbol{\pi}(v)=j,\, \boldsymbol{\pi}(w)=i\}.\end{aligned}$$

It remains to show that the probabilities of A_1,A_2,A_3 are given by (8.2) for all $i,j\in[n]$. The event $p_{T'}(n)=w$ implies that $\boldsymbol{\sigma}(w)=L<K$. Therefore,

A_1 occurs precisely when $K = j$, $L = i$, and $\boldsymbol{\sigma}(w) = i$. Then,

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

Next, $p_{T'}(v) = n$ implies that $\boldsymbol{\sigma}(v) \geq K$ and thus $\boldsymbol{\pi}(v) = \boldsymbol{\sigma}(v) + 1$. It then follows that A_2 occurs when $K = i$, $\boldsymbol{\sigma}(v) = j - 1$, and $X_{j-1} = 1$. Therefore,

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

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

$$A_{3,k} = \{p_{T'}(v) = w,\, \boldsymbol{\pi}(v) = j,\, \boldsymbol{\pi}(w) = i,\, K = k\}.$$

In computing the probabilities $\mathbf{P}(A_{3,k})$ we use that $(T, \boldsymbol{\sigma})$ is uniformly random in RD_{n-1}. If $K > j$, then both $\boldsymbol{\sigma}(v) = \boldsymbol{\pi}(v)$ and $\boldsymbol{\sigma}(w) = \boldsymbol{\pi}(w)$; in addition, $p_{T'}(v) = w$ only if $p_T(v) = w$. Therefore, if $k > j$, then

$$\begin{aligned}
\mathbf{P}(A_{3,k}) &= \mathbf{P}(K = k)\,\mathbf{P}(p_T(v) = w,\, \boldsymbol{\sigma}(v) = j,\, \boldsymbol{\sigma}(w) = i) \\
&= \frac{1}{n(n-1)(n-2)(j-1)}.
\end{aligned}$$

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

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

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

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

Altogether, we have shown that condition *iii)* is satisfied and so the proof is complete. $\square$

8.3 Large degrees in RRTs

The aim in this section is to bound the convergence rate of the law of [2]

$$Z_m^{(n)} \stackrel{\mathcal{L}}{=} \#\{v \in [n] : \mathrm{d}_{\mathrm{RT}_n}(v) \geq m\}$$

to a suitable Poisson random variable. Our tool in this section is the Chen-Stein method as stated in Proposition 8.3.1 below. Given probability measures μ and ν, a coupling of μ and ν is a pair (X, Y) of random variables (either real or vector-valued) with $X \sim \mu$ and $Y \sim \nu$.

Let $I = (I_a,\, a \in \mathcal{A})$ be a collection of $\{0,1\}$-valued random variables. Let μ be the law of $W = \sum_{a\in\mathcal{A}} I_a$ and for $a \in \mathcal{A}$ let ν_a be the conditional law of W given that $I_a = 1$, so

$$\nu_a(B) = \mathbf{P}\left(W \in B \,|\, I_a = 1\right).$$

Proposition 8.3.1 ([40, Theorem 3.7])**.** *Let $I = (I_a,\, a \in \mathcal{A})$ be a collection of $\{0,1\}$-valued random variables. For each $a \in \mathcal{A}$ fix a coupling (W, W_a) of μ and ν_a. Then with $\lambda = \mathbf{E}[W]$, we have*

$$\mathrm{d}_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda)) \leq \min\{\lambda^{-1}, 1\} \sum_{a\in\mathcal{A}} \mathbf{E}[I_a]\, \mathbf{E}[|W - (W_a - 1)|].$$

If the variables $I = (I_a,\, a \in \mathcal{A})$ are exchangeable, then for any fixed $a \in \mathcal{A}$ and coupling (W, W_a) of μ and ν_a. Then

$$\mathrm{d}_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda)) \leq \mathbf{E}[|W - (W_a - 1)|]. \tag{8.4}$$

[2] By Fact 8.1.2, considering either R_n or RT_n in the definition of $Z_m^{(n)}$ is equivalent.

Now, for the remainder of the section, fix m and let $I = (I_v,\, v \in [n])$ have $I_v = \mathbf{1}_{[\mathrm{d}_{\mathrm{RT}_n}(v) \geq m]}$; in that case, $W = \sum_{i\in[n]} I_v = Z_m^{(n)}$. The next proposition, which states that the random variables $(I_1, \ldots, I_n)$ are 'nearly' negatively correlated, is an important input to the proof of Theorem 8.1.6.

Proposition 8.3.2. *For any $c \in (0,2)$ there exists $\alpha = \alpha(c) > 0$ such that uniformly for $m = m(n) < c \ln n$ and distinct $v, w \in [n]$,*

$$\mathbf{E}\left[I_v I_w\right] - \mathbf{E}\left[I_v\right]\mathbf{E}\left[I_w\right] \leq O(2^{-2m-\alpha \log n}).$$

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

The proof of Proposition 8.3.2 appears in Appendix A; we make precise the upper bound for α in Proposition 8.3.2 as this is crucial to Corollary 8.1.7. We note that a weaker version of Proposition 8.3.2, without explicit error bounds, was proved in [3, Proposition 4.2].

Additionally, we note in passing that the degree sequence of RRTs $(\mathrm{d}_{R_n}(v),\, v \in [n])$ is negative orthant dependent; for a definition see [34]. This fact can be proven by induction from the two-vertex case $(\mathrm{d}_{R_n}(v), \mathrm{d}_{R_n}(w))$, which, in turn, follows essentially from the negative orthant dependency of multinomial distributions, see e.g. [31, Lemma 1]. As a consequence, for all $v, w \in [n]$,

$$\mathbf{P}\left(\mathrm{d}_{R_n}(v) \geq m,\, \mathrm{d}_{R_n}(w) \geq m\right) - \mathbf{P}\left(\mathrm{d}_{R_n}(v) \geq m\right)\mathbf{P}\left(\mathrm{d}_{R_n}(w) \geq m\right) \leq 0. \quad (8.5)$$

In a slight abuse of notation let us denote by μ the law of $(I_1, \ldots, I_n)$ and denote by $\nu = \nu_n$ the conditional law of $(I_1, \ldots, I_n)$ given that $I_n = 1$. Now, let $(I, J) = ((I_v,\, v \in [n]), (J_v,\, v \in [n])$ be a coupling of μ and $\nu = \nu_n$ and write $W_n = \sum_{v\in[n]} J_v$, we get

$$\mathbf{E}\left[|W - (W_n - 1)|\right] \leq \mathbf{E}\left[I_n\right] + \sum_{v\in[n-1]} \mathbf{E}\left[|I_v - J_v|\right]. \quad (8.6)$$

To apply Proposition 8.3.1 with as tight as possible bounds, one has to analyze couplings of μ and ν. For example, suppose one can provide an explicit coupling such that $J_v - I_v \leq 0$ almost surely, for all $v \in [n-1]$. It would then follow that

$$\mathbf{E}[I_n I_v] - \mathbf{E}[I_n]\,\mathbf{E}[I_v] = -\mathbf{E}[I_n]\,\mathbf{E}[|J_v - I_v|] \leq 0;$$

which would be the corresponding inequality to (8.5). Although we do not claim the bounds in Proposition 8.3.2 are optimal, it seems that the property in (8.5) is lost when randomizing the vertex labels of R_n to obtain RT_n.

Nevertheless, Proposition 8.3.2 suggests we can provide couplings of μ and ν for which $I_v - J_v \geq 0$ for all $v \in [n-1]$ with high probability. The next proposition is the key ingredient in using Proposition 8.3.1 to prove Theorem 8.1.6.

