Limiting behaviour of combinatorial Markov chains

Down-up chains on multifurcating trees



Frederik Sørensen

St Edmund Hall

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2021

To Teddy Hall

To the Middle Common Room

Floreat Aula

Acknowledgements

This dissertation marks the end of almost four years as a DPhil student in Oxford. And what incredible four years it has been. This place, the students, the academics, the atmosphere and all of the experiences I have been fortunate to live have made an impact on me that is difficult to overstate.

Nothing and no one encapsulates this more than Teddy Hall and its staff, fellows and students. This place and the extraordinarily friendly and talented people I have been fortunate enough to meet here (and many of whom I am now lucky enough to call my close friends) have truly become my home away from home. Throughout my time in Oxford, Teddy Hall has been the center my life evolved around - I have met most of my network here, spent most of my evenings here, attended countless formal dinners, and generally been involved in all aspects and on all levels of college life. Especially the Middle Common Room will forever hold a special place in my heart.

I would also like to extend my sincerest gratitude to the entire Section of Biostatistics at the University of Copenhagen, and in particular to Esben Budtz-Jørgensen and Julie Lyng Forman, who — in addition to providing me with a workspace when I fled the U.K. in the early days of the pandemic — made it possible for me to keep one foot in the world of applied biostatistics whilst I continued chasing down a more theoretical avenue during my DPhil.

A very special thanks goes out to my supervisor, Matthias Winkel. Not only for suggesting the research topic that ultimately ended up constituting my entire DPhil, but also for providing excellent supervision, and always taking time and being willing to discuss new methods, ideas and odd technicalities. Your huge insight has been massively appreciated.

Last but not least, I would also like to thank my examiners, James Martin and Philippe Marchal, for an enjoyable discussion during my viva and for their insightful comments that ultimately lead to an improved dissertation.

Funding. This research project is funded by the Lundbeck Foundation (R263-2017-3677) and the Engineering and Physical Sciences Research Council (1930434). I would also like to thank Augustinusfonden, Christian og Otillia Brorsons Rejselegat, Knud Højgaards Fond, Louis Hansen Fonden, and the Anglo-Danish Society for providing crucial funding during the first year of my DPhil.

Abstract

In this dissertation, we propose to study a general notion of a down-up Markov chain for multifurcating trees with n labelled leaves. We study in detail down-up chains associated with the (α, γ)-growth process of Chen et al. (2009), in several different ways. Firstly, in the special case of $\gamma = 1 - \alpha$, the stable tree growth process, we generalize a Markov chain by Aldous on binary trees, and show that an unrooted version of the chain has relaxation time of exactly order n^2, whilst the relaxation time for a rooted version is at least of order n^2 and at most of order $n^{3-\alpha}$. Secondly, for general choice of γ, we generalise and further develop previous work by Forman et al. (2018, 2020) in the binary special cases to construct a Markov chain on multifurcating trees with stationary distribution induced by the (α, γ)-growth process. The technique we deploy utilizes the construction of a growth process and a down-up Markov chain on trees with planar structure. Our construction ensures that natural projections of the down-up chain are Markov chains in their own right. We establish label dynamics that at the same time preserve the labelled (α, γ) distribution and keep the branch points between the k smallest labels for order n^2 time steps for all $k \geq 2$. We conjecture the existence of diffusive scaling limits generalising the "Aldous diffusion" by Forman et al. (2018+) as a continuum-tree-valued process and the "algebraic α-Ford tree evolution" by Löhr et al. (2018+) and by Nussbaumer and Winter (2020) as a process in a space of algebraic trees.

Contents

Acknowledgements **i**

Abstract **iii**

Contents **v**

1 Introduction **1**

1.1 Motivation and context 1

1.2 A brief overview of framework and results 4

1.2.1 The exchangeable stable chain: generalizing the Aldous chain 5

1.2.2 The (α, γ)-chain: a multifurcating chain with label swapping 7

1.3 Structure 13

2 A down-up chain with persistent labels on multifurcating trees **15**

2.1 Introduction 15

2.2 Preliminaries 24

2.2.1 Urn schemes 24

2.2.2 Chinese Restaurant Processes 26

2.2.3 Trees 27

2.2.4 Tree growth processes 29

2.2.5 Down-up chains 31

2.3 A semi-planar growth process 33

2.4 A semi-planar down-up chain 38

2.5 Projecting to a non-planar down-up chain 46
2.6 A Markov Chain on decorated trees 49
2.7 Sampling semi-planar (α, γ)-trees 60
2.8 Consistency in stationarity 70

3 Relaxation and mixing times for down-up chains on multifurcating trees 79

3.1 Bounds for the relaxation and mixing times 79
3.2 Preliminaries 81
3.2.1 State spaces and Markov chains 82
3.2.2 Mixing times 84
3.3 On the reversibility of down-up chains 86
3.4 Properties of the exchangeable stable chain 89
3.5 Lower bound for the relaxation time 97
3.6 Upper bound for the relaxation time 99

A Non-reversibility of the (α, γ)-chain 111

Bibliography 115

Chapter 1

Introduction

Tree structures arise in a plethora of different settings. They are of interest in their own right in mathematics, statistics and computer science, and have applications in areas such as physics and the biomedical sciences. In particular, there is an increasingly wide-spread appearance of large tree structures, called phylogenetic trees, arising from genetic data of various types, relating to different species or different specimens of the same species, first and foremost of humans. This poses several interesting and challenging questions from a statistical and mathematical point of view. Before introducing the processes that we will be studying in greater detail, we will therefore give a very broad contextual overview of how the research presented in this dissertation fits into the wider literature.

1.1 Motivation and context

Fundamentally, one can either choose to estimate phylogenetic trees encountered in the biomedical sciences with or without relying on a parametric statistical model quantifying the likelihood of observing a specific phylogenetic tree. As a general statistical observation a non-parametric approach can be advantageous if one does not believe that there exists a parametric model that could adequately describe the data, but such models usually also comes with downsides such as worse statistical power compared to parametric models that describe data well. It is therefore of paramount importance to develop parametric models for phylogenetic data with powerful theoretical properties. This dissertation seeks to contribute to this work.

Specifically, simulating a sequence of large random trees is computationally expensive due to the sheer number of possible trees (as an example, there are more than 200 million possible (rooted) trees with just 10 labelled leaves). One way to approach the problem is to employ Markov chain Monte Carlo methods (for a relevant introduction, see [29]) to explore a subset of possible trees by generating dependent samples. We direct the reader to [19, 49, 52, 53] for additional phylogenetic context. In this framework one relies on the existence of Markov chains that not only have a well-described stationary distribution but also approach this limiting distribution quickly. The simplest of such Markov chains relocate exactly one leaf of the tree in each transition of the Markov chain, whilst other processes relocate entire subtrees in each transition. In the simple case, such relocation can be obtained as the outcome of a two-step procedure:

(i) a down-step where we delete the leaf we wish to relocate, and

(ii) an up-step where we re-insert the leaf based on some rule.

Such processes are naturally Markovian, and — as is common also for other discrete structures in combinatorial probability — we call them *down-up chains*.

We will focus on the development of down-up chains with pre-specified target stationary distributions arising from random tree growth processes, where we at step n of the process construct a random tree of size n conditionally given the previous tree in the sequence. Rémy introduced a random tree growth process for uniform binary leaf-labelled trees [44] and subsequently Aldous studied a down-up chain with a uniform deletion step and an up-step governed by Rémy's uniform growth process. The relaxation and mixing times of the Aldous chain, which quantify how quickly the chain approaches stationarity, were analyzed by Aldous [1] and Schweinsberg [48]. To further demonstrate the relevance of Markov chains such as the Aldous down-up chain in a DNA inference context, see [54].

The overarching purpose of this dissertation is to construct and analyse generalisations of Aldous's tree-valued down-up Markov chain where we let the up-step be governed by a two-parameter growth process, the (α, γ)-growth process [12], which generalizes a host of other growth processes [2, 7, 12, 14, 20, 37, 43, 30]. This endeavour includes basic structural

properties of the Markov chain including equilibrium distributions and asymptotic properties such as mixing times and relaxation times as studied by Aldous and Schweinsberg in their special case. In particular, we are interested in answering the following question: if our target distribution on trees with n leaves arises from the random growth process, can we construct a down-step by reverse engineering the up-step, meaning that if we start from the target stationary distribution on trees with n leaves, can we delete a leaf in a random way that ensures that we end up with the target distribution with $n-1$ leaves? This question is not specifically related to trees; there are many mathematical processes with labels that are constructed incrementally using some kind of up-step (e.g. urn schemes, Chinese Restaurant processes, etc.), and there are similar down-up chains on such structures with similar properties, see e.g. [28, 41, 45].

Rather than having to simulate these processes for a substantial amount of time, one could hope that there was some "limit process", which may be easier to simulate than the large trees, approximating the Markov chain well. Whilst genetic data are naturally discrete and all populations in the real world are finite, there is often a lot to be gained by considering continuum limits, particularly when "finite" really is "very large". For univariate observations, many summary statistics are extremely well-approximated by the normal distribution and related distributions, by the Central Limit Theorem. In fact, there are very powerful invariance principles that mean that the normal distribution is a universal limit of many summary statistics composed of observations that are not themselves well-modelled by the normal distribution. Similarly there are invariance principles for trees that mean that in many random tree growth models, the large random trees approach (with appropriate rescaling) a universal limit distribution, which is Aldous's continuum random tree [1] or related continuum tree structures [37, 12, 31, 32, 34, 30]. From a modelling perspective, this often reduces a wide choice of discrete models to a small parametric family of continuum limit models, which serve as good approximations of the discrete models. In particular, restricting the class of models to a small parametric family of discrete random tree growth models becomes less restrictive by these invariance principles.

In a sequence of papers, [25, 26, 22, 23], Forman, Pal, Rizzolo and Winkel have carried out collaborative efforts, and as have Löhr, Mytnik and Winter in [36], to understand

the limiting evolution of Aldous's tree-valued down-up chain, ultimately constructing a continuum-tree-valued Markov process with continuous paths as conjectured by David Aldous. We will lay the foundation for studying asymptotics in a similar fashion to the approach by Forman, Pal, Rizzolo and Winkel. Key ingredients in this endeavour will be more elementary combinatorial structures such as partition- and composition-valued Chinese Restaurant Processes (see e.g. [42] and [43]) describing relevant subparts of the tree structure. It is also worth mentioning other Markov chains on tree spaces such as the subtree prune and regraft chain of Swofford and Olsen [52] and the Aldous-Broder algorithm [5, 8, 11] for which Evans and Winter in [17] and Evans, Pitman and Winter in [18], respectively, studied continuum tree variations which evolve discontinuously by displacing subtrees.

1.2 A brief overview of framework and results

We will now give a concise overview of the mathematical framework as well as the main results of this dissertation.

Throughout this exposition we will for brevity write $[n] := \{1,\ldots,n\}$ and $[n]_0 := [n] \cup \{0\}$. An $[n]$-tree is a (non-planar) rooted, multifurcating tree with $n \in \mathbb{N}$ leaves labelled by $1,\ldots,n$, an additional vertex of degree 1 called the root which we will sometimes label by 0, and in which the degree of any other vertex, called a branch point, will be at least 3. We will denote the space of such trees by $\mathbb{T}_{[n]}$, but will sometimes restrict our attention to binary trees, i.e. elements of $\mathbb{T}_{[n]}$ where all branch points have exactly degree 3, also called n-leaf cladograms, and denote this space by $\mathbb{T}^{\mathrm{bin}}_{[n]}$.

The *Aldous chain* [7, 48] is a Markov chain on $\mathbb{T}^{\mathrm{bin}}_{[n]}$ with a transition kernel that can be decomposed into a down-step followed by an up-step defined as follows:

Down-step: Delete a leaf uniformly at random and contract the parent branch point away.

Up-step: Insert a new leaf with the same label by selecting an edge uniformly at random, inserting a new branch point on that edge and attaching the new leaf onto that branch point.

It is natural to extend the above type of Markov chain by introducing alternative ways of deleting and inserting leaves, including allowing insertions into branch points thereby allowing multifurcating trees. We will refer to any Markov chain on $\mathbb{T}_{[n]}$ (or a subset thereof) as a *down-up chain* if its transition kernel can be decomposed into a down-step where a random (but not necessarily uniformly distributed) leaf is deleted, and an up-step where a new leaf is inserted into a randomly (but again not necessarily uniformly) picked edge or branch point. In addition, we will allow relabelling of the tree between the down- and up-step. For brevity, we will throughout refer to the collection of edges and branch points of $\boldsymbol{t} \in \mathbb{T}_{[n]}$ as insertable parts of $\boldsymbol{t}$, and denote them by $\mathrm{ins}(\boldsymbol{t})$.

Our focus is on down-up chains where the up-step is governed by the two-parameter (α, γ)-growth process, $0 \leq \gamma \leq \alpha \leq 1$, introduced by Chen, Ford and Winkler in [12]. This is a Markov chain characterized as follows. For each $n \in \mathbb{N}$, conditionally given the nth step of the chain taking the value $\boldsymbol{t} \in \mathbb{T}_{[n]}$, we attach a leaf labelled by $n+1$ to an insertable part of $\boldsymbol{t}$ with probabilities proportional to the following weights: each leaf edge has weight $1-\alpha$, each internal edge (including the edge adjacent to the root) has weight γ and each branch point of degree k has weight $(k-2)\alpha - \gamma$. We refer to the collection of Markov kernels from $\mathbb{T}_{[n]}$ to $\mathbb{T}_{[n+1]}$, $n \in \mathbb{N}$, as a growth rule. The (α, γ)-growth process, denoted by $(T_n)_{n\in\mathbb{N}}$, generalizes many other growth processes in the literature: for $\gamma = 1-\alpha$ we obtain Marchal's [37] stable tree growth process, and for $\gamma = \alpha$ this is exactly Ford's α-model for binary trees [20]; both of which are already generalizations of Rémy's uniform growth process [44] obtained by setting $\gamma = \alpha = \frac{1}{2}$.

In this dissertation we construct and analyze two different types of down-up chains, both with an up-step governed by the (α, γ)-growth rule. This naturally splits into two chapters, which we will now outline the main results of in greater detail.

1.2.1 The exchangeable stable chain: generalizing the Aldous chain

Specifically, we wish to construct a down-up chain on $\mathbb{T}_{[n]}$ with stationary distribution equal to that of T_n, the nth step of the (α, γ)-growth process. Our approach will be to construct a down-step that reverse engineers the up-step driven by the (α, γ)-chain in the

following sense: if $T_n^{\downarrow}$ denotes T_n after having applied the down-step, we wish for that to have the same distribution as T_{n-1}.

This turns out to be a rather complicated endeavour for general γ. However, in the case of the stable growth process, i.e. where $\gamma = 1 - \alpha$, we are assisted by the fact that the leaf labels are exchangeable, meaning that the distribution of T_n, the nth step of the $(\alpha, 1-\alpha)$-growth process, is invariant under permutations of the leaf labels. Circumventing the problem of reverse engineering the up-step entirely, we can utilize the exchangeability property in this special case, allowing us to mimic and generalize the down-up chain of Aldous by reinserting the deleted leaf according to the nth step of the $(\alpha, 1-\alpha)$-growth rule — even if the deleted leaf is not labelled by n.

Definition 1 (Exchangeable stable chain, Definition 3.2.1)**.** Let $\alpha \in [0, 1]$. The *exchangeable stable chain on* $\mathbb{T}_{[n]}$ is a down-up chain on $\mathbb{T}_{[n]}$ where the two steps of the transition kernel can be characterized as follows:

Down: Delete a leaf chosen uniformly at random.

Up: Insert the deleted leaf according to the stable growth rule (i.e. the $(\alpha, 1-\alpha)$-growth rule).

It is easy to see that the stationary distribution of the exchangeable stable chain on $\mathbb{T}_{[n]}$ is indeed the distribution of T_n, the nth step of the $(\alpha, 1-\alpha)$-growth process. In addition to the above, we also consider an unlabelled version, the *unlabelled stable chain*, and an unrooted version, the *unrooted exchangeable stable chain*, where we allow for the root to be relocated like any leaf during transitions of the chain.

We will in Section 3.3 show that all three of these chains are reversible, allowing us to access a variety of different techniques to lower and upper bound the relaxation and mixing times of the chains. In all three cases, we can show that the relaxation and mixing time of the chains on $\mathbb{T}_{[n]}$ are at least of order n^2. However, it turns out that the existence of a root that is not allowed to be relocated has a rather big effect on the upper bounds that we can produce for the relaxation time using the argument of Schweinsberg [48]. Hence we obtain the following two results:

Theorem 2 (Theorem 3.1.1). *The relaxation time of the unrooted exchangeable stable chain is exactly of order n^2.*

Theorem 3 (Theorem 3.1.2). *The relaxation time of the exchangeable stable chain and the unlabelled stable chain are at least of order n^2 and at most of order $n^{3-\alpha}$.*

Using standard links between the bounds for the relaxation and mixing times of any reversible Markov chains immediately leads to the following results on the mixing times:

Corollary 4 (Corollary 3.1.3). *The mixing time of the unrooted exchangeable stable chain is at least of order n^2 and at most of order $n^3 \log(n)$, whilst the mixing time of the exchangeable stable chain and the unlabelled stable chain are at least of order n^2 and at most of order $n^{4-\alpha} \log(n)$.*

The gap between the lower and upper bounds of the relaxation times of the rooted versions of the chains seems to be an artefact of the methods of proof rather than a fundamental difference in how quickly the chains approach stationarity. Additionally, it is well-known that coupling arguments can be more efficient in giving upper bounds for the mixing time compared to bounds using the relaxation time. Indeed, Aldous already gives an argument for a tighter upper bound of n^3 for the mixing time of the unrooted exchangeable stable chain with $\alpha = \frac{1}{2}$ (i.e. the Aldous chain) in [7], and further mentions analogies to a different Markov chain with a mixing time of order n^2. Additional evidence for the mixing and relaxation times being of order n^2 can be found from the diffusive scaling limits of [36] and [39]. We therefore pose the following conjecture:

Conjecture 5 (Conjecture 3.1.4). *The relaxation and mixing times of the exchangeable stable chain and the unlabelled stable chain, as well as the mixing time of the unrooted exchangeable stable chain, are all exactly of order n^2.*

1.2.2 The (α, γ)-chain: a multifurcating chain with label swapping

We will now return to the problem of defining a Markov chain on $\mathbb{T}_{[n]}$ with the nth step of the (α, γ)-growth process as the stationary distribution, this time for general choices of $0 \leq \gamma \leq \alpha \leq 1$. In order to do so, we will develop a refined version of the (α, γ)-growth process on what we call *semi-planar trees*. Through imposing an ordering of all the subtrees (but not distinguishing between the two subtrees containing the lowest labels) within

a branch point, these trees mimic binary trees, in spite of being multifurcating, thus allowing us to utilize the ideas developed for binary trees in [21, 24] [1].

Elaborating on this further, we will every time we create a branch point in the (α, γ)-growth process start an ordered Chinese Restaurant Process (see e.g. [42, 43]) to determine the order of any additional subtrees of the branch point. This apparatus and a corresponding down-up chain with an up-step governed by a semi-planar version of the (α, γ)-growth process is developed in full detail in Section 2.3.

By use of the intertwining criterion [46], forgetting the order of the subtrees within a branch point leads to the following down-up chain on $\mathbb{T}_{[n]}$. In order to describe this, we will need the notion of a spinal bush from a leaf $i \in [n]$ in a tree $\boldsymbol{t} \in \mathbb{T}_{[n]}$. This is the collection of subtrees rooted at a branch point on the ancestral line from i towards the root in $\boldsymbol{t}$, and we will enumerate them in top-to-bottom spinal order from the leaf i.

Definition 6 (Definition 2.1.1)**.** Fix $n \in \mathbb{N}$. The (α, γ)-*chain on* $\mathbb{T}_{[n]}$, $(T_n(m))_{m \in \mathbb{N}_0}$, is a Markov chain on $\mathbb{T}_{[n]}$ with a transition kernel characterized as follows:

(i) Conditional on $T_n(m) = \boldsymbol{t}$, select leaf I uniformly at random.

(ii) Let c_v denote the number of children of v, the parent of leaf I in $\boldsymbol{t}$.

- If $c_v = 2$, let $\tilde{I} = \max\{I, a, b\}$ where

$$a = \min \left\{ \begin{matrix} \text{leaf labels in the first spinal bush} \\ \text{on the ancestral line from leaf } I \end{matrix} \right\},$$

$$b = \min \left\{ \begin{matrix} \text{leaf labels in the second spinal bush} \\ \text{on the ancestral line from leaf } I \end{matrix} \right\}.$$

- If $c_v > 2$, let $i_1, \ldots, i_{c_v}$ denote the smallest leaf labels in the c_v subtrees of v, enumerated in increasing order, and say that $I = i_j$ for some $j \in [c_v]$. Define the

[1] [21] is a pre-print version of [24]. The pre-print version of the paper treats a down-up chain with an up-step governed by Ford's α-growth rule, for general $\alpha \in [0, 1]$, whilst the published version only treats the case where $\alpha = \frac{1}{2}$.

conditional distribution of $\tilde{I}$ by

$$\mathrm{P}\left(\tilde{I}=i_{j'} \;\middle|\; T_n(m)=\boldsymbol{t}, I=i_j\right)=\begin{cases}\frac{\alpha-\gamma}{(c_v-2)\alpha-\gamma} & \text{if } j\leq 2, j'=3\\ \frac{\alpha}{(c_v-2)\alpha-\gamma} & \text{if } j\leq 2, 3<j'\leq c_v\\ \frac{\alpha}{(c_v-2)\alpha-\gamma} & \text{if } 2<j<j'\leq c_v\\ \frac{(j-2)\alpha-\gamma}{(c_v-2)\alpha-\gamma} & \text{if } 2<j=j'\leq c_v\end{cases}$$

(iii) Swap I and $\tilde{I}$, and subsequently delete the leaf now labelled by $\tilde{I}$.

(iv) Relabel $\{\tilde{I}+1,\ldots,n\}$ by $\{\tilde{I},\ldots,n-1\}$ using the increasing bijection.

(v) Perform an up-step according to the (α,γ)-growth process.

Examples of transitions of the (α,γ)-chain are illustrated in Figure 1.1. It is not at all obvious that the above down-step reverse engineers the up-step of the (α,γ)-chain in the sense outlined previously, and builds on the substantial apparatus we build up around semi-planar trees developed in the early sections of Chapter 2. However, once we show that this is indeed the case we get the following conclusion:

Theorem 7 (Theorem 2.1.2)**.** *Fix $n\in\mathbb{N}$. Let $(T_n(m))_{m\in\mathbb{N}_0}$ denote the (α,γ)-chain and let T_n denote the nth step of the (α,γ)-growth process. Then $T_n(m)\xrightarrow{D}T_n$ as $m\to\infty$ for any initial distribution, where $\xrightarrow{D}$ denotes convergence in distribution.*

It is important to note that the (α,γ)-chain is not a generalization of the exchangeable stable chain, even in the case where $\gamma=1-\alpha$. Whilst the unlabelled versions of the two chains coincide, the label dynamics are purposefully and fundamentally different in the two chains. The importance of the above result is that, remarkably, the stationary distribution of the two chains is the same.

The remaining results of this part of the thesis lay the groundwork for constructing continuum limit processes for multifurcating trees that [25, 26, 22, 23] do in the binary case, but those processes are beyond the scope of this thesis. The construction relies on certain projections of our Markov chain involving reduced subtrees to again be Markovian, but this is generally a rather delicate feature.

Before further explaining this, let us specify the projections of interest. Fix $1 \leq k \leq n$. For any $\boldsymbol{t} \in \mathbb{T}_{[n]}$ we can take the subtree spanned by the root and the leaves labelled by $[k]$, which after reduction (i.e. removing branch points of degree 2) will give us a tree $\boldsymbol{s} \in \mathbb{T}_{[k]}$. Now, to each insertable part $x \in \mathrm{ins}(\boldsymbol{s})$ we can associate an integer y_x which is the number of leaves of the original tree $\boldsymbol{t}$ from which x is the first insertable part of the reduced subtree encountered on the ancestral line to the root in $\boldsymbol{t}$. This gives us a tree shape $\boldsymbol{s}$ and a vector of masses $\mathbf{y} = (y_x)_{x \in \mathrm{ins}(\boldsymbol{s})}$ satisfying that $\sum_{x \in \mathrm{ins}(\boldsymbol{s})} y_x = n$, and we call $(\boldsymbol{s}, \mathbf{y})$ a decorated $[k]$-tree of mass n. We denote the space of these trees by $\mathbb{T}^{\bullet n}_{[k]}$ and the surjective projection map from $\mathbb{T}_{[n]}$ to $\mathbb{T}^{\bullet n}_{[k]}$ by $\pi^{\bullet n}_{[k]}$.

Let us now investigate the behaviour of the exchangeable stable chain compared to the (α, γ)-chain when observing the projection onto decorated trees, in a simple example where we set $k = 2$ and note that there is a unique element of $\boldsymbol{s} \in \mathbb{T}_{[2]}$, with exactly four insertable parts — two leaf edges, one branch point and one internal edge (the root edge). We have the following result:

Consider first the (α, γ)-chain, $(T_n(m))_{m \in \mathbb{N}_0}$, and its projection to $\mathbb{T}^{\bullet n}_{[2]}$, such that $\pi^{\bullet n}_{[2]}(T_n(m)) =: (\boldsymbol{s}, \mathbf{Y}_n(m))$ where the vector $\mathbf{Y}_n(m) := \left(Y_n^1(m), Y_n^2(m), Y_n^{\mathrm{bp}}(m), Y_n^0(m)\right)$ is enumerated in the obvious way. Then it follows from arguments from [40] that

$$\left(\frac{1}{n}\mathbf{Y}_n(\lfloor n^2 t \rfloor)\right)_{0 \leq t \leq S_n} \xrightarrow{D} (\mathbf{Y}(t))_{0 \leq t \leq S}$$

as $n \to \infty$, where $\xrightarrow{D}$ denotes convergence in distribution and $\mathbf{Y}$ is a Wright-Fisher diffusion with parameters $(-\alpha, -\alpha, 2\alpha - 1, 1 - \alpha)$, a process that generalizes the usual multi-type Wright-Fisher process to 'mutation' parameters that can be negative. Here S_n is the first time we select leaf 1 or 2 for deletion where $Y_n^1(\cdot)$ or $Y_n^2(\cdot)$ are equal to 1, respectively, and S is the first time one of the first two components of the Wright-Fisher diffusion vanishes. A more general version of this result can be found in Proposition 2.1.6.

In the (α, γ)-chain this stopping only occurs at step $S_n = m$ if either leaf 1 or leaf 2 are the only contributor to $Y_n^1(m)$ or $Y_n^2(m)$, respectively (i.e. $Y_n^1(m) = 1$ or $Y_n^2(m) = 1$), and we select that leaf in the down-step of the chain. One such example is depicted in Figure 1.1. In this case, the branch point of the projected tree is first contracted away in a down-step at this time. But in the case of the exchangeable stable chain the branch point

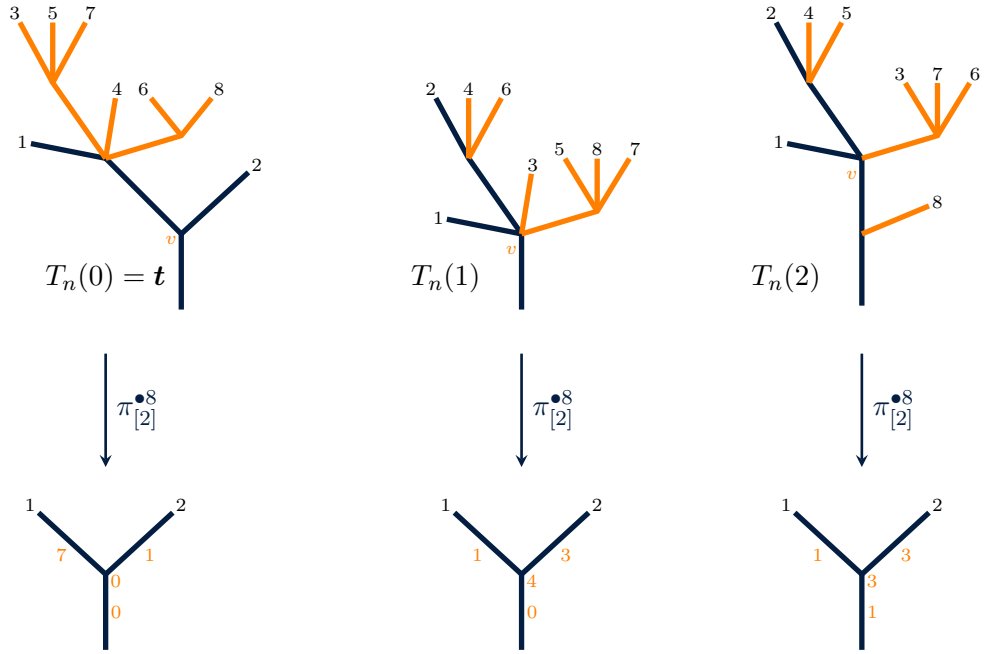


Figure 1.1: Illustration of the label dynamics of the (α, γ)-chain on $\mathbb{T}_{[8]}$ including projections to $[2]$-decorated trees of mass 8. Starting from $T_n(0) = \boldsymbol{t}$ obtain $T_n(1)$ by selecting $i = 2$ and finding $\tilde{I} = 2$, as the smallest label in the only subtree on the ancestral line from i towards the root is 1. Hence we delete $\tilde{I} = 2$, and subsequently relabel $\{3, \dots, 8\}$ by $\{2, \dots, 7\}$ using the increasing bijection. We then insert leaf 8 into the branch point of degree 3 with probability $\frac{\alpha-\gamma}{7-\alpha}$ to obtain $T_n(1)$. For the next transition we select leaf 3, and we find the smallest labels in the four subtrees of v to be $\{1, 2, 3, 5\}$, of which 3 has rank 3. With probability $\frac{\alpha}{2\alpha-\gamma}$ we have $\tilde{I} = 5$, we swap the labels $i = 3$ and $\tilde{I} = 5$ and subsequently delete the leaf now labelled by $\tilde{I} = 5$, and finish the down-step by relabelling $\{6, 7, 8\}$ by $\{5, 6, 7\}$ using the increasing bijection. In the up-step we insert leaf 8 into the root edge with probability $\frac{\gamma}{7-\alpha}$ to obtain $T_n(2)$.

separating the root and leaves 1 and 2 will be relocated far too often (after $O(n)$ steps), implying that we cannot bear any hope for a similar convergence result for the exchangeable stable chain. Another related and subtle difference between the exchangeable stable chain and the (α, γ)-chain is in regards to retaining the Markov property when projecting the chains onto decorated trees. Importantly, this property is retained in the case of the (α, γ)-chain (if we start from a suitable initial distribution) but not for the exchangeable stable chain.

Remark 8. The down-side of the more complicated label dynamics of the (α, γ)-chain compared to the exchangeable stable chain is that we lose the reversibility property of the chain. This result is included in Appendix A.

With the above example in mind, we will now vary $k \in [n]$ and project the (α, γ)-chain

on $\mathbb{T}_{[n]}$ to the space of decorated $[k]$-trees of mass n. We will show that the projected chain is Markovian if it runs in stationarity, and characterize its transition kernel without referencing the stationary (α, γ)-chain (see Proposition 2.6.7). This was done in the binary case (i.e. for $\gamma = \alpha$) in [21, 24] by combining two well-known results for when a function of a Markov chain is again Markovian: the Kemeny-Snell criterion [33] and the intertwining criterion [46]. We will use the same techniques in conjunction with one another by considering an intermediate Markov chain.

To make some of these statements precise, let K_n denote the transition kernel of the (α, γ)-chain and let T_n denote the nth step of the (α, γ)-growth process. Now, define the following natural Markov kernel from $\mathbb{T}^{\bullet n}_{[k]}$ to $\mathbb{T}_{[n]}$:

$$\Pi^{\bullet n}_k(\boldsymbol{t}^\bullet, \cdot) = \mathrm{P}\left(T_n \in \cdot \;\middle|\; \pi^{\bullet n}_{[k]}(T_n) = \boldsymbol{t}^\bullet\right),$$

for $\boldsymbol{t}^\bullet \in \mathbb{T}^{\bullet n}_{[k]}$. This Markov kernel samples full $[n]$-trees from a decorated $[k]$-tree of mass n, see Figure 1.1 for examples. To define a down-up chain on $\mathbb{T}^{\bullet n}_{[k]}$, let us abuse notation so that K_n and $\Pi^{\bullet n}_k$ denote both the kernel and the corresponding Markovian matrix, whilst $\pi^{\bullet n}_{[k]}$ is both the projection map and the matrix $\pi^{\bullet n}_{[k]} = \left(1_{\pi^{\bullet n}_{[k]}(\boldsymbol{t})=\boldsymbol{t}^\bullet}\right)_{\boldsymbol{t}\in\mathbb{T}_{[n]}, \boldsymbol{t}^\bullet\in\mathbb{T}^{\bullet n}_{[k]}}$.