Proposition 8.3.3. *Let $c \in (1,2)$. There is $\beta = \beta(c) > 0$ such that for any $m = m(n) > c \ln n$ there exists a coupling $(I, J) = ((I_1, \ldots, I_n), (J_1, \ldots, J_n))$ of μ and ν, in which for all $v \in [n-1]$,*

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

The coupling of Proposition 8.3.3 is based on the Robin-Hood pruning and its proof is the content of Section 8.4. Next, we provide the proofs of Theorem 8.1.6 (assuming Proposition 8.3.3), followed by the proofs of Corollary 8.1.7 and Theorem 8.1.8.

Proof of Theorem 8.1.6 assuming Proposition 8.3.3

Fix $1 < c < c' < 2$ and let $c \ln n < m = m(n) < c' \ln n$. We apply the Chen-Stein method to $Z_m^{(n)} \stackrel{\mathcal{L}}{=} \sum_{v \in [n]} I_v$. First, we use the coupling $(I, J) = ((I_1, \ldots, I_n), (J_1, \ldots, J_n))$ of μ and ν given in Proposition 8.3.3. By (8.4) and

(8.6), we have

$$\mathrm{d_{TV}}\left(Z_m^{(n)}, \mathrm{Poi}(\mathbf{E}\left[\lambda_{n,m}\right])\right) \leq \mathbf{E}\left[I_n\right] + \sum_{v\in[n-1]} \mathbf{E}\left[|I_v - J_v|\right].$$

It thus remains to show that the terms in the bound above are $O(2^{-m+(1-\alpha)\log n})+ O(n^{-\beta})$, where $\alpha = \alpha(c') \in (0,1)$ and $\beta = \beta(c) > 0$ are defined as in Propositions 8.3.2 and 8.3.3. First, by Lemma 8.1.5 and the fact that $\alpha < 1$ gives

$$\mathbf{E}\left[I_n\right] = 2^{-m}(1+o(1)) = O(2^{-m+(1-\alpha)\log n}).$$

Second, for any $v \in [n-1]$,

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

The terms in the last line are bounded by Proposition 8.3.2 and Proposition 8.3.3, respectively; giving

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

In the last line we also use that Lemma 8.1.5 implies that $\mathbf{E}\left[I_n\right]^{-1} = O(2^m)$. □

Proof of Corollary 8.1.7

Fix $c' \in (1, \log e)$ and let $\alpha = \alpha(c')$ be as in Theorem 8.1.6. Simple computations using the upper bound in Proposition 8.3.2 for α show that $(1-\alpha)\log e < c'$. Thus, we can chose $c \in ((1-\alpha)\log e, c')$.

Let $m = m(n)$ be such that $c \ln n < m < c' \ln n$. By the choice of c and c', we have that, as $n \to \infty$, $(1-\alpha)\log n - m < 0$; while by Lemma 8.1.5,

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

The result then follows by Theorem 8.1.6 and the central limit theorem of Poisson variables, see e.g. [35, Exercise 3.4.4].

□

Proof of Theorem 8.1.8

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 $Z \stackrel{\mathcal{L}}{=} \mathrm{Poi}(\lambda_{m,n})$.

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

$$\mathbf{P}\left(\Delta_n < \lfloor \log n \rfloor - i\right) = \mathbf{P}\left(Z_m^{(n)} = 0\right) \leq \mathbf{P}\left(Z = 0\right) + \mathrm{d_{TV}}(Z_m^{(n)}, Z). \tag{8.7}$$

We deal with the two terms on the right-hand side of (8.7) 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, Lemma 8.1.5 gives $\gamma > 0$ such that $\lambda_{n,m} = 2^{i+\varepsilon_n} + o(n^{-\gamma} \ln n)$. Consequently,

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

For the second term in (8.7), Theorem 8.1.6 gives $\alpha, \beta > 0$ such that

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

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

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

similarly, for C large enough,

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

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

□

8.4 Proof of Proposition 8.3.3: An auxiliary coupling

In this section we fix $n \in \mathbb{N}$, $c \in (1,2)$ and $m = m(n) > c\ln n$. Let $(T, \boldsymbol{\sigma})$ be uniformly random in $\mathcal{RD}_{n-1}$. Consider (K, L, X) an RH_n-set and let (K', L', X') be distributed as an RH_n-set conditioned to satisfy $\sum_{i=K}^{n-1} X'_i \geq m$. Now, write

$$(T', \boldsymbol{\pi}) = \mathrm{rh}((T, \boldsymbol{\sigma}), (K, L, X)) = \mathrm{RH}(T, \boldsymbol{\sigma}), \tag{8.8}$$

$$(T'_m, \boldsymbol{\pi}) = \mathrm{rh}((T, \boldsymbol{\sigma}), (K', L', X')). \tag{8.9}$$

By Fact 8.2.3, $(T'_m, \boldsymbol{\pi})$ is a conditional version of $\mathrm{RH}_n(T, \boldsymbol{\sigma})$ given that $\mathrm{d}_{\mathrm{RH}_n(T,\boldsymbol{\sigma})}(n) \geq m$. Consequently, any coupling of (K, L, X) and (K', L', X') yields a coupling of the measures μ and ν, defined in Section 8.3, by setting $I_v = \mathbf{1}_{[\mathrm{d}_{T'}(v) \geq m]}$ and $J_v = \mathbf{1}_{[\mathrm{d}_{T'_m}(v) \geq m]}$ for all $v \in [n]$. With this notation, our goal is to couple (K, L, X) and (K', L', X') in such a way that for some $\beta = \beta(c) > 0$,

$$\mathbf{P}\left(\mathrm{d}_{T'}(v) < m \leq \mathrm{d}_{T'_m}(v)\right) = O(n^{-1-\beta}). \tag{8.10}$$

We start with some straightforward lemmas. For any integer $n - m \leq k < n$, let $X^k = (X_i^k, i \in [n-1])$ be a conditional version of X given that $\sum_{i=k}^{n-1} X_i \geq m$. The following observation is quite standard but we include a proof for completeness. For $a = (a_1, \ldots, a_d)$ and $b = (b_1, \ldots, b_d) \in \{0,1\}^d$, $a \leq b$ 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 $a \in S$ implies $b \in S$.

Lemma 8.4.1. *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. By Strassen's theorem [57], it suffices to prove that X^k dominates stochastically X. That is, for all monotone subsets $S \in \{0,1\}^{n-1}$,

$$\mathbf{P}\left(X^k \in S\right) \geq \mathbf{P}\left(X \in S\right). \tag{8.11}$$

Note that $S_k = \{a \in \{0,1\}^{n-1} : a_1 + \ldots + a_k \geq m\}$ is a monotone subset of $\{0,1\}^{n-1}$. By Harris inequality, for any monotone subset $S \in \{0,1\}^{n-1}$,

$$\mathbf{P}\left(X \in S \cap S_k\right) \geq \mathbf{P}\left(X \in S_k\right) \mathbf{P}\left(X \in S\right).$$

Dividing the above inequality by $\mathbf{P}\left(X \in S_k\right)$ yields (8.11); thus, completing the proof. □

Lemma 8.4.2. *There exists a coupling of (K, L) and (K', L') such that $K' \leq K$ and $L' \leq L$.*

Proof. Let U_1, U_2 have uniform distributions on $[0,1]$. Set $K = \lceil nU_1 \rceil$, and $L = \lceil (K-1)U_2 \rceil$. Independently of U_1 and U_2, let $X = (X_1, \ldots, X_{n-1})$ be independent with $X_i \stackrel{\mathcal{L}}{=} \text{Bernoulli}(1/i)$. Then (K, L, X) is an RH_n-set.

Next, for each $k \in [n]$, let $p_k = \mathbf{P}\left(K = k \,|\, \sum_{i=k}^{n-1} X_i \geq m\right)$ and set

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

and $L' = \lceil (K'-1)U_2 \rceil$.

The random variable K' has the correct law by construction. Moreover, conditionally given $K' = k$,

$$L' \stackrel{\mathcal{L}}{=} \begin{cases} \text{Unif}(k-1) & \text{if } k > 1, \\ 0 & \text{if } k = 1; \end{cases}$$

note that the distribution of L conditionally given $K = k$ has the same expression. Therefore, for all $l \leq k - 1$ we have

$$\begin{aligned}
\mathbf{P}\left(K' = k, L' = l\right) &= \mathbf{P}\left(L' = l | K' = k\right)\mathbf{P}\left(K' = k\right) \\
&= \frac{\mathbf{P}\left(L = l,\ K = k\right)\mathbf{P}\left(K = k,\ \sum_{i=k}^{n-1} X_i \geq m\right)}{\mathbf{P}\left(K = k\right)\mathbf{P}\left(\sum_{i=k}^{n-1} X_i \geq m\right)} \\
&= \frac{\mathbf{P}\left(K = k,\ L = l,\ \sum_{i=K}^{n-1} X_i \geq m\right)}{\mathbf{P}\left(\sum_{i=K}^{n-1} X_i \geq m\right)} \\
&= \mathbf{P}\left(K = k,\ L = l \,\middle|\, \sum_{i=K}^{n-1} X_i \geq m\right);
\end{aligned}$$

the third equality holds by the independence between X and (K, L). It follows that (K', L') has the correct law.

Finally, since X is independent of K, for each $k \in [n]$,

$$\mathbf{P}\left(K' = k\right) = \frac{\mathbf{P}\left(K = k, \sum_{i=k}^{n-1} X_i \geq m\right)}{\mathbf{P}\left(\sum_{i=K}^{n-1} X_i \geq m\right)} = \left[n\mathbf{P}\left(\sum_{i=K}^{n-1} X_i \geq m\right)\right]^{-1}\mathbf{P}\left(\sum_{i=k}^{n-1} X_i \geq m\right).$$

This chain of equalities show that p_k is proportional to $\mathbf{P}\left(\sum_{i=k}^{n-1} X_i \geq m\right)$, which is decreasing in k. Consequently, if $K' = j$ then

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

in other words, $K \geq j = K'$. In turn, $L = \lceil (K-1)U_2 \rceil \geq \lceil (K'-1)U_2 \rceil = L'$. $\square$

Proposition 8.4.3. *There exists a coupling of (K, L, X) and (K', L', X') such that $K \geq K'$, $L \geq L'$ and $X_i \leq X_i'$ for all $i \in [n-1]$.*

Proof. First, couple $((K, L), (K', L'))$ as in Lemma 8.4.2 and also let $X = (X_1, \ldots X_{n-1})$ be as in the proof of that lemma. For each $1 \leq k < n$ fix a vector X^k coupled with X according to Lemma 8.4.1.

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^i are independent of (K', L'). Since (K', L') are determined by uniform variables U_1, U_2 independent of X, the existence of such joint coupling is straightforward.

Next, for each $i \in [n-1]$ write $X'_i = X_i^{K'}$ and let $X' = (X'_1, \dots, X'_{n-1})$. Now it remains to show that, (K', L', X') has the correct law.

For any $(k, l, x) \in \mathcal{C}_n$ and $j \le n-1$, we use the independence of X^j from (K', L') to obtain,

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

That the coupling $((K, L, X), (K', L', X'))$ has the desired properties follows from Lemmas 8.4.1 and 8.4.2. □

Under the coupling of Proposition 8.4.3 we obtain necessary conditions for $\mathrm{d}_{T'}(v) < m \le \mathrm{d}_{T'_m}(v)$ to hold.

Lemma 8.4.4. *Using variables as in Proposition 8.4.3 and the trees defined in (8.8) and (8.9). For any $v \in [n-1]$,*

$$\{\mathrm{d}_{T'}(v) < m \le \mathrm{d}_{T'_m}(v)\} \subset \{L' = \boldsymbol{\sigma}(v)\} \cap \{\mathrm{d}_T(v) \ge m-1\}.$$

Proof. From the properties of the coupling in Proposition 8.4.3,

$$\sum_{i=K}^{n-1} X_i \, \mathbf{1}_{[v = p_T(\boldsymbol{\sigma}^{-1}(i))]} \le \sum_{i=K'}^{n-1} X'_i \, \mathbf{1}_{[v = p_T(\boldsymbol{\sigma}^{-1}(i))]}. \tag{8.12}$$

Consequently, using Fact 8.2.3 we have that $\mathrm{d}_{T'_m}(v) - \mathrm{d}_{T'}(v) \le \mathbf{1}_{[L'=\boldsymbol{\sigma}(v)]}$. On the other hand, if $\{\mathrm{d}_{T'}(v) < m \le \mathrm{d}_{T'_m}(v)\}$ holds, then it follows that $\mathrm{d}_{T'_m}(v) - \mathrm{d}_{T'}(v) > 0$ and so it is necessary that $\{L' = \boldsymbol{\sigma}(v)\}$ holds.

Finally, $\{\mathrm{d}_{T'}(v) < m \le \mathrm{d}_{T'_m}(v)\}$ implies that

$$m \le \mathrm{d}_{T'_m}(v) = \mathrm{d}_T(v) + \mathbf{1}_{[L'=\boldsymbol{\sigma}(v)]} - \sum_{i=K'}^{n-1} X'_i \mathbf{1}_{[v=p_T(\boldsymbol{\sigma}^{-1}(i))]} \le \mathrm{d}_T(v) + 1;$$

or equivalently, that $\{\mathrm{d}_T(v) \ge m - 1\}$. □

The next tail bounds the degree of vertices in RRTs are obtained using standard estimates for binomial variables.

Lemma 8.4.5. *Fix $c > 1$. There exists $\beta = \beta(c) > 0$ such that uniformly over $m > c \ln n$, and for each $i \in [n]$,*

$$\mathbf{P}\left(\mathrm{d}_{R_n}(i) > m\right) = O(n^{-\beta}).$$

Proof. Let $(B_k,\, k \ge 1)$ be independent Bernoulli variables with mean $1/k$ respectively. By the construction of R_n we have that $\mathrm{d}_{R_n}(i) \overset{\mathcal{L}}{=} \sum_{k=i}^{n} B_k \le \sum_{k=1}^{n} B_k$. Therefore,

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

We use the following version of Bernstein inequalities (see, e.g. [52] Theorem 2.8, (2.5)). For a sum S of $\{0,1\}$-valued variables and $\varepsilon > 0$,

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

Since $\mathbf{E}\left[\sum_{k=1}^{n} B_k\right] = \ln n + O(1) < c \ln n$, we can use the above inequality with $\varepsilon = c - 1 + o(1)$ and set $\beta = \frac{3\varepsilon^2}{2(3+\varepsilon)}$. □

Proof of Proposition 8.3.3. Fix $c \in (1,2)$. Let $m = m(n) > c \ln n$ and $\beta = \beta(c) > 0$ be as in Lemma 8.4.5. Let us use (T', T'_m) as defined in (8.8) and

(8.9) with $((K, L, X), (K', L', X'))$ coupled as in Proposition 8.4.3. Set $I_v = \mathbf{1}_{[\mathrm{d}_{T'}(v) \geq m]}$ and $J_v = \mathbf{1}_{[\mathrm{d}_{T'_m}(v) \geq m]}$ for all $v \in [n]$. We now show that $(I, J) = ((I_1, \dots, I_n), (J_1, \dots J_n))$ is a coupling of the measures μ and ν which satisfies

$$\mathbf{P}(I_n < J_n) = \mathbf{P}\left(\mathrm{d}_{T'}(v) < m \leq \mathrm{d}_{T'_m}(v)\right) = O(n^{-1-\beta}).$$