Definition 9 (Decorated (α, γ)-chain, Definition 2.1.3)**.** Fix $n > k \in \mathbb{N}$. The *decorated (α, γ)-chain on* $\mathbb{T}^{\bullet n}_{[k]}$, is a Markov chain on $\mathbb{T}^{\bullet n}_{[k]}$ with transition kernel

$$K^{\bullet n}_k := \Pi^{\bullet n}_k K_n \pi^{\bullet n}_{[k]}. \tag{1.1}$$

By combining the Kemeny-Snell and intertwining criteria mentioned earlier we obtain another main result of this thesis:

Theorem 10 (Theorem 2.1.4)**.** *Fix $n \geq 2$. Let $(T_n(m))_{m\in\mathbb{N}_0}$ denote an (α, γ)-chain. For $1 \leq k < n$, let $T^{\bullet n}_{[k]}(m) := \pi^{\bullet n}_{[k]}(T_n(m))$, $m \in \mathbb{N}_0$, denote the projection of the chain onto decorated $[k]$-trees of mass n. If $T_n(0) \sim \Pi^{\bullet n}_k(\boldsymbol{t}^\bullet, \cdot)$ for some $\boldsymbol{t}^\bullet \in \mathbb{T}^{\bullet n}_{[k]}$, then $\left(T^{\bullet n}_{[k]}(m)\right)_{m\in\mathbb{N}_0}$ is a decorated (α, γ)-chain on $\mathbb{T}^{\bullet n}_{[k]}$ started from $\boldsymbol{t}^\bullet$.*

Varying k we can now obtain a consistent system of projected Markov chains that are all running in stationarity:

Corollary 11 (Consistency in stationarity, Corollary 2.1.5). *Fix $n \geq 2$. Let $(T_n(m))_{m\in\mathbb{N}_0}$ denote an (α,γ)-chain running in stationarity. For $1 \leq k < n$, let $T^{\bullet n}_{[k]}(m) := \pi^{\bullet n}_{[k]}(T_n(m))$, $m \in \mathbb{N}_0$, denote the projection of the chain onto decorated $[k]$-trees of mass n. Then each of the $\left(T^{\bullet n}_{[k]}(m)\right)_{m\in\mathbb{N}_0}$ is a decorated (α,γ)-chain on $\mathbb{T}^{\bullet n}_{[k]}$ running in stationarity.*

This gives rise to the following conjecture, which strengthens results in [36] for $\gamma = \alpha$, and in [39] for $\alpha = \gamma = \frac{1}{2}$. We denote by $\mathcal{T}^{\alpha,\gamma}$ the limit in the Gromov-Hausdorff-Prokhorov topology of weighted $\mathbb{R}$-trees, see [31, 38] of the unlabelled (α,γ)-growth process suitably represented as $\mathbb{R}$-trees with edge length $n^{-\gamma}$, see [12].

Conjecture 12 (Conjecture 2.1.8). *The unlabelled (α,γ)-chain with n^2 steps per time unit and edges scaled by n^γ has a scaling limit that is a path-continuous Markov process in the Gromov-Hausdorff-Prokhorov space , whose stationary distribution is that of $\mathcal{T}^{\alpha,\gamma}$.*

1.3 Structure

The dissertation is structured as follows. Chapters 2 and 3 are largely self-contained chapters. Chapter 2 was submitted to *Random Structures and Algorithms* on September 22nd 2020, and the material of Chapter 3 is currently being prepared for publication as an additional research paper.

A more detailed overview is the following:

- Chapter 2 sets up most of the framework that we will work within. In particular, Sections 2.3 and 2.4 introduce the semi-planar trees and the apparatus to efficiently and precisely deal with them, as well as a semi-planar version of the (α,γ)-chain further feeding into a down-up chain with a desirable stationary distribution.

  In Section 2.5 we perform the first of many projection arguments, showing that the projection of the semi-planar (α,γ)-chain is the (non-planar) (α,γ)-chain, and we prove Theorem 7.

  Section 2.6 introduces down-up chains on decorated trees. And after studying technical aspects of projections of the stationary distribution of the (α,γ)-chain to decorated tree spaces in Section 2.7, we combine the intertwining and Kemeny-Snell criteria to show that projections of suitable (α,γ)-chains onto the decorated tree spaces

are again Markovian, with the transition kernel previously defined. The proofs of Theorem 10 and Corollary 11 are the very final proofs of the chapter, utilizing almost the entire machinery that we will have developed up to that point.

- Chapter 3 properly defines the notions of relaxation and mixing times, and investigates the relaxation times of down-up chains where the up-step is governed by the stable tree growth model of [37]. As part of this investigation, we will provide an important discussion regarding the reversibility of the chains introduced here. We prove Theorems 2 and 3, and Corollary 4 by combining lower bounds deduced in Section 3.5 with upper bounds from Section 3.6.

- Appendix A is a small side note that proves that none of the down-up chains in Chapter 2 are reversible, thus excluding them from the investigation in Chapter 3.

Chapter 2

A down-up chain with persistent labels on multifurcating trees

This chapter was submitted to Random Structures and Algorithms on September 22nd 2020. Apart from minor formatting edits, and incorporating the abstract and bibliography into the overall abstract and bibliography of this dissertation, it is presented here in the state in which it was submitted.

2.1 Introduction

Let $[n] := \{1, \dots, n\}$. An $[n]$-tree is a (non-planar) rooted, multifurcating tree with $n \in \mathbb{N}$ leaves labelled by $1, \dots, n$, an additional vertex of degree 1 called the root, and in which the degree of any other vertex, called a branch point, will be at least 3. We will denote the space of such trees by $\mathbb{T}_{[n]}$. Occasionally, we will only allow binary trees, also called n-leaf cladograms, and will denote this space by $\mathbb{T}^{\text{bin}}_{[n]}$. The *Aldous chain* [7, 48] is a Markov chain on $\mathbb{T}^{\text{bin}}_{[n]}$ with a transition kernel consisting of a down-step followed by an up-step defined as follows:

Down-step: Delete a leaf uniformly at random and contract the parent branch point away.

Up-step: Insert a new leaf with the same label by selecting an edge uniformly at random, inserting a new branch point on that edge and attaching the new leaf onto that branch point.

It is natural to extend the above type of Markov chain by introducing alternative ways of deleting and inserting leaves as well as by extending it to the space of multifurcating trees. We will refer to any Markov chain on $\mathbb{T}_{[n]}$ (or a subset thereof) as a *down-up chain* if its transition kernel can be decomposed into a down-step where a random (but not necessarily uniformly distributed) leaf is deleted, and an up-step where a new leaf is inserted into a randomly picked edge or branch point.

Noting that the up-step described above is identical to one step of Rémy's growth process [44], a specific problem is to define new down-up chains where the up-step is governed by other growth processes from the literature. For our purposes Ford's α-model for binary trees [20], extending Rémy's growth process [44], and the (α, γ)-growth rule for multifurcating trees [12], of which both Ford's α-model and Marchal's stable tree growth process [37] are special cases, are of highest importance. The (α, γ)-growth process for $0 \leq \gamma \leq \alpha \leq 1$, is characterized by, for each $n \in \mathbb{N}$, obtaining an element of $\mathbb{T}_{[n+1]}$ from an element $\boldsymbol{t} \in \mathbb{T}_{[n]}$ by inserting a new leaf labelled by $n+1$ into an *insertable part* of $\boldsymbol{t}$, denoted by $\mathrm{ins}(\boldsymbol{t})$, where an insertable part $x \in \mathrm{ins}(\boldsymbol{t})$ is picked with probability proportional to an associated weight, w_x, with

$$w_x = \begin{cases} 1-\alpha & \text{if } x \text{ is a leaf edge,} \\ \gamma & \text{if } x \text{ is an internal edge,} \\ (c_x - 1)\,\alpha - \gamma & \text{if } x \text{ is a branch point with } c_x \text{ children.} \end{cases} \tag{2.1}$$

It is rather natural to construct a down-step such that the stationary distribution of the down-up chain is that induced by the growth process that governs the up-step. To be more precise, let $(T_n)_{n\in\mathbb{N}}$ denote a growth process. We are then looking to devise a down-step, such that when carrying out the down-step from T_n we obtain $T_n^{\downarrow}$ with the property that $T_n^{\downarrow} \overset{D}{=} T_{n-1}$ for each $n \geq 2$. Combining the down-step with the up-step from the growth process yields a stationary down-up Markov chain, $(T_n(m))_{m\in\mathbb{N}_0}$, if $T_n(0) \overset{D}{=} T_n$. If the leaf labels are exchangeable, i.e. if for any permutation σ of $[n]$ the probability of obtaining an $[n]$-tree, $\boldsymbol{t}$, is the same as the probability of obtaining $\boldsymbol{t}$ where leaf i is labelled by $\sigma(i)$ for each $i \in [n]$, a down-step with a simple uniform deletion of the leaves will suffice. However, in the absence of exchangeability the situation is more complicated.

In [21, 24] a modified version of the Aldous chain, the α-chain, was introduced, where the up-step is governed by Ford's α-model. For $\alpha = \frac{1}{2}$, Ford's α-model is the same as Rémy's growth process, and so the leaf labels are exchangeable, but this is not true for $0 < \alpha \neq \frac{1}{2} < 1$, so a more complicated down-step was introduced:

(i) Down-step (selection): Select leaf i uniformly at random.

(ii) Down-step (local search): Defining a and b to be the smallest leaf labels in the first and second subtree on the ancestral line from i, respectively, let $\tilde{\imath} := \max\{i, a, b\}$.

(iii) Down-step (swap and delete): Swap i and $\tilde{\imath}$, delete leaf $\tilde{\imath}$, and contract away the parent branch point.

(iv) Down-step (relabel): Relabel the leaves using the increasing bijection $[n]\backslash\{\tilde{\imath}\} \to [n-1]$ to obtain an element of $\mathbb{T}^{\mathrm{bin}}_{n-1}$.

(v) Up-step: Insert leaf n according to Ford's α-growth rule [20] (the (α, γ)-growth rule with $\gamma = \alpha$).

In the exchangeable case where $\alpha = \frac{1}{2}$, the relabelling in step (iv) is unnecessary and one can therefore adjust (v) accordingly and just insert the deleted leaf again, which yields the uniform chain from [21].

Irrespective of the exchangeability of the leaf labels, the above down-step serves a key purpose in a broader context which we will now outline. It is well known that the stationary distribution of the Aldous chain, the uniform distribution on the space of binary trees [7], converges when suitably scaled to that of the Brownian Continuum Random Tree [1], as the number of leaves tends to infinity. Aldous and Schweinsberg [7, 48] showed that the relaxation time of the Aldous chain is of order n^2. Aldous conjectured that running the chain n^2 times faster than the number of leaves tend to infinity would result in a "diffusion on continuum trees" [6].

In a sequence of papers [25, 26, 22, 23] the authors have carried out a programme to construct this "Aldous diffusion". Parts of this construction arise by studying projections of the uniform chain onto decorated trees. To be more precise, consider for an $[n]$-tree

the reduced subtree spanned by the root and the leaves labelled by 1 and 2, where all vertices of degree 2 have been contracted away, and consider how many of the n leaves are sitting in subtrees of the branch point and on each of the edges of that reduced tree (for a more precise description of this, see Definition 2.6.1). This will, for a binary tree, yield three masses summing to n. For each $m \in \mathbb{N}_0$, let $T_n(m)$ denote the mth step of the uniform chain on $\mathbb{T}_{[n]}$, let $Y_n^0(m), Y_n^1(m), Y_n^2(m)$ denote the proportion of these masses when projecting $T_n(m)$ to a decorated [2]-tree of mass n. In [40] a notion of a Wright-Fisher diffusion with negative mutation rates was developed, and it was shown that

$$\big(Y_n^0(\lfloor n^2 t \rfloor), Y_n^1(\lfloor n^2 t \rfloor), Y_n^2(\lfloor n^2 t \rfloor)\big)_{t\geq 0} \xrightarrow{D} \big(Y^0(t), Y^1(t), Y^2(t)\big)_{t\geq 0} \tag{2.2}$$

as $n \to \infty$, stopped the first time one of the last two coordinates hits 0, where $\xrightarrow{D}$ denotes convergence in distribution and (Y^0, Y^1, Y^2) is a Wright-Fisher diffusion with mutation rates $\big(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\big)$. We extend this result in Proposition 2.1.6. If we delete leaves uniformly at random, we will end up deleting the branch point due to one of the masses associated with the leaf edges dropping to 0. In the limit, the Wright-Fisher process does not provide any guidance as to how to continue the process. But with the down-step of the uniform chain we have a mechanism by which we can identify another branch point and continue the process. The two key aspects that make this possible, are the label swapping as well as the relabelling of the remaining leaf labels.

The aim of this paper is to study and generalize down-up chains with label swapping to the space of multifurcating trees. Specifically, we will study chains where the up-step is governed by the (α, γ)-growth process [12], which also encompasses the prime special case of the stable tree growth [37]. Whilst any (α, γ)-growth process with $0 < \gamma \leq \alpha < 1$ has a continuum random tree as its scaling limit, the stable tree growth is of particular motivational interest due to the universality of the stable continuum random tree [37], and the fact that the stable tree growth induces exchangeable leaf labels. Hence a down-up chain with an up-step governed by the (α, γ)-growth process will be analogous to the Aldous and uniform chains, in terms of devising a "diffusive limit on continuum trees". To envisage the construction of such a process on continuum trees it is of key importance to devise a down-step that allows the construction beyond a disappearing branch point, when

projecting it to the space of decorated trees. This is the key purpose of this paper.

Definition 2.1.1. Fix $n \in \mathbb{N}$. The (α, γ)-*chain on* $\mathbb{T}_{[n]}$, $(T_n(m))_{m\in\mathbb{N}_0}$, is a Markov chain on $\mathbb{T}_{[n]}$ with a transition kernel characterized as follows:

(i) Conditional on $T_n(m) = \boldsymbol{t}$, select leaf I uniformly at random.

(ii) Let c_v denote the number of children of v, the parent of leaf I in $\boldsymbol{t}$.

- If $c_v = 2$, let $\tilde{I} = \max\{I, a, b\}$ where

$$a = \min\left\{\begin{matrix}\text{leaf labels in the first spinal bush} \\ \text{on the ancestral line from leaf } I\end{matrix}\right\},$$

$$b = \min\left\{\begin{matrix}\text{leaf labels in the second spinal bush} \\ \text{on the ancestral line from leaf } I\end{matrix}\right\}.$$

- If $c_v > 2$, let $i_1, \ldots, i_{c_v}$ denote the smallest leaf labels in the c_v subtrees of v, enumerated in increasing order, and say that $I = i_j$ for some $j \in [c_v]$. Define the conditional distribution of $\tilde{I}$ by

$$\mathrm{P}\left(\tilde{I} = i_{j'} \;\middle|\; T_n(m) = \boldsymbol{t}, I = i_j\right) = \begin{cases} \frac{\alpha-\gamma}{(c_v-2)\alpha-\gamma} & \text{if } j \le 2, j' = 3 \\ \frac{\alpha}{(c_v-2)\alpha-\gamma} & \text{if } j \le 2, 3 < j' \le c_v \\ \frac{\alpha}{(c_v-2)\alpha-\gamma} & \text{if } 2 < j < j' \le c_v \\ \frac{(j-2)\alpha-\gamma}{(c_v-2)\alpha-\gamma} & \text{if } 2 < j = j' \le c_v \end{cases}$$

(iii) Swap I and $\tilde{I}$ and delete $\tilde{I}$.

(iv) Relabel $\{\tilde{I}+1, \ldots, n\}$ by $\{\tilde{I}, \ldots, n-1\}$ using the increasing bijection.

(v) Perform an up-step according to the (α, γ)-growth process.

Formal definitions of the terminology and operations used in Definition 2.1.1 can be found in Sections 2.2.4 and 2.2.5.

Theorem 2.1.2. *Fix $n \in \mathbb{N}$. Let $(T_n(m))_{m\in\mathbb{N}_0}$ denote the (α, γ)-chain and let T_n denote the nth step of the (α, γ)-growth process. Then $T_n(m) \xrightarrow{D} T_n$ as $m \to \infty$ for any initial distribution, where $\xrightarrow{D}$ denotes convergence in distribution.*

The significance of Theorem 2.1.2 is that the unique invariant distribution for the (α, γ)-chain is exactly the distribution induced by the (α, γ)-growth process, proving that the down-step does indeed counteract the up-step governing the (α, γ)-growth process.

We now turn our focus to the investigation of projections of the (α, γ)-chain. Specifically, we will project the (α, γ)-chain on $\mathbb{T}_{[n]}$ to the space of decorated $[k]$-trees of mass n, a special case of which was introduced prior to (2.2). We will show that the projected chain is Markovian and characterize its transition kernel without referencing the (α, γ)-chain. There are several well-known results on when a function of a Markov chain is again Markovian. As in [24], we will utilize both the Kemeny-Snell criterion [33] and the intertwining criterion [46], and we will need to use them in conjunction with one another by considering an intermediary Markov chain.

In general, a decorated $[k]$-tree of mass n is a $[k]$-tree $\boldsymbol{s} \in \mathbb{T}_{[k]}$, called the *tree shape*, where all branch points and edges have an affiliated non-negative integer mass such that any leaf edge has mass at least 1 and the sum of all masses is n. An $[n]$-tree naturally gives rise to such a decorated tree, by letting the $[k]$-tree be the reduced subtree of the $[n]$-tree spanned by the root and the leaves labelled by $[k]$ where all branch points of degree 2 are contracted away. Comparing this reduced subtree to the original $[n]$-tree, the mass associated with an insertable part, i.e. an edge or a branch point, now arise by counting all the leaves for which the insertable part is the first insertable part of the reduced $[k]$-tree that is reached on the ancestral line from the leaf. We denote the surjective map that yields a decorated $[k]$-tree of mass n from an $[n]$-tree by $\pi^{\bullet n}_{[k]} \colon \mathbb{T}_{[n]} \to \mathbb{T}^{\bullet n}_{[k]}$. We will discuss this in more detail in Section 2.6.

In the following, let K_n denote the transition kernel of the (α, γ)-chain from Definition 2.1.1. Define the following natural Markov kernel from the space $\mathbb{T}^{\bullet n}_{[k]}$ of decorated $[k]$-trees of mass n to $\mathbb{T}_{[n]}$:

$$\Pi^{\bullet n}_{k}(\boldsymbol{t}^{\bullet}, \cdot) = \mathrm{P}\left(T_n \in \cdot \;\middle|\; \pi^{\bullet n}_{[k]}(T_n) = \boldsymbol{t}^{\bullet}\right), \tag{2.3}$$

where T_n is the nth step of the (α, γ)-growth process.

In order to introduce a down-up chain on $\mathbb{T}^{\bullet n}_{[k]}$, let us abuse notation so that K_n and $\Pi^{\bullet n}_k$ denote both the kernel and the corresponding Markovian matrix, whilst $\pi^{\bullet n}_{[k]}$ is both the projection map and the matrix $\pi^{\bullet n}_{[k]} = \left(1_{\pi^{\bullet n}_{[k]}(\boldsymbol{t})=\boldsymbol{t}^\bullet}\right)_{\boldsymbol{t}\in\mathbb{T}_{[n]},\boldsymbol{t}^\bullet\in\mathbb{T}^{\bullet n}_{[k]}}$.

Definition 2.1.3 (Decorated (α,γ)-chain)**.** Fix $n > k \in \mathbb{N}$. The *decorated (α,γ)-chain on* $\mathbb{T}^{\bullet n}_{[k]}$, is a Markov chain on $\mathbb{T}^{\bullet n}_{[k]}$ with transition kernel

$$K^{\bullet n}_k := \Pi^{\bullet n}_k K_n \pi^{\bullet n}_{[k]}. \tag{2.4}$$

The decorated (α,γ)-chain on $\mathbb{T}^{\bullet n}_{[k]}$ can be described in an autonomous way without referencing the transition kernel of the (α,γ)-chain. This is done in Proposition 2.6.7. By combining the Kemeny-Snell and intertwining criteria mentioned earlier we obtain another main result of this paper:

Theorem 2.1.4. *Fix $n \geq 2$. Let $(T_n(m))_{m\in\mathbb{N}_0}$ denote an (α,γ)-chain. For $1 \leq k < n$, let $T^{\bullet n}_{[k]}(m) := \pi^{\bullet n}_{[k]}(T_n(m))$, $m \in \mathbb{N}_0$, denote the projection of the chain onto decorated $[k]$-trees of mass n. If $T_n(0) \sim \Pi^{\bullet n}_k(\boldsymbol{t}^\bullet, \cdot)$ for some $\boldsymbol{t}^\bullet \in \mathbb{T}^{\bullet n}_{[k]}$, then $\left(T^{\bullet n}_{[k]}(m)\right)_{m\in\mathbb{N}_0}$ is a decorated (α,γ)-chain on $\mathbb{T}^{\bullet n}_{[k]}$ started from $\boldsymbol{t}^\bullet$.*

Theorem 2.1.4 makes it possible to vary k and obtain a consistent system of projected Markov chains all running in stationarity, provided that the (α,γ)-chain itself is running in stationarity.

Corollary 2.1.5 (Consistency in stationarity)**.** *Fix $n \geq 2$. Let $(T_n(m))_{m\in\mathbb{N}_0}$ denote an (α,γ)-chain running in stationarity. For $1 \leq k < n$, let $T^{\bullet n}_{[k]}(m) := \pi^{\bullet n}_{[k]}(T_n(m))$, $m \in \mathbb{N}_0$, denote the projection of the chain onto decorated $[k]$-trees of mass n. Then each of the $\left(T^{\bullet n}_{[k]}(m)\right)_{m\in\mathbb{N}_0}$ is a decorated (α,γ)-chain on $\mathbb{T}^{\bullet n}_{[k]}$ running in stationarity.*

We can extend the Wright-Fisher convergence result of (2.2) to the (α,γ)-chain projected to decorated $[k]$-trees of mass n:

Proposition 2.1.6. *Fix $k \geq 2$. For each $n \geq k$, let $\left(T^{\bullet n}_{[k]}(m)\right)_{m\in\mathbb{N}_0}$ denote the decorated (α,γ)-chain on $\mathbb{T}^{\bullet n}_{[k]}$ started from $\boldsymbol{t}^\bullet_n \in \mathbb{T}^{\bullet n}_{[k]}$, respectively, all having tree shape $\boldsymbol{s} \in \mathbb{T}_{[k]}$. Let $Y^x_n(m)$ denote the mass associated with the insertable part $x \in \mathrm{ins}(\boldsymbol{s})$ in $T^{\bullet n}_{[k]}(m)$. For each $n \geq k$, let S_n be the first time a leaf is selected for deletion where the mass associated to the*

corresponding leaf edge, $Y_n^e(\cdot)$*, is exactly* 1 *for any leaf edge* $e \in \mathrm{edge}(\boldsymbol{s})$*. If* $(Y_n^x(0))_{x\in\mathrm{ins}(\boldsymbol{s})}$ *converges in distribution as* $n \to \infty$ *then*

$$\left(\left(\frac{1}{n}Y_n^x(\lfloor n^2 t\rfloor)\right)_{x\in\mathrm{ins}(\boldsymbol{s})}\right)_{0\le t\le S_n} \xrightarrow{D} \left((Y^x(t))_{x\in\mathrm{ins}(\boldsymbol{s})}\right)_{0\le t\le S} \quad as \quad n\to\infty, \tag{2.5}$$

where $(Y^x)_{x\in\mathrm{ins}(\boldsymbol{s})}$ *is a Wright-Fisher diffusion with weights* $(\tilde{w}_x)_{x\in\mathrm{ins}(\boldsymbol{s})}$ *and* S *is the first time one of the components of the Wright-Fisher diffusion corresponding to the external edges vanishes. Here* $\tilde{w}_x = w_x - 1$ *whenever* x *is an external edge,* $\tilde{w}_x = w_x$ *otherwise, and* w_x *is defined by* (2.1)*.*

Remark 2.1.7. The generalization of the usual multi-type Wright-Fisher process to accept negative 'mutation' parameters can be found in [40]. The result follows by combining arguments from [40] with well-known results from Chapter 10 of [16].

Let $0 < \gamma \le \alpha < 1$. Let $\mathbb{T}_n^\circ$ denote the set of rooted multifurcating trees with n unlabelled leaves and consider the natural projection from $\mathbb{T}_{[n]}$ to $\mathbb{T}_n^\circ$. Then the (α,γ)-chain satisfies the Kemeny-Snell criterion, and the projected $\mathbb{T}_n^\circ$-valued Markov chain evolves by simply (i) selecting a uniform leaf, (ii) deleting it, and (iii) using the growth rule of the unlabelled (α,γ)-Markov branching model [12] to insert a new leaf. The stationarity of this can also be seen directly from the sampling consistency of this Markov branching model. In Corollary 3 of [12] it was shown that, when suitably represented as discrete $\mathbb{R}$-trees with edge lengths $n^{-\gamma}$, these stationary distributions converges weakly, in the Gromov-Hausdorff topology to the distribution of a continuum random tree $\mathcal{T}^{\alpha,\gamma}$. By Theorem 11 of [31] this can be strengthened to Gromov-Hausdorff-Prokhorov convergence of weighted $\mathbb{R}$-trees. In the light of (2.5) we make the following conjecture:

Conjecture 2.1.8. *The unlabelled* (α,γ)*-chain with* n^2 *steps per time unit and edges scaled by* n^γ *has a scaling limit that is a path-continuous Markov process in the Gromov-Hausdorff-Prokhorov space of weighted* $\mathbb{R}$*-trees, whose stationary distribution is that of* $\mathcal{T}^{\alpha,\gamma}$*.*

In the binary case where $\gamma = \alpha$, this conjecture strengthens results by Löhr et al. [36], in the case $\alpha = \frac{1}{2}$, and by Nussbaumer and Winter [39], who use a state space of algebraic

trees with a different topology that, in some sense, disregards the edge lengths.

The main new technique of this paper is the *semi-planar* (α, γ)*-growth process* and a down-up chain on what we will call *semi-planar trees*, which will be two auxiliary processes that we use to derive the results stated above. Crucially, both the (α, γ)-growth process and the (α, γ)-chain arise as projections of their semi-planar counterpart. We define the space of semi-planar trees to be trees where each branch point is equipped with a left-to-right ordering of its subtrees (as a planar tree), where we insist that the two subtrees with the smallest leaf labels are placed leftmost. The processes we define on the space of semi-planar trees are furthermore of mathematical interest in their own right. The semi-planar (α, γ)-growth process is defined similarly to the (α, γ)-growth process, but the weight $(c-1)\alpha-\gamma$ in a branch point with c children is split up so that there is weight α to the left of any subtree not leftmost, and weight $\alpha-\gamma$ to the right of the rightmost subtree. (α, γ)-growth processes are known to have intimate links to random partitions induced by Chinese restaurant processes [4, 42] and ordered Chinese restaurant processes [43], see e.g. [9, 12, 43], and the semi-planar trees provide exactly the right structure to further hook onto this link efficiently.

Compared to a non-planar tree, the additional structure will allow us to define a local search (as in step (ii) of the α-chain) that coincides with the local search from the α-chain in the cases where the leaf i selected for deletion in step (i) is located in a binary branch point. The key aspect of the down-step is that projecting the (α, γ)-chain onto the space of (non-planar) $[n]$-trees using the natural surjective projection yields exactly the (α, γ)-chain of Definition 2.1.1, if the initial distribution is picked according to an appropriate Markov kernel similar to (2.3).

The paper is structured as follows:

Section 2 contains a brief overview of notation and structures relied on throughout the exposition. Most notable are urn schemes, (un)ordered Chinese Restaurant Processes, basic tree growth processes and a general setup for tree-valued down-up chains. We will also cover notation for operations on trees such as deletion, insertion and swapping of leaves.

These fundamental notions will be updated to various tree spaces in later sections.

In Section 3 we develop the notion of semi-planar trees, construct and investigate properties of a semi-planar (α, γ)-growth process, and in Section 4 we go on to defining the semi-planar down-up chain and prove that it has a suitable stationary distribution.

The projection of the semi-planar down-up chain onto non-planar trees is carried out in Section 5, and Theorem 2.5.4 yields that this is the (α, γ)-chain of Definition 2.1.1. This section outlines the methodology that will be used in later sections.

In Section 6, an in-depth description of the behaviour of the semi-planar down-up chain projected to the space of decorated trees is provided.

Section 7 provides a selection of key results characterizing the link between projections of the semi-planar down-up chain and the semi-planar down-up chain, and Section 8 pulls together the threads from Section 5, 6, and 7, and proves Theorem 2.1.4.

Acknowledgements. This research is funded by the Lundbeck Foundation (R263-2017-3677) and the Engineering and Physical Sciences Research Council (1930434).

2.2 Preliminaries

2.2.1 Urn schemes

Urn schemes are intimately connected with Chinese Restaurant Processes and other exchangeable processes, and this connection will play a crucial role throughout this exposition. We will here give a brief account of properties of the generalized *Pólya urn*.

Definition 2.2.1 (Generalized Pólya urn)**.** Fix $k \in \mathbb{N}$ and $\mathbf{w} = (w_j)_{j\in[k]}$ where $w_j \geq 0$ for each $j \in [k]$ and $\sum_{j=1}^{k} w_j > 0$. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables taking values in $[k]$ such that

$$\forall j \in [k]\colon \quad \mathrm{P}\left(X_{n+1} = j \mid N_1^n = n_1, \ldots, N_k^n = n_k\right) = \frac{w_j + n_j}{\sum_{i\in[k]} w_i + n}, \tag{2.6}$$

where $N_j^n = \sum_{j'=1}^n 1_{\{X_{j'}=j\}}$ for each $j \in [k]$ and $n \in \mathbb{N}_0$. We call $(X_n)_{n\in\mathbb{N}}$ a *generalized Pólya urn with initial weights* $\mathbf{w}$.

More loosely, we will use the term *urn scheme* for any countable sequence of random variables governed by (2.6). Before summarizing some of the properties of the above urn scheme, let us briefly recall a few distributions related to urn schemes:

Definition 2.2.2 (Dirichlet-multinomial distribution)**.** Fix $k \in \mathbb{N}$, $n \in \mathbb{N}$, and a vector $\mathbf{w} = (w_j)_{j\in[k]}$ where $w_j > 0$. Let $\mathbf{N} = (N_1, \ldots, N_k)$ be a random variable taking values on $\{(n_1, \ldots, n_k) \in \mathbb{N}_0^k \mid \sum_{j=1}^k n_j = n\}$. If

$$\mathrm{P}\left(\mathbf{N} = (n_1, \ldots, n_k)\right) = \frac{n!\,\Gamma\left(\sum_{j=1}^k w_j\right)}{\Gamma\left(n + \sum_{j=1}^k w_j\right)} \prod_{j=1}^k \frac{\Gamma(n_j + w_j)}{n_j!\,\Gamma(w_j)} \tag{2.7}$$

we say that $\mathbf{N}$ has a *Dirichlet-multinomial distribution with n trials and weights* $\mathbf{w}$, and denote this by $\mathbf{N} \sim \mathrm{DM}^n(\mathbf{w})$. For $k = 2$, we will write $N_1 \sim \mathrm{DM}^n(w_1, w_2)$, since $N_2 = n - N_1$. This distribution is also known as the *Beta-binomial distribution*, denoted by $\mathrm{BetaBin}^n(w_1, w_2)$.

We now summarize a couple of properties which will be of value later:

Proposition 2.2.3. *Fix $k \in \mathbb{N}$ and $\mathbf{w} = (w_1, \ldots, w_k)$ where $w_j > 0$ for each $j \in [k]$, and let $X = (X_n)_{n\in\mathbb{N}}$ be a generalized Pólya urn with weights $\mathbf{w}$. Furthermore, for each $j \in [k]$ and $n \in \mathbb{N}$ let N_j^n be as in Definition 2.2.1. Then*

(i) X is exchangeable,

(ii) $(N_1^n, \ldots, N_k^n) \sim \mathrm{DM}^n(\mathbf{w})$,

(iii) $N_j^n \sim \mathrm{DM}^n\left(w_j, \sum_{i\neq j} w_i\right) = \mathrm{BetaBin}^n\left(w_j, \sum_{i\neq j} w_i\right)$ for each $j \in [k]$,

(iv) $\mathrm{P}\left(X_1 = j \mid N_1^n = n_1, \ldots, N_k^n = n_k\right) = \frac{n_j}{n}$ for any $j \in [k]$.

2.2.2 Chinese Restaurant Processes

We briefly recall the concept of Chinese Restaurant Processes.