First, using Lemma 8.4.4, we have

$$\mathbf{P}\left(\mathrm{d}_{T'}(v) < m \leq \mathrm{d}_{T'_m}(v)\right) \leq \sum_{j=1}^{n-1} \mathbf{P}\left(\boldsymbol{\sigma}(v) = j,\, L' = j,\, \mathrm{d}_T(v) \geq m - 1\right).$$

On the other hand, $\boldsymbol{\sigma}$ is uniformly random in $\mathcal{S}_{n-1}$ and L' is indpendent of $\boldsymbol{\sigma}$. Therefore, uniformly for each $j \in [n-1]$,

$$\begin{aligned}
\mathbf{P}\left(\boldsymbol{\sigma}(v) = j,\, L' = j,\, \mathrm{d}_T(v) \geq m - 1\right) &= \frac{1}{n-1} \mathbf{P}\left(\mathrm{d}_T(v) \geq m - 1 \,|\, \boldsymbol{\sigma}(v) = j\right) \mathbf{P}\left(L' = j\right) \\
&= \frac{1}{n-1} \mathbf{P}\left(\mathrm{d}_{R_{n-1}}(j) \geq m - 1\right) \mathbf{P}\left(L' = j\right) \\
&\leq \mathbf{P}\left(L' = j\right) O(n^{-1-\beta});
\end{aligned}$$

the second inequality, since $\boldsymbol{\sigma}(T) \stackrel{\mathcal{L}}{=} R_{n-1}$ and the last one by Lemma 8.4.5. Therefore, we get for all $v \in [n-1]$,

$$\begin{aligned}
\mathbf{P}\left(I_v < J_v\right) &\leq \sum_{j=1}^{n-1} \mathbf{P}\left(\boldsymbol{\sigma}(v) = j,\, L' = j,\, \mathrm{d}_T(v) \geq m - 1\right) \\
&= O(n^{-1-\beta}) \sum_{j=1}^{n-1} \mathbf{P}\left(L' = j\right) = O(n^{-1-\beta}).
\end{aligned}$$

□

8.5 Coalescents as recursively decorated trees

Coalescent processes are essentially defined as Na and Rapoport described 'static' trees. We will first explain the definition of coalescents using chains of forests and decorated trees. Following, we define Kingman's coalescent in terms of such chains and briefly note its connection with increasing binary trees.

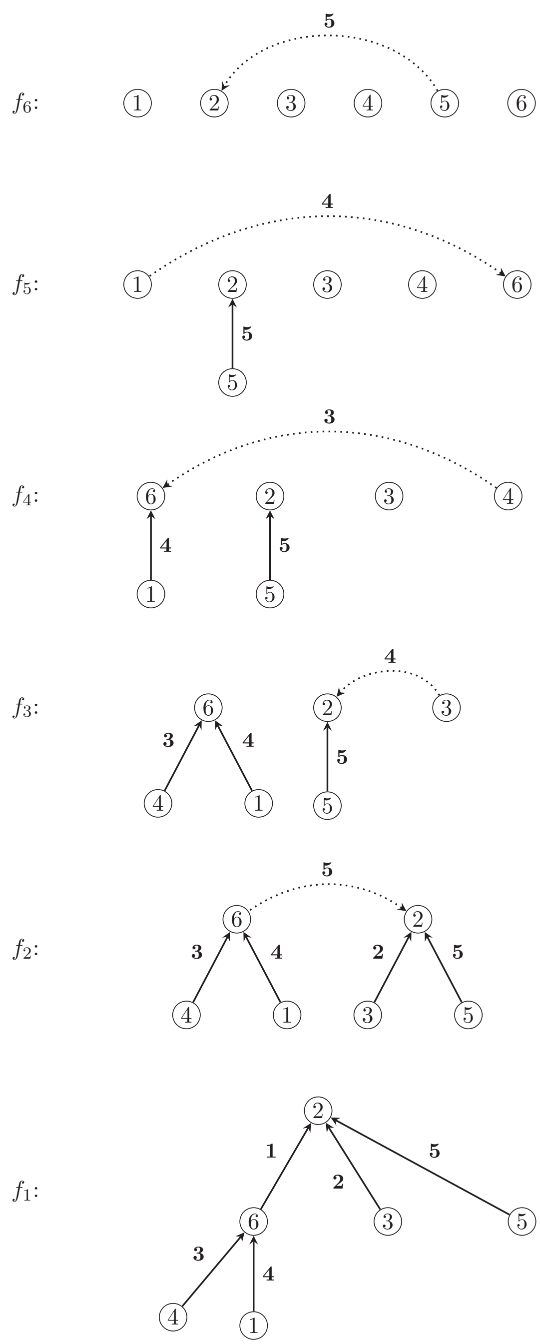


Figure 8–3: An example of an n-chain with $n = 6$. The edge labelling ρ_n is presented with numbers in bold.

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 $C = (f_n, \ldots, f_1)$ 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 8–3.

The relation of n-chains with coalescents is the following. For an n-chain $(f_n, \ldots, f_1)$, 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.

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 $C = (f_n, \ldots, f_1)$, for each $e \in E(t_C)$, let

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

We next define a vertex labeling $\sigma_C : V(t_C) \to [n]$. For each $uv \in E(t_C)$, let

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

and let $\sigma_C(r(t_C)) = 1$. The pair $(t_C, \sigma_C) \in \mathcal{RD}_n$ contains all the information to recover the original n-chain C.

Proposition 8.5.1. *Let $\mathcal{CF}_n$ be the set of n-chains and $\Upsilon : \mathcal{CF}_n \to \mathcal{RD}_n$ be defined as follows. For an n-chain $C = (f_n, \ldots, f_1)$, let $\Upsilon(C) = (t_C, \sigma_C)$. Then Υ is a bijection.*

Proof. First, we show that $\mathcal{CF}_n$ and $\mathcal{RD}_n$ have the same cardinality and that Υ is well defined.

To count the number of n-chains, consider constructing $(f_n, \ldots, f_1)$ by deciding which edge to add from f_k to f_{k-1}. Since there are k trees in f_k, when we have chosen $(f_1, \ldots, f_k)$, there are $k(k-1)$ possible directed edges to add. Therefore, $|\mathcal{CF}_n| = n!(n-1)!$.

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

Now we show that Υ is injective and, thus it is a bijection between $\mathcal{CF}_n$ and $\mathcal{RD}_n$. Consider two distinct n-chains $C = (f_n, \ldots, f_1)$ and $C' = (f'_n, \ldots, f'_1)$, then $k = \min\{i : f_i \neq f'_i\}$ is well defined. If $k = 1$ then $t_C \neq t_{C'}$ and clearly, $\Upsilon(C) \neq \Upsilon(C')$. Otherwise, the edges $e \in E(f_{k-1}) \setminus E(f_k)$ and $e' \in E(f'_{k-1}) \setminus E(f'_k)$ are distinct. However $t_C = t_{C'}$, it thus follows that $e = uv \in f'_k$ and so $\sigma_C(u) = k > \sigma_{C'}(u)$. This shows that Υ is injective, completing the proof. $\square$

Kingman's coalescent is characterized by the property that the merging probability of any pair of components is independent of the components' sizes. The following definition describes Kingman's n-coalescent as a random n-chain $\mathbf{C} = (F_n, \ldots, F_1)$. This construction has been previously exploited to study high-degree vertices in RRTs [3] and is closely related to the 'union-find' algorithm used in computer science (see e.g. [78]).

For an n-chain $(f_n, \ldots, f_1)$ and $1 \le i \le n$, list the trees of f_i in increasing order of their smallest-labeled vertex as $t^{(i)}_1, \ldots, t^{(i)}_i$. Independently for each $1 < i \le n$ let $\{a_i, b_i\} \subset \{\{a, b\} : 1 \le a < b \le i\}$ be uniformly chosen at random; in addition, let ξ_i be independent Bernoulli random variables with mean $1/2$.

Definition 8.5.2. *Kingman's n-coalescent is defined as $\mathbf{C} = (F_n, \ldots, F_1)$ constructed as follows. For $1 < i \leq n$, F_{i-1} is obtained from F_i by adding an edge between $r(T^{(i)}_{a_i})$ and $r(T^{(i)}_{b_i})$. If $\xi_i = 1$ then direct the edge towards $r(T^{(i)}_{a_i})$; otherwise direct it towards $r(T^{(i)}_{b_i})$. The forest F_{i-1} consists of the new tree and the remaining $i-2$ unaltered trees from F_i.*

Proposition 8.5.3. *Let $\mathbf{C} = (F_n, \ldots, F_1)$ be a Kingman's coalescent and let $\mathrm{RT}_n \in \mathcal{RD}_n$ be uniformly random. Then $T_{\mathbf{C}} \stackrel{\mathcal{L}}{=} \mathrm{RT}_n$.*

Proof. Let $\mathbf{C} = (F_n, \ldots, F_1)$ be a Kingsman's coalescent. Then for any fixed $(f_n, \ldots, f_1) \in \mathcal{CF}_n$,

$$\mathbf{P}\left((F_n, \ldots, F_1) = (f_n, \ldots, f_1)\right) = \prod_{k=1}^{n-1} \mathbf{P}\left(F_k = f_k | (F_n, \ldots, F_{k+1}) = (f_n, \ldots, f_{k+1})\right).$$

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 $(f_n, \ldots, f_1)$,

$$\mathbf{P}\left((F_n, \ldots, F_1) = (f_n, \ldots, f_1)\right) = [(n-1)!n!]^{-1}.$$

By Proposition 8.5.1, $T_{\mathbf{C}} \in \mathcal{RD}_n$ and it has a uniform distribution, since the bijection preserves the uniform measure of $\mathbf{C}$. □

Kingman's n-coalescent is usually represented by extended binary trees with n external vertices and an increasing labeling on the $n-1$ internal vertices. The role of internal vertices is as follows. For each $k \in [n-1]$, consider the two sets of leaves in the subtrees of internal vertex labeled k; these two sets are merged at the $(n-k)$-th step of the coalescent. The labeling of the external vertices represent the same elements as the vertex set of t_C. The correspondence between recursive trees and increasing binary trees is mentioned, e.g., in [37, exercise II.33].

8.6 Futher research on the tree growth process

The Robin-Hood pruning yields an interesting process $((T_n, \boldsymbol{\sigma}_n),\, n \geq 1)$ which has connections to mathematical models of social and economic networks and raises challenging theoretical questions.

By Theorem 8.1.3 and Proposition 8.1.1, $\boldsymbol{\sigma}(T_n) \overset{\mathcal{L}}{=} R_n$ for all $n \geq 1$. The novelty of this process is that T_n is not necessarily obtained from T_{n-1} by a simple addition of a new edge and vertex. Rather, the Robin-Hood pruning is a fairly complex dynamic of trees. About half of the time the newly added vertex will simply attach to a uniformly random vertex, as in the standard construction of RRTs. While from time to time, a large proportion of edges will be rewired towards the newly added vertex, drastically reshaping the structure of the tree.

Note, for example, by Fact 8.2.3,

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

while, for any $a \in [0,1)$,

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

In the context of random networks, the Robin-Hood pruning has an interpretation in terms of 'trends'. That is, a new vertex brings in a new idea to the network and that rewires the interests or connections of established individuals in the network. The decoration $\boldsymbol{\sigma}_n$ 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 and, although there exists a bijection between these binary trees and n-chains, it is not clear how the Robin-Hood pruning process would have a significant interpretation in terms of the genealogical information.

Regardless of the perspective we use to motivate the process $((T_n, \boldsymbol{\sigma}_n),\, n \geq 1)$, 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 in both $(R_n,\, n \geq 1)$ and $(T_n,\, n \geq 1)$.
   - How often does the maximum degree change?
   - Is this the same in both processes?

Appendix A: Proof of Proposition 8.3.2

The proof mimics that of [3, Proposition 4.2], but requires little more care as we wish to obtain explicit error bounds. By Proposition 8.5.3 we can work with the tree $T^{(n)}$ constructed from Kingman's coalescent in Section 8.5.

Recall that Kingman's coalescent consists of a chain $\mathbf{C} = (F_n, \ldots, F_1)$ and that $T^{(n)}$ is the unique tree contained in F_1. 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

$$\mathcal{S}_n(v) = \{2 \leq j \leq n : T_j(v) \in \{T^{(j)}_{a_j}, T^{(j)}_{b_j}\}\};$$

this set keeps record of the times when the tree containing v merges. Finally, for each $2 \le j \le n$, we say that ξ_j is *favourable* 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 ξ_j favours v, then $r(T_j(v))$ increases its degree by one in the process; otherwise $r(T_j(v))$ attaches to the root of the other merging tree and the degree of $r(T_j(v))$ remains unchanged for the rest of the process. Since all vertices start the process as roots, $\mathrm{d}_{T^{(n)}}(v)$ is equal to the length of the first streak of favourable times for v. Moreover, $(\xi_j,\, j \in [n-1])$ are independent and distributed as Bernoulli(1/2). Therefore we have the following distributional equivalence.

Fact 8.6.1. *Let D be a random variable with distribution* Geo(1/2) *independent of $\mathcal{S}_n(v)$, then*

$$\mathrm{d}_{T^{(n)}}(v) \stackrel{\mathcal{L}}{=} \min\{D, |\mathcal{S}_n(v)|\}.$$

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

Lemma 8.6.2. *If $c \in (0,2)$ and $0 < \varepsilon \le 1 - c/2$. Writing $a = 1 - \varepsilon - c/2$, we have*

$$\mathbf{P}\left(|\mathcal{S}_n(v) \setminus [n^a]| > c \ln n\right) \le O(1) n^{-\varepsilon^2/(\varepsilon + c/2)}.$$

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 $\mathbf{P}\left(j \in \mathcal{S}_n(v)\right) = 2/j$. Since the merging trees are chosen independently at each time, we have that for any $a \in [0,1)$ we have

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

where the variables $B_1, \ldots B_n$ are independent Bernoulli variables with $\mathbf{E}\left[B_i\right] = 2/i$, respectively. The desired bound is then a straightforward application of

Bernstein's inequalities (see, e.g. [52], Theorem 2.8 and (2.6)). For a sum S of $\{0,1\}$-valued variables, we have $\mathbf{P}\left(S \leq \mathbf{E}\left[S\right] - t\right) \leq \exp\{-t^2/2\mathbf{E}\left[S\right]\}$. In this case, $S = \sum_{i=n^a}^{n} B_i$ and

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

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

Proposition 8.6.3. *If $c \in (0,2)$ and $m < c\ln n$, then for $\varepsilon = (2-c)^2/4$,*

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

Proof of Proposition 8.6.3. It follows from Lemma 8.6.1 that

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

The upper bound on $\mathbf{P}\left(\mathrm{d}_{T^{(n)}}(1) \geq m\right)$ is then trivial, while the lower bound follows by Lemma 8.6.2 using $\varepsilon = 1 - c/2$ and that $\mathcal{S}_n(v) = \mathcal{S}_n(v) \setminus [1]$. $\square$

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

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

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

$$\mathbf{P}\left(\mathrm{d}_{T^{(n)}}(v) \geq m,\, \mathrm{d}_{T^{(n)}}(w) \geq m\right) \geq 2^{-2m}\mathbf{P}\left((\mathcal{S}_n(v), \mathcal{S}_n(w)) \in \mathcal{G}_m\right). \tag{8.13}$$