Definition 2.2.4. Fix α and θ such that either $0 \leq \alpha \leq 1$ and $\theta > -\alpha$, or $\alpha < 0$ and $\theta = l_{\max}\alpha$ for some $l_{\max} \in \mathbb{N}$. The seating plan for the customers in a *Chinese Restaurant Process* with parameters (α, θ), an (α, θ)-CRP, is as follows:

- The first customer sits at the first table.
- If the first n customers are seated at L tables (where $L \leq l_{\max}$ in the case where $\alpha < 0$ and $\theta = l_{\max}\alpha$) with $N_1, \ldots, N_L$ customers seated at each table, customer $n+1$ will sit at
  - table l with probability $\frac{N_l - \alpha}{n+\theta}$, for each $l \in [L]$, and
  - at a new table, $L+1$, with probability $\frac{L\alpha+\theta}{n+\theta}$.

The table-seating of the first n customers naturally induces a partition of $[n]$ by saying that j and j' belong to the same block of the partition if and only if customers j and j' are seated at the same table in the Chinese Restaurant. It is well-known that in an (α, θ)-CRP the number of customers seated at the first table, N_1, after having seated n customers in total, satisfies that $N_1 \sim \text{BetaBin}^n(1-\alpha, \alpha+\theta)$. For a more extensive account of this process, see [42].

We can define an *ordered Chinese Restaurant Process* with parameters (α, θ) for $0 \leq \alpha \leq 1$ and $\theta \geq 0$, denoted by $\text{oCRP}(\alpha, \theta)$, in a similar fashion. The seating dynamic is the same but we now associate a left-right ordering to the tables. For any number of tables $L \in \mathbb{N}$, conditional on customer $n+1$ sitting at a new table, $L+1$, place this table

- to the left of the leftmost table with probability $\frac{\alpha}{L\alpha+\theta}$,
- between any pair of neighbouring tables with probability $\frac{\alpha}{L\alpha+\theta}$, and
- to the right of the rightmost table with probability $\frac{\theta}{L\alpha+\theta}$.

This construction not only gives rise to a random partition of $[n]$ but also a random permutation of $[L]$, Σ, corresponding to the ordering of the L tables seating n customers.

It holds that

$$\mathrm{P}\left(\Sigma = \sigma\right) = \left(\frac{\theta}{\alpha}\right)^{R(\sigma)} \frac{\Gamma\left(\frac{\theta}{\alpha}\right)}{\Gamma\left(\frac{\theta}{\alpha} + L\right)} \tag{2.8}$$

for every permutation σ of $[L]$, where

$$R(\sigma) = \#\left\{j \in [L] \;\middle|\; \sigma(j) > \sigma(j') \text{ for all } 1 \leq j' < j\right\}$$

is the number of record values in σ. Whenever we wish to refer to the distribution of a random partition induced by an (α, θ)-CRP or (α, θ)-oCRP with n customers, we will denote these distributions by $\mathrm{CRP}^n(\alpha, \theta)$ and $\mathrm{oCRP}^n(\alpha, \theta)$, respectively.

We have the following result:

Proposition 2.2.5 (Proposition 6 of [43])**.** *For each $n \in \mathbb{N}$, let $\mathbf{N}_n = (N_1^n, \ldots, N_k^n)$ denote the vector of the number of customers seated in an (α, θ)-*oCRP *with n customers, enumerated in the table order of appearance. It holds that the corresponding sequence $(\mathbf{N}_n)_{n\in\mathbb{N}}$ is regenerative, i.e. that given the first part of $\mathbf{N}_n$ is n_1, then the remaining parts of $\mathbf{N}_n$ sum to $n - n_1$ and have the same distribution as $\mathbf{N}_{n-n_1}$. Furthermore,*

$$\mathrm{P}\left(\mathbf{N}_n = (n_1, \ldots, n_k)\right) = \prod_{i=0}^{k-1} q_{\alpha,\theta}\left(n - \sum_{i'=1}^{i} n_{i'}, n_{i+1}\right) \tag{2.9}$$

where the decrement matrix, $q_{\alpha,\theta}$, is given by

$$q_{\alpha,\theta}(n, m) = \binom{n}{m} \frac{(n-m)\alpha + m\theta}{n} \frac{\Gamma(m-\alpha)}{\Gamma(1-\alpha)} \frac{\Gamma(n-m+\theta)}{\Gamma(n+\theta)}. \tag{2.10}$$

For more information regarding (α, θ)-oCRP and their role in regenerative tree growth processes, see [43].

2.2.3 Trees

Let $\mathbb{T}_A$ denotes the space of (non-planar) rooted, multifurcating trees with leaves labelled by the elements of A, and call an element of $\mathbb{T}_A$ an A-tree. For each $j \in A$ we will refer

to the leaf labelled by j as *leaf* j, whilst denoting the root vertex by $\emptyset$. Let $\mathrm{edge}(\boldsymbol{t})$ and $\mathrm{vert}(\boldsymbol{t})$ denote the set of edges and vertices of $\boldsymbol{t} \in \mathbb{T}_A$, respectively.

Notation 2.2.6. Fix a finite set A and $\boldsymbol{t} \in \mathbb{T}_A$.

(i) Let $e \in \mathrm{edge}(\boldsymbol{t})$ be an edge between $u, v \in \mathrm{vert}(\boldsymbol{t})$. If u is closer to the root, $\emptyset$, than v, in the graph distance, we say that u is the *parent* of v, and that v is the *child* of u. For brevity, we will sometimes write $u \prec_{\boldsymbol{t}} v$ if u is the parent of v in $\boldsymbol{t}$. We will write e_{uv} if there is any ambiguity about which edge we are discussing, and in cases where we do not wish to assign a name to the parent of v, we will simply denote the edge between v and its parent by $e_{v\downarrow}$. For the unique edge connected to root we will write $e_{\emptyset\uparrow}$, and, to ease notation, we shall simply refer to a leaf edge by the label of the associated leaf.

(ii) If $v_1, v_2 \in \mathrm{vert}(\boldsymbol{t})$ have the same parent, we say that they are *siblings*.

(iii) We say that a vertex that does not have any children is a *leaf*, and will refer to an edge connected to a leaf as a *leaf edge* or an *external edge*. It is worth noting that the root is not a leaf, and importantly, we do not consider the edge $e_{\emptyset\uparrow}$ external. We will refer to any vertex apart from the root that is not a leaf as a *branch point*, and refer to the collection of these by $\mathrm{bp}(\boldsymbol{t})$. To ease notation later on, we will define *insertable parts* of the tree as $\mathrm{ins}(\boldsymbol{t}) := \mathrm{bp}(\boldsymbol{t}) \cup \mathrm{edge}(\boldsymbol{t})$.

(iv) The *ancestral line* from a vertex $v \in \mathrm{vert}(\boldsymbol{t})$ is the collection of vertices $v_0, \ldots, v_l$, and edges, $e_1, \ldots, e_l$, such that $v_0 = v$, $v_l = \emptyset$, whilst for all $j \in [l]$ it holds that v_j is the parent of v_{j-1} with $e_j = e_{v_{j-1}v_j}$.

Now fix $\boldsymbol{t} \in \mathbb{T}_A$ and some $i \in A$. If $v_0, \ldots, v_l$ and $\left(e_{v_{j-1}v_j}\right)_{j\in[l]}$ denotes the ancestral line from leaf i, this naturally splits $\boldsymbol{t}$ into the ancestral line and a collection of *spinal bushes* each consisting of a collection of *subtrees*. To make this precise, note how for each $j \in [l-1]$ there is a collection of subtrees, $\boldsymbol{b}_j = \left(\boldsymbol{t}_{j,j'}\right)_{j'\in[c_{v_j}-1]}$, rooted at v_j, such that each of them only contains descendants of v_j that are not also descendants of v_{j-1}. Here c_{v_j} denotes the number of children of v_j. We call this collection of subtrees a *spinal bush*, and will refer to the spinal bush located at v_j as the *jth spinal bush on the ancestral line from leaf i* for each $j \in [l-1]$. Note how the ancestral line from i and the collection of the

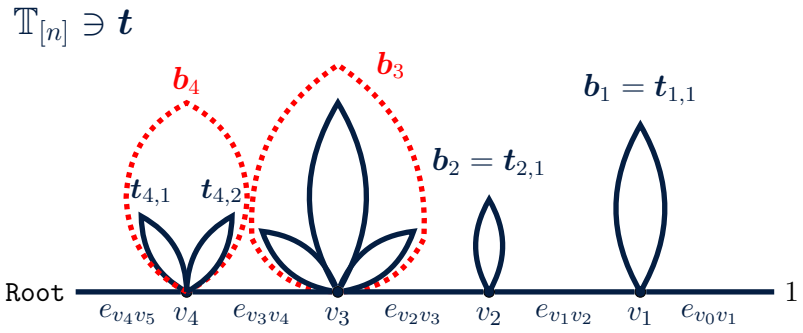


Figure 2.1: Illustration of the spinal decomposition from leaf 1 *of an* $[n]$*-tree,* $\boldsymbol{t}$*. There are four branch points separating leaf* 1 *from the root,* $v_1, \dots, v_4$*. Branch points* v_1 *and* v_2 *are binary, and so there is just a single subtree (constituting its own spinal bush) rooted at each of them that does not contain any edges or vertices from the ancestral line. Both of the branch points* v_3 *and* v_4 *are multifurcating with the spinal bushes rooted at these vertices containing three and two subtrees that do not contain any edges or vertices from the ancestral line, respectively. Together with the ancestral line, the spinal bushes exhaust* $\boldsymbol{t}$*.*

spinal bushes are disjoint apart from the root-vertices of the subtrees in the spinal bushes, and exhaust $\boldsymbol{t}$, and are thus referred to as the *spinal decomposition of* $\boldsymbol{t}$ *from* i.

2.2.4 Tree growth processes

In order to define tree growth processes we first need to introduce various operations on trees:

Definition 2.2.7 (Insertion of a leaf)**.** Let $A \subset \mathbb{N}$ be a finite set, fix $\boldsymbol{s} \in \mathbb{T}_A$, $j \in \mathbb{N} \setminus A$, and $x \in \mathrm{ins}(\boldsymbol{s})$. The operation of inserting leaf j to x in $\boldsymbol{s}$ is the map $(\boldsymbol{s}, x, j) \overset{\varphi}{\mapsto} \boldsymbol{t} \in \mathbb{T}_{A\cup\{j\}}$ where $\boldsymbol{t}$ is defined as follows.

- If $x = e_{vw} \in \mathrm{edge}(\boldsymbol{s})$, $\boldsymbol{t}$ is obtained by adding new vertices v' and j to $\mathrm{vert}(\boldsymbol{s})$, and replacing the edge e_{vw} with the three edges $e_{vv'}, e_{v'w}$, and $e_{jv'}$.
- If $x = v \in \mathrm{bp}(\boldsymbol{s})$, $\boldsymbol{t}$ is obtained by adding a new vertex, j, to $\mathrm{vert}(\boldsymbol{s})$, and an edge, e_{jv}, to $\mathrm{edge}(\boldsymbol{s})$.

Definition 2.2.8 (Deletion of a leaf)**.** Let A be a finite set and fix $\boldsymbol{t} \in \mathbb{T}_A$. The operation of deleting leaf $j \in A$ from $\boldsymbol{t}$ is the map

$$(\boldsymbol{t}, j) \overset{\mathring{\rho}}{\mapsto} \boldsymbol{s} \in \mathbb{T}_{A\setminus\{j\}}$$

where $\boldsymbol{s}$ is obtained in the following way. Let v be the parent of j and let c_v be the number of children of v. If $c_v \geq 3$, $\boldsymbol{s}$ is obtained by removing j from $\mathrm{vert}(\boldsymbol{t})$ and e_{jv} from $\mathrm{edge}(\boldsymbol{t})$. If $c_v = 2$, let w be the sibling of j, v' the parent of v, and obtain $\boldsymbol{s}$ by removing j and v from $\mathrm{vert}(\boldsymbol{t})$ and $e_{wv}, e_{vv'}$, and e_{jv} from $\mathrm{edge}(\boldsymbol{t})$ and subsequently adding $e_{wv'}$ to the latter. We denote by ρ the map that, after applying $\mathring{\rho}$, subsequently relabels $A \setminus \{j\}$ by $[\#A - 1]$ using the increasing bijection.

If we wish to delete all leaves of $\boldsymbol{t} \in \mathbb{T}_A$ labelled by the elements of $B \subsetneq A$, we will denote this by $\mathring{\rho}(\boldsymbol{t}, B)$ and $\rho(\boldsymbol{t}, B)$, respectively, obtaining the resulting tree by successively applying the above definition to the elements of B.

Throughout this exposition we will need the following notion from [43]:

Definition 2.2.9 (Tree growth process)**.** A Markov chain, $(T_n)_{n\in\mathbb{N}}$, is a *tree growth process* if T_n takes values in $\mathbb{T}_{[n]}$ and $\rho\left(T_{n+1}, n+1\right) = T_n$ for each $n \in \mathbb{N}$.

Equivalently, one could define a tree growth process in terms of a sequence of Markov kernels, $G = (G_n)_{n\in\mathbb{N}}$, where G_n is a Markov kernel from $\mathbb{T}_{[n]}$ to $\mathbb{T}_{[n+1]}$. We will refer to G as a *growth rule*. The main tree growth process of interest to us is the following:

Definition 2.2.10 ((α,γ)-growth process)**.** Fix $0 \leq \gamma \leq \alpha \leq 1$. Let T_1 be the unique element of $\mathbb{T}_1$. For any $n \in \mathbb{N}$ construct T_{n+1} conditional on $T_n = \boldsymbol{s} \in \mathbb{T}_{[n]}$, by, for each $x \in \mathrm{ins}(\boldsymbol{s})$, setting

$$w_x = \begin{cases} 1-\alpha & \text{if } x \in \mathrm{edge}\left(\boldsymbol{s}\right) \text{ is external,} \\ \gamma & \text{if } x \in \mathrm{edge}\left(\boldsymbol{s}\right) \text{ is internal,} \\ \left(c_x - 1\right)\alpha - \gamma & \text{if } x \in \mathrm{bp}(\boldsymbol{s}) \text{ with } c_x \text{ children,} \end{cases}$$

and then defining the conditional distribution of T_{n+1} by

$$\mathrm{P}\left(T_{n+1} = \varphi\left(\boldsymbol{s}, x, n+1\right) \mid T_n = \boldsymbol{s}\right) = \frac{w_x}{n-\alpha}.$$

The sequence $(T_n)_{n\in\mathbb{N}}$ is referred to as the (α,γ)-*growth process*.

The (α,γ)-growth process was first introduced in [12], and encompasses several other well-known growth processes. For $\gamma = 1-\alpha$ and $\frac{1}{2} \leq \alpha < 1$ the above growth process is

Marchal's stable tree growth [37]. For $0 \leq \gamma = \alpha \leq 1$ it is identical to Ford's α-model [20]. And, as mentioned previously, a special case of Ford's α-model is Rémy's uniform binary tree growth process [44], which arises from choosing $\gamma = \alpha = \frac{1}{2}$. One property of the latter growth process is that T_n will be uniformly distributed on $\mathbb{T}_{[n]}^{\text{bin}}$, the space of binary trees with n leaves [7], for each $n \in \mathbb{N}$. We will refer to the growth rule governing the (α, γ)-growth process as the (α, γ)-growth rule.

2.2.5 Down-up chains

Let us now introduce a general notion of down-up chain for trees that contains not only the Aldous down-up chain [7] but also the down-up chain proposed in [21]. The definition of down-up chains given here is in accordance with down-up and up-down chains on more general state spaces studied in detail in [10, 27, 28, 41].

Definition 2.2.11 (Down-up chains)**.** Fix $n \in \mathbb{N}$, $\boldsymbol{t} \in \mathbb{T}_{[n]}$, some growth rule, G, and a *transformation function* $f \colon \mathbb{T}_{[n]} \times [n] \to \mathbb{T}_{[n]} \times [n]$. The (f, G)*-down-up chain,* $(T_n(m))_{m \in \mathbb{N}_0}$*, started at* $\boldsymbol{t}$ is a Markov chain on $\mathbb{T}_{[n]}$ with $T_n(0) = \boldsymbol{t}$ and with the following transition kernel. For each $m \in \mathbb{N}_0$ obtain $T_n(m+1)$ from $T_n(m)$ by

(i) applying the transformation function, f, to $T_n(m)$ and a random variable $I \sim \text{Unif}([n])$, thereby defining

$$\left(T_n'(m), \tilde{I}\right) = f\left(T_n(m), I\right),$$

(ii) deleting the leaf labelled $\tilde{I}$ from $T_n'(m)$, and relabel $\{\tilde{I}+1, \ldots, n\}$ by $\{\tilde{I}, \ldots, n-1\}$ using the increasing bijection, i.e. defining

$$T_n^{\downarrow}(m) = \rho\left(T_n'(m), \tilde{I}\right) \in \mathbb{T}_{[n-1]},$$

(iii) and inserting leaf n to $T_n^{\downarrow}(m)$ according to G_{n-1} to obtain $T_n(m+1)$.

We will only work with the above type of down-up chains, but there are examples (see e.g. [21]) of down-up chains where there is no relabelling. In that case, ρ is replaced by $\mathring{\rho}$ in (ii), and (iii) is replaced with a step where the leaf $\tilde{I}$ is inserted directly into $T_n^{\downarrow}(m)$. This is particularly useful when the growth rule induces exchangeable leaf labels, which

is the case for Rémy's uniform tree growth and Marchal's stable tree growth, but not for other (α, γ)-growth processes.

Let us focus on the role of the transformation map f. In the above we require I to be uniform on $[n]$. As in step (i), say that $\left(\boldsymbol{s}', \tilde{I}\right) := f(\boldsymbol{s}, I)$. Even though I is uniformly distributed on the leaves of $\boldsymbol{s}$, it will be dependent on f whether or not $\tilde{I}$ is uniformly distributed on $\boldsymbol{s}'$. This will allow us to make non-uniform deletions, which will be of vital importance.

In the basic cases, f will simply be the identity function, in which case we will call the chain an (id, G)-down-up chain and then specify G. For our purposes, it will be important to use more complicated function, in which case we will characterize the output of f explicitly in terms of operations on $\boldsymbol{s}$ and the value of I. Furthermore, if G is the (α, γ)-growth rule we will denote the associated chains by $(f, (\alpha, \gamma))$-down-up chain, and then specify f.

We note that the (id, G)-down-up chain without relabelling, where G is the *uniform growth rule*, corresponds to Aldous' down-up chain [7]. To prepare a generalization of down-up chains related to the (α, γ)-growth process we need the following definition:

Definition 2.2.12 (Label swapping)**.** Let A be a finite set. The operation of *swapping labels $i \in A$ and $j \in A$ in $\boldsymbol{s} \in \mathbb{T}_A$* is the map

$$(\boldsymbol{s}, i, j) \overset{\tau}{\mapsto} \boldsymbol{t} \in \mathbb{T}_A$$

where, if we let v_i and v_j denote the parents of i and j, respectively, $\boldsymbol{t}$ is obtained from $\boldsymbol{s}$ by replacing e_{iv_i} and e_{jv_j} with e_{iv_j} and e_{jv_i}, respectively, in $\mathrm{edge}(\boldsymbol{s})$.

Let us discuss two examples in the binary setting studied in [24] and [21]. To this end, let $\boldsymbol{s} \in \mathbb{T}_{[n]}^{\mathrm{bin}}$, and define a transformation function, f, by setting $f(\boldsymbol{s}, i) = (\tau(\boldsymbol{s}, i, \tilde{\imath}), \tilde{\imath})$, where

$$a := \min\{\text{leaves in 1st subtree on the ancestral line from } i \text{ in } \boldsymbol{s}\},$$
$$b := \min\{\text{leaves in 2nd subtree on the ancestral line from } i \text{ in } \boldsymbol{s}\},$$

and $\tilde{\imath} := \max\{i, a, b\}$. Then the (f, G)-down-up chain where G is the uniform growth rule

is exactly the *uniform chain* on $\mathbb{T}^{\text{bin}}_{[n]}$ without relabelling described in [24]. If instead G is the α-growth rule, then the (f, G)-down-up chain with relabelling is the α-*chain* of [21].

Theorem 2.2.13 (Theorem 1 of [21]). *Let $(T_n(m))_{m\in\mathbb{N}_0}$ denote the uniform chain on $\mathbb{T}^{\text{bin}}_{[n]}$ and let T_n denote the nth step of the uniform growth process. Then the unique invariant distribution of the uniform chain is that of T_n, i.e. the uniform distribution on $\mathbb{T}^{\text{bin}}_{[n]}$.*

2.3 A semi-planar growth process

In order to introduce an appropriate down-step in the multifurcating case we start by introducing a refined version of the (α, γ)-growth process. In the following we will, for any $c \in \mathbb{N}$, let $\mathbb{S}_{[c]}$ denote the space of permutations of $[c]$. We will require the existence of an element of $\mathbb{S}_\emptyset$, and we will use the convention that this element is denoted by 0. Additionally, we will use the notation that $\mathbb{S}_{[c]} \ni \sigma = (l_j)_{j\in[c]}$ if $\sigma(j) = l_j$ for each $j \in [c]$. Recall one of the many ways of defining a planar tree:

Definition 2.3.1 (Planar Trees). Let A be a finite set and fix $\boldsymbol{t} \in \mathbb{T}_A$. Let $\boldsymbol{\sigma}^* = (\sigma^*_v)_{v\in\text{bp}(\boldsymbol{t})}$ be a collection of permutations, such that for each $v \in \text{bp}(\boldsymbol{t})$, $\sigma^*_v \in \mathbb{S}_{[c_v]}$, where c_v is the number of children of v. We call $\hat{\boldsymbol{t}}^* = (\boldsymbol{t}, \boldsymbol{\sigma}^*)$ a *planar tree* with *tree shape* $\boldsymbol{t}$, and denote by $\widehat{\mathbb{T}}^*_A$ the space of planar trees with leaves labelled by A.

Let us introduce some vocabulary for planar trees that will enable us to work efficiently with these objects. Let $\hat{\boldsymbol{t}}^* = (\boldsymbol{t}, \boldsymbol{\sigma}^*) \in \widehat{\mathbb{T}}^*_A$ be given, fix $v \in \text{bp}(\boldsymbol{t})$, and let $\hat{\boldsymbol{t}}^{*v}_1, \ldots, \hat{\boldsymbol{t}}^{*v}_{c_v}$ denote the planar subtrees rooted at v, enumerated in the order of least elements of the associated sets of leaf labels. For any $1 \le i \ne j \le c_v$ we say that $\hat{\boldsymbol{t}}^{*v}_i$ is to the *right* of $\hat{\boldsymbol{t}}^{*v}_j$ if $\sigma^*_v(i) > \sigma^*_v(j)$, and we use the term *right neighbour* if $\sigma^*_v(i) = \sigma^*_v(j) + 1$. From the above it is also clear what we mean by the terms *rightmost*, *left* and *leftmost* which are defined in the obvious fashion.

Any planar version of the (α, γ)-growth process will have to have a total weight of $\alpha - \gamma$ in a branch point with two children, and then iteratively adding a weight of α for every subsequent leaf being inserted into that branch point. Hence, in a branch point with c children there are only $c - 1$ natural weights to use. One way to incorporate this dynamic, is to designate the original two subtrees of the branch point a special role, and only allow insertions to the right of those subtrees. These two subtrees are, by definition of a growth

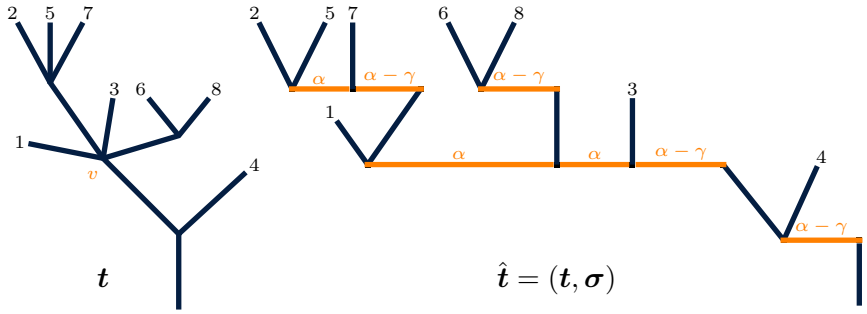


Figure 2.2: Illustration of a semi-planar tree with the allocation of weights within the branch points. The horizontal orange lines are expansions of the branch points illustrating the $c_v - 1$ insertable parts associated with branch point v. The only non-trivial element of $\boldsymbol{\sigma}$ is σ_v characterized by $\sigma_v(1) = 2$ and $\sigma_v(2) = 1$, and the only other semi-planar tree with tree shape $\boldsymbol{t}$, would come from swapping the position of the subtrees labelled by $\{3\}$ and $\{6, 8\}$, such that v would have the associated permutation $\sigma'_v = (1, 2)$.

process, easily identifiable as the two subtrees with the lowest leaf labels. If we insist that these two subtrees are located leftmost in the branch point, we can limit our attention to the following state space:

Definition 2.3.2 (Semi-planar Tree)**.** Fix $\boldsymbol{t} \in \mathbb{T}_A$ for a finite set A. Let $\boldsymbol{\sigma} = (\sigma_v)_{v \in \mathrm{bp}(\boldsymbol{t})}$ be a collection of permutations, such that for each $v \in \mathrm{bp}(\boldsymbol{t})$, σ_v is a permutation of $[c_v - 2]$ where c_v is the number of children of v. We call $\hat{\boldsymbol{t}} = (\boldsymbol{t}, \boldsymbol{\sigma})$ a *semi-planar tree*, and denote by $\widehat{\mathbb{T}}_A$ the space of semi-planar trees with leaves labelled by A.

For an illustration of an element of $\widehat{\mathbb{T}}_{[8]}$, see Figure 2.2. The space of semi-planar trees, $\widehat{\mathbb{T}}_A$, arises from $\widehat{\mathbb{T}}^*_A$ as follows:

(i) Define $\mathbb{S}$ to be the subset of planar trees containing elements $\hat{\boldsymbol{s}}^* \in \widehat{\mathbb{T}}^*_A$ such that $\hat{\boldsymbol{s}}^* = (\boldsymbol{s}, \boldsymbol{\sigma}^*)$ where for all $v \in \mathrm{bp}(\boldsymbol{s})$ and $l \in \{1, 2\}$ we have $\sigma^*_v(l) \in \{1, 2\}$.

(ii) Form a quotient space of $\mathbb{S}$ by identifying $\hat{\boldsymbol{t}}^*_1 = (\boldsymbol{t}, \boldsymbol{\sigma}^*)$ with $\hat{\boldsymbol{t}}^*_2 = (\boldsymbol{t}, \boldsymbol{\sigma}^{**})$ if $\sigma^*_v(l) = \sigma^{**}_v(l)$ for all $v \in \mathrm{bp}(\boldsymbol{t})$ and $3 \le l \le c_v$, where c_v is the number of children of v.

(iii) Construct $\widehat{\mathbb{T}}_A$ as the image space of the function $\mathbb{S} \ni \hat{\boldsymbol{t}}^* = (\boldsymbol{t}, \boldsymbol{\sigma}^*) \mapsto (\boldsymbol{t}, \boldsymbol{\sigma})$ where σ_v is a permutation of $[c_v - 2]$ such that $\sigma_v(l) = \sigma_v(l + 2) - 2$ for each $l \in [c_v - 2]$ and each $v \in \mathrm{bp}(\boldsymbol{t})$.

For each element $\hat{\boldsymbol{t}} = (\boldsymbol{t}, \boldsymbol{\sigma}) \in \widehat{\mathbb{T}}_A$ we will systematically pick a representative $\hat{\boldsymbol{t}}^* = (\boldsymbol{t}, \boldsymbol{\sigma}^*) \in \widehat{\mathbb{T}}^*_A$ by setting $\sigma^*_v(1) = 1$, $\sigma^*_v(2) = 2$ and $\sigma^*_v(l) = \sigma_v(l-2)+2$ for all $3 \le l \le c_v$ for each $v \in \mathrm{bp}(\boldsymbol{t})$, where c_v denotes the number of children of v.

It is clear that there is a surjective projection map from the space of semi-planar (or planar) trees to the space of trees, $\hat{\pi}\colon \widehat{\mathbb{T}}_A \to \mathbb{T}_A$ for any $A \subseteq \mathbb{N}$, defined by

$$\hat{\pi}\left(\hat{\boldsymbol{t}}\right) = \hat{\pi}\left(\boldsymbol{t}, \boldsymbol{\sigma}\right) = \boldsymbol{t}. \tag{2.11}$$

We will continue to use all the previous terminology such as edge set, set of branch points, and ancestral path for a semi-planar tree, $\hat{\boldsymbol{t}}$, by which we will mean the corresponding versions in $\hat{\pi}(\hat{\boldsymbol{t}})$. Before introducing a semi-planar version of the (α, γ)-growth process, let us update the notion of inserting a leaf to the semi-planar setting, recalling that φ denotes the insertion map of Definition 2.2.7.

Definition 2.3.3 (Insertion of a leaf in $\widehat{\mathbb{T}}_A$)**.** Let $A \subset \mathbb{N}$ be a finite set, $\hat{\boldsymbol{s}} = (\boldsymbol{s}, \boldsymbol{\sigma}) \in \widehat{\mathbb{T}}_A$, $j \in \mathbb{N}$, $x \in \mathrm{ins}(\boldsymbol{s})$ and $l \in \mathbb{N}_0$. The operation of inserting leaf j to x in location l in $\boldsymbol{s}$ is the map $(\boldsymbol{s}, x, l, j) \overset{\hat{\varphi}}{\mapsto} \hat{\boldsymbol{t}} \in \widehat{\mathbb{T}}_{A\cup\{j\}}$ defined as follows.

- If $x = e \in \mathrm{edge}(\boldsymbol{s})$, only $l = 0$ is allowed and $\hat{\boldsymbol{t}} = \left(\varphi\left(\boldsymbol{s}, e, j\right), \boldsymbol{\sigma}^{\varphi}\right)$ where $\sigma^{\varphi}_{v'} = 0$ if v' is the parent of leaf j in $\varphi(\boldsymbol{s}, e, j)$, and $\sigma^{\varphi}_{v} = \sigma_v$ for all other $v \in \mathrm{bp}(\varphi(\boldsymbol{s}, e, j))$;

- If $x = v' \in \mathrm{bp}(\boldsymbol{s})$ with $c \geq 2$ children, $l \in [c-1]$ is allowed and $\hat{\boldsymbol{t}} = \left(\varphi\left(\boldsymbol{s}, v', j\right), \boldsymbol{\sigma}^{\varphi}\right)$, where

$$\sigma^{\varphi}_{v'}(l') = \begin{cases} \sigma_{v'}(l') + 1\left(\sigma_{v'}(l') \geq l\right) & \text{for all } l' \in [c-2] \\ l & \text{for } l' = c-1, \end{cases}$$

  and $\sigma^{\varphi}_{v} = \sigma_v$ for all $v \in \mathrm{bp}(\boldsymbol{s}) \setminus \{v'\}$.

Just as in the previous section, we will write $\hat{\varphi}(\hat{\boldsymbol{s}}, x, j, l)$ in both of the above cases, and then specify $x \in \mathrm{edge}(\boldsymbol{s})$ or $x \in \mathrm{bp}(\boldsymbol{s})$ for the leaf insertion into the edge or vertex, respectively. Definition 2.3.3 simply states, that whenever we insert a new leaf into a vertex, we can insert it to the right of the rightmost or in between any two children, with the exeption of the two leftmost, of that vertex. We are now ready to define a semi-planar version of the (α, γ)-growth process:

Definition 2.3.4 (Semi-planar (α, γ)-growth process)**.** Fix $0 \leq \gamma \leq \alpha \leq 1$. Let $\widehat{T}_1$ be the unique element of $\widehat{\mathbb{T}}_1$. For any $n \in \mathbb{N}$ construct $\widehat{T}_{n+1}$ conditional on $\widehat{T}_n = \hat{\boldsymbol{t}}$, by setting

$$w_{x,l} = \begin{cases} 1-\alpha & \text{if } x \in \mathrm{edge}(\hat{\boldsymbol{t}}) \text{ is external and } l = 0, \\ \gamma & \text{if } x \in \mathrm{edge}(\hat{\boldsymbol{t}}) \text{ is internal and } l = 0, \\ \alpha & \text{if } x \in \mathrm{bp}(\hat{\boldsymbol{t}}) \text{ with } c_x \text{ children and } l \in [c_x - 2], \\ \alpha - \gamma & \text{if } x \in \mathrm{bp}(\hat{\boldsymbol{t}}) \text{ with } c_x \text{ children and } l = c_x - 1, \end{cases} \tag{2.12}$$

and then defining the conditional distribution of $\widehat{T}_{n+1}$ by

$$\mathrm{P}\left(\widehat{T}_{n+1} = \hat{\varphi}\left(\hat{\boldsymbol{t}}, x, j, l\right) \,\middle|\, \widehat{T}_n = \hat{\boldsymbol{t}}\right) = \frac{w_{x,l}}{n - \alpha} \tag{2.13}$$

for each $x \in \mathrm{ins}\left(\hat{\boldsymbol{t}}\right)$ and $l \in [c_x - 1]$ for $x \in \mathrm{bp}(\hat{\boldsymbol{t}})$ and $l = 0$ for $x \in \mathrm{edge}(\hat{\boldsymbol{t}})$. The sequence $\left(\widehat{T}_n\right)_{n \in \mathbb{N}}$ is referred to as the *semi-planar* (α, γ)*-growth process.*

The weights specified in (2.12) are such that the projection of the semi-planar (α, γ)-growth process from $\widehat{\mathbb{T}}_{[n]}$ onto $\mathbb{T}_{[n]}$ exactly yields the (non-planar) (α, γ)-growth process from Definition 2.2.10.