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

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

Lemma 8.6.4. *For $a \in (0,1)$, $\mathbf{P}(\tau > n^a) \leq 4n^{-a}$.*

Proof. Vertices in $T^{(n)}$ are exchangeable, so we can take $v = 1, w = 2$; these vertices belong to distinct trees in F_j for all $j \geq \tau$. 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 $2 < k \leq n$,

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

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

$$\frac{\mathbf{P}(\tau \leq k)}{\mathbf{P}(\tau \leq k+1)} = \mathbf{P}(\tau \leq k | \tau \leq k+1) = \mathbf{P}(\{a_{k+1}, b_{k+1}\} \neq \{1,2\}) = 1 - \frac{2}{(k+1)k}.$$

Next, for k larger enough,

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

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

Lemma 8.6.5. *If $c \in (0,2)$ and $m < c \ln n$, then for any $\gamma < \frac{1}{4}(1 - c + \sqrt{1 + 2c - c^2})$,*

$$\mathbf{P}((\mathcal{S}_n(v), \mathcal{S}_n(w)) \in \mathcal{G}_m) \geq 1 - o(n^{-\gamma}).$$

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

$$\mathbf{P}((\mathcal{S}_n(v), \mathcal{S}_n(w)) \notin \mathcal{G}_m) \leq \mathbf{P}(\tau > n^a) + 2\mathbf{P}(|\mathcal{S}_n(v) \setminus [n^a]| < c \ln n). \quad (8.14)$$

Before, establishing (8.14), we note that the terms in the right-hand side of (8.14) are bounded by Lemmas 8.6.4 and 8.6.2, respectively. Since such bounds depend on the choice of ε, we can use

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

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 $1-\varepsilon-c/2 = \varepsilon^2/(\varepsilon+\frac{c}{2})$.

We now proceed to establish equation (8.14). At step τ, exactly one of v and w is favoured by ξ_τ. Thus, at least one of v or w gets its degree fixed for the remainder of the process. Therefore,

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

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

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

from which (8.14) follows. □

Proof of Proposition 8.3.2. Fix $c \in (0,2)$, $m = m(n) < c\ln n$ and let I_v, J_v be defined as in Proposition 8.3.2. By Proposition 8.5.3, if follows that $\mathbf{E}[I_v] = \mathbf{P}\left(\mathrm{d}_{T^{(n)}}(v) \geq m\right)$ and

$$\begin{aligned}
\mathbf{E}[I_v]\,\mathbf{E}[J_v] = \mathbf{E}[I_v I_n] &= \mathbf{P}\left(\mathrm{d}_{T^{(n)}}(v) \geq m,\, \mathrm{d}_{T^{(n)}}(n) \geq m\right) \\
&= 2^{-2m}\mathbf{P}\left((\mathcal{S}_n(v), \mathcal{S}_n(n)) \in \mathcal{G}_m\right);
\end{aligned}$$

the last equality by (8.13). Lemmas 8.6.5 and 8.6.3 then gives that for $\alpha < \frac{1}{4}(1 - c + \sqrt{1 + 2c - c^2})$,

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

$\square$

Part III

Conclusions and further research

CHAPTER 9
Application to network connectivity

In this section we present the problem of cutting down trees, also known as tree destruction: given a rooted tree, find the minimum number of uniformly random edge removals required to disconnect the root from the rest of the graph. More formally, cutting down a rooted tree t is performed as follows. Start with $t_0 = t$, select a uniformly random edge in $E(t)$ and deleted it from t. This yields a graph with two components, let t_1 be the tree component that contains the root. Sequentially, for $i \geq 1$, let t_i be the tree component containing the root, after deleting a uniformly chosen random edge in t_{i-1}, until the remaining tree consists only of the root.

Cutting processes on Cayley and recursive trees were first studied by Meir and Moon in the 1970s [63, 65], and a deeper study for recursive trees emerged in the early 2000s [69]. Surprisingly, the process of cutting recursive trees is associated with the Bolthausen-Sznitman coalescent in 2005 [41]. Since then, destruction of a wide range of random trees has been studied; such as, recursive trees [47, 33], deterministic trees [51], 'very simple' trees [70], Galton-Watson trees [2], binary search trees and split trees in general [46, 45]. For a general class of random trees, [12] describes the tree destuction process using an auxiliary 'cut-tree'. More recently, the same process has been analysed through the lens of percolation, and additionally, sizes of clusters have been studied [11, 54, 8].

For a recursive tree T_n, let J_n be the number of edges that we need to remove to disconnect the root in the cutting process described above. Meir and

Moon were the first to analyze the mean of J_n, showing that $\mathbf{E}[J_n] \sim n/\ln n$ [65]; Panholzer obtained a law of large numbers.

Proposition 9.0.1 ([69]). *In probability, $\frac{\ln n}{n} J_n \to 1$.*

In [69], Panholzer notes that, using the method of moments, it is not possible to obtain a limiting distribution of a centered, scaled version of J_n. Instead, Drmota, Iksanov, Moehle and Roesler proved the following convergence in distribution.

Theorem 9.0.2 ([33]). *The sequence*

$$Y_n = \frac{(\ln n)^2}{n} J_n - \ln n - \ln \ln n,$$

converges weakly to a stable random variable Y with characteristic function

$$\varphi_Y(\lambda) = \exp\{i\lambda \ln|\lambda| - \pi|\lambda|/2\}.$$

The proof of Theorem 9.0.2 in [33] uses an analytic combinatorics approach, and [47] gives a probabilistic proof. Finally, Holmgren generalizes this result to split trees, and in particular, to linear PA trees [45].

9.1 Targeted cutting down

A natural extension of the destruction of trees to targeted cuttings would be the following. For a rooted tree t on n vertices, list vertices $(w_i, \in [n])$ such that $\mathrm{d}_t(w_1) \geq \cdots \geq \mathrm{d}_t(w_n)$, and break ties among same-degree vertices uniformly at random. The targeted cutting down process is performed as follows. Sequentially remove vertices $w_1, w_2, \ldots$, each time keeping only the tree containing the root. Continue such pruning until the root is selected to be removed. At this point, we say that the tree has been destroyed.

For a recursive trees T_n, let $\mathcal{K}_n$ be the number of vertices we need to remove in order to destroy T_n using targeted cuttings and let $D = D^{(n)} = \mathrm{d}_{T_n}(1)$ be the degree of the root in T_n. Then, $\mathcal{K}_n$ is upper-bounded by

$$Z^{(n)}_{\geq D} := \#\{v \in [n],\, \mathrm{d}_{T_n}(v) \geq \mathrm{d}_{T_n}(1)\}.$$

The methodology developed in this manuscript allows us to obtain concentration of $Z^{(n)}_{\geq D}$ around $n^{1-\ln 2}$, up to the exponent term, and convergence of its mean value.

Proposition 9.1.1. *Let $\gamma = 1 - \ln 2$. Then, in probability*

$$\frac{\ln Z^{(n)}_{\geq D}}{\ln n} \to \gamma.$$

Proof. To simplify notation, let $Y_a = Z^{(n)}_{\geq(\ln 2 + a)\log n}$ for any constant $a \geq -\ln 2$. Note that if $\left|\frac{D}{\ln n} - 1\right| \leq \varepsilon/2\ln 2$, then $Y_{\varepsilon/2} \leq Z^{(n)}_{D} \leq Y_{-\varepsilon/2}$. Since D is concentrated around $\ln n$ (see (4.1) and the remarks afterwards), it follows that the proof is complete if we show that, for all $\varepsilon > 0$ sufficiently small,

$$\begin{aligned}
&\mathbf{P}\left(Z^{(n)}_{\geq D} \notin (n^{\gamma-\varepsilon}, n^{\gamma+\varepsilon}),\, \tfrac{D}{\ln n} \in (1 - \tfrac{\varepsilon}{2\ln 2}, 1 + \tfrac{\varepsilon}{2\ln 2})\right) \\
\leq &\mathbf{P}\left(Y_{\varepsilon/2} \leq n^{\gamma-\varepsilon}\right) + \mathbf{P}\left(Y_{-\varepsilon/2} \geq n^{\gamma+\varepsilon}\right) \to 0.
\end{aligned} \tag{9.1}$$

Now, write $\mu_a = \mathbf{E}[Y_a]$ and let θ_a such that $\mu_a\theta_a = n^{\gamma-2a}$. By Proposition 6.2.1, $\theta_a = n^{-a}(1 + o(1))$, and if $a < \gamma$, then $\mu_a \to \infty$ and $\mathbf{E}[Y_a^2] = \mathbf{E}[Y_a]^2(1 + o(1))$ and consequently, $\mathbf{Var}[Y_a] = o(\mathbf{E}[Y_a]^2)$. The last two observations yield the following. First, if $a < 0$, then for n large enough, $\theta_a \geq 1$ and by Chebyshev,

$$\mathbf{P}\left(Y_a \geq n^{\gamma-2a}\right) \leq \frac{\mathbf{Var}[Y_a]}{\mu_a^2(\theta_a - 1)^2} \to 0; \tag{9.2}$$

while if $a \in (0, \gamma)$, we have $\theta_a \to 0$ and in particular, we can apply Paley-Zygmund's inequality to obtain,

$$\mathbf{P}\left(Y_a \geq n^{\gamma-2a}\right) \geq (1 - \theta_a)^2 \frac{\mathbf{E}[Y_a]^2}{\mathbf{E}[Y_a^2]} \to 1. \tag{9.3}$$

Together, (9.2) for $a = -\varepsilon/2$ and (9.3) for $a = \varepsilon/2$ yield (9.1). $\square$

The usual cutting process and the targeted cutting process are not directly comparable. However, Proposition 9.1.1 shows that far fewer deletions are needed when targeting the deletions to the highest-degree vertices, contrary to random edge-deletions. This phenomenon resembles that observed in previous studies on preferential attachment models and scale-free networks [22, 23, 17]. It would be interesting to perform a thorough analysis between the qualitative differences between the two cutting down processes.

CHAPTER 10
Conclusions

Our findings filled a gap in the knowledge of recursive trees, comparatively with linear PA trees. In contrast to linear PA trees, at no point a vertex with maximum degree is established for the rest of the process. Instead, the vertices attaining the maximum degree change as the tree growth process evolves and at no stage there is deterministically a unique vertex with maximum degree.

More precisely, we highlight two of our results. First, we provide all the possible limiting distributions of the number of vertices attaining maximum degree (along distinct subsequences), see Proposition 5.1.3. Second, we describe the joint law of vertices listed in decreasing order of degrees using a marked point process where the marks provide the (scaled) depths of such vertices, see Theorem 5.1.2.

Recursive trees, and increasing trees in general, can be studied through a wide range of probabilistic and analytic tools, several of which were reviewed in Chapter 3. However, our results are obtained by exploiting two distinct approaches to their construction; namely, Kingman's coalescent and what we call the Robin-Hood pruning; see Sections 4.1 and 8.2, respectively. It is worth noting that the Robin-Hood pruning presents a novel growing procedure for growing networks; a discussion about the lines of research that this new process opens is given in Section 8.6.

Recursive trees, having the uniform distribution on increasing trees, lend themselves to the alternative constructions exploited in this manuscript. It remains a challenge to adapt these techniques to a broader class of increasing tree distributions.

Finally, Chapter 9 present a problem on targeted attacks on recursive trees. Much about this targeted attack process is left unknown. In particular, we would like to obtain better estimates on the distribution of $\mathcal{K}_n$, and it would also be interesting to study the process that keeps track of the sizes of the trees pruned out during such deletion process, in the same spirit of the works in [11, 12].

References

[1] Louigi Addario-Berry. Partition functions of discrete coalescents: From cayley's formula to frieze's $\zeta(3)$ limit theorem. In *XI Symposium on Probability and Stochastic Processes: CIMAT, Mexico, November 18-22, 2013*, pages 1–45. Springer International Publishing, Cham, 2015.

[2] Louigi Addario-Berry, Nicolas Broutin, and Cecilia Holmgren. Cutting down trees with a Markov chainsaw. *Ann. Appl. Probab.*, 24(6):2297–2339, 2014.

[3] Louigi Addario-Berry and Laura Eslava. High degree of random recursive trees. `arXiv:1507.05981`.

[4] 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.

[5] Krishna B. Athreya and Samuel Karlin. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist.*, 39:1801–1817, 1968.

[6] Nicolas Bacaër. *A short history of mathematical population dynamics*. Springer-Verlag London, Ltd., London, 2011.

[7] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[8] Erich Baur. Percolation on random recursive trees. *Random Structures & Algorithms*, 48(4):655–680, 2016.

[9] Nathanaël Berestycki. *Recent progress in coalescent theory*, volume 16 of *Ensaios Matemáticos [Mathematical Surveys]*. Sociedade Brasileira de Matemática, Rio de Janeiro, 2009.

[10] François Bergeron, Philippe Flajolet, and Bruno Salvy. Varieties of increasing trees. In *CAAP '92 (Rennes, 1992)*, volume 581 of *Lecture Notes in Comput. Sci.*, pages 24–48. Springer, Berlin, 1992.

[11] Jean Bertoin. Sizes of the largest clusters for supercritical percolation on random recursive trees. *Random Structures & Algorithms*, 44(1):29–44, 2014.

[12] Gabriel Berzunza. The cut-tree of large trees with small heights. `arXiv:1509.01141`, 2016.

[13] Shankar Bhamidi. Universal techniques to analyze preferential attachment tree and networks: Global and local analysis. `http://www.unc.edu/ bhamidi/preferent.pdf`, To be submitted to Probability Surveys.

[14] J. D. Biggins. The first- and last-birth problems for a multitype age-dependent branching process. *Advances in Appl. Probability*, 8(3):446–459, 1976.

[15] J. D. Biggins. The growth and spread of the general branching random walk. *Ann. Appl. Probab.*, 5(4):1008–1024, 11 1995.

[16] Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[17] Béla Bollobás and Oliver Riordan. Robustness and vulnerability of scale-free random graphs. *Internet Math.*, 1(1):1–35, 2003.

[18] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, 18(3):279–290, 2001.

[19] Graham Brightwell and Malwina Luczak. Vertices of high degree in the preferential attachment tree. *Electron. J. Probab.*, 17:43 pp., 2012.

[20] N. Broutin, L. Devroye, E. McLeish, and M. de la Salle. The height of increasing trees. *Random Structures Algorithms*, 32(4):494–518, 2008.

[21] Sébastien Bubeck, Luc Devroye, and Gabor Lugosi. Finding adam in growing random trees. `arXiv:1411.3317`, preprint (2015), to appear in Random Structures and Algorithms.

[22] Reuven Cohen, Keren Erez, Daniel ben Avraham, and Shlomo Havlin. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.*, 85:4626–4634, 2000.

[23] Reuven Cohen, Keren Erez, Daniel ben Avraham, and Shlomo Havlin. Breakdown of the internet under intentional attack. *Phys. Rev. Lett.*, 86:3682–3685, 2001.

[24] Daryl J. Daley and David Vere-Jones. *An introduction to the theory of point processes. Vol. I*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2003. Elementary theory and methods.

[25] Daryl J. Daley and David Vere-Jones. *An introduction to the theory of point processes. Vol. II.* Probability and its Applications (New York). Springer, New York, second edition, 2008. General theory and structure.

[26] Luc Devroye. Branching processes in the analysis of the heights of trees. *Acta Inform.*, 24(3):277–298, 1987.