In the semi-planar growth process, one can think of the additional structure of a branch point as a small comb tree rooted at the right end and with a leaf for every subtree, where naturally the two leftmost leaves connect to the same branch point, and where internal edges between branch points have weight α and the edge to the root has weight $\alpha - \gamma$ (see Figure 2.2). In particular, we do not allow any insertions between or to the left of the two leftmost subtrees. If $\widehat{T}_n = \hat{\boldsymbol{t}} \in \widehat{\mathbb{T}}_{[n]}$, it thus holds for any $v \in \mathrm{bp}(\hat{\boldsymbol{t}})$ that if $i_1, \ldots, i_c$ denotes the least label in each of the subtrees rooted at v, enumerated in the order of least element, then i_1 and i_2 are always located in the two leftmost subtrees.

This is not the first definition of a binary tree growth process within the (α, γ)-chain. In [12], the authors use a colouring scheme of the edges to make the translation between the multifurcating and the binary setting, whereas we work with a semi-planar structure for the multifurcating trees.

Let us now investigate the left-to-right ordering of the subtrees in a branch point further. For each $v \in \mathrm{bp}(\hat{\boldsymbol{t}})$, let $i_1, \ldots, i_{c_v}$ denote the smallest leaf label in each of the c_v

subtrees rooted at v, enumerated in the order of least element, and let $\hat{\boldsymbol{t}}_1, \ldots, \hat{\boldsymbol{t}}_{c_v}$ denote the corresponding subtrees of $\hat{\boldsymbol{t}}$. Then, by definition of semi-planar trees, $\hat{\boldsymbol{t}}_1$ and $\hat{\boldsymbol{t}}_2$, are located leftmost in v. The semi-planar (α, γ)-growth process assigns a random location to each of the remaining $c_v - 2$ subtrees within the branch point, by inducing a permutation of $[c_v - 2]$, σ_v, associated with v. Letting Σ_v denote the random permutation induced by the growth process associated with v, and disregarding the two leftmost subtrees, we have that $\hat{\boldsymbol{t}}_l$ is the $\Sigma_v(l-2)$ leftmost subtree in v, for each $3 \leq l \leq c_v$.

We can characterize Σ_v in the following way:

Lemma 2.3.5. *Let $\widehat{T}_n$ be the nth step of the semi-planar (α, γ)-growth process, and fix $\boldsymbol{t} \in \mathbb{T}_{[n]}$. Then $(\Sigma_v)_{v \in \mathrm{bp}(\boldsymbol{t})}$ is conditionally independent given $\hat{\pi}\left(\widehat{T}_n\right) = \boldsymbol{t}$, and*

$$\forall v \in \mathrm{bp}(\boldsymbol{t}): \quad \mathrm{P}\left(\Sigma_v = \sigma \;\middle|\; \hat{\pi}\left(\widehat{T}_n\right) = \boldsymbol{t}\right) = p^{c_v-2}_{\alpha,\alpha-\gamma}(\sigma), \tag{2.14}$$

where $p^{c_v-2}_{\alpha,\alpha-\gamma}$ is the probability function of the random permutation of $[c_v - 2]$ induced by the random order of the tables in an $(\alpha, \alpha-\gamma)$-oCRP with $c_v - 2$ tables, described by (2.8).

Proof. The assertion that $(\Sigma_v)_{v \in \mathrm{bp}\,\boldsymbol{t}}$ is conditionally independent given $\hat{\pi}\left(\widehat{T}_n\right) = \boldsymbol{t}$ is a consequence of the independence of the leaf insertions. Now fix $v \in \mathrm{bp}(\boldsymbol{t})$, and observe how the subtrees rooted at v are analogous to the tables in an ordered Chinese Restaurant. Firstly, note that there is nothing to prove for $c_v = 2$, so fix $c_v = 3$. By construction the two subtrees with the smallest labels will be located leftmost in v and there will be a single subtree to the right of them. That subtree will have weight $\alpha - \gamma$ to the right and α to the left. This corresponds exactly to the $(\alpha, \alpha-\gamma)$-oCRP with 1 table. For any other $c_v > 3$, let us disregard the two leftmost subtrees, only focus on the remaining $c_v - 2$, and consider how these have been placed by the growth process. Each insertion corresponds to a new customer in the ordered Chinese Restaurant analogy, and each subtree corresponds to a table. Amongst these subtrees, there will be weight α to the left of the leftmost subtree as well as in between any neighbouring pair. Furthermore there will be weight $\alpha - \gamma$ to the right of the rightmost subtree. This is exactly the same weight specification as in an $(\alpha, \alpha-\gamma)$-oCRP with $c_v - 2$ tables, which finishes the proof. ∎

Lemma 2.3.5 naturally characterizes a conditional distribution on $\widehat{\mathbb{T}}_{[n]}$. Fix some $\boldsymbol{t} \in \mathbb{T}_{[n]}$ and, for each $v \in \mathrm{bp}(\boldsymbol{t})$, let c_v denote the number of children of v. Then

$$\begin{aligned} \mathrm{P}\left(\widehat{T}_n = \hat{\boldsymbol{t}} \;\middle|\; \hat{\pi}\left(\widehat{T}_n\right) = \boldsymbol{t}\right) &= \prod_{v \in \mathrm{bp}(\boldsymbol{t})} \mathrm{P}\left(\Sigma_v = \sigma_v \;\middle|\; \hat{\pi}\left(\widehat{T}_n\right) = \boldsymbol{t}\right) \\ &= \prod_{v \in \mathrm{bp}(\boldsymbol{t})} p_{\alpha,\alpha-\gamma}^{c_v-2}\left(\sigma_v\right). \end{aligned} \tag{2.15}$$

2.4 A semi-planar down-up chain

The motivation for introducing the semi-planar version of the (α, γ)-growth model was fundamentally to get an ordering of the subtrees within a multifurcating branch point. This was done to enable a generalization of the α-chain described in the introduction, by generalizing the local search from a designated leaf to a multifurcating setting where there might be more than one subtree to search in the first branch point on the ancestral line.

To avoid any ambiguity, we will update the notions of swapping and deleting leaves, respectively, before defining a semi-planar down-up chain. The sole reason for defining the label swapping map $\hat{\tau}$ on the space of planar trees, $\widehat{\mathbb{T}}_A^*$, rather than directly on $\widehat{\mathbb{T}}_A$, is that certain label swaps will bring us outside the latter state space. However, the definition can easily be applied to elements of $\widehat{\mathbb{T}}_A$ by using the identification described just after Definition 2.3.2. See Figure 2.3 for an illustration. Recall that τ denotes the label swapping map on $\mathbb{T}_{[n]}$ from Definition 2.2.12.

Definition 2.4.1 (Label swapping in $\widehat{\mathbb{T}}_A^*$)**.** Let A be a finite set and fix $\hat{\boldsymbol{t}}^* = (\boldsymbol{t}, \boldsymbol{\sigma}^*) \in \widehat{\mathbb{T}}_A^*$. The operation of *swapping labels* is a map $\left(\hat{\boldsymbol{t}}^*, i, j\right) \overset{\hat{\tau}}{\mapsto} (\tau(\boldsymbol{t}, i, j), \boldsymbol{\sigma}^{*\tau})$ for any pair $i, j \in [n]$ that satisfies that if $v \in \mathrm{bp}(\boldsymbol{t})$ is the parent of i, and $i_1, \ldots, i_{c_v}$ denotes the smallest leaf label of the subtrees of v in $\boldsymbol{t}$, enumerated in increasing order, then $j \in \{i_1, \ldots, i_{c_v}\}$ if $c_v > 2$ or $j \in \{i_1, i_2, b\}$ if $c_v = 2$ where b is the smallest leaf label in the second spinal bush on the ancestral line from i. If $c_v > 2$, $i = i_{l_i}$ and $j = i_{l_j}$ some $l_i, l_j \in [c_v]$, then for each

$l \in [c_v]$ let

$$\sigma_v^{*\tau}(l) = \begin{cases} \sigma_v^*(l_i) & \text{if } l = l_j, \\ \sigma_v^*(l+1) & \text{if } l_i \leq l < l_j, \\ \sigma_v^*(l-1) & \text{if } l_j < l \leq l_i, \\ \sigma_v^*(l) & \text{otherwise,} \end{cases}$$

and $\sigma_{v'}^{*\tau} = \sigma_{v'}^*$ for all other $v' \in \mathrm{bp}(\boldsymbol{t})$. If $c_v \leq 2$, let w denote the grandparent of i, and define σ_w^* analogously to the above, with the slight alteration that l_i and l_j denotes the ranks of the subtrees containing i and j, respectively, in w, in the order of least elements.

The above definition seems technical, but it simply states that all subtrees in v_i and v_j keep the positions, with the subtree containing j in v_j taking the position of the subtree containing i in v_i and vice versa. The label swapping will be used prior to deleting a leaf in the semi-planar down-up chain defined in this section. Swaps like this will occur in the down-up chain, but the subsequent deletion of a leaf will make sure we end up with a semi-planar tree again.

Definition 2.4.2 (Deletion of a leaf in $\widehat{\mathbb{T}}_A^*$)**.** Let A be a finite set and fix $\hat{\boldsymbol{t}}^* \in \widehat{\mathbb{T}}_A^*$. The

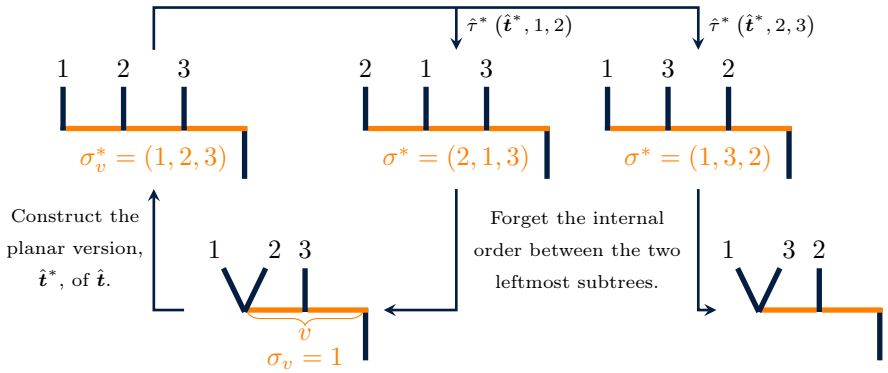


Figure 2.3: Swapping labels in $\widehat{\mathbb{T}}_{[3]}^*$. The orange "edges" is an illustrative expansion of the branch point v so that the left-to-right ordering of the subtrees of v becomes apparent. Thus, all semi-planar and planar trees depicted above have the same tree shape, but have different permutations associated to v. Label swapping in $\hat{\boldsymbol{t}}$ (bottom left) occurs by constructing the planar version, $\hat{\boldsymbol{t}}^*$ (top left), and then swapping the labels in the planar version. Swapping labels 1 and 2 yields a planar tree corresponding to a semi-planar tree since the two subtrees with the lowest labels are located leftmost. This is not the case if we swap labels 2 and 3, and thus the bottom right is not a semi-planar tree, i.e. not an element of $\widehat{\mathbb{T}}_{[3]}$.

operation of deleting leaf $j \in A$ from $\hat{\boldsymbol{t}}^*$ is the map

$$\left(\hat{\boldsymbol{t}}^*, j\right) = (\boldsymbol{t}, \boldsymbol{\sigma}^*, j) \overset{\mathring{\hat{\rho}}^*}{\mapsto} \left(\mathring{\rho}\left(\boldsymbol{t}, j\right), \boldsymbol{\sigma}^{*\rho}\right) \in \widehat{\mathbb{T}}^*_{A\setminus\{j\}} \tag{2.16}$$

where $\sigma_u^{*\rho} = \sigma_u^*$ for all $u \neq v$, v is the parent of j in $\hat{\boldsymbol{t}}$, and, if $c_v > 2$,

$$\sigma_v^{*\rho}(i) = \sigma_v^*(i) - 1\left(\sigma_v^*(i) > \sigma_v^*(i_j)\right) \tag{2.17}$$

for all $i \in [c_v - 1]$, where c_v is the number of children of v in $\hat{\boldsymbol{t}}$, and i_j is the rank of j amongst the smallest leaf labels of the subtrees rooted at v in $\hat{\boldsymbol{t}}$. As in Definition 2.2.8, we denote by $\hat{\rho}^*$ the map that, after applying $\mathring{\hat{\rho}}^*$, subsequently relabels $A \setminus \{j\}$ by $[\#A - 1]$ using the increasing bijection.

Of course, (2.17) is not needed in the case where $c_v = 2$ since the branch point will completely disappear upon the deletion of leaf j, and the corresponding permutation will thus not be included in $\boldsymbol{\sigma}^{*\rho}$. The above definition states, that if $\hat{\boldsymbol{t}}_l^*$ and $\hat{\boldsymbol{t}}_r^*$ are the neighbouring subtrees to the left and right, respectively, of leaf j in $\hat{\boldsymbol{t}}^*$ before deletion, then $\hat{\boldsymbol{t}}_l^*$ is the left neighbour of $\hat{\boldsymbol{t}}_r^*$ after the deletion.

Definition 2.4.3 (Semi-planar (α, γ)-chain)**.** Fix $n \geq 2$. The *semi-planar* (α, γ)*-chain* on $\widehat{\mathbb{T}}_{[n]}$ is an $(f, (\alpha, \gamma))$-down-up chain where $f\colon \widehat{\mathbb{T}}_{[n]} \times [n] \to \widehat{\mathbb{T}}^*_{[n]} \times [n]$ is defined by setting $f(\hat{\boldsymbol{t}}, i) = \left(\hat{\tau}^*\left(\hat{\boldsymbol{t}}, i, \tilde{\imath}\right), \tilde{\imath}\right)$, with $\tilde{\imath} = \max\{i, a, b\}$, where we by l_i denote the rank of i amongst the least leaf labels of the c_v subtrees of the parent v of i in $\hat{\boldsymbol{t}}$, set

$$a = \min\left\{\begin{matrix}\text{leaf labels in the first spinal bush} \\ \text{on the ancestral line from leaf } i\end{matrix}\right\},$$

and

$$b = \begin{cases} \min\left\{\begin{matrix}\text{leaf labels in the second spinal bush} \\ \text{on the ancestral line from leaf } i\end{matrix}\right\} & \text{if } c_v = 2, \\ \min\left\{\begin{matrix}\text{leaf labels in the first subtree not} \\ \text{containing } a \text{ to the right of leaf } i\end{matrix}\right\} & \text{if } c_v > 2,\ l_i \leq 2, \\ \min\left\{\begin{matrix}\text{leaf labels in the first subtree not} \\ \text{containing } a \text{ to the left of leaf } i\end{matrix}\right\} & \text{if } c_v > 2,\ l_i > 2, \end{cases}$$

with the convention that $b = 0$ if $c_v = 2$ and the ancestral line from leaf i has only one spinal bush.

Firstly, recall that we can apply $\hat{\tau}$, and hence f, to elements of $\widehat{\mathbb{T}}_{[n]}$ (instead of $\widehat{\mathbb{T}}^*_{[n]}$) by using the canonical way of translating a semi-planar tree to a planar tree described just after Definition 2.3.2. However, it is not clear a priori that carrying out the down-step of the above chain started from a semi-planar tree again yields one such. This is needed for the above to be well-defined as a Markov chain on $\widehat{\mathbb{T}}_{[n]}$. The following result shows that this is indeed the case:

Lemma 2.4.4. *Let A be a finite set and fix $\hat{\boldsymbol{t}} \in \widehat{\mathbb{T}}_A$. For any $i \in A$ it holds that $\mathring{\hat{\rho}}^*\left(\hat{\tau}^*(\hat{\boldsymbol{t}}, i, \tilde{\imath}), \tilde{\imath}\right) \in \widehat{\mathbb{T}}_{A\setminus\{\tilde{\imath}\}}$, where $\tilde{\imath}$ is defined as in Definition 2.4.3.*

Proof. We have to check that for every branch point in $\mathring{\hat{\rho}}\left(\hat{\tau}^*(\hat{\boldsymbol{t}}, i, \tilde{\imath}), \tilde{\imath}\right)$, the two subtrees with the smallest leaf labels, respectively, are located leftmost. From the definition of the semi-planar (α, γ)-chain it is clear that the only two branch points in $\hat{\boldsymbol{t}}$ where any changes can take place are the parent v and potentially the grandparent v' of leaf i in $\hat{\boldsymbol{t}}$, since all other branch points will remain unchanged as $i \leq \tilde{\imath}$ and $\tilde{\imath}$ is the smallest leaf in its subtree. We now split up into cases based on the location of leaf i in $v \in \mathrm{bp}(\hat{\boldsymbol{t}}^*)$, where $\hat{\boldsymbol{t}}^*$ is the canonical representative of $\hat{\boldsymbol{t}}$ in $\widehat{\mathbb{T}}^*_A$ described just after Definition 2.3.2.

If i is not one of the two leftmost subtrees in v, there is no option of swapping with the smallest leaf label in either of the two leftmost subtrees, as they are both smaller than i by definition of a semi-planar tree.

Now consider the case where i is placed leftmost in v, with v having more than two children. In $\hat{\boldsymbol{t}}^*$, i will then be placed leftmost or second leftmost in v. In either case $\tilde{\imath} = b$ will be the smallest label in the third leftmost subtree of v in $\hat{\boldsymbol{t}}^*$. $\hat{\tau}^*(\hat{\boldsymbol{t}}, i, \tilde{\imath}) \in \widehat{\mathbb{T}}^*_A$ will not correspond to an element of $\widehat{\mathbb{T}}_A$, but consider the ordering of the subtrees of v in this planar tree. i will have retained its rank (1 or 2) amongst the least labels of the subtrees of v, and will now be located in the third leftmost subtree. Upon the deletion of $\tilde{\imath}$ in $\hat{\tau}^*(\hat{\boldsymbol{t}}, i, \tilde{\imath})$ the two leftmost subtrees will again be the ones containing the two smallest of the least leaf labels of the subtrees of v.

The last case is when i is placed leftmost in v, with v only having two children. Here, b will always be found in v', the parent of v, and no matter which swap occurs, the branch

point v will disappear upon deletion of $\tilde{\imath}$. Hence we only need to check that the new ordering of subtrees in v' is acceptable. If $\tilde{\imath} \in \{a, i\}$ there is nothing to prove, as this means that the two subtrees with the lowest labels in v' will not be affected by the swapping of i and $\tilde{\imath}$ and subsequent deletion of $\tilde{\imath}$. If $\tilde{\imath} = b$, this means that b was the smallest leaf label (not in a subtree of v) in v' and that a was even smaller. Hence, by definition of the (α, γ)-growth rule, the smallest labels in the two leftmost subtrees of v' after swapping and deleting, respectively, will be a and i.

In all cases we conclude that $\mathring{\hat{\rho}}^* \left(\hat{\tau}(\hat{\boldsymbol{t}}, i, \tilde{\imath}), \tilde{\imath}\right) \in \widehat{\mathbb{T}}_{A\setminus\{\tilde{\imath}\}}$. ■

We will approach the proof of Theorem 2.4.6 via the following technical result. The statement of the lemma as well as the proof technique is an adaptation of Lemma 4 and the proof of Theorem 1 in [21], designed to work in the case where $\alpha = \gamma = \frac{1}{2}$. It is possible to prove that the stationary distribution of a down-up with the transition kernel described by Proposition 2.5.2 is the distribution of the the nth step of the (α, γ)-growth rule without any reference to a semi-planar structure, thus making this lemma obsolete, in the case where $\gamma = 1 - \alpha$ for $\alpha \in [0, 1]$. However, this falls outside the scope of this exposition.

Lemma 2.4.5. *Fix $n \in \mathbb{N}$ and let $\left(\widehat{T}_m\right)_{1 \le m \le n}$ be the first n steps of the semi-planar (α, γ)-growth process. Fix $1 \le i \le \tilde{\imath} \le n$ and define*

$$E_{i,\tilde{\imath}} = \left\{\tilde{\imath} = \max\{i, a, b\} \text{ in } \widehat{T}_n\right\}, \tag{2.18}$$

where a and b are as in Definition 2.4.3. Then $E_{i,\tilde{\imath}} \perp\!\!\!\perp \widehat{T}_{\tilde{\imath}-1}$ and, conditional on the event $E_{i,\tilde{\imath}}$, $\hat{\rho}^*\left(\hat{\tau}^*\left(\widehat{T}_n, i, \tilde{\imath}\right), \tilde{\imath}\right)$ and $\widehat{T}_{n-1}$ have the same distribution.

Proof. We will prove the first assertion by splitting into two cases. The overall proof technique will be to translate $E_{i,\tilde{\imath}}$ to restrictions regarding the growth process and to use this characterization to obtain the result.

Case A: $i < \tilde{\imath}$. In this case, $E_{i,\tilde{\imath}}$ is equivalent to one set of the following disjoint events, some of them illustrated in Figure 2.4, where we will denote the parent branch point of i in $\widehat{T}_{\tilde{\imath}-1}$ by v and the number of children of v in $\widehat{T}_{\tilde{\imath}-1}$ by c_v.

- if $c_v = 2$ then either

  (i) $\tilde{\imath}$ is inserted into v (as right neighbour of i), and

  (ii) $\tilde{\imath}+1, \dots, n$ are *not* inserted on the leaf edge i or as right neighbour i in v in $\widehat{T}_{\tilde{\imath}}, \dots, \widehat{T}_{n-1}$, respectively,

  or

  (i) $\tilde{\imath}$ is inserted on the parent edge of i, $e_{v\downarrow}$, and

  (ii) $\tilde{\imath}+1, \dots, n$ are *not* inserted on the leaf edge i, into the branch point v, nor on the edge $e_{v\downarrow}$ in $\widehat{T}_{\tilde{\imath}}, \dots, \widehat{T}_{n-1}$, respectively;

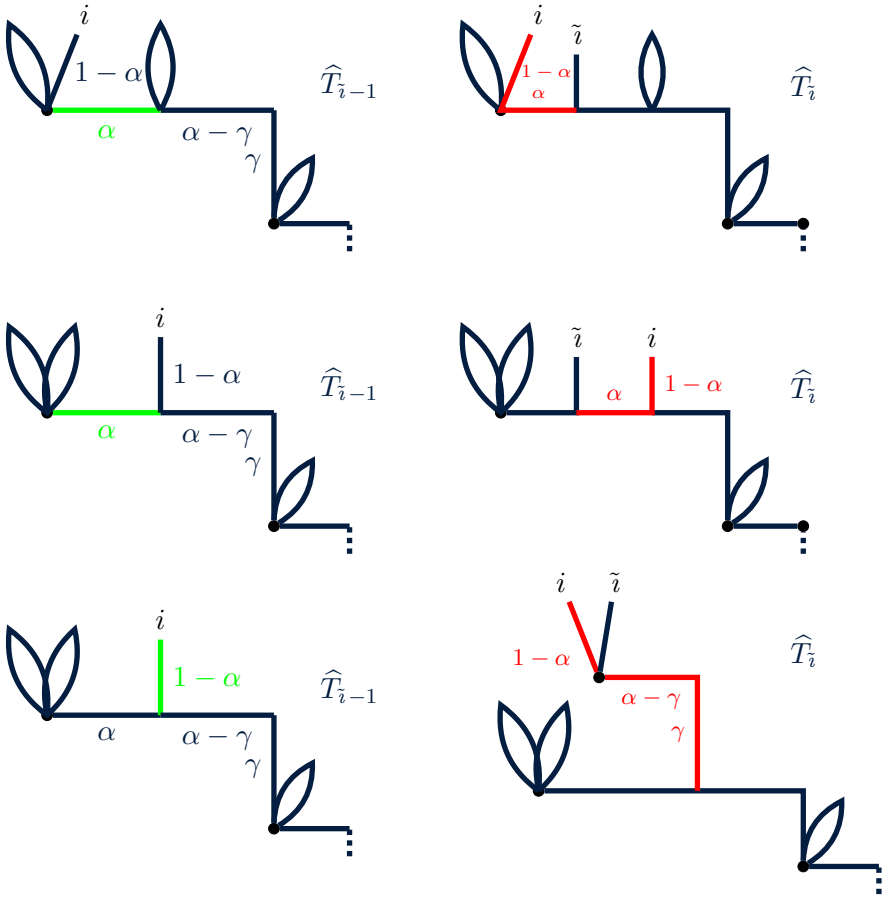


Figure 2.4: Illustration of three of the characterizations of the event $E_{i,\tilde{\imath}}$ in the case where $i < \tilde{\imath}$ and i is located in a multifurcating branch point, from the proof of Lemma 2.4.5. The horizontal lines in each illustration correspond to the expansion of branch points described just after Definition 2.3.4. Each row is an insertion scenario, where the left column is an illustration of $\widehat{T}_{\tilde{\imath}-1}$ where the insertion location of $\tilde{\imath}$ is colored green, while the right column is an illustration of $\widehat{T}_{\tilde{\imath}}$ where the edges colored in red are the ones where insertions of $\tilde{\imath}+1, \dots, n$ are disallowed.

- if $c_v > 2$ and i is leftmost in v, then

  (i) $\tilde{\imath}$ is inserted as the right neighbour of i in v, and

  (ii) $\tilde{\imath}+1, \ldots, n$ are *not* inserted on leaf edge i or between i and the subtree containing $\tilde{\imath}$ in the branch point v in $\widehat{T}_{\tilde{\imath}}, \ldots, \widehat{T}_{n-1}$, respectively.

  This situation is depicted in the top row of Figure 2.4.

- if $c_v > 2$ and i is not leftmost in v, then

  (i) $\tilde{\imath}$ is inserted as the left neighbour of i in v, and

  (ii) $\tilde{\imath}+1, \ldots, n$ are *not* inserted on leaf edge i or between i and the subtree containing $\tilde{\imath}$ in the branch point v in $\widehat{T}_{\tilde{\imath}}, \ldots, \widehat{T}_{n-1}$, respectively.

  This situation is depicted in the middle row of Figure 2.4.

In addition to the above, there is, regardless of which of the above scenarios we are in, the option to

(i) insert $\tilde{\imath}$ on the leaf edge i in $\widehat{T}_{\tilde{\imath}-1}$, and

(ii) *not* insert $\tilde{\imath}+1, \ldots, n$ on the first two edges nor in the first branch point on the ancestral line from leaf i in $\widehat{T}_{\tilde{\imath}}, \ldots, \widehat{T}_{n-1}$, respectively,

which is illustrated in the bottom row of Figure 2.4.

Case B: $i = \tilde{\imath}$. By the same reasoning as earlier, $E_{i,\tilde{\imath}}$ is equivalent to

- *not* inserting $\tilde{\imath}+1, \ldots, n$ on the first two edges, nor in the parent branch point of i, on the ancestral line from i in $\widehat{T}_{\tilde{\imath}}, \ldots, \widehat{T}_{n-1}$, if $i = \tilde{\imath}$ was inserted into an edge of $\widehat{T}_{\tilde{\imath}-1}$,

or

- *not* inserting $\tilde{\imath}+1, \ldots, n$ on the leaf edge $i = \tilde{\imath}$ or as the left neighbour of $i = \tilde{\imath}$ in the first branch point on the ancestral line from i in $\widehat{T}_{\tilde{\imath}}, \ldots \widehat{T}_{n-1}$, respectively, if $i = \tilde{\imath}$ is inserted into a branch point of $\widehat{T}_{\tilde{\imath}-1}$.

We have now characterized the event $E_{i,\tilde{\imath}}$ in terms of making or not making insertions, respectively, to specific edges with weights summing to 1 in all cases. This implies that

the probability of making or not making these insertions, respectively, is the same in all cases, and resolving the telescoping products, we find

$$\begin{aligned}
&\mathrm{P}\left(\left\{\widehat{T}_{\tilde{\imath}-1} = \hat{\rho}^*(\hat{\boldsymbol{t}}, [n] \setminus [\tilde{\imath}-1])\right\} \cap E_{i,\tilde{\imath}}\right) \\
&= \begin{cases} \frac{1}{n-1-\alpha}\mathrm{P}\left(\widehat{T}_{\tilde{\imath}-1} = \hat{\rho}^*\left(\hat{\boldsymbol{t}}, [n] \setminus [\tilde{\imath}-1]\right)\right) & \text{if } 1 \leq i < \tilde{\imath} \leq n \\ \frac{\tilde{\imath}-2-\alpha}{n-1-\alpha}\mathrm{P}\left(\widehat{T}_{\tilde{\imath}-1} = \hat{\rho}^*\left(\hat{\boldsymbol{t}}, [n] \setminus [\tilde{\imath}-1]\right)\right) & \text{if } 2 \leq i = \tilde{\imath} \leq n \end{cases}
\end{aligned}$$

for all $\hat{\boldsymbol{t}} \in \widehat{\mathbb{T}}_{[n]}$, showing that $\widehat{T}_{\tilde{\imath}-1} \perp\!\!\!\perp E_{i,\tilde{\imath}}$ for all fixed $1 \leq i \leq \tilde{\imath} \leq n$.

For the second assertion we start by noting that

$$\mathrm{P}\left(\widehat{T}_{\tilde{\imath}-1} \in \cdot \;\middle|\; E_{i,\tilde{\imath}}\right) = \mathrm{P}\left(\widehat{T}_{\tilde{\imath}-1} \in \cdot\right)$$

due to the independence proven above. Once again using the characterization of $E_{i,\tilde{\imath}}$, we note that we can construct $\hat{\rho}^*\left(\hat{\tau}^*\left(\widehat{T}_n, i, \tilde{\imath}\right), \tilde{\imath}\right)$ from $\widehat{T}_{\tilde{\imath}-1}$ under the conditional law of $E_{i,\tilde{\imath}}$: The leaf $\tilde{\imath}$ is never attached and so the "forbidden edges", characterized above and coloured red in Figure 2.4, do not exist. But every leaf that was previously labelled $\tilde{\imath}+1, \ldots, n$ will still be attached according to the (α, γ)-growth rule since the total weight of the forbidden edges are exactly 1 in all cases. This finishes the proof. ■

Theorem 2.4.6. *Fix $n \in \mathbb{N}$. Let $\left(\widehat{T}_n(m)\right)_{m \in \mathbb{N}_0}$ denote the semi-planar (α, γ)-chain and let $\widehat{T}_n$ denote the nth step of the semi-planar (α, γ)-growth process. Then it holds that $\widehat{T}_n(m) \xrightarrow{D} \widehat{T}_n$ for $m \to \infty$, where $\xrightarrow{D}$ denotes convergence in distribution. Furthermore, it holds that the projection of $\widehat{T}_n$ to the space of (non-planar) $[n]$-trees has the same distribution as the nth step of the (α, γ)-growth process.*

Proof. Fix $n \in \mathbb{N}$. That the semi-planar (α, γ)-chain on $\widehat{\mathbb{T}}_{[n]}$ has a unique stationary distribution is easily seen, since the state space is finite and there is a unique, recurrent communicating class for

- $\alpha = 1$ and all $\gamma \in (0, 1)$ (only insertions in branch points or internal edges),
- $\gamma = 0$ and all $\alpha \in (0, 1)$ (only insertions in branch points or leaf edges),
- $\gamma = \alpha$ and all $\alpha \in [0, 1]$ (binary trees),

- $\alpha = 1$ and $\gamma = 0$ (only insertions in branch points),

whilst the Markov chain is irreducible otherwise. Thus we are done if we can show that $\widehat{T}_n(1) \overset{D}{=} \widehat{T}_n(0) := \widehat{T}_n$, which we will do by utilizing Lemma 2.4.5. Let I be a uniform random variable on $[n]$, independent of everything else, let f denote the transformation function from Definition 2.4.3 and note that for all $\hat{\boldsymbol{t}} \in \widehat{\mathbb{T}}_{n-1}$

$$\begin{aligned}
& \mathrm{P}\left(\hat{\rho}^*\left(f\left(\widehat{T}_n, I\right)\right) = \hat{\boldsymbol{t}}\right) \\
= & \sum_{1 \le i \le \tilde{\imath} \le n} \mathrm{P}\left(\hat{\rho}^*\left(\hat{\tau}^*\left(\widehat{T}_n, i, \tilde{\imath}\right), \tilde{\imath}\right) = \hat{\boldsymbol{t}} \;\middle|\; E_{i,\tilde{\imath}} \cap \{I = i\}\right) \mathrm{P}\left(E_{i,\tilde{\imath}} \cap \{I = i\}\right) \\
= & \sum_{1 \le i \le \tilde{\imath} \le n} \mathrm{P}\left(\widehat{T}_{n-1} = \hat{\boldsymbol{t}}\right) \mathrm{P}\left(E_{i,\tilde{\imath}} \cap \{I = i\}\right) \\
= & \mathrm{P}\left(\widehat{T}_{n-1} = \hat{\boldsymbol{t}}\right)
\end{aligned}$$

where we have used the independence of I, the definition of f as well as Lemma 2.4.5. This shows that $\hat{\rho}^*\left(f\left(\widehat{T}_n, I\right)\right) \overset{D}{=} \widehat{T}_{n-1}$, and since $\widehat{T}_n(1)$ is constructed from the former using the same growth rule used to construct $\widehat{T}_n$ from the latter, this finishes our proof. ■

2.5 Projecting to a non-planar down-up chain

In the previous section we constructed a semi-planar down-up chain on $\widehat{\mathbb{T}}_{[n]}$ with the distribution of $\widehat{T}_n$ as its stationary distribution, where $\widehat{T}_n$ denotes the nth step of the semi-planar (α, γ)-growth process of Definition 2.3.4. Recalling that $\hat{\pi}\colon \widehat{\mathbb{T}}_{[n]} \to \mathbb{T}_{[n]}$ denotes the projection map from the space of semi-planar trees to that of non-planar trees, we have already noted that $\hat{\pi}(\widehat{T}_n) \overset{D}{=} T_n$, where the latter denotes the nth step of the (α, γ)-growth process of Definition 2.2.10. Now let us construct a non-planar down-up chain on $\mathbb{T}_{[n]}$ from the semi-planar version.