[27] Luc Devroye. Applications of the theory of records in the study of random trees. *Acta Inform.*, 26(1-2):123–130, 1988.

[28] Luc Devroye. *Branching Processes and Their Applications in the Analysis of Tree Structures and Tree Algorithms*, pages 249–314. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

[29] Luc Devroye. Universal limit laws for depths in random trees. *SIAM J. Comput.*, 28(2):409–432, 1999.

[30] Luc Devroye, Omar Fawzi, and Nicolas Fraiman. Depth properties of scaled attachment random recursive trees. *Random Structures Algorithms*, 41(1):66–98, 2012.

[31] 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.

[32] Michael Drmota. *Random trees*. SpringerWienNewYork, Vienna, 2009. An interplay between combinatorics and probability.

[33] Michael Drmota, Alex Iksanov, Martin Moehle, and Uwe Roesler. A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. *Random Structures & Algorithms*, 34(3):319–336, 2009.

[34] Devdatt Dubhashi and Desh Ranjan. Balls and bins: a study in negative dependence. *Random Structures Algorithms*, 13(2):99–124, 1998.

[35] Richard Durrett. *Probability: theory and examples*. Duxbury Press, Belmont, CA, second edition, 1996.

[36] Laura Eslava. Depth of vertices with high degree in random recursive trees. `arXiv:1611.07466`.

[37] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[38] J. L. Gastwirth and P. K. Bhattacharya. Two probability models of pyramid or chain letter schemes demonstrating that their promotional claims are unreliable. *Oper. Res.*, 32(3):527–536, 1984.

[39] 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.

[40] Christina Goldschmidt. The chen-stein method for convergence of distributions. `https://www.stats.ox.ac.uk/ goldschm/chen-stein.ps.gz`, 2000.

[41] Christina Goldschmidt and James B. Martin. Random recursive trees and the Bolthausen-Sznitman coalescent. *Electron. J. Probab.*, 10:no. 21, 718–745 (electronic), 2005.

[42] E.J. Gumbel. Les valeurs extrêmes des distributions statistiques. *Ann. Inst. H. Poincaré*, 5(2):115–158, 1935.

[43] J. M. Hammersley. Postulates for subadditive processes. *Ann. Probability*, 2:652–680, 1974.

[44] Theodore E. Harris. *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin; Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.

[45] Cecilia Holmgren. Random records and cuttings in split trees: extended abstract. In *Fifth Colloquium on Mathematics and Computer Science*, Discrete Math. Theor. Comput. Sci. Proc., AI, pages 269–281. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2008.

[46] Cecilia Holmgren. Random records and cuttings in binary search trees. *Combin. Probab. Comput.*, 19(3):391–424, 2010.

[47] Alex Iksanov and Martin Möhle. A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Comm. Probab.*, 12:28–35, 2007.

[48] Peter Jagers. *Branching processes with biological applications*. Wiley-Interscience [John Wiley & Sons], London-New York-Sydney, 1975. Wiley Series in Probability and Mathematical Statistics—Applied Probability and Statistics.

[49] Svante Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Process. Appl.*, 110(2):177–245, 2004.

[50] Svante Janson. Asymptotic degree distribution in random recursive trees. *Random Structures Algorithms*, 26(1-2):69–83, 2005.

[51] Svante Janson. Random cutting and records in deterministic and random trees. *Random Structures Algorithms*, 29(2):139–179, 2006.

[52] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[53] Kumar Joag-Dev and Frank Proschan. Negative association of random variables, with applications. *Ann. Statist.*, 11(1):286–295, 1983.

[54] Z. Kalay and E. Ben-Naim. Fragmentation of random trees. *Journal of Physics A: Mathematical and Theoretical*, 48(4), 2015.

[55] David G. Kendall. Branching processes since 1873. *J. London Math. Soc.*, 41:385–406. (1 plate), 1966.

[56] J. F. C. Kingman. The first birth problem for an age-dependent branching process. *Ann. Probability*, 3(5):790–801, 1975.

[57] Torgny Lindvall. On Strassen's theorem on stochastic domination. *Electron. Comm. Probab.*, 4:51–59 (electronic), 1999.

[58] Malwina Luczak and Peter Winkler. Building uniformly random subtrees. *Random Structures Algorithms*, 24(4):420–443, 2004.

[59] Hosam M. Mahmoud. Limiting distributions for path lengths in recursive trees. *Probab. Engrg. Inform. Sci.*, 5(1):53–59, 1991.

[60] Hosam M. Mahmoud. Distances in random plane-oriented recursive trees. *J. Comput. Appl. Math.*, 41(1-2):237–245, 1992. Asymptotic methods in analysis and combinatorics.

[61] Hosam M. Mahmoud. *Pólya urn models*. Texts in Statistical Science Series. CRC Press, Boca Raton, FL, 2009.

[62] Hosam M. Mahmoud and R. T. Smythe. Asymptotic joint normality of outdegrees of nodes in random recursive trees. *Random Structures Algorithms*, 3(3):255–266, 1992.

[63] A. Meir and J. W. Moon. Cutting down random trees. *J. Austral. Math. Soc.*, 11:313–324, 1970.

[64] A. Meir and J. W. Moon. On nodes of out-degree one in random trees. In *Combinatorics (Eger, 1987)*, volume 52 of *Colloq. Math. Soc. János Bolyai*, pages 405–416. North-Holland, Amsterdam, 1988.

[65] A. Meir and J.W. Moon. Cutting down recursive trees. *Mathematical Biosciences*, 21(3):173 – 181, 1974.

[66] Tamás F. Móri. The maximum degree of the Barabási-Albert random tree. *Combin. Probab. Comput.*, 14(3):339–348, 2005.

[67] Hwa Sung Na and Anatol Rapoport. Distribution of nodes of a tree by degree. *Math. Biosci.*, 6:313–329, 1970.

[68] Olle Nerman. On the convergence of supercritical general (c-m-j) branching processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57(3):365–395, 1981.

[69] Alois Panholzer. Destruction of recursive trees. In *Mathematics and computer science. III*, Trends Math., pages 267–280. Birkhäuser, Basel, 2004.

[70] Alois Panholzer. Cutting down very simple trees. *Quaest. Math.*, 29(2):211–227, 2006.

[71] Alois Panholzer and Helmut Prodinger. Level of nodes in increasing trees revisited. *Random Structures Algorithms*, 31(2):203–226, 2007.

[72] Société philomatique de Paris Extraits. *De la loi de multiplication et de la durée des familles [electronic resource] : probabilités, par M.J. Bienaymé.* Cosson Paris, 1845.

[73] Jim Pitman. Coalescent random forests. *J. Comb. Theory A.*, 85:165–193, 1999.

[74] Boris Pittel. On growing random binary trees. *J. Math. Anal. Appl.*, 103(2):461–480, 1984.

[75] Boris Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures Algorithms*, 5(2):337–347, 1994.

[76] Shi, Zhan. Random walks and trees. *ESAIM: Proc.*, 31:1–39, 2011.

[77] Robert T. Smythe and Hosam M. Mahmoud. A survey of recursive trees. *Teor. Ĭmovīr. Mat. Stat.*, (51):1–29, 1994.

[78] Thomas A. Standish. *Data Structure Techniques.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1980.

[79] Jerzy Szymański. On a nonuniform random recursive tree. In *Random graphs '85 (Poznań, 1985)*, volume 144 of *North-Holland Math. Stud.*, pages 297–306. North-Holland, Amsterdam, 1987.

[80] Jerzy Szymański. On the maximum degree and the height of a random recursive tree. In *Random graphs '87 (Poznań, 1987)*, pages 313–324. Wiley, Chichester, 1990.

IEEE@IEEE: Institute of Electrical and Electronics Engineers, Inc. CDMA@CDMA: code-division multiple access CTAN@CTAN: comprehensive TEX archive network