Fix $n \in \mathbb{N}$. Let $\hat{K}_n$ denote the transition kernel of the semi-planar (α, γ)-chain on $\widehat{\mathbb{T}}_{[n]}$ and define a $\hat{\pi}$-induced Markov kernel from $\mathbb{T}_{[n]}$ to $\widehat{\mathbb{T}}_{[n]}$ by

$$\hat{\Pi}_n\left(\boldsymbol{t}, \{\hat{\boldsymbol{t}}\}\right) := \mathrm{P}\left(\widehat{T}_n = \hat{\boldsymbol{t}} \;\middle|\; \hat{\pi}\left(\widehat{T}_n\right) = \boldsymbol{t}\right) = \prod_{v \in \mathrm{bp}(\boldsymbol{t})} p_{\alpha,\alpha-\gamma}^{c_v - 2}\left(\sigma_v\right), \tag{2.19}$$

for each $\hat{\boldsymbol{t}} \in \widehat{\mathbb{T}}_{[n]}$ and $\boldsymbol{t} \in \mathbb{T}_{[n]}$, where c_v denotes the number of children of a branch point

$v \in \mathrm{bp}(\boldsymbol{t})$ and σ_v is the permutation of $[c_v - 2]$ associated with the same branch point in $\hat{\boldsymbol{t}}$. The latter expression stems from Lemma 2.3.5 and (2.15). Let us abuse notation, so that $\hat{\Pi}_n$ also denotes the induced Markovian matrix $\hat{\Pi}_n := \left(\hat{\Pi}_n\left(\boldsymbol{t}, \{\hat{\boldsymbol{t}}\}\right)\right)_{\boldsymbol{t}\in\mathbb{T}_{[n]}, \hat{\boldsymbol{t}}\in\widehat{\mathbb{T}}_{[n]}}$, whilst $\hat{\pi} := \left(1_{\hat{\pi}(\hat{\boldsymbol{t}})=\boldsymbol{t}}\right)_{\hat{\boldsymbol{t}}\in\widehat{\mathbb{T}}_{[n]}, \boldsymbol{t}\in\mathbb{T}}$ denotes the matrix linking a semi-planar tree to its non-planar counterpart. We now give an alternative definition of the (α, γ)-chain on $\mathbb{T}_{[n]}$ which we identify with Definition 2.1.1 in Proposition 2.5.2:

Definition 2.5.1 ((α, γ)-chain)**.** Fix $n \in \mathbb{N}$. The (α, γ)-*chain on* $\mathbb{T}_{[n]}$ is a Markov chain on $\mathbb{T}_{[n]}$ with transition kernel

$$K_n := \hat{\Pi}_n \hat{K}_n \hat{\pi}. \tag{2.20}$$

The transition kernel K_n is defined so that the diagram in Figure 2.5 commutes, and a more detailed explanation of the transitions of the (α, γ)-chain, $(T_n(m))_{m\in\mathbb{N}_0}$, is consequently the following:

(i) Fix $m \in \mathbb{N}_0$. If $T_n(m) = \boldsymbol{t} \in \mathbb{T}_{[n]}$, pick a semi-planar tree, $\hat{\boldsymbol{t}} \in \widehat{\mathbb{T}}_{[n]}$, according to $\hat{\Pi}_n(\boldsymbol{t}, \cdot)$ defined in (2.19).

(ii) Make one transition of the semi-planar (α, γ)-chain started from $\hat{\boldsymbol{t}}$, to obtain $\hat{\boldsymbol{t}}'$.

(iii) Project $\hat{\boldsymbol{t}}'$ to the space of non-planar trees, i.e. set $T_n(m+1) = \hat{\pi}(\hat{\boldsymbol{t}}')$.

The following result ensures that the two ways of defining the (α, γ)-chain (Definitions 2.1.1 and 2.5.1) are compatible.

Proposition 2.5.2. *Definition 2.1.1 and Definition 2.5.1 are equivalent.*

Proof. Before we consider the down-step itself, note that it does not make any difference if we perform the up-step for the semi-planar (α, γ)-chain and then project to the non-planar trees, or if we project to the non-planar trees immediately after the down-step, and then perform the up-step by using the non-planar (α, γ)-growth process, see the brief discussion following Definition 2.3.4. This ensures, that we in the following only have to argue that the down-step outlined in Steps 1 and 2 is the correct characterization.

So fix $\boldsymbol{t} \in \mathbb{T}_{[n]}$ and sample $\widehat{T}$ from $\hat{\Pi}_n(\boldsymbol{t}, \cdot) = \mathrm{P}\left(\widehat{T}_n \in \cdot \;\middle|\; \hat{\pi}\left(\widehat{T}_n\right) = \boldsymbol{t}\right)$.

If i is in a binary branch point in $\boldsymbol{t}$, the same is the case in $\widehat{T}$. Hence the characterization of $\tilde{I}$ is trivial, as it is exactly the same here as in Definition 2.4.3.

If i is in a multifurcating branch point with $c_v > 2$ children in $\boldsymbol{t}$, we need to split up into cases based on the location of i within the branch point in $\widehat{T}$. So let v denote the parent branch point of i, let $i_1, \dots, i_{c_v}$ denote the smallest leaf labels in the c_v subtrees of v, enumerated in increasing order, and say that $i = i_j$ for some $j \in [c_v]$. This is all deterministic based on $\boldsymbol{t}$ and i. Furthermore, recall the definition of a and b from Definition 2.4.3.

If i is placed leftmost in v in $\widehat{T}$, i.e. if $j \in [2]$, then the b of Definition 2.3.4 is the smallest leaf to the right of the two leftmost subtrees, which will always be the largest of i, a and b, implying that $\tilde{\imath} = \max\{i, a, b\} = b$. Hence, conditional on $T_n(m) = \boldsymbol{t}$ and $I = i$, the probability of $\tilde{I} = i_{j'}$ for some $2 < j' \leq c_v$ is the probability that $\widehat{T}$ has the subtree containing $i_{j'}$ placed to the immediate right of i in v. Considering how these subtrees were formed by the growth rule in the ordered Chinese Restaurant analogy outlined in Lemma 2.3.5, this is the probability having table $j' - 2$, enumerated in the order of least element, placed leftmost in an $(\alpha, \alpha - \gamma)$-oCRP with $c_v - 2$ tables.

If i is not places leftmost in v in $\widehat{T}$, i.e. if $2 < j \leq c_v$, b is the smallest leaf in the tree to the left of i in v. In this case we always find that $\tilde{\imath} = \max\{i, b\}$ as a by definition is equal to i_1. Consequently, conditional on $T_n(m) = \boldsymbol{t}$ and $I = i$, the probability of $\tilde{I} = i = i_j$ stems from the probability that $\widehat{T}$ has one of the subtrees containing $i_1, \dots, i_{j-1}$, respectively, placed to the immediate left of i in v. Hence in the ordered Chinese Restaurant analogy, this corresponds to not opening a new table to the immediate left of table $j - 2$, enumerated in the order of least element, in an $(\alpha, \alpha - \gamma)$-oCRP with $c_v - 2$ tables. Similarly the probability of getting $\tilde{I} = i_{j'}$ for any $j < j' \leq c_v$ corresponds to having table $j' - 2$ to the immediate left of table $j - 2$ in the same ordered Chinese Restaurant. This defines the conditional distribution of $\tilde{I}$ outlined in Definition 2.1.1. ■

The following result now follows immediately from Theorem 2.4.6.

Proposition 2.5.3. *Fix $n \in \mathbb{N}$ and let T_n denote the nth step of the (α, γ)-growth process. The (α, γ)-chain on $\mathbb{T}_{[n]}$ started from T_n is stationary.*

Proof. First note that the distribution of T_n is pushed forward to the distribution of $\widehat{T}_n$ by $\hat{\Pi}_n$, where $\hat{\Pi}_n$ is defined by (2.19) and $\widehat{T}_n$ denotes the nth step of the semi-planar

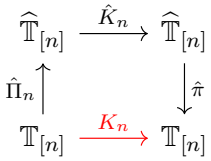


Figure 2.5: Commutative diagram to illustrate the construction of the transition kernel for the (α, γ)-chain.

(α, γ)-growth process. By Theorem 2.4.6 the semi-planar (α, γ)-chain started from $\widehat{T}_n$ is stationary, so if $\left(\widehat{T}_n(m)\right)_{m\in\mathbb{N}_0}$ denotes the semi-planar (α, γ)-chain started from $\widehat{T}_n$, then $\widehat{T}_n(1) \overset{D}{=} \widehat{T}_n(0) \overset{D}{=} \widehat{T}_n$. Lastly, by construction of the semi-planar (α, γ)-growth process, we have that $\hat{\pi}(\widehat{T}_n) \overset{D}{=} T_n$, and so $T_n(1) \overset{D}{=} T_n(0) \overset{D}{=} T_n$. ■

Proof of Theorem 2.1.2. This now follows directly from Propositions 2.5.2 and 2.5.3 by observing that the chain is either irreducible or have exactly the same unique, recurrent communicating classes as outlined in the proof of Theorem 2.4.6. ■

Proposition 2.5.3 is a fairly easy result to obtain due to the construction of the (α, γ)-chain on $\mathbb{T}_{[n]}$ and the results regarding the semi-planar (α, γ)-chain at our current disposal. However, the link between the two down-up chains is somewhat deeper than what this result eludes to.

Theorem 2.5.4. *Fix $n \in \mathbb{N}$. Let $\left(\widehat{T}_n(m)\right)_{m\in\mathbb{N}_0}$ be a semi-planar (α, γ)-chain, and let $T_n(m) := \hat{\pi}\left(\widehat{T}_n(m)\right)$ denote the projection of the chain onto $\mathbb{T}_{[n]}$. If $\widehat{T}_n(0) \sim \hat{\Pi}_n(\boldsymbol{t}, \cdot)$ for some $\boldsymbol{t} \in \mathbb{T}_{[n]}$, then $(T_n(m))_{m\in\mathbb{N}_0}$ is the (α, γ)-chain started from $\boldsymbol{t}$ with stationary distribution being that of T_n, the nth step of the (α, γ)-growth process.*

Theorem 2.5.4 is substantially deeper result than both Propositions 2.5.2 and 2.5.3, but we will defer the proof as we have yet to develop the methodology to prove it, and as it is a special case of Proposition 2.6.7 and Theorem 2.8.4.

2.6 A Markov Chain on decorated trees

So far we have introduced a down-up Markov chain on semi-planar trees, and shown that there is a corresponding Markov chain on the space of (non-planar) trees. We now continue our endeavour to define down-up chains on various spaces of trees that arise as projections

of $[n]$-trees. To make this precise, we introduce the notion of decorated trees, as well as an intermediary known as collapsed trees, which arise from affiliating every element of $\mathrm{ins}(\boldsymbol{t})$, the collection of edges and branch points of the tree $\boldsymbol{t} \in \mathbb{T}_{[n]}$, with an integer or a set of integers, respectively.

Definition 2.6.1 (Decorated Trees)**.** Let A be a finite set, and fix some $\#A \leq n \in \mathbb{N}$. Let $\boldsymbol{s} \in \mathbb{T}_A$ be given, and let $\mathbf{y} = (y_x)_{x \in \mathrm{ins}(\boldsymbol{s})}$ be a collection of non-negative integers such that $\sum_x y_x = n$ where $y_x \geq 1$ for every external $x \in \mathrm{edge}(\boldsymbol{s})$. We call $\boldsymbol{t}^\bullet = (\boldsymbol{s}, \mathbf{y})$ a *decorated A-tree of mass n*, and we denote the space of such trees by $\mathbb{T}_A^{\bullet n}$, and will refer to $\boldsymbol{s}$ as the *tree shape* of $\boldsymbol{t}$.

For any finite sets $A \subseteq B$ with $\#B =: n$ we can easily obtain a decorated A-tree of mass n from any $\boldsymbol{t} \in \mathbb{T}_B$ by projection onto the subtree spanned by the root and the leaves with labels in A. Let us explain this more formally. In the remainder of this exposition we will use the notation, $\boldsymbol{s} \subseteq \boldsymbol{t} \in \mathbb{T}_B$, to mean that there exists a set $A \subseteq B$ such that $\rho(\boldsymbol{t}, B \setminus A) = \boldsymbol{s}$. Consequently, we will insist that $\mathrm{vert}(\boldsymbol{s}) \subseteq \mathrm{vert}(\boldsymbol{t})$ in the usual sense, and hence it is meaningful to talk about a branch point of $\boldsymbol{s}$ being on the ancestral line from a leaf in $\boldsymbol{t}$. Note however that, unless we are in the degenerate situation where $A = B$ or where all leaves labelled by $B \setminus A$ have a parent branch point of $\boldsymbol{s}$ in $\boldsymbol{t}$, we will have $\mathrm{edge}(\boldsymbol{s}) \nsubseteq \mathrm{edge}(\boldsymbol{t})$.

So fix $\boldsymbol{t} \in \mathbb{T}_B$, and define $\boldsymbol{s} := \mathring{\rho}(\boldsymbol{t}, B \setminus A)$, where we recall that the latter means that we delete all leaves with labels of B not in A from $\boldsymbol{t}$, see Definition 2.2.8. Recalling that $u \prec_{\boldsymbol{t}} v$ denotes that u is the parent of v in $\boldsymbol{t}$, we now define the *decoration function* $g \colon B \to \boldsymbol{s} \subseteq \boldsymbol{t}$ by setting

- $g(i) = v \in \mathrm{bp}(\boldsymbol{s})$ if i is in a subtree of v in $\boldsymbol{t}$ which only contains leaves labelled by elements of $B \setminus A$, and
- $g(i) = e_{v_0 v_m} \in \mathrm{edge}(\boldsymbol{s})$ if there exist $m \geq 2$, vertices $v_0, v_m \in \mathrm{vert}(\boldsymbol{s})$ and branch points $v_1, \ldots, v_{m-1} \in \mathrm{bp}(\boldsymbol{t}) \setminus \mathrm{bp}(\boldsymbol{s})$ such that for all $j \in [m]$ it holds that $v_{j-1} \prec_{\boldsymbol{t}} v_j$, $v_0 \prec_{\boldsymbol{s}} v_m$, and that there is a $j \in [m-1]$ such that i is in a subtree of v_j in $\boldsymbol{t}$ which only contains leaves labelled by elements of $B \setminus A$.

Then define

$$y_x = \sum_{i=1}^{n} \delta_{xg(i)} = \#g^{-1}(\{x\}) = \#\{i \in [n] \mid g(i) = x\} \tag{2.21}$$

for each $x \in \mathrm{ins}(\boldsymbol{s})$, where δ_{xy} is the Kronecker delta, and define $\mathbf{y} = (y_x)_{x\in\mathrm{ins}(\boldsymbol{s})}$. Finally, define $\boldsymbol{t}^\bullet = (\boldsymbol{s}, \mathbf{y})$.

For each $k \leq n \in \mathbb{N}$, we will denote the map $\mathbb{T}_{[n]} \ni \boldsymbol{t} \mapsto \boldsymbol{t}^\bullet = (\boldsymbol{s}, \mathbf{y}) \in \mathbb{T}^{\bullet n}_{[k]}$, described above, by $\pi^{\bullet n}_{[k]}$. This construction yields another family of trees as well:

Definition 2.6.2 (Collapsed Trees)**.** Let A be a finite set, and fix some $\#A \leq n \in \mathbb{N}$. With g being the decoration function, we denote the map that sends $\boldsymbol{t} \in \mathbb{T}_{[n]}$ to its associated *collapsed A-tree of mass n* by π^{*n}_A:

$$\pi^{*n}_A(\boldsymbol{t}) = \left(\boldsymbol{s}, \left(g^{-1}(\{x\})\right)_{x\in\mathrm{ins}(\boldsymbol{s})}\right), \tag{2.22}$$

where $\boldsymbol{s} = \rho(\boldsymbol{t}, [n] \setminus A)$. We denote the space of such trees by

$$\mathbb{T}^{*n}_A := \left\{ \pi^{*n}_A(\boldsymbol{t}) \;\middle|\; \boldsymbol{t} \in \mathbb{T}_{[n]} \right\}. \tag{2.23}$$

The aim of this section is similar to that of the previous one, where we studied a Markov chain on $\mathbb{T}_{[n]}$ by using a kernel to lift an element $\boldsymbol{t} \in \mathbb{T}_{[n]}$ to $\widehat{\mathbb{T}}_{[n]}$. Paraphrasing, we sampled the order of the subtrees within each branch point of $\boldsymbol{t}$, then carried out both the down- and up-step, respectively, of the semi-planar (α, γ)-chain on $\widehat{\mathbb{T}}_{[n]}$, and finally projected back to $\mathbb{T}_{[n]}$.

Constructing a Markov chain on decorated trees corresponding to the semi-planar (α, γ)-chain on $\widehat{\mathbb{T}}_{[n]}$, is done by the following procedure. Fix $\boldsymbol{t}^\bullet = (\boldsymbol{s}, \mathbf{y}) \in \mathbb{T}^{\bullet n}_{[k]}$. Similar to the construction of $\hat{\Pi}_n$ in (2.19), we will construct the kernel $\hat{\Pi}^{\bullet n}_k$, by setting

$$\forall \boldsymbol{t}^\bullet \in \mathbb{T}^{\bullet n}_{[k]},\ \hat{\boldsymbol{t}} \in \widehat{\mathbb{T}}_{[n]}\colon \quad \hat{\Pi}^{\bullet n}_k(\boldsymbol{t}^\bullet, \{\hat{\boldsymbol{t}}\}) := \mathrm{P}\left(\widehat{T}_n = \hat{\boldsymbol{t}} \;\middle|\; \hat{\pi}^{\bullet n}_{[k]}\left(\widehat{T}_n\right) = \boldsymbol{t}^\bullet\right), \tag{2.24}$$

where $\widehat{T}_n$ is the nth step of the semi-planar (α, γ)-growth process, and $\hat{\pi}^{\bullet n}_{[k]} := \pi^{\bullet n}_{[k]} \circ \hat{\pi}$ denotes the projection from semi-planar n-trees to decorated $[k]$-trees of mass n, for brevity.

Define a transition kernel on $\mathbb{T}_{[k]}^{\bullet n}$ by

$$K_k^{\bullet n} := \hat{\Pi}_k^{\bullet n} \hat{K}_n \hat{\pi}_{[k]}^{\bullet n}, \tag{2.25}$$

where $\hat{K}_n$ is the transition kernel for the semi-planar (α, γ)-chain and we have used the same notation as in (2.20) for the composition of kernels.

Definition 2.6.3 (Decorated (α, γ)-chain)**.** Fix $n > k \in \mathbb{N}$ and let $K_k^{\bullet n}$ denote the transition kernel defined by (2.25). The *decorated* (α, γ)*-chain on* $\mathbb{T}_{[k]}^{\bullet n}$ is a Markov chain on $\mathbb{T}_{[k]}^{\bullet n}$ with transition kernel $K_k^{\bullet n}$.

Remark 2.6.4. This is equivalent to the Definition 2.1.3.

The goal of the remainder of this exposition is two-fold. On one hand we wish to show that if we project semi-planar (α, γ)-chains of Definition 2.4.3 to the space of decorated $[k]$-trees of mass n, we obtain the decorated (α, γ)-chain of Definition 2.6.3. On the other hand we wish to describe the transition kernel of the decorated (α, γ)-chain without referencing the semi-planar transition kernel. It turns out that the latter is much easier, and so we will take on this task first.

So consider the behaviour of the semi-planar (α, γ)-chain projected to decorated trees, i.e. for $k < n$ consider observing $\left(\hat{\pi}_{[k]}^{\bullet n}\left(\widehat{T}_n(m)\right)\right)_{m \in \mathbb{N}_0}$, and recall that the semi-planar (α, γ)-chain is characterized by composing two kernels in a down- and an up-step. The focus in this section is to specify similar steps on the space of decorated trees, so that doing a down- or up-step, respectively, in the semi-planar (α, γ)-chain and then projecting to a decorated tree, is the same as projecting to a decorated tree and then performing a down- or up-step, respectively. It turns out that the down-step is significantly more complicated than the up-step, and so we start with doing the latter.

Definition 2.6.5 ((α, γ)-growth process for decorated trees)**.** Fix $k \in \mathbb{N}$, fix $\boldsymbol{s} \in \mathbb{T}_{[k]}$, and let $T_k^{\bullet k}$ be the unique element of $\mathbb{T}_{[k]}^{\bullet k}$ with tree shape $\boldsymbol{s}$. For any $k \leq n \in \mathbb{N}$ construct

$T_k^{\bullet(n+1)}$ conditional on $T_k^{\bullet n} = (\boldsymbol{s}, \mathbf{y}) \in \mathbb{T}_{[k]}^{\bullet n}$ by, for each $x \in \mathrm{ins}(\boldsymbol{s})$, setting

$$\mathrm{P}\left(T_k^{\bullet(n+1)} = (\boldsymbol{s}, \mathbf{y} + \mathbf{1}_x) \;\middle|\; T_k^{\bullet n} = (\boldsymbol{s}, \mathbf{y})\right) = \begin{cases} \frac{y_x - \alpha}{n-\alpha} & \text{if } x \in \mathrm{edge}(\boldsymbol{s}) \text{ is external,} \\ \frac{y_x + \gamma}{n-\alpha} & \text{if } x \in \mathrm{edge}(\boldsymbol{s}) \text{ is internal,} \\ \frac{y_x + (c-1)\alpha - \gamma}{n-\alpha} & \text{if } x \in \mathrm{bp}(\boldsymbol{s}) \text{ with } c \text{ children,} \end{cases}$$

for all decorated trees $(\boldsymbol{s}, \mathbf{y}) \in \mathbb{T}_{[k]}^{\bullet n}$ where $\mathbf{1}_x = (\delta_{xx'})_{x' \in \mathrm{ins}(\boldsymbol{s})}$. We will call $(T_k^{\bullet n})_{n \geq k}$ *the decorated (α, γ)-growth process started from $T_k^{\bullet k} = \boldsymbol{s}$.*

We immediately note that the above is simply a Pólya urn scheme with initial weights labelled by the elements of $\mathrm{ins}(\boldsymbol{s})$. Specifically, the initial weights will be $1 - \alpha$ for an external edge, γ for an internal edge, and $(c-1)\alpha - \gamma$ for a branch point with c children in $\boldsymbol{s}$. This ensures that if $(T_n)_{n \in \mathbb{N}}$ is the (α, γ)-growth process then

$$\mathrm{P}\left(\pi_{[k]}^{\bullet n}(T_n) = \cdot \;\middle|\; T_k = \boldsymbol{s}\right) = \mathrm{P}\left(T_k^{\bullet n} = \cdot\right)$$

for all $n \geq k$. Additionally it holds that

$$\mathrm{P}\left(\hat{\pi}_{[k]}^{\bullet(n+1)}\left(\widehat{T}_{n+1}\right) \in \cdot \;\middle|\; \widehat{T}_n = \hat{\boldsymbol{t}}\right) = \mathrm{P}\left(T_k^{\bullet(n+1)} \in \cdot \;\middle|\; T_k^{\bullet n} = \hat{\pi}_{[k]}^{\bullet n}(\hat{\boldsymbol{t}})\right) \tag{2.26}$$

for all $\hat{\boldsymbol{t}} \in \widehat{\mathbb{T}}_{[n]}$. Consider again observing $\left(\hat{\pi}_{[k]}^{\bullet n}\left(\widehat{T}_n(m)\right)\right)_{m \in \mathbb{N}_0}$ and let $\widehat{T}_n^{\downarrow}(m)$ denote the semi-planar chain after the down-step. Then (2.26) ensures that we end up with the same distribution on the space of decorated $[k]$-trees of mass n by either

(i) projecting $\widehat{T}_n^{\downarrow}(m)$ to the space of decorated $[k]$-trees of mass $n-1$, and then performing the up-step from the (α, γ)-growth process for decorated trees, or

(ii) performing the up-step from the semi-planar (α, γ)-growth process, and then projecting the resulting tree to the space of decorated $[k]$-trees of mass n.

What is more technical is how to construct a down-step on the space of decorated trees that is consistent with the down-step defined in the semi-planar (α, γ)-chain. For intuition, let us consider a semi-planar (α, γ)-chain, $\left(\widehat{T}_n(m)\right)_{m \in \mathbb{N}_0}$, and assume that, before carrying

out the down-step, $\widehat{T}_n(m) = \hat{\boldsymbol{t}}$ and $\hat{\pi}^{\bullet n}_{[k]}(\hat{\boldsymbol{t}}) = (\boldsymbol{s}, \mathbf{y}) \in \mathbb{T}^{\bullet n}_{[k]}$. Now observe how there are three separate scenarios for the deletion of a leaf in the down-step:

- We select a leaf with label $i > k$ in $\hat{\boldsymbol{t}}$ that contributes to exactly one of the decorated masses y_x for some $x \in \mathrm{ins}(\boldsymbol{s})$, in which case we reduce that mass by one, but the tree shape $\boldsymbol{s}$ and all other decorating masses stay unchanged.

- We initially select a leaf with label $i \in [k]$ in $\hat{\boldsymbol{t}}$, search for a and b as defined in Definition 2.4.3, and define $\tilde{\imath} = \max\{i, a, b\}$:

  - If $\tilde{\imath} \leq k$ we will swap i and $\tilde{\imath}$, delete $\tilde{\imath}$, and relabel $\{\tilde{\imath}+1, \ldots, k+1\}$ by $\{\tilde{\imath}, \ldots, k\}$ using the increasing bijection. Depending on which y_x the leaf labelled by $k+1$ contributed to in $(\boldsymbol{s}, \mathbf{y})$, this can result in the tree shape changing. The decorating masses need to change accordingly.

  - If $\tilde{\imath} > k$ we will retain the same tree shape, irrespective of the location of i in $\hat{\boldsymbol{t}}$ prior to the swap. Indeed, if i was located in a multifurcating branch point of $\hat{\boldsymbol{t}}$, our search rule guarantees that $\tilde{\imath}$ is found in a subtree of the same branch point. If i was located in a binary branch point of $\hat{\boldsymbol{t}}$, $\tilde{\imath}$ will either be the smallest label in the other subtree of that branch point, or be the smallest leaf in the second spinal bush on the ancestral line from i, but due to the deletion of the binary branch point after the swap, this will not change the tree shape of the projected tree. See Figure 2.6 for an illustration.

From the above description it is clear, that the most complicated scenario of the down-step is when $\tilde{\imath} \leq k$, whereas the other scenarios are easier to deal with. For an example of one of down-steps, see Figure 2.6.

In the following definition, we will for fixed $n \in \mathbb{N}$ and $\boldsymbol{t} \in \mathbb{T}_{[n]}$ with $\pi^n_{[k]}(\boldsymbol{t}) = (\boldsymbol{s}, \mathbf{y}) \in \mathbb{T}^{\bullet n}_{[k]}$ let w_x denote the insertion weight associated with $x \in \mathrm{ins}(\boldsymbol{s})$ from Definition 2.2.10. To ease notation, we will for $(\boldsymbol{s}, \mathbf{y}) \in \mathbb{T}^{\bullet n}_{[k]}$ set

$$\tilde{y}_x = \begin{cases} y_x - 1 & \text{if } x \in \mathrm{edge}(\boldsymbol{s}) \text{ is external,} \\ y_x & \text{otherwise,} \end{cases}$$

for all $x \in \mathrm{ins}(\boldsymbol{s})$.

Definition 2.6.6 (Resampling a label)**.** Fix $\boldsymbol{t}^\bullet = (\boldsymbol{s}, \mathbf{y}) \in \mathbb{T}^{\bullet(n-1)}_{[k-1]}$. By *resampling leaf k in $\boldsymbol{t}^\bullet$* we mean obtaining an element $\boldsymbol{t}^\bullet_{\mathrm{re}} \in \mathbb{T}^{\bullet(n-1)}_{[k]}$ from $\boldsymbol{t}^\bullet$ in the following way:

1. Select $x \in \mathrm{ins}\,(\boldsymbol{s})$ with probability $\frac{\tilde{y}_x}{n-k}$, and define $\boldsymbol{s}^{\mathrm{re}} = \varphi(\boldsymbol{s}, x, k)$.

2. Define $\boldsymbol{t}^\bullet_{\mathrm{re}} = (\boldsymbol{s}^{\mathrm{re}}, \mathbf{y}^{\mathrm{re}})$, where $\mathbf{y}^{\mathrm{re}} = \left(y^{\mathrm{re}}_{x'}\right)_{x' \in \mathrm{ins}(\boldsymbol{s}^{\mathrm{re}})}$ is constructed by setting $y^{\mathrm{re}}_{x'} = y_{x'}$ for all $x' \in \mathrm{ins}(\boldsymbol{s}) \setminus \{x\}$ and splitting up y_x in the following way:

   - If $x \in \mathrm{edge}(\boldsymbol{s})$, let v denote the branch point created in the previous step. Let $(Y_k, Y_v, Y_{v\downarrow}, Y_{v\uparrow}) \sim \mathrm{DM}^{\tilde{y}_x - 1}(1-\alpha, \alpha-\gamma, \gamma, w_x)$, set $y^{\mathrm{re}}_{v\downarrow} = Y_{v\downarrow}, y^{\mathrm{re}}_{v} = Y_v$, $y^{\mathrm{re}}_k = Y_k + 1$, and implicitly define $y_{v\uparrow}$ by $\tilde{y}_{v\uparrow} = Y_{v\uparrow}$.
   - If $x \in \mathrm{bp}(\boldsymbol{s})$, let $(Y_k, Y_x) \sim \mathrm{DM}^{y_x - 1}(1-\alpha, w_x)$, set $y^{\mathrm{re}}_x = y_x - Y_x - 1$ and $y^{\mathrm{re}}_k = Y_k + 1$.

We will use the above resampling mechanism to give an autonomous description of the transition kernel of the decorated (α, γ)-chain from Definition 2.6.3. The arguments used in the proof of the following proposition are similar to the ones used to argue the spinal

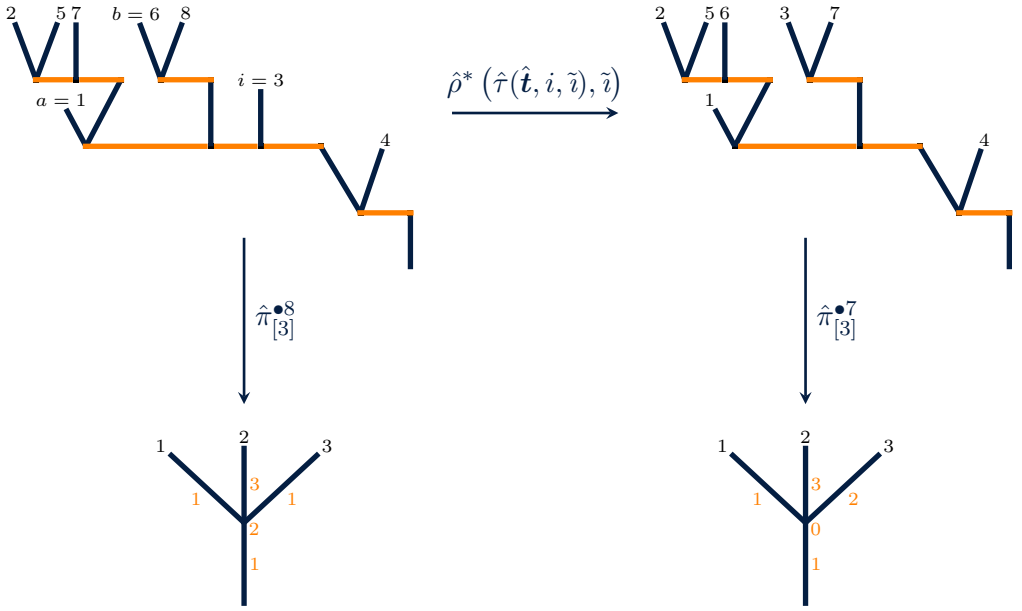


Figure 2.6: An illustration of the down-step in the semi-planar (α, γ)-chain on $\widehat{\mathbb{T}}_{[8]}$. Starting from $\hat{\boldsymbol{t}} \in \widehat{\mathbb{T}}_{[8]}$ (top left), leaf $i = 3$ is selected for deletion, whereby $a = 1$ and $b = 6$ so that $\tilde{\imath} = \max\{i, a, b\} = 6$. Hence the leaves labelled by 3 and 6, respectively, are swapped, and subsequently the leaf labelled by 6 is deleted, and $(7, 8)$ is relabelled by $(6, 7)$, yielding the top right semi-planar tree. In the bottom row the projection of the two trees onto $\mathbb{T}^{\bullet 8}_{[3]}$ and $\mathbb{T}^{\bullet 7}_{[3]}$, respectively, are depicted with the decorated masses marked in orange.

decomposition for the (α, γ)-growth model in [12]. In the following, recall that $q_{\alpha,\theta}$ denotes the decrement matrix of the (α, θ)-oCRP defined by (2.10).

Proposition 2.6.7. *Fix $1 \leq k < n \in \mathbb{N}$. The transition kernel of the decorated (α, γ)-chain on $\mathbb{T}^{\bullet n}_{[k]}$, $(T^{\bullet n}_k(m))_{m\in\mathbb{N}_0}$, can be characterized as follows:*

1. *Conditional on $T^{\bullet n}_{[k]}(m) = (\boldsymbol{s}, \mathbf{y}) \in \mathbb{T}^{\bullet n}_{[k]}$, select $x \in \mathrm{ins}(\boldsymbol{s})$ with probability $\frac{y_x}{n}$.*

2.a *Unless $x \in \mathrm{edge}(\boldsymbol{s})$ is external with $y_x = 1$, reduce y_x by 1.*

2.b *If $x = i \in \mathrm{edge}(\boldsymbol{s})$ is external with $y_x = 1$, let $v \in \mathrm{bp}(\boldsymbol{s})$ denote the parent of x in $\boldsymbol{s}$, and let c_v denote the number of children of v in $\boldsymbol{s}$.*

   - *If $c_v > 2$, let $N \sim \mathrm{BetaBin}^{y_v}(\alpha, (c_v - 2)\alpha - \gamma)$.*
     - *Given $N = n_g > 0$, sample Y_{new} from $q_{\alpha,\alpha}(n_g, \cdot)$, and replace y_x with Y_{new} and y_v with $y_v - Y_{new}$.*
     - *If $N = 0$, perform the down-step of the (α, γ)-chain on $\mathbb{T}_{[k]}$ from $\boldsymbol{s}$ (step (ii) of Definition 2.1.1) to form $\boldsymbol{t}^{\bullet}_{\downarrow} \in \mathbb{T}^{\bullet(n-1)}_{[k-1]}$, and resample leaf k in $\boldsymbol{t}^{\bullet}_{\downarrow}$.*
   - *If $c_v = 2$ do the following:*
     - *if $y_v > 0$, let $Y_{new} \sim q_{\alpha,\alpha-\gamma}(y_v, \cdot)$, and replace y_x with Y_{new} and y_v with $y_v - Y_{new}$.*
     - *if $y_v = 0$ and $y_{e_{v\downarrow}} > 0$, define $\boldsymbol{s}' = \varphi\left(\rho(\boldsymbol{s}, i), e_{v\downarrow}, i\right)$, and let $N_b \sim q_{\gamma,\gamma}(y_{e_{v\downarrow}}, \cdot)$. Conditional on the event $\{N_b = n_b\}$, sample $N_t - 1 \sim \mathrm{BetaBin}^{n_b-1}(1-\alpha, \alpha-\gamma)$, replace y_i with N_t, y_v with $N_b - N_t$, and $y_{e_{v\downarrow}}$ with $y_{e_{v\downarrow}} - N_b$.*
     - *if $y_v = y_{e_{v\downarrow}} = 0$, perform the down-step of the (α, γ)-chain on $\mathbb{T}_{[k]}$ from $\boldsymbol{s}$ (step (ii) of Definition 2.1.1) to form $\boldsymbol{t}^{\bullet}_{\downarrow} \in \mathbb{T}^{\bullet(n-1)}_{[k-1]}$, and resample leaf k in $\boldsymbol{t}^{\bullet}_{\downarrow}$.*

3. *Perform an up-step using the Markov kernel of the decorated (α, γ)-growth process from Definition 2.6.5.*

Proof. Before we consider the down-step itself, note that it does not make any difference if we perform the up-step for the semi-planar (α, γ)-chain and then project to the decorated trees, or if we project to the decorated trees immediately after the down-step, and then perform the up-step by using the decorated (α, γ)-growth process, see the brief discussion

following Definition 2.6.5. This ensures, that we in the following only have to argue that the down-step outlined in Step 1 and 2 is the correct characterization.

Consider the decorated (α, γ)-chain started at $\boldsymbol{t}^\bullet = (\boldsymbol{s}, \mathbf{y}) \in \mathbb{T}^{\bullet n}_{[k]}$. Now

(i) sample $\widehat{T}$ from $\mathrm{P}\left(\widehat{T}_n \in \cdot \;\middle|\; \hat{\pi}^{\bullet n}_{[k]}\left(\widehat{T}_n\right) = \boldsymbol{t}^\bullet\right)$,

(ii) in the down-step of the semi-planar (α, γ)-chain from $\widehat{T}$, condition on $I = i$ and $\tilde{I} = \tilde{\imath}$, delete $\tilde{\imath}$, and relabel $\tilde{\imath}+1, \ldots, n$ by $\tilde{\imath}, \ldots, n-1$ using the increasing bijection, obtaining $\widehat{T}^{\downarrow}$.

Furthermore let $T^{\bullet n}_{[k]}(m)^{\downarrow}$ denote $T^{\bullet n}_{[k]}(m)$ after the down-step. Our aim is to prove that

$$\mathrm{P}\left(T^{\bullet n}_{[k]}(m)^{\downarrow} \in \cdot \;\middle|\; T^{\bullet n}_{[k]}(m) = \boldsymbol{t}^\bullet\right) = \mathrm{P}\left(\hat{\pi}^{\bullet(n-1)}_{[k]}\left(\widehat{T}^{\downarrow}\right) \in \cdot\right) \tag{2.27}$$

Fundamentally, we now split up into cases based on $\boldsymbol{t}^\bullet$ and i where we recall the notation $B_x = g^{-1}(\{x\})$ for $x \in \mathrm{ins}(\boldsymbol{s})$ where g is the decoration function from $\hat{T}$ to $\boldsymbol{s}$. Since $\hat{T}$ is random, the sets $(B_x)_{x \in \mathrm{ins}(\boldsymbol{s})}$ form a random partition of $[n]$. However, we note that in all cases Step 1 is consistent with how we select leaf I in the semi-planar (α, γ)-chain, since $I \sim \mathrm{Unif}([n])$ immediately implies

$$\mathrm{P}\left(I \in B_x = g^{-1}(\{x\})\right) = \frac{y_x}{n}$$

for all $x \in \mathrm{ins}(\boldsymbol{s})$.

In the following we will let a, b and $\tilde{\imath} = \max\{i, a, b\}$ be defined as in Definition 2.4.3, and will let $(S^{\downarrow}, Y^{\downarrow}) := \hat{\pi}^{\bullet(n-1)}_{[k]}\left(\widehat{T}^{\downarrow}_n\right)$ be the decorated $[k]$-tree of mass $n-1$ obtained by projecting the semi-planar (α, γ)-chain after the down-step.

Case A1: $i \in B_x$ for external $x \in \mathrm{edge}(\boldsymbol{s})$, $y_x > 1$. Note how in this case $i \in B_x$ implies that $\tilde{\imath} = \max\{i, a, b\} \in B_x$ as well, implying that $\tilde{\imath} > k$. Hence the tree shape does not change during the down-step, and $(S^{\downarrow}, Y^{\downarrow})$ will have the property that $S^{\downarrow} = \boldsymbol{s}$, $Y^{\downarrow}_x = y_x - 1$, and $Y^{\downarrow}_{x'} = y_{x'}$ for all $\mathrm{ins}(\boldsymbol{s}) \ni x' \neq x$. This corresponds to Step 2.a where x is an external edge.

Case A2: $i \in B_x$ for internal $x \in \mathrm{edge}(\boldsymbol{s})$. Noting that $i > k$, the definition of a and b implies that $\tilde{\imath} \in B_x$ as well. Hence this case is similar to Case A1, and corresponds to Step 2.a where x is an internal edge.

Case A3: $i \in B_x$ **for** $x \in \mathrm{bp}(\boldsymbol{s})$**.** This is completely analogous to Case A2, and corresponds to Step 2.a where x is a branch point.

In all of the following cases, $v \in \mathrm{bp}(\boldsymbol{s})$ refers to the parent branch point of i in $\boldsymbol{s}$, c_v will refer to the number of children of v in $\boldsymbol{s}$, and we construct $\hat{\boldsymbol{s}} = \hat{\rho}\left(\widehat{T}, [n] \setminus [k]\right)$ which implies that $\hat{\pi}(\hat{\boldsymbol{s}}) = \boldsymbol{s}$.

Case B1: $i \in B_x$ **for external** $x \in \mathrm{edge}(\boldsymbol{s}), y_x = 1, y_v > 0, \tilde{\imath} > k$. Firstly, we note that the restrictions on i and y_x implies that $i \leq k$. Now consider the semi-planar structure in the branch point v, and note how $y_x = 1$ and $y_v > 0$ implies that $\max\{i, a, b\} = \max\{i, b\}$. Hence this case covers the situations where (1) $c_v > 2$ and a leaf label larger than k is found to the left (or right, if i is placed leftmost in v) of i in $\widehat{T}$, and (2) $c_v = 2$, in which case we are guaranteed to find a label larger than k to the right of i in $\widehat{T}$. In order to specify the decorated tree after the down-step, we only need to characterize the number of leaves found in the subtree of v containing $\tilde{\imath}$ in $\widehat{T}$.

From the decoration, it is clear that y_v leaves have been inserted to form subtrees of v, different from the ones already in $\hat{\boldsymbol{s}}$. For $c_v > 2$, with i not located leftmost in v, these subtrees will have grown in the following way. Initially there will be weight α to the immediate left of i, and an accumulated weight of $(c_v - 2)\alpha - \gamma$ in the rest of the branch point. Thus, the numbers of insertions happening to the immediate left of i versus everywhere else in v, respectively, will follow a Pólya urn scheme with these initial weights. Now, say that $n_g > 0$ leaves have been inserted in the "gap" between i and the left neighbour of i in $\hat{\boldsymbol{s}}$. These n_g leaves will form subtrees of v, and the location and sizes of these subtrees will be governed by an (α, α)-oCRP. Specifically, b will be located in the first subtree to the left of i in $\widehat{T}$, so denoting the size of this subtree by Y_{new}, we observe that Y_{new} is distributed as the size of the rightmost table in an (α, α)-oCRP with n_g customers. Hence swapping i and $\tilde{\imath}$, deleting $\tilde{\imath}$, relabelling $\{\tilde{\imath}+1, \ldots, n\}$ by $\{\tilde{\imath}, \ldots, n-1\}$ using the increasing bijection, and projecting to decorated $[k]$-trees of mass $n-1$, implies that we obtain $S^{\downarrow} = \boldsymbol{s}$, $Y_i^{\downarrow} = Y_{\mathrm{new}}$, $Y_v^{\downarrow} = y_v - Y_{\mathrm{new}}$ and $Y_x^{\downarrow} = y_x$ for all $x \in \mathrm{ins}\,(\boldsymbol{s}_{\downarrow}) \setminus \{v\}$.

The argument above is exactly the same if i is placed leftmost in v in $\widehat{T}$, as we then search for b to the right of leaf i instead of to the left, but still encounter an (α, α)-oCRP.

If $c = 2$, the argument is the same, except that we are guaranteed to find something to the right of i in v in $\widehat{T}$. Contrary to the above, the location and sizes of the subtrees of

v in $\widehat{T}$ with smallest leaf label larger than k is governed by an $(\alpha, \alpha-\gamma)$-oCRP, but the remainder of the argument carries over.

Case B2: $i \in B_x$ **for external** $x \in \text{edge}(\boldsymbol{s}), y_x = 1, y_v > 0, \tilde{\imath} \leq k$. Following the same reasoning as above, this case only contains the event where $n_g = 0$ leaves with labels larger than k have been inserted in the gap between i and the left (or right if i is placed leftmost in v) neighbour of i in v, implying that we have $\tilde{\imath} = \max\{i, b\} \leq k$. By swapping i and $\tilde{\imath}$ and subsequently deleting $\tilde{\imath}$, we end up removing one of the leaves that defines the tree shape $\boldsymbol{s}$. Consider how $\widehat{T}$ was grown from any semi-planar representative $\hat{\boldsymbol{s}}$ of $\boldsymbol{s}$ using the semi-planar (α, γ)-growth process. The sets $(B_x)_{x \in \text{ins}(\boldsymbol{s})}$ form a random partition of $[n]$ induced by a Pólya urn scheme, and it follows from Proposition 2.2.3 that the probability of seeing $k+1 \in B_x$ is exactly $\frac{\tilde{y}_x}{n-k}$ for any $x \in \text{ins}(\boldsymbol{s})$. In $\widehat{T}$, swap i and $\tilde{\imath}$, delete $\tilde{\imath}$ and relabel $\tilde{\imath}+1, \ldots, n$ by $\tilde{\imath}, \ldots, n-1$ using the increasing bijection, and consider the tree shape and decoration masses of the projected tree (these operations below are well-defined as no edges or vertices of $\boldsymbol{s}$ apart from the leaf edge $\tilde{\imath}$ are deleted). Say that $k+1 \in B_{X_0}$ for a random $X_0 \in \text{ins}(\boldsymbol{s})$, the tree shape after projecting to decorated trees will thus be $\varphi\left(\rho(\tau(\boldsymbol{s}, i, \tilde{\imath}), \tilde{\imath}), X_0, k\right)$, i.e.

(i) if $X_0 = v \in \text{bp}(\boldsymbol{s})$ the new tree shape will have k as a child of v, and

(ii) if $X_0 = e \in \text{edge}(\boldsymbol{s})$, k will split the edge e up into a new branch point and three edges.

Further considerations of the correspondence between urn schemes and inserting the leaves $B_{X_0} \setminus \{k+1\}$ into $\varphi(\boldsymbol{s}, X_0, k+1) = \hat{\pi}\left(\hat{\rho}(\widehat{T}, [n] \setminus [k+1])\right)$, or more precisely into the insertable parts of the projected tree specified above, shows that resampling $\tilde{\imath}$ exactly corresponds to deleting $\tilde{\imath}$, relabelling using the increasing bijection, and subsequently projecting to decorated $[k]$-trees of mass $n-1$.

Case B3: $i \in B_x$ **for external** $x \in \text{edge}(\boldsymbol{s})$, $y_x = 1$, $y_v = 0$, $\tilde{\imath} > k$. This case is where we have a binary branch point, and there is mass on the edge $e_{v\downarrow}$ in $\boldsymbol{s}$, but is otherwise similar to Case B1, with the notable difference that the leaf insertions into $e_{v\downarrow} \in \text{edge}(\boldsymbol{s})$ are governed by a (γ, γ)-oCRP. In this case, $\tilde{\imath} = b$ and b will be the smallest leaf in the second spinal bush on the ancestral line from leaf i in $\widehat{T}$. Letting N_b denote the random number of leaves in this spinal bush, by Proposition 2.2.5 we have that $N_b \sim q_{\gamma,\gamma}(e_{v\downarrow}, \cdot)$.

Further considerations of the insertions, as in Case B2, show that conditional on $N_b = n_b$, the number of leaves in each of the subtrees of of the spinal bush will be governed by a Chinese Restaurant process with parameters $(\alpha, \alpha - \gamma)$ with $n_b - 1$ customers. This is also an immediate consequence of the spinal decomposition of an (α, γ)-tree (see Lemma 11 of [12]).

Case B4: $i \in B_x$ **for external** $x \in \mathrm{edge}(\boldsymbol{s})$, $y_x = 1$, $y_v = 0$, $\tilde{\imath} \leq k$. This case is where we have a binary branch point, and there is no mass on $e_{v\downarrow}$ in $\boldsymbol{s}$, and so $\tilde{\imath} < k$ and the argument presented in Case B2 can be used verbatim. ■

With an autonomous description of the transition kernel for the decorated (α, γ)-chain on $\mathbb{T}^{\bullet n}_{[k]}$ we turn to the more complicated task of proving that the projection of the semi-planar (α, γ)-chain on $\widehat{\mathbb{T}}_{[n]}$ onto $\mathbb{T}^{\bullet n}_{[k]}$, $\left(\hat{\pi}^{\bullet n}_{[k]}\left(\widehat{T}_n(m)\right)\right)_{m\in\mathbb{N}_0}$, is a Markov chain on $\mathbb{T}^{\bullet n}_{[k]}$ with the transition kernel characterized in Proposition 2.6.7.

2.7 Sampling semi-planar (α, γ)-trees

Similar to how we characterized $\hat{\Pi}$ in (2.19), we now wish to characterize $\hat{\Pi}^{\bullet n}_k$. In order to do this, let us define the concept of internal structures which will allow us to efficiently sample non-planar or semi-planar $[n]$-trees from a collapsed or decorated tree of mass n.

So fix $\boldsymbol{t} \in \mathbb{T}_{[n]}$ and let $\pi^{\bullet n}_{[k]}(\boldsymbol{t}) = \boldsymbol{t}^\bullet = (\boldsymbol{s}, \mathbf{y})$ for some $\boldsymbol{s} \in \mathbb{T}_{[k]}$, and let g be the decorating function. There is a partition of all edges and vertices of $\boldsymbol{t}$ indexed by the elements of $\mathrm{ins}(\boldsymbol{s})$. Define for each $x \in \mathrm{ins}(\boldsymbol{s})$ the set of all vertices of $\boldsymbol{t}$ that are not vertices of $\boldsymbol{s}$, and from which x is the "first" edge or vertex of $\boldsymbol{s}$, respectively, that is reached on the ancestral line. More precisely, for each $v \in \mathrm{bp}(\boldsymbol{s})$ define

$$\mathring{V}^v_{\boldsymbol{s},\boldsymbol{t}} = \left\{ v' \in \mathrm{vert}(\boldsymbol{t}) \;\middle|\; \begin{array}{c} v' = v \text{ or } \exists j \in g^{-1}(\{v\})\ \exists v_0, \ldots, v_m \in \mathrm{vert}(\boldsymbol{t}): \\ v_0 = v, v_m = j, v' = v_i \text{ for some } i \in [m], \\ \text{and } v_{i-1} \prec_{\boldsymbol{t}} v_i \text{ for all } i \in [m] \end{array} \right\},$$

and

$$\mathring{E}^v_{\boldsymbol{s},\boldsymbol{t}} = \left\{ e' \in \mathrm{edge}(\boldsymbol{t}) \;\middle|\; \exists u, u' \in \mathring{V}^v_{\boldsymbol{s},\boldsymbol{t}} \colon e' = e_{uu'} \right\},$$

whilst for each $e \in \mathrm{edge}(\boldsymbol{s})$ with, say, $e = e_{uv}$, define

$$\mathring{V}^e_{\boldsymbol{s},\boldsymbol{t}} = \left\{ v' \in \mathrm{vert}(\boldsymbol{t}) \;\middle|\; \begin{array}{c} \exists j \in g^{-1}(\{e\})\ \exists v_0, \ldots, v_m \in \mathrm{vert}(\boldsymbol{t}) \colon \\ v_0 = u, v_m = j, v' = v_i \text{ for some } i \in [m], \\ \text{and } v_{i-1} \prec_{\boldsymbol{t}} v_i \text{ for all } i \in [m] \end{array} \right\}.$$

and

$$\mathring{E}^e_{\boldsymbol{s},\boldsymbol{t}} = \left\{ e' \in \mathrm{edge}(\boldsymbol{t}) \;\middle|\; \begin{array}{c} e' = e_{uv} \in \mathrm{edge}(\boldsymbol{t}) \text{ if } \mathring{V}^e_{\boldsymbol{s},\boldsymbol{t}} = \emptyset, \text{or} \\ \exists u', v' \in \mathring{V}^e_{\boldsymbol{s},\boldsymbol{t}} \colon e' = e_{u'v'} \text{ or } e' = e_{uu'} \text{ or } e' = e_{v'v} \end{array} \right\}.$$

We note that $\mathring{V}^x_{\boldsymbol{s},\boldsymbol{t}}$ either is empty, contains only $\{x\}$ if $x \in \mathrm{bp}(\boldsymbol{s})$, or will contain at least one leaf for each $x \in \mathrm{ins}(\boldsymbol{s})$, and in the latter case we continue to label these leaves by their corresponding label in $\boldsymbol{t}$. Additionally, the only situation where $\mathring{V}^x_{\boldsymbol{s},\boldsymbol{t}}$ is empty, but $\mathring{E}^x_{\boldsymbol{s},\boldsymbol{t}}$ is not, is when $x \in \mathrm{edge}(\boldsymbol{s})$ is internal and $g^{-1}(\{x\}) = \emptyset$, in which case $\mathring{E}^x_{\boldsymbol{s},\boldsymbol{t}} = \{x\}$. Likewise, whenever $x \in \mathrm{bp}(\boldsymbol{t})$ and $g^{-1}(\{x\}) = \emptyset$, $\mathring{V}^x_{\boldsymbol{s},\boldsymbol{t}} = \{x\}$ and $\mathring{E}^x_{\boldsymbol{s},\boldsymbol{t}} = \emptyset$.

The above sets constitute a partition of $\boldsymbol{t}$, i.e. a partition of both the vertex and edge sets of $\boldsymbol{t}$, labelled by the elements of $\boldsymbol{s}$, but it is only almost true that $(\mathring{V}^x_{\boldsymbol{s},\boldsymbol{t}}, \mathring{E}^x_{\boldsymbol{s},\boldsymbol{t}})$ constitutes a tree for each $x \in \mathrm{ins}(\boldsymbol{s})$. To ease stating later results, we amend this by adding some missing parts:

Definition 2.7.1 (Internal Structure)**.** Fix $\boldsymbol{t} \in \mathbb{T}_{[n]}$ and $\pi^{\bullet n}_{[k]}(\boldsymbol{t}) = (\boldsymbol{s}, \mathbf{y})$. For each $x \in \mathrm{ins}(\boldsymbol{s})$, make the following alterations to $\mathring{V}^x_{\boldsymbol{s},\boldsymbol{t}}$ and $\mathring{E}^x_{\boldsymbol{s},\boldsymbol{t}}$ defined above:

- If $x = e_{uv} \in \mathrm{edge}(\boldsymbol{s})$ is external such that $u \prec_{\boldsymbol{s}} v$ then
  - (i) add $\emptyset$ to $\mathring{V}^x_{\boldsymbol{s},\boldsymbol{t}}$, and
  - (ii) replace $e_{uu'}$ with $e_{\emptyset u'}$ in $\mathring{E}^x_{\boldsymbol{s},\boldsymbol{t}}$, with $u' \in \mathring{V}^x_{\boldsymbol{s},\boldsymbol{t}}$ uniquely characterised by $u \prec_{\boldsymbol{t}} u'$.
- If $x = e_{uv} \in \mathrm{edge}(\boldsymbol{s})$ is internal such that $u \prec_{\boldsymbol{s}} v$ then
  - (i) add $\emptyset$ and 1 to $\mathring{V}^x_{\boldsymbol{s},\boldsymbol{t}}$, and
  - (ii) if $\mathring{V}^x_{\boldsymbol{s},\boldsymbol{t}} = \emptyset$ replace the edge e_{uv} with $e_{\emptyset 1}$, and otherwise replace the two edges $e_{uu'}$ and $e_{v'v}$ with $e_{\emptyset u'}$ and $e_{v'1}$ in $\mathring{E}^x_{\boldsymbol{s},\boldsymbol{t}}$, where $u', v' \in \mathring{V}^x_{\boldsymbol{s},\boldsymbol{t}}$ uniquely characterized by $u \prec_{\boldsymbol{t}} u'$ and $v' \prec_{\boldsymbol{t}} v$.

- If $x = v \in \mathrm{bp}(\boldsymbol{s})$ with c children in $\boldsymbol{s}$ then

  (i) add $\emptyset$ and $1, \ldots, c$ to $\mathring{V}^x_{\boldsymbol{s},\boldsymbol{t}}$, and

  (ii) add the edges $e_{\emptyset v}, e_{v1}, \ldots, e_{vc}$ to $\mathring{E}^x_{\boldsymbol{s},\boldsymbol{t}}$.

In all three cases, replace each of the leaf labels in the resulting tree with their respective ranks to obtain $V^x_{\boldsymbol{s},\boldsymbol{t}}$ and $E^x_{\boldsymbol{s},\boldsymbol{t}}$, and define the *internal structure of x in $\boldsymbol{t}$* to be the tree $\textsc{int}(x) := \left(V^x_{\boldsymbol{s},\boldsymbol{t}}, E^x_{\boldsymbol{s},\boldsymbol{t}}\right)$.

We note that if $x \in \mathrm{edge}(\boldsymbol{s})$ is external then $\textsc{int}(x) \in \mathbb{T}_{[y_x]}$, if $x \in \mathrm{edge}(\boldsymbol{s})$ is internal then $\textsc{int}(x) \in \mathbb{T}_{[y_x+1]}$, whilst if $x \in \mathrm{bp}(\boldsymbol{s})$ with c children then $\textsc{int}(x) \in \mathbb{T}_{[c+y_x]}$. Additionally, note how all internal structures above could have been defined equivalently from any semi-planar tree, $\hat{\boldsymbol{t}}$, such that $\hat{\pi}(\hat{\boldsymbol{t}}) = \boldsymbol{t}$, and even turned into semi-planar trees, simply by letting each vertex in the internal structure inherit the permutation from $\hat{\boldsymbol{t}}$ as all ranks of subtrees are preserved (note how this is well-defined by the addition of the leaves $\{1, \ldots, c\}$ to $V^v_{\boldsymbol{s},\boldsymbol{t}}$ before relabelling). In most cases, it will be completely clear what $\boldsymbol{t}$ and $\boldsymbol{s}$ are in the above setup, but whenever that is not the case, we will write $\textsc{int}^{\boldsymbol{t}}_k(x)$ to refer to the internal structure of $x \in \mathrm{ins}(\boldsymbol{s})$ in $\boldsymbol{t} \in \mathbb{T}_{[n]}$, where $\boldsymbol{s} = \rho\left(\boldsymbol{t}, [n] \setminus [k]\right)$.

Before moving on to characterizing the (random) internal structure of a (random) decorated tree, we will need the following growth procedures, which are only slightly different from the (α, γ)-growth process, and arise from the internal structure of internal edges and branch points, respectively.

Definition 2.7.2 (Internal (α, γ)-growth process)**.** Fix $0 < \gamma \leq \alpha \leq 1$. Let T^γ_0 be the unique element of $\mathbb{T}_{[1]}$. For any $n \in \mathbb{N}_0$ construct T^γ_{n+1} conditional on $T^\gamma_n = \boldsymbol{t} \in \mathbb{T}_{[n+1]}$, by

$$w_x = \begin{cases} 1 - \alpha & \text{if } x \in \mathrm{edge}\left(\boldsymbol{s}\right) \text{ is external and } x \neq e_{1\downarrow} \\ \gamma & \text{if } x \in \mathrm{edge}\left(\boldsymbol{s}\right) \text{ is internal or } x = e_{1\downarrow} \\ \left(c_x - 1\right)\alpha - \gamma & \text{if } x \in \mathrm{bp}(\boldsymbol{s}) \text{ with } c_x \text{ children} \end{cases}$$

and then defining the conditional distribution of T^γ_{n+1} by

$$\mathrm{P}\left(T^\gamma_{n+1} = \varphi\left(\boldsymbol{s}, x, n+1\right) \;\middle|\; T^\gamma_n = \boldsymbol{s}\right) = \frac{w_x}{n + \gamma},$$

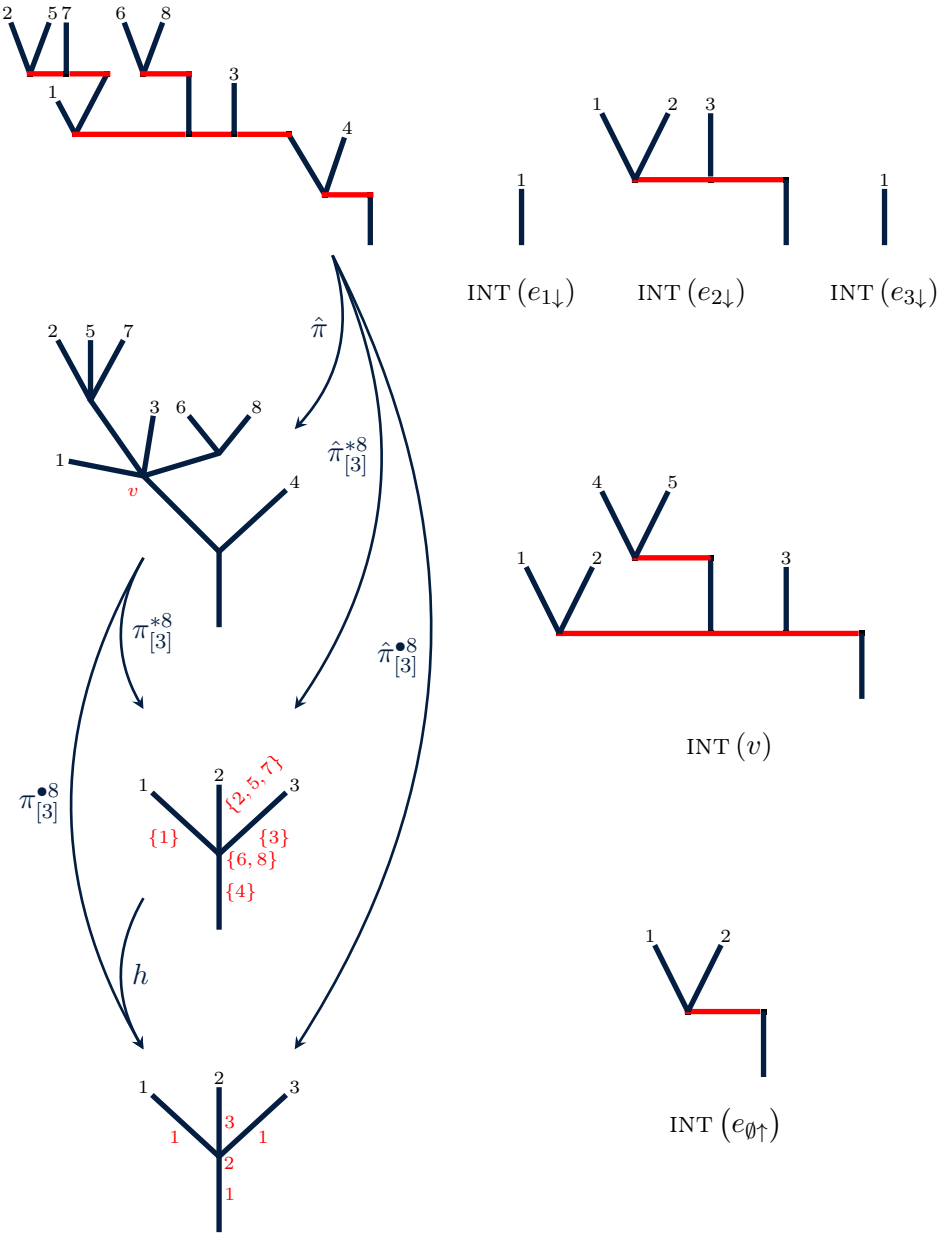


Figure 2.7: An illustration of the internal structures of an element of $\widehat{\mathbb{T}}_{[8]}$.

for each $x \in \mathrm{ins}(\boldsymbol{t})$. We refer to the sequence $(T_n^{\gamma})_{n\in\mathbb{N}_0}$ as the *internal* (α, γ)*-growth process.*

Definition 2.7.3 (c-order branch point (α, γ)-growth process)**.** Fix $0 \leq \gamma < \alpha \leq 1$ and $c \geq 2$. Let $T_0^{c-\mathrm{bp}}$ be the unique element of $\mathbb{T}_{[c]}$ with only one branch point. For any $n \in \mathbb{N}_0$

construct $T_{n+1}^{c-\mathrm{bp}}$ conditional on $T_n^{c-\mathrm{bp}} = \boldsymbol{t} \in \widehat{\mathbb{T}}_{[c+n]}$, by setting

$$w_x = \begin{cases} 1-\alpha & \text{if } x \in \operatorname{edge}(\boldsymbol{s}) \setminus \{e_{1\downarrow}, \ldots, e_{c\downarrow}\} \text{ is external} \\ \gamma & \text{if } x \in \operatorname{edge}(\boldsymbol{s}) \text{ is internal and } x \neq e_{\emptyset\uparrow} \\ (c_x - 1)\,\alpha - \gamma & \text{if } x \in \mathrm{bp}(\boldsymbol{s}) \text{ with } c_x \text{ children} \\ 0 & \text{if } x \in \{e_{\emptyset\uparrow}, e_{1\downarrow}, \ldots, e_{c\downarrow}\} \end{cases}$$

and then defining the conditional distribution of $T_{n+1}^{c-\mathrm{bp}}$ by

$$\mathrm{P}\left(T_{n+1}^{c-\mathrm{bp}} = \varphi(\boldsymbol{s}, x, n+1) \;\middle|\; T_n^{c-\mathrm{bp}} = \boldsymbol{s}\right) = \frac{w_x}{n + (c-1)\alpha - \gamma},$$

for each $x \in \mathrm{ins}(\boldsymbol{t})$. The sequence $\left(T_n^{c-\mathrm{bp}}\right)_{n\in\mathbb{N}_0}$ is referred to as the *c-order branch point* (α, γ)*-growth process.*

As the growth processes defined above are very similar to the (α, γ)-growth process described in Section 2.3, a wide range of results for the (α, γ)-growth process can can be modified to fit the internal growth processes of Definitions 2.7.2 and 2.7.3. It should also be clear that we can define a semi-planar version of the above growth procedures, just as we did for the (α, γ)-growth process. For the c-order branch point process, this entails sampling an initial distribution to provide the order of the initial c leaves in the first branch point, as described by (2.14). We will refer to these processes as the *semi-planar internal* (α, γ)*-growth process* and the *semi-planar c-order branch point* (α, γ)*-growth process.*

From the spinal decomposition of the (α, γ)-growth process [12], it is clear that there is an equally intimate link between (un)ordered Chinese Restaurant Processes and the internal (α, γ)-process. The only difference is that the ordered Chinese Restaurant involved in the spinal decomposition outlined in Lemma 11 of [12] is a (γ, γ)-oCRP rather than a $(\gamma, 1-\alpha)$-oCRP. This (γ, γ)-oCRP is the same as the one appearing in Proposition 2.6.7. From similar considerations we get the following characterization of the first split in the c-order branch point (α, γ)-growth process:

Lemma 2.7.4. *Fix* $0 < \gamma < \alpha \leq 1$ *and* $c \geq 2$. *Let* $\left(\widehat{T}_n^{c-bp}\right)_{n\in\mathbb{N}_0}$ *denote the semi-planar c-order branch point* (α, γ)*-process. For each* $n \in \mathbb{N}$ *and* $l \in [c-1]$ *let* Y_l^n *de-*

note the number of leaves with labels larger than c that have been inserted in the l'th gap (enumerated from left to right) in the branch point closest to the root in $\widehat{T}_n^{c-bp}$. In addition, let $\widehat{S}_l^n = \left(\widehat{S}_{l,1}^n,\dots,\widehat{S}_{l,M_l}^n\right)$, enumerated from left to right, denote the subtrees in the l'th gap of that same branch point, let X_l^n denote the associated ordered partition relabelled using the increasing bijection $\{c+1,\dots,c+n\}\mapsto\{1,\dots,n\}$, and let $\widehat{S}_{l,j}^{n,re}$ denote $\widehat{S}_{l,j}^n$ relabelled by $\{1,\dots,\#\widehat{S}_{l,j}^n\}$ using the increasing bijection for each $j\in[M_l], l\in[c-1]$. Then, $(Y_1^n,\dots,Y_{c-1}^n)\sim \mathrm{DM}^n(\alpha,\dots,\alpha,\alpha-\gamma)$, and conditionally given $\left(Y_1^n,\dots,Y_{c-1}^n\right)=(y_1,\dots,y_{c-1})$, $(X_l^n)_{l\in[c-1]}$ are independent and

$$\forall 1\le l<c-1\colon\ X_l^n\sim \mathrm{oCRP}^{y_l}(\alpha,\alpha)\quad \text{and}\quad X_{c-1}^n\sim \mathrm{oCRP}^{y_{c-1}}(\alpha,\alpha-\gamma)\,.$$

Further conditioning on $\left(\#\widehat{S}_{l,1}^n,\dots,\#\widehat{S}_{l,M_l}^n\right)=(n_{l,1},\dots,n_{l,M_l})$ yields that $\widehat{S}_{l,1}^{n,re},\dots,\widehat{S}_{l,M_l}^{n,re}$ are independent with $\widehat{S}_{l,j}^{n,re}\overset{D}{=}\widehat{T}_{n_{l,j}}$ for each $j\in[M_l], l\in[c-1]$.

Proof. All of the above assertions follows from (by now) standard growth process arguments, as one notes that the only possible insertions in these processes are either into the branch point described above, or into a subtree of that branch point with a least label larger than k. In the semi-planar tree, consider the insertion of these leaves into the "gaps" of the branch point between the subtrees with least leaf label smaller than k. There are $c-1$ such gaps, of which $c-2$ has weight α and the rightmost one has weight $\alpha-\gamma$. Hence the numbers of leaves being inserted into the gaps follow a Pólya urn scheme with initial weights $\mathbf{w}=(\alpha,\dots,\alpha,\alpha-\gamma)\in[0,1]^{c-1}$, and Proposition 2.2.3 directly yields that $\left(Y_1^n,\dots,Y_{c-1}^n\right)\sim\mathrm{DM}^n(\mathbf{w})$. When conditioning on which gap a new leaf is inserted into, the location within that gap is clearly governed by an oCRP, the distribution of which is detailed by the same argument as in Lemma 2.3.5. And lastly, when conditioning on which leaves end up in the same subtree of v, each of these subtrees will after relabelling be a semi-planar (α,γ)-tree with the corresponding number of leaves. ■

Corollary 2.7.5. *Fix $0<\gamma<\alpha\le 1$ and $c\ge 2$. Let $\left(T_n^{c-bp}\right)_{n\in\mathbb{N}_0}$ denote the (non-planar) c-order branch point (α,γ)-process. If $S_1^n,\dots,S_L^n$ denotes the subtrees with least label larger than c in the branch point closest to the root of T_n^{c-bp}, enumerated in the order of least leaf*

label, then

$$\mathrm{P}\left(L=l, S_1^n=\boldsymbol{s}_1,\ldots,S_m^n=\boldsymbol{s}_l\right)=\mathrm{P}\left(T'_1=\boldsymbol{s}_1^{re},\ldots,T'_k=\boldsymbol{s}_l^{re}\right)\\ \times\mathrm{P}\left((N_1,\ldots,N_L)=(n_1,\ldots,n_l)\right)$$

where $\boldsymbol{s}_{l'}^{re}$ *is* $\boldsymbol{s}_{l'}$ *relabelled by* $[n_{l'}]$ *using the increasing bijection and* $T'_{l'}\overset{D}{=}T_{n_{l'}}$ *for each* $l'\in[l]$, $\left(T'_{l'}\right)_{l'\in[l]}$ *are independent, and* $(N_1,\ldots,N_L)$ *are the numbers of customers sitting at the* L *tables of an* $(\alpha,(c-2)\alpha-\gamma)$-CRP *with* n *customers.*

We will now prove some useful versions of Lemma 2.4.5 and Theorem 2.4.6 for the semi-planar (α,γ)-growth process with a special tree as a starting point $\hat{\boldsymbol{t}}_0\in\mathbb{T}_{[n_0]}$ with associated weights $\mathbf{w}_0=\left(w^0_{x,l}\right)_{x\in\mathrm{ins}(\hat{\boldsymbol{t}}_0),0\le l\le c_x-1}$, where we use the convention that $c_x=1$ for $x\in\mathrm{edge}(\hat{\boldsymbol{t}}_0)$ and $1\le l\le c_x-1$ if $x\in\mathrm{bp}(\hat{\boldsymbol{t}})$ with c_x children. Crucially, we allow these weights to be different from the weight specification of the growth process. The result that we will now outline have wider implications for label swapping in more general structures that induce partitions and come about from a growth procedure, such as ordered Chinese Restaurant Processes. However, we will not investigate this further in this paper.

In the following we need to be careful about 'where' the modified weights appear after the insertion of a leaf according to the growth rule. Hence for any $\hat{\boldsymbol{t}}\in\widehat{\mathbb{T}}_{[n]}$ consider how $\hat{\boldsymbol{t}}'=\hat{\varphi}\left(\hat{\boldsymbol{t}},n+1,x,l\right)$ is formed by carrying out one step of the semi-planar (α,γ)-growth rule, where $x\in\mathrm{ins}(\hat{\boldsymbol{t}})$ and $0\le l\le c_x-1$. We will organize the weights of $\hat{\boldsymbol{t}}'$ such that

$$w'_{x',l'}=\begin{cases}\alpha & x'=x\in\mathrm{bp}(\hat{\boldsymbol{t}}),l'=l\\ w_{x,l-1} & x'=x\in\mathrm{bp}(\hat{\boldsymbol{t}}),l\le l'\le c_{x'}-1\\ \alpha-\gamma & x'=(n+1)\downarrow\in\mathrm{bp}(\hat{\boldsymbol{t}}'),x\in\mathrm{edge}(\hat{\boldsymbol{t}}),l'=1\\ \gamma & x=e_{uv}\in\mathrm{edge}(\hat{\boldsymbol{t}}),x'=e_{u(n+1)\downarrow},l'=0\\ w_{x,0} & x=e_{uv}\in\mathrm{edge}(\hat{\boldsymbol{t}}),x'=e_{(n+1)\downarrow v},l'=0\\ 1-\alpha & x'=n+1\in\mathrm{edge}(\hat{\boldsymbol{t}}'),l'=0\\ w_{x,l} & \text{otherwise.}\end{cases}\tag{2.28}$$

for every $x'\in\mathrm{ins}(\hat{\boldsymbol{t}}')$ and $0\le l'\le c_{x'}-1$. This is illustrated in Figure 2.8.

The *semi-planar* (α,γ)*-growth process started from* $\hat{\boldsymbol{t}}_0$ is simply the same growth process

as in Definition 2.3.4, started from $\hat{\boldsymbol{t}}_0 \in \widehat{\mathbb{T}}_{[n_0]}$, i.e. where

$$\mathrm{P}\left(\widehat{T}_1 = \varphi\left(\hat{\boldsymbol{t}}_0, x, l, n_0 + 1\right)\right) = \frac{w^0_{x,l}}{\sum_{x,l} w^0_{x,l}} \tag{2.29}$$

and we iteratively use the weight scheme defined in (2.28) to form T_{n+1} from T_n for every $n \in \mathbb{N}$.

We note that the above specification of the weights is consistent with the original definition of the semi-planar (α, γ)-growth process, if the weights $\mathbf{w}_0$ above are the same as those specified in Definition 2.3.4. The importance of organizing the weights as in (2.28) is that we can handle explicitly starting a semi-planar (α, γ)-growth process from a tree with associated weights different from the ones coming out of the semi-planar (α, γ)-growth process. This is the case for the semi-planar internal (α, γ)-growth process as well as the semi-planar c-order branch points (α, γ)-growth process. We have the following result, where we recall that f is the transformation function from Definition 2.4.3.

Proposition 2.7.6. *Fix $n_0 \in \mathbb{N}$ and $\hat{\boldsymbol{t}}_0 \in \mathbb{T}_{[n_0]}$ with associated non-negative weights $\mathbf{w}_0 = \left(w^0_{x,l}\right)_{x \in \mathrm{ins}(\hat{\boldsymbol{t}}_0), 0 \le l \le c_x - 1}$ such that $\sum_{x,l} w^0_{x,l} > 0$. Let $\left(\widehat{T}_m\right)_{1 \le m \le n}$ be the first n steps of the semi-planar (α, γ)-growth process started from $\hat{\boldsymbol{t}}_0$. Fix $n_0 + 1 \le i \le \tilde{\imath} \le n_0 + n$ and define $E_{i,\tilde{\imath}} = \left\{\tilde{\imath} = \max\{i, a, b\} \text{ in } \widehat{T}_n\right\}$, with a and b from Definition 2.4.3. Then*

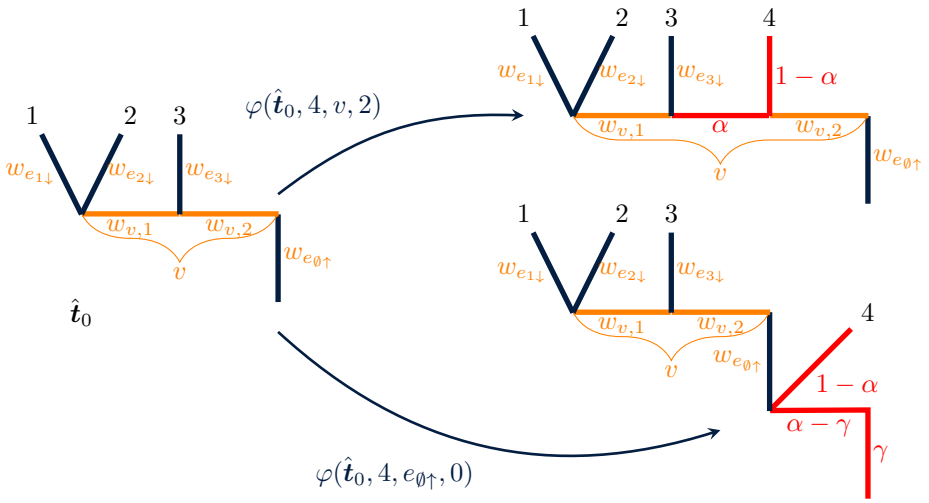


Figure 2.8: Illustration of the structuring of weights following a leaf insertion into a tree with pre-specified weights that might not coincide with the usual weights of the (α, γ)-growth process, as described by (2.28). The weights appearing as a result of the leaf insertion are marked in red.

$E_{i,\tilde{\imath}} \perp\!\!\!\perp \widehat{T}_{\tilde{\imath}-1}$ and, conditional on the event $E_{i,\tilde{\imath}}$, $\hat{\rho}^*\left(\hat{\tau}^*\left(\widehat{T}_n, i, \tilde{\imath}\right), \tilde{\imath}\right)$ and T_{n-1} have the same distribution. Furthermore, if $I \sim \mathrm{Unif}\left([n_0+n] \setminus [n_0]\right)$, then

$$\mathrm{P}\left(\hat{\rho}^*\left(f\left(\widehat{T}_n, I\right)\right) \in \cdot\right) = \mathrm{P}\left(\widehat{T}_{n-1} \in \cdot\right) \tag{2.30}$$

for any $n \in \mathbb{N}$.

Proof. The proof of Lemma 2.4.5 can used almost verbatim here, as we note that the restrictions on i and $\tilde{\imath}$ ensure none of the insertable parts of the trees $\widehat{T}_{\tilde{\imath}-1}, \ldots, \widehat{T}_n$ that contribute to the characterization of $E_{i,\tilde{\imath}}$ have $\mathbf{w}_0$ as associated weights. This is a consequence of the way we have organized the weights in the growth process, combined with the definition of a and b from Definition 2.4.3. ■

Note how Proposition 2.7.6 generalises Lemma 2.4.5. This generalisation is useful, as have have the following corollaries, which are easily checked to satisfy the assumptions of Proposition 2.7.6.

Corollary 2.7.7. *Let $\left(\widehat{T}_m^{\gamma}\right)_{m\in[n]}$ be the first n steps of the semi-planar internal (α,γ)-growth process. Fix $2 \le i \le \tilde{\imath} \le n$ and, with a and b as in Definition 2.4.3, define $E_{i,\tilde{\imath}}^{\gamma} = \left\{\tilde{\imath} = \max\{i,a,b\} \text{ in } \widehat{T}_n^{\gamma}\right\}$. Then $E_{i,\tilde{\imath}}^{\gamma} \perp\!\!\!\perp \widehat{T}_{i-1}^{\gamma}$ and, conditional on the event $E_{i,\tilde{\imath}}^{\gamma}$, $\hat{\rho}\left(\hat{\tau}\left(\widehat{T}_n^{\gamma}, i, \tilde{\imath}\right), \tilde{\imath}\right)$ and $\widehat{T}_{n-1}^{\gamma}$ have the same distribution.*
Furthermore, if $I \sim \mathrm{Unif}\left(\{2, \ldots, n+1\}\right)$, then

$$\mathrm{P}\left(\hat{\rho}^*\left(f\left(\widehat{T}_n^{\gamma}, I\right)\right) \in \cdot\right) = \mathrm{P}\left(\widehat{T}_{n-1}^{\gamma} \in \cdot\right) \tag{2.31}$$

for any $n \in \mathbb{N}$.

Corollary 2.7.8. *Fix $c \ge 2$. Let $\left(\widehat{T}_m^{c-bp}\right)_{m\in[n]}$ be the first n steps of the semi-planar c-order branch point (α,γ)-growth process started from $\hat{\boldsymbol{s}} \in \widehat{\mathbb{T}}_{[c]}$ with $\#\,\mathrm{bp}(\hat{\boldsymbol{s}}) = 1$. Fix $c < i \le \tilde{\imath} \le c+n$ and define $E_{i,\tilde{\imath}}^{c-bp} = \left\{\tilde{\imath} = \max\{i,a,b\} \text{ in } \widehat{T}_n^{c-bp}\right\}$, with a and b as in Definition 2.4.3. Then $E_{i,\tilde{\imath}}^{c-bp} \perp\!\!\!\perp \widehat{T}_{i-1}^{c-bp}$ and, conditional on the event $E_{i,\tilde{\imath}}^{c-bp}$, $\hat{\rho}\left(\hat{\tau}\left(\widehat{T}_n^{c-bp}, i, \tilde{\imath}\right), \tilde{\imath}\right)$ and $\widehat{T}_{n-1}^{c-bp}$ have the same distribution. Furthermore, if $I \sim \mathrm{Unif}\left(\{c+1, \ldots, c+n\}\right)$, then*

$$\mathrm{P}\left(\hat{\rho}^*\left(f\left(\widehat{T}_n^{c-bp}, I\right)\right) \in \cdot\right) = \mathrm{P}\left(\widehat{T}_{n-1}^{c-bp} \in \cdot\right) \tag{2.32}$$

for any $n \in \mathbb{N}$.

With the above machinery in place we can now characterize the decorated trees in the following way:

Lemma 2.7.9 (Spatial Markov property)**.** *Fix $n \in \mathbb{N}$. Let $(T_m)_{m\in[n]}$, $(T_m^{\gamma})_{m\in[n]}$, and $\left(T_m^{c-bp}\right)_{m\in[n]}$ denote the first n steps of the (α,γ)-, internal (α,γ)-, and c-order branch point (α,γ)-growth processes, respectively. Fix $\boldsymbol{t} \in \mathbb{T}_{[n]}$, and define $\boldsymbol{t}^{\bullet} = \pi_{[k]}^{\bullet n}(\boldsymbol{t}) = (\boldsymbol{s},\mathbf{y}) \in \mathbb{T}_{[k]}^{\bullet n}$ and $\boldsymbol{t}^{*} = \pi_{[k]}^{*n}(\boldsymbol{t}) = \left(\boldsymbol{s}, (g^{-1}(\{x\}))_{x\in\mathrm{ins}(\boldsymbol{s})}\right)) \in \mathbb{T}_{[k]}^{*n}$. Then*

(i) $(\textsc{int}(x))_{x\in\mathrm{ins}(\boldsymbol{s})}$ *are conditionally independent given* $\pi_{[k]}^{\bullet n}(T_n) = \boldsymbol{t}^{\bullet}$,

(ii) $(\textsc{int}(x))_{x\in\mathrm{ins}(\boldsymbol{s})}$ *are conditionally independent given* $\pi_{[k]}^{*n}(T_n) = \boldsymbol{t}^{*}$,

(iii) $(\textsc{int}(x))_{x\in\mathrm{ins}(\boldsymbol{s})}$ *and* $\left(g^{-1}(\{x\})\right)_{x\in\mathrm{ins}(\boldsymbol{s})}$ *are conditionally independent given* $\pi_{k}^{\bullet n}(T_n) = \boldsymbol{t}^{\bullet}$,

(iv) $\mathrm{P}\left(\textsc{int}(x) = \cdot \mid \pi_{k}^{\bullet n}(T_n) = \boldsymbol{t}^{\bullet}\right) = \mathrm{P}(T_{y_x}^{\gamma} = \cdot)$ *for each internal* $x \in \mathrm{edge}(\boldsymbol{s})$,

(v) $\mathrm{P}\left(\textsc{int}(x) = \cdot \mid \pi_{k}^{\bullet n}(T_n) = \boldsymbol{t}^{\bullet}\right) = \mathrm{P}(T_{y_x} = \cdot)$ *for each external* $x \in \mathrm{edge}(\boldsymbol{s})$, *and*

(vi) $\mathrm{P}\left(\textsc{int}(x) = \cdot \mid \pi_{k}^{\bullet n}(T_n) = \boldsymbol{t}^{\bullet}\right) = \mathrm{P}(T_{y_x}^{c-bp} = \cdot)$ *for each* $x \in \mathrm{bp}(\boldsymbol{s})$ *with c children.*

Proof. Firstly note that the conditional independence in (ii) follows directly from the independence of the insertions of new leaves in the (α,γ)-growth process. This observation directly implies (i). For (iii), fix $x \in \mathrm{ins}(\boldsymbol{s})$, and note that the probability of $\textsc{int}(x)$ being equal to a specific tree is independent of the labels, by virtue of the growth process. Next, (iv) and (v) follows by noting how T_n is produced from $T_k = \boldsymbol{s}$, since the latter is a part of what we condition on. For any $x \in \mathrm{edge}(\boldsymbol{s})$, note how $\textsc{int}_k^{T_k}(x)$ will be the unique element of $\mathbb{T}_{[1]}$, where the edge has associated weight γ if x is an internal edge and $1-\alpha$ if x is a leaf edge, respectively. Now observe how each $\textsc{int}(x)$ grows according to either the (α,γ)- or the internal (α,γ)-growth process, according to the initial weight of the leaf. For (vi), note how for fixed $x \in \mathrm{bp}(\boldsymbol{s})$ with $y_x > 0$ we can construct $\textsc{int}(x)$ using the c-order branch point (α,γ)-growth process. The insertion of a leaf into the branch point, x, corresponds to opening up a new table in a CRP, and for every such insertion the weight at x increases by α. By construction of $\textsc{int}(x)$ all of these subtrees will be located

at the branch point closest to the root in $\textsc{int}(x)$, and each of the subtrees evolves as a (α, γ)-growth process. Hence we obtain the above characterization. ■

In addition to the above characterization we have the result below, which will play a role in the relabelling of a decorated $[k]$-tree later on, where we recall that $\tilde{y}_x$, as in Definition 2.6.6, is equal to $y_x - 1$ for external $x \in \text{edge}(\boldsymbol{s})$ and equal to y_x otherwise.

Lemma 2.7.10. *Let $(T_m)_{1 \le m \le n}$ denote the first n elements of the (α, γ)-growth process. Fix $\boldsymbol{t} \in \mathbb{T}_{[n]}$ and define $\pi^{\bullet n}_{[k]}(\boldsymbol{t}) = (\boldsymbol{s}, \mathbf{y}) \in \mathbb{T}^{\bullet n}_{[k]}$. Then*

$$\mathrm{P}\left(T_{k+1} = \varphi(\boldsymbol{s}, x, k+1) \;\middle|\; \pi^{\bullet n}_{[k]}(T_n) = (\boldsymbol{s}, y)\right) = \frac{\tilde{y}_x}{n-k}$$

for each $x \in \text{ins}(\boldsymbol{s})$.

Proof. Note how T_n is constructed from $T_k = \boldsymbol{s}$ in the (α, γ)-growth process, if we are solely interested in the number of leaves being inserted into each $x \in \text{ins}(\boldsymbol{s})$. Then this is simply a generalized Pólya urn scheme with weights corresponding to the weight associated with an insertable part x in $\boldsymbol{s}$. In this analogy we are interested in the colour of the first ball drawn from this urn scheme where we have a total of $n - k$ draws. According to Proposition 2.2.3 (iv) this is exactly $\frac{\tilde{y}_x}{n-k}$ for each $x \in \text{ins}(\boldsymbol{s})$. ■

The above proposition states that, if we follow an (α, γ)-growth process and condition on a specific decorated tree, the probability that we originally inserted leaf $k+1$ in a specific location of the tree shape associated with the decorated tree, only depends on the total number of insertions we made into that part of the tree shape.

2.8 Consistency in stationarity

Having obtained an autonomous description of the decorated (α, γ)-chain, we now turn our focus to proving that the projection of the (α, γ)-chain on $\widehat{\mathbb{T}}_{[n]}$ onto the space of decorated $[k]$-trees of mass n, $\left(\hat{\pi}^{\bullet n}_{[k]}\left(\widehat{T}_n(m)\right)\right)_{m \in \mathbb{N}_0}$, is indeed the decorated (α, γ)-chain on $\mathbb{T}^{\bullet n}_{[k]}$. For brevity we have used the notation that $\hat{\pi}^{\bullet n}_{[k]} := \pi^{\bullet n}_{[k]} \circ \hat{\pi}$. As noted in [21], this is an instance of a more general question: under what conditions is a surjective function of a Markov chain Markovian in its own right? To be more precise, let $(X_m)_{m \in \mathbb{N}_0}$ be a Markov chain on some finite state space $\mathbb{X}$, and let $\lambda \colon \mathbb{X} \to \mathbb{Y}$ be a function. Then the question is, if

$Y_m = \lambda(X_m)$, $m \in \mathbb{N}_0$, is a Markov chain. We will use the following two well-known criteria.

Proposition 2.8.1. *Let $(X_m)_{m\in\mathbb{N}_0}$ be a Markov chain on a finite state space $\mathbb{X}$ with transition kernel K. Assume that $\lambda\colon \mathbb{X} \to \mathbb{Y}$ is a surjective function.*

(i) If for each $x_1, x_2 \in \mathbb{X}$ such that $\lambda(x_1) = \lambda(x_2)$ it holds that

$$K\left(x_1, \lambda^{-1}(\{y\})\right) = K\left(x_2, \lambda^{-1}(\{y\})\right) \qquad \text{for any } y \in \mathbb{Y}, \tag{2.33}$$

define the transition kernel, Q, by $Q(y_1, \{y_2\}) = K(x, \lambda^{-1}(\{y_2\}))$ for any $x \in \lambda^{-1}(\{y_1\})$.

(ii) Let q be a stationary distribution of $(X_m)_{m\in\mathbb{N}_0}$, and further assume that $X_0 \sim \Lambda(Y_0, \cdot)$ for some $\mathbb{Y}$-valued random variable Y_0, where $\Lambda(y, \cdot) = q(\cdot \mid \lambda = y)$. If for each $y \in \mathbb{Y}$ it holds that

$$\mathrm{P}\left(X_1 = x \mid \lambda(X_0) = y_0, \lambda(X_1) = y\right) = \Lambda\left(y, \{x\}\right) \tag{2.34}$$

for all $x \in \mathbb{X}$ with $\lambda(x) = y$, define $Q = \Lambda K \lambda$ using the same notation as in (2.20).

In either case $(\lambda(X_m))_{m\in\mathbb{N}_0}$ is a Markov chain with transition kernel, Q, and initial distribution given by $\lambda(X_0)$.

In the above, (i) is referred to as the *Kemeny-Snell criterion* [33], whilst (ii) is known as the *intertwining criterion* [46]. The goal for this section is to combine the two criteria to prove that $\left(\pi^{\bullet n}_{[k]}\left(T_n(m)\right)\right)_{m\in\mathbb{N}_0}$ is a Markov chain with transition kernel described in Proposition 2.6.7. Illustrated in Figure 2.9, we seek to do this by splitting this up, so rather than projecting all the way to decorated trees we will use collapsed trees as an intermediary. In doing this we note that $\pi^{\bullet n}_{[k]} = h \circ \pi^{* n}_{[k]}$, where h denotes the natural projection from $\mathbb{T}^{* n}_{[k]}$ to $\mathbb{T}^{\bullet n}_{[k]}$, i.e. $\boldsymbol{t}^* = \left(\boldsymbol{s}, (B_x)_{x\in\mathrm{ins}(\boldsymbol{s})}\right) \overset{h}{\mapsto} \left(\boldsymbol{s}, (\#B_x)_{x\in\mathrm{ins}(\boldsymbol{s})}\right) = \boldsymbol{t}^\bullet$.

We therefore construct an auxiliary Markov chain on $\mathbb{T}^{* n}_{[k]}$ analogously to the way we constructed the decorated (α, γ)-chain on $\mathbb{T}^{\bullet n}_{[k]}$ in Definition 2.6.3. Define a Markov kernel from $\mathbb{T}^{* n}_{[k]}$ to $\widehat{\mathbb{T}}_{[n]}$ by

$$\forall \boldsymbol{t}^* \in \mathbb{T}^{* n}_{[k]},\ \hat{\boldsymbol{t}} \in \widehat{\mathbb{T}}_{[n]}\colon \quad \hat{\Pi}^{* n}_k(\boldsymbol{t}^*, \{\hat{\boldsymbol{t}}\}) := \mathrm{P}\left(\widehat{T}_n = \hat{\boldsymbol{t}} \;\middle|\; \hat{\pi}^{* n}_{[k]}\left(\widehat{T}_n\right) = \boldsymbol{t}^*\right), \tag{2.35}$$

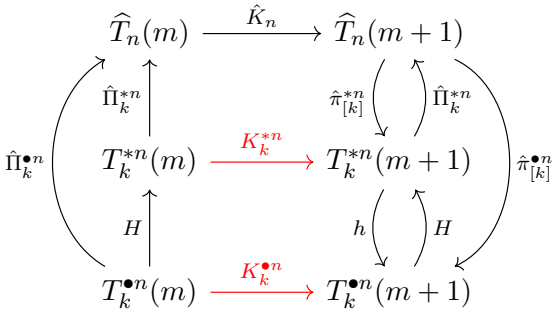


Figure 2.9: Commutative diagram illustrating the construction of $K_k^{\bullet n}$ described in Proposition 2.6.7 and the proof tactic to show that the projection of the semi-planar (α, γ)-chain is indeed a Markov chain with that same transition kernel.

where $\widehat{T}_n$ is the nth step of the semi-planar (α, γ)-growth process, and $\hat{\pi}_{[k]}^{*n} := \pi_{[k]}^{*n} \circ \hat{\pi}$ denotes the projection from semi-planar n-trees to collapsed $[k]$-trees of mass n, for brevity. Now define the *collapsed* (α, γ)*-chain on* $\mathbb{T}_{[k]}^{*n}$ be a Markov chain on $\mathbb{T}_{[k]}^{*n}$ with transition kernel

$$K_k^{*n} := \hat{\Pi}_k^{*n} \hat{K}_n \hat{\pi}_{[k]}^{*n}, \tag{2.36}$$

Lastly, we define a Markov kernel from $\mathbb{T}_{[k]}^{\bullet n}$ to $\mathbb{T}_{[k]}^{*n}$ by

$$\forall \boldsymbol{t}^\bullet \in \mathbb{T}_{[k]}^{*n},\ \boldsymbol{t}^* \in \mathbb{T}_{[k]}^{*n}: \quad H(\boldsymbol{t}^\bullet, \{\boldsymbol{t}^*\}) := \mathrm{P}\left(\pi_{[k]}^{*n}\left(\widehat{T}_n\right) = \boldsymbol{t}^* \;\middle|\; \hat{\pi}_{[k]}^{\bullet n}\left(\widehat{T}_n\right) = \boldsymbol{t}^\bullet\right). \tag{2.37}$$

The goal is thus to show that both the upper and lower part of the diagram in Figure 2.9 commutes, respectively. We will deploy the Kemeny-Snell criterion to show that the lower diagram commutes, whilst the intertwining criterion will be used for the upper part of the diagram. We will start by proving that the projection from collapsed trees to decorated trees satisfies the Kemeny-Snell criterion.

Lemma 2.8.2. *Let $(T_n(m))_{m\in\mathbb{N}_0}$ denote the (α, γ)-chain. Then*

$$\begin{aligned} &\mathrm{P}\left(\pi_{[k]}^{*n}\left(T_n(1)\right) \in h^{-1}(\{\boldsymbol{t}^\bullet\}) \;\middle|\; \pi_{[k]}^{*n}\left(T_n(0)\right) = \boldsymbol{t}_1^*\right) \\ = \;&\mathrm{P}\left(\pi_{[k]}^{*n}\left(T_n(1)\right) \in h^{-1}(\{\boldsymbol{t}^\bullet\}) \;\middle|\; \pi_{[k]}^{*n}\left(T_n(0)\right) = \boldsymbol{t}_2^*\right) \end{aligned}$$

for all $\boldsymbol{t}^\bullet \in \mathbb{T}_{[k]}^{\bullet n}$, whenever $h(\boldsymbol{t}_1^*) = h(\boldsymbol{t}_2^*)$.

Proof. Say that $\boldsymbol{t}_1^* = \left(\boldsymbol{s}_1, \left(B_x^1\right)_{x\in\mathrm{ins}(\boldsymbol{s}_1)}\right)$ and $\boldsymbol{t}_2^* = \left(\boldsymbol{s}_2, \left(B_x^2\right)_{x\in\mathrm{ins}(\boldsymbol{s}_2)}\right)$. As $h(\boldsymbol{t}_1^*) = h(\boldsymbol{t}_2^*)$

we easily see that $\boldsymbol{s}_1 = \boldsymbol{s}_2 =: \boldsymbol{s}$ and $\#B_x^1 = \#B_x^2$, ensuring that $\mathrm{P}\left(I \in B_x^1\right) = \mathrm{P}\left(I \in B_x^2\right)$, for every $x \in \mathrm{ins}(\boldsymbol{s})$.

Now observe that there is a unique bijection $\tau\colon [n] \to [n]$ which is increasing on B_x^1 for each $x \in \mathrm{ins}\left(\boldsymbol{s}_1\right)$ such that $B_x^2 = \tau\left(B_x^1\right)$ for every $x \in \mathrm{ins}\left(\boldsymbol{s}\right)$, and use this bijection to obtain $\tau(\boldsymbol{t})$ from $\boldsymbol{t} \in \mathbb{T}_{[n]}$. Then it holds that

$$\left\{\tau(\boldsymbol{t}) \;\middle|\; \pi_{[k]}^{*n}\left(\boldsymbol{t}\right) = \boldsymbol{t}_1^*\right\} = \left(\pi_{[k]}^{*n}\right)^{-1}\left(\boldsymbol{t}_2^*\right)$$

and that $\mathrm{P}\left(T_n = \boldsymbol{t}\right) = \mathrm{P}\left(T_n = \tau(\boldsymbol{t})\right)$ for each $\boldsymbol{t} \in \mathbb{T}_{[n]}$ that satisfies $\pi_{[k]}^{*n}(\boldsymbol{t}) = \boldsymbol{t}_1^*$. Crucially, the rank of each element of any set B_x is preserved under this relabelling. Note how the internal structure, $\textsc{int}(x)$, is completely preserved under τ by Lemma 2.7.9 (iii), for each $x \in \mathrm{ins}\left(\boldsymbol{s}\right)$, and that the distribution of each $\textsc{int}(x)$ is described by Lemma 2.7.9(iv)-(vi). This immediately implies that $\pi_{[k]}^{\bullet n}\left(T_n(1)\right) = \pi_{[k]}^{\bullet n}\left(T_n^\tau(1)\right)$ (not just in distribution!), where $T_n(\cdot)$ and $T_n^\tau(\cdot)$ are both (non-planar) (α,γ)-chains but $T_n(0) = \boldsymbol{t}$ and $T_n^\tau(0) = \tau(\boldsymbol{t})$. This proves that the Kemeny-Snell criterion is satisfied. ■

With the bottom diagram of Figure 2.9 commuting, we only need to prove that the upper part of the same diagram commutes as well. This is a slightly more intricate argument, where we will aim to show that the intertwining criterion is satisfied with $(X_m)_{m\in\mathbb{N}_0} = \left(\widehat{T}_n(m)\right)_{m\in\mathbb{N}_0}$ and $\lambda = \hat{\pi}_{[k]}^{*n} = \pi_{[k]}^{*n} \circ \hat{\pi}\colon \widehat{\mathbb{T}} \to \mathbb{T}_{[k]}^{*n}$.

Lemma 2.8.3. *Let* $\left(\widehat{T}_n(m)\right)_{m\in\mathbb{N}_0}$ *be the semi-planar* (α,γ)*-chain with* $\widehat{T}_n(0) \sim \hat{\Pi}_k^{*n}(\boldsymbol{t}_0^*, \cdot)$ *for some* $\boldsymbol{t}_0^* \in \mathbb{T}_{[k]}^{*n}$*. Then*

$$\mathrm{P}\left(\widehat{T}_n(1) = \hat{\boldsymbol{t}} \;\middle|\; \hat{\pi}_{[k]}^{*n}\left(\widehat{T}_n(0)\right) = \boldsymbol{t}_0^*, \hat{\pi}_{[k]}^{*n}\left(\widehat{T}_n(1)\right) = \boldsymbol{t}_1^*\right) = \hat{\Pi}_k^{*n}(\boldsymbol{t}_1^*, \{\hat{\boldsymbol{t}}\})$$

for all $\hat{\boldsymbol{t}} \in \widehat{\mathbb{T}}_{[n]}$*, whenever* $\boldsymbol{t}_0^*, \boldsymbol{t}_1^* \in \mathbb{T}_{[k]}^{*n}$ *are such that*

$$\mathrm{P}\left(\hat{\pi}_{[k]}^{*n}\left(\widehat{T}_n(0)\right) = \boldsymbol{t}_0^*, \hat{\pi}_{[k]}^{*n}\left(\widehat{T}_n(1)\right) = \boldsymbol{t}_1^*\right) > 0.$$

Essentially, the following proof will follow the structure of Proposition 2.6.7, meaning that we will split up the proof into cases based on the location and value of I in $\hat{\pi}_{[k]}^{*n}\left(\widehat{T}_n\right)$:

Proof. Consider the collapsed tree $\boldsymbol{t}_0^* = \left(\boldsymbol{s}, (B_x)_{x\in\mathrm{ins}(\boldsymbol{s})}\right) \in \mathbb{T}_{[k]}^{*n}$, the event that the projection of the initial state of the semi-planar (α,γ)-chain onto $\mathbb{T}_{[k]}^{*n}$ yields the collapsed tree

$\boldsymbol{t}_0^*$, $\left\{\hat{\pi}_{[k]}^{*n}\left(\widehat{T}_n(0)\right) = \boldsymbol{t}_0^*\right\}$, and the internal structure as characterized by the spatial Markov property in Lemma 2.7.9. In light of this result it is sufficient to show that the internal structure is preserved under the down-step if we further condition on the selected leaf I, the deleted leaf $\tilde{I}$, and in some cases additional information about label sets, ensuring that we uniquely characterize the collapsed tree after the down-step. Hence, we

(i) sample $\widehat{T}$ from $\mathrm{P}\left(\widehat{T}_n \in \cdot \;\middle|\; \hat{\pi}_{[k]}^{*n}\left(\widehat{T}_n\right) = \boldsymbol{t}_0^*\right)$,

(ii) condition on $I = i$ and $\tilde{I} = \tilde{\imath}$ (and in some cases additional information), delete $\tilde{\imath}$, and relabel $\tilde{\imath}{+}1, \ldots, n$ by $\tilde{\imath}, \ldots, n-1$ using the increasing bijection, obtaining $\widehat{T}^{\downarrow}$, and

(iii) project to $\mathbb{T}_{[k]}^{*(n-1)}$ to obtain T^*, and study the internal structures in $\widehat{T}^{\downarrow}$.

Even though the statement of the lemma formally involves the up-step, we will as an intermediary result show that

$$\begin{aligned} & \mathrm{P}\left(\widehat{T}_n^{\downarrow} = \hat{\boldsymbol{t}} \;\middle|\; \hat{\pi}_{[k]}^{*n}\left(\widehat{T}_n(0)\right) = \boldsymbol{t}_0^*, \hat{\pi}_{[k]}^{*(n-1)}\left(\widehat{T}^{\downarrow}\right) = \boldsymbol{t}_1^{*\downarrow}\right) \\ = \;& \mathrm{P}\left(\widehat{T}_{n-1} = \hat{\boldsymbol{t}} \;\middle|\; \hat{\pi}_{[k]}^{*(n-1)}\left(\widehat{T}^{\downarrow}\right) = \boldsymbol{t}_1^{*\downarrow}\right) \end{aligned} \tag{2.38}$$

where $\boldsymbol{t}_1^{*\downarrow}$ denotes $\boldsymbol{t}_1^*$ with label n removed. By definition of the down-up chain the statement of the lemma easily follows from (2.38), since the up-step is simply defined by adding $\{n\}$ to one of the B_x's according to the same probabilities as in the decorated (α, γ)-growth process, Definition 2.6.5. Now, from the spatial Markov Property (Lemma 2.7.9) we need mainly to focus on the internal structures of $\boldsymbol{t}_0^*$ in $\widehat{T}$ that are affected by the down-step.

We now split up into cases based on $\boldsymbol{t}_0^*$, i and $\tilde{\imath}$:

Case A1: $i \in B_x$ **for external** $x \in \mathrm{edge}(\boldsymbol{s})$**,** $y_x > 1$**.** Note how in this case $I \in B_x$ implies that $\tilde{I} = \max\{I, a, b\} \in B_x$ as well, where a and b are defined as in the semi-planar (α, γ)-chain. This implies that the tree shape does not change during the down-step. Hence conditioning on $I = i$ and $\tilde{I} = \tilde{\imath}$ in $\boldsymbol{t}_0^*$, is exactly the same as conditioning on the corresponding random variables $I' = \textsc{rank}_{B_x}(i)$ and $\tilde{I}' = \textsc{rank}_{B_x}(\tilde{\imath})$ in $\textsc{int}(x)$. Now, recall that $\textsc{int}(x) \overset{D}{=} \widehat{T}_{\#B_x}$, and so Lemma 2.4.5 ensures that, under this conditioning $\textsc{int}(x)$ after the down move is distributed as $\widehat{T}_{\#B_x - 1}$. Since this is the only internal

structure affected in the down-step all the other internal structures will still have the same distribution, and they will all still be conditionally independent as in Lemma 2.7.9.

Case A2: $i \in B_x$ **for internal** $x \in \text{edge}(\boldsymbol{s})$**.** This is very similar to Case A1, except that we have $I > k$ in this case. As above, note that conditioning on $I = i$ and $\tilde{I} = \tilde{\imath}$, both being elements of B_x, is equivalent to conditioning on the corresponding random variables $I' = \text{RANK}_{B_x}(i)$ and $\tilde{I}' = \text{RANK}_{B_x}(\tilde{\imath})$ in $\text{INT}(x)$. Note that $1 < I' \leq \tilde{I}' \leq \#B_x$, and recall that $\text{INT}(x) \overset{D}{=} \widehat{T}^{\gamma}_{\#B_x}$. Combining these observations with Corollary 2.7.7 yields that $\text{INT}(x)$ after the down-step has the same distribution as $\widehat{T}^{\gamma}_{\#B_x - 1}$.

Case A3: $i \in B_x$ **for** $x \in \text{bp}(\boldsymbol{s})$**.** This is the most complicated scenario due to the complexity of the internal structure of a branch point. But similarly to Cases A1 and A2, we do not change the tree shape since $I > k$ ensures that $\tilde{I} \in B_x$ as well. Hence we only need to argue that this internal structure has the correct (conditional) distribution after the down step, as in the previous cases. Similar to these cases, conditioning on $I = i$ and $\tilde{I} = \tilde{\imath}$ in $\hat{\boldsymbol{t}}^*_0$ is equivalent to conditioning on the corresponding random variables $I' = c + \text{RANK}_{B_x}(i)$ and $\tilde{I}' = c + \text{RANK}_{B_x}(\tilde{\imath})$ in $\text{INT}(x)$, where c is the number of children of x in $\boldsymbol{s}$.

From Lemma 2.7.9(vi) we obtain that $\text{INT}(x)$ has the same distribution as the $\#B_x$'th step of a c-order branch point (α, γ)-growth process started from $\rho(\text{INT}(x), [c + \#B_x] \setminus [c])$. Hence Corollary 2.7.8 immediately yields the results.

Since no other ranked internal structures are altered in the above situation, and since we have now also conditioned on $\tilde{\imath}$, we can easily relabel all the other label sets $B_{x'}$ for $x' \neq x$ if applicable.

In all of the following cases, v refers to the parent branch point of i in $\boldsymbol{s}$.

Case B1: $i \in B_x$ **for external** $x \in \text{edge}(\boldsymbol{s}), y_x = 1, y_v > 0, \tilde{\imath} > k$. Firstly, we note that the restrictions on i and y_x imply that $i \leq k$. Now consider the semi-planar structure in the branch point v, and note how $y_x = 1$ and $y_v > 0$ imply that $\max\{i, a, b\} = \max\{i, b\}$, where a and b are defined as in the semi-planar (α, γ)-chain. Hence this case covers the events where (1) i is located in a multifurcating branch point of $\widehat{T}$ and the local search finds a label larger than k to the left (or right, if i is placed leftmost in v) of i, and (2) i is located in a bifurcating branch point where there is a decorated mass in the branch point. The two situations are completely analogous so we will only provide the argument

for the multifurcating case. Here, two internal structures are affected: $\textsc{int}(x)$ and $\textsc{int}(v)$. So condition on $I = i$ and $\tilde{I} = \tilde{\imath} \in B_v$. By Lemma 2.7.9, conditional on the subtree of v in $\widehat{T}$ containing $\tilde{\imath}$ having leaves labelled by $C_{\tilde{\imath}}$, we will first show that $\textsc{int}_k^{T^\downarrow}(x)$ has the distribution of $\widehat{T}_{\#C_{\tilde{\imath}}}$. Consider how $\textsc{int}(v)$ is changed. First, conditional on $I = i, \tilde{I} = \tilde{\imath}$ and $C_{\tilde{\imath}}$, we will insert the leaves labelled by the elements of $C_{\tilde{\imath}} \setminus \{\tilde{\imath}\}$ into the subtree of v with lowest label $\tilde{\imath}$, and no other leaves will go into this subtree, so that what we extract as $\textsc{int}(x)$ in $\widehat{T}^\downarrow$ is indeed distributed like $\widehat{T}_{\#C_{\tilde{\imath}}}$. Then, to form $\textsc{int}(v)$ in $\widehat{T}$, we would still insert the leaves labelled by the elements of $B_v \setminus C_{\tilde{\imath}}$ according to the same growth procedure that formed $\textsc{int}(v)$ in $\widehat{T}$ to form the equivalent in $\widehat{T}^\downarrow$; the leaves cannot (in this case) be inserted into either of the subtrees with smallest label i and $\tilde{\imath}$, respectively, nor in the gap between them. This proves that both the relevant internal structures have the correct conditional distribution after the down-step, given $I = i, \tilde{I}$ and $C_{\tilde{\imath}}$.

Case B2: $i \in B_x$ **for external** $x \in \text{edge}(\boldsymbol{s}), y_x = 1, y_v > 0, \tilde{\imath} \leq k$. Following the same reasoning as above, this case only contains the event where the local search in $\widehat{T}$ does not find a label larger than k in a multifurcating branch point to the left (or right) of i. Hence $\tilde{\imath} \in [k]$, and by swapping i and $\tilde{\imath}$ and subsequently deleting $\tilde{\imath}$, we end up removing one of the leaves that defines the tree shape $\boldsymbol{s}$. The leaves $\tilde{\imath}+1, \ldots, k$ will be relabelled by $\tilde{\imath}, \ldots, k-1$ using the increasing bijection, thus forming $\boldsymbol{s}' = \rho\left(\tau(\boldsymbol{s}, i, \tilde{\imath}), \tilde{\imath}\right)$, and k will reappear in the following way.

(i) If $k+1 \in B_{v'}$ for some $v' \in \text{bp}(\boldsymbol{s})$, the new tree shape will have k as a child of v', i.e. the new tree shape is $\varphi\left(\boldsymbol{s}', v', k\right)$.

(ii) If $k+1 \in B_{e'}$ for some $e' \in \text{edge}(\boldsymbol{s})$, k will split the edge up into a new branch point and three edges, i.e. the new tree shape is $\varphi\left(\boldsymbol{s}', e', k\right)$.

Consequently, we need to focus on three internal structures in $\widehat{T}$: $\textsc{int}(e_{\tilde{\imath}\downarrow})$, $\textsc{int}(v)$ (where v is the parent of i in $\widehat{T}$), and $\textsc{int}(x')$ where $k+1 \in B_{x'}$.

The first of these remains the same, but becomes the internal structure on the leaf edge i after the down-step.

From arguments similar to Proposition 2.7.6 we obtain that $\textsc{int}(v)$ in $\widehat{T}^\downarrow$ has the same distribution as the y_v'th step of a semi-planar $c-1$-order branch point (α, γ)-process.

Now observe how the relabelling naturally partitions $B_{x'}$ into either two blocks (if x' is a branch point) or four blocks (if x' is an edge), and which of the two situations we

are in is determined solely by $\boldsymbol{t}_0^*$. So if $x' = e_{v'w} \in \mathrm{edge}(\boldsymbol{s})$ condition on $B'_{e_{k\downarrow}}, B'_{w'}, B'_{e_{v'w'}}$, and $B'_{e_{w'w}}$ partitioning $B_{x'}$. On the other hand, if $x' \in \mathrm{bp}(\boldsymbol{s})$ condition on $B'_{e_{k\downarrow}}$ and $B'_{x'}$ partitioning $B_{x'}$. Considering how insertions of leaves were made into $\rho\left(T_n, [n] \setminus [k+1]\right)$ to obtain T_n, we deploy Lemma 2.7.9 which yields that the internal structures corresponding to the branch points and edges outlined above have the correct distributions.

Case B3: $i \in B_x$ **for external** $x \in \mathrm{edge}(\boldsymbol{s})$, $y_x = 1$, $y_v = 0$, $\tilde{\imath} > k$. This case is where we have a binary branch point, and there is decorated mass on the edge below, but is otherwise similar to Case B1.

Case B4: $i \in B_x$ **for external** $x \in \mathrm{edge}(\boldsymbol{s})$, $y_x = 1$, $y_v = 0$, $\tilde{\imath} \leq k$. This case is where we have a binary branch point, and there is no decorated mass on the edge below, but is otherwise similar to Case B2.

Since the above cases are exhaustive, and that we in all cases obtain the correct conditional distributions, this proves (2.38). ■

With Lemma 2.8.2 and Lemma 2.8.3 in place, we are now ready to prove the final theorem of this exposition, from which Corollary 2.1.5 also follows easily.

Theorem 2.8.4. *Fix $n \geq 2$. Let $\left(\widehat{T}_n(m)\right)_{m\in\mathbb{N}_0}$ denote a semi-planar (α,γ)-chain. For $1 \leq k \leq n$, let $T^{\bullet n}_{[k]}(m) := \hat{\pi}^{\bullet n}_{[k]}\left(\widehat{T}_n(m)\right)$, $m \in \mathbb{N}_0$, denote the projection of the chain onto decorated $[k]$-trees of mass n. If $\widehat{T}_n(0) \sim \hat{\Pi}^{\bullet n}_k(\boldsymbol{t}^\bullet, \cdot)$ for some $\boldsymbol{t}^\bullet \in \mathbb{T}^{\bullet n}_{[k]}$, then $\left(T^{\bullet n}_{[k]}(m)\right)_{m\in\mathbb{N}_0}$ is a decorated (α,γ)-chain on $\mathbb{T}^{\bullet n}_{[k]}$ started from $\boldsymbol{t}^\bullet$.*

Remark 2.8.5. Note that the case $k = n$ is essentially Theorem 2.5.4.

Proof. This follows from the diagram in Figure 2.9 commuting by Lemmas 2.8.2 and 2.8.3. Sampling $\widehat{T}_n(0)$ from $\hat{\Pi}^{\bullet n}_k(\boldsymbol{t}^\bullet, \cdot)$ corresponds to first sampling a collapsed tree $T^{*n}_k(0)$ from H, defined in 2.37, and subsequently sampling $\widehat{T}_n(0)$ from $\hat{\Pi}^{*n}_k\left(T^{*n}_k(0), \cdot\right)$ by definition of the kernels. To obtain $T^{\bullet n}_{[k]}(1)$, one can now perform one step of the collapsed (α,γ)-chain and then project the result to $\mathbb{T}^{\bullet n}_{[k]}$ using the function h. From Lemma 2.8.3, the intertwining criterion, Proposition 2.8.1(ii), is satisfied for semi-planar (α,γ)-chain, the function $\hat{\pi}^{*n}_{[k]}$ and the Markov kernel $\hat{\Pi}^{*n}_{[k]}$, and hence we obtain the collapsed (α,γ)-chain on $\mathbb{T}^{*n}_{[k]}$ with initial distribution given by $T^{*n}_k(0)$. Finally, the Kemeny-Snell criterion, Proposition 2.8.1(i), and Lemma 2.8.2 ensures that projecting the collapsed (α,γ)-chain

with initial state $T_k^{*n}(0)$ to $\mathbb{T}_{[k]}^{\bullet n}$ using the function h yields a Markov chain on $\mathbb{T}_{[k]}^{\bullet n}$ with transition kernel characterized in Definition 2.6.3 started from $\boldsymbol{t}^\bullet$. This proves the theorem. ■

Proof of Theorem 2.1.4. This now follows immediately from Theorem 2.8.4 and Theorem 2.5.4. ■

Chapter 3

Relaxation and mixing times for down-up chains on multifurcating trees

Having constructed a variety of different down-up chains on multifurcating tree spaces in Chapter 2 with explicit stationary distributions, we now turn our attention to the question of how quickly these Markov chains will approach their limiting distribution. Specifically, we set out to investigate the relaxation and mixing times of the (non-planar) down-up chain of Chapter 2 as well as related Markov chains. A natural element of the investigation turns out to be related to the question of reversibility of the Markov chains in question, and a discussion of this aspect will guide which Markov chains we will focus on when discussing the mixing and relaxation times.

3.1 Bounds for the relaxation and mixing times

Combining results from [7] and [48] shows that a down-up chain on binary unrooted trees with n leaves where we in the down-step simply delete a leaf uniformly at random, and in the up-step reinsert the deleted leaf into a uniformly chosen edge, has a relaxation time of precisely order n^2 and a mixing time that can be bounded from below by something of order n^2 and from above by something of order n^3.

In this chapter we will generalize these results to certain down-up chains on multifurcating trees. Postponing precise definitions to the next section, our main focus will be the *exchangeable stable chain* where we in the down-step delete a leaf uniformly at random, and in the up-step reinsert it according to the stable growth rule (or equivalently the $(\alpha, 1-\alpha)$-growth rule), as well as the *unrooted exchangeable stable chain* where we treat the root as a leaf and hence allow for it to be deleted and reinserted in a transition of the down-up chain.

In Sections 3.3 we will show that these two down-up chains fits the framework of methods used in [7] and [48], and we will subsequently deploy those methods in Sections 3.5 and 3.6 to achieve the following result:

Theorem 3.1.1. *The relaxation time of the unrooted exchangeable stable chain is exactly of order n^2.*

As we will discuss later, it seems that the presence of a root that cannot be moved in the (unrooted) exchangeable stable chain requires different arguments to obtain a similar upper bound. Importantly, the argument used by Schweinsberg has natural analogues but it only appears to provide a weaker upper bound.

Theorem 3.1.2. *The relaxation time of the exchangeable stable chain is at least of order n^2 and at most of order $n^{3-\alpha}$.*

From well-established bounds on the mixing time using the relaxation time and the stationary distribution of the Markov chain, we obtain the following bounds on the mixing times:

Corollary 3.1.3. *The mixing time of the unrooted exchangeable stable chain is at least of order n^2 and at most of order $n^3 \log(n)$, whilst the mixing time of the exchangeable stable chain and the unlabelled stable chain are at least of order n^2 and at most of order $n^{4-\alpha} \log(n)$.*

We believe the gap between the lower and upper bounds for the relaxation time of the exchangeable stable chain and the unlabelled stable chain above to be artefacts of the method of proof rather than anything else. And from results on scaling limits in e.g. [36] we also believe this to be the case for the mixing time. In fact, Aldous already showed

that in the case of the unrooted exchangeable stable chain with $\alpha = \frac{1}{2}$, the mixing time is at most of order n^3, giving a tighter upper bound than we provide in that special case. This leads us to conjecture the following:

Conjecture 3.1.4. *The relaxation and mixing times of the exchangeable stable chain and the unlabelled stable chain as well as the mixing time of the unrooted exchangeable stable chain are all exactly of order n^2.*

The chapter is structured as follows:

We will start with a preliminary section by giving formal definitions of the down-up chains and state spaces we will study and will give a brief and very targeted overview of definitions and results regarding relaxation and mixing times of Markov chains on finite state spaces.

In Section 3.3 we will cover reversibility results related to various chains that we are studying, which ultimately is the deciding factor in focusing on the unrooted and rooted exchangeable stable chains.

Section 3.4 is a collection of miscellaneous results for the exchangeable stable chain, including an explicit characterization of the transition probabilities, which we will compare with the transition probabilities of the binary Aldous chain of [7].

Finally, Section 3.5 proves the lower bounds of Theorems 3.1.1 and 3.1.2, whilst Section 3.6 proves the upper bounds followed by a discussion on the link between the relaxation and mixing times of the chains in question.

3.2 Preliminaries

The foundation for the results of this chapter naturally splits into two parts. In order to be precise about the terminology and which Markov chains we are studying, we will begin a short recap of the core definitions of Chapter 2 which will be relevant to our current endeavour. Subsequently, we will provide a brief — and indeed highly selective — overview of the definitions and theory for mixing and relaxation times of general Markov chains on finite state spaces.

3.2.1 State spaces and Markov chains

Throughout this exposition we will for any $n \geq 2$ work with two kinds of trees:

- Rooted, leaf-labelled trees with n leaves labelled by $[n] := \{1, \dots, n\}$, the space of which we denote by $\mathbb{T}_{[n]}$. We denote an element of $\mathbb{T}_{[n]}$ by $\boldsymbol{t}$.

- Rooted, unlabelled trees with n leaves, the space of which we denote by $\mathbb{T}_n^\circ$. We similarly denote its elements by $\boldsymbol{t}^\circ$.

It is clear that there is a surjective projection map from $\mathbb{T}_{[n]}$ to $\mathbb{T}_n^\circ$, which we shall denote by π_n°.

Now fix $0 \leq \gamma \leq \alpha \leq 1$. We recall that the (α, γ)-growth process, see Definition 2.2.10, is a Markov chain on $\cup_{n\in\mathbb{N}}\mathbb{T}_{[n]}$ where, conditional on the nth step of the chain taking the value $\boldsymbol{t} \in \mathbb{T}_{[n]}$, we attach a leaf labelled by $n+1$ to an insertable part (recall that this is the collection of branch points and edges of $\boldsymbol{t}$) with the following probabilities: each leaf edge has probability $\frac{1-\alpha}{n-\alpha}$, each internal edge has probability $\frac{\gamma}{n-\alpha}$ and each branch point of degree k have probability $\frac{(k-2)\alpha-\gamma}{n-\alpha}$.

It turns out that the complicated label dynamics of the (α, γ)-chain, see Definition 2.1.1, makes it rather difficult to analyze both the relaxation and mixing time of the down-up chain, as the chain is not reversible, see Appendix A. Instead, we study two generalizations of the Aldous chain: the *exchangeable stable chain* where we rely on the exchangeability of the leaf labels in the $(\alpha, 1-\alpha)$-growth process, and the *unrooted exchangeable stable chain* where we also allow for the root to be relocated in a transition.

Definition 3.2.1 (Exchangeable stable chain)**.** The *exchangeable stable chain on* $\mathbb{T}_{[n]}$ is a down-up chain on $\mathbb{T}_{[n]}$ where the two steps of the transition kernel can be characterized as follows:

Down: Delete a leaf chosen uniformly at random.

Up: Insert the leaf according to the stable growth rule (i.e. the $(\alpha, 1-\alpha)$-growth rule).

We will always denote such a chain by $(T_n^{\mathrm{ex}}(m))_{m\in\mathbb{N}_0}$.

Remark 3.2.2. In the sense of Definition 2.2.11 this is the $(\mathrm{id}, (\alpha, 1-\alpha))$-down-up chain without relabelling.

For convenience we will, in order to define the unrooted exchangeable stable chain, for each $n \in \mathbb{N}$ denote the root of a tree in $\mathbb{T}_{[n]}$ by 0.

Definition 3.2.3 (Unrooted exchangeable stable chain)**.** The *unrooted exchangeable stable chain on* $\mathbb{T}_{[n]}$ is a down-up chain on $\mathbb{T}_{[n]}$, where the two steps of the transition kernel can be characterized as follows:

Down: Delete a leaf or the root, i.e. one of the leaves labeled by $\{0, \dots, n\}$, chosen uniformly at random.

Up: Insert the leaf (or in the case where we delete the root, insert the root as if it was a leaf) according to the stable growth rule (i.e. the $(\alpha, 1-\alpha)$-growth rule).

We will always denote such a chain by $(T_n^{\mathrm{un}}(m))_{m \in \mathbb{N}_0}$.

Remark 3.2.4. Note that Definition 3.2.3 is well-defined since we do not distinguish between internal and leaf edges during the up-step in the stable growth process.

Lastly, we can remove the labels altogether and define the unlabelled projection of the chain $(\pi_n^\circ(T_n(\cdot)))$ which is also a Markov chain (this follows by a Kemeny-Snell type argument — see Proposition 2.8.1(i) and Theorem 2.8.4). We can also directly describe the transition mechanism of the unlabelled chain, as it is straightforward to see that the swapping of labels during a down-move in the (α, γ)-chain does not alter the shape of the resulting tree of the transition.

Definition 3.2.5 (Unlabelled (α, γ)-chain)**.** The *unlabelled* (α, γ)*-chain on* $\mathbb{T}_{[n]}^\circ$ is a down-up chain on $\mathbb{T}_{[n]}^\circ$, where the two steps of the transition kernel can be characterized as follows:

Down: Delete a leaf chosen uniformly at random.

Up: Insert a leaf according to the (α, γ)-growth rule.

We will always denote such a chain by $(T_n^\circ(m))_{m \in \mathbb{N}_0}$.

Remark 3.2.6. The unlabelled (α, γ)-chain defined above could equivalently have been defined in terms of a function f in line with Definition 2.2.11.

3.2.2 Mixing times

Let us now briefly highlight key notions regarding mixing times of Markov chains. Since we will apply the results of this section to different Markov chains we will use the general notation that, for each $n \in \mathbb{N}$, $(X_n(m))_{m\in\mathbb{N}}$ is a Markov chain on a finite state space $\mathcal{X}_n$ and denote its stationary distribution by μ_n (if it exists and is unique). From [35], we define the *maximal m-step total variation distance* to the stationary distribution of the chain

$$d_n(m) = \max_{x\in\mathcal{X}_n} \max_{A\subseteq\mathcal{X}_n} |\mathrm{P}\left(X_n(m) \in A \mid X_n(0) = x\right) - \mu_n(A)| \tag{3.1}$$

and the maximal m-step total variation distance between distributions of the same chain started from different initial distributions

$$\bar{d}_n(m) = \max_{x,x'} \max_{A} |\mathrm{P}(X_n(m) \in A \,|\, X_n(0) = x) - \mathrm{P}\left(X_n(m) \in A \,|\, X_n(0) = x'\right)|.$$

Lemmas 4.11 and 4.12 of [35] yield that $d_n(m_1 + m_2) \le d_n(m_1)\bar{d}_n(m_2) \le 2d_n(m_1)d_n(m_2)$ for any $m_1, m_2 \in \mathbb{N}$ and so

$$d_n(jm) \le 2d_n(m)d_n\left((j-1)m\right) \le \cdots \le 2^j d_n(m)^j \qquad \text{for any} \qquad j, m \in \mathbb{N}. \tag{3.2}$$

Defining the *mixing time* by

$$\tau_n^{\mathrm{mix}} := \tau_n^{\mathrm{mix}}\left(\frac{1}{4}\right) \qquad \text{where} \qquad \tau_n^{\mathrm{mix}}(\varepsilon) := \inf\left\{m \in \mathbb{N} \mid d_n(m) \le \varepsilon\right\},$$

(3.2) yields that $d_n(j\tau_n^{\mathrm{mix}}) \le 2^{-j}$ for any $j \in \mathbb{N}$.

Now, for a *reversible* Markov chain, let $\lambda_{n,2}$ be the second largest eigenvalue of the transition matrix of the down-up chain, and define the *relaxation time* $\tau_n^{\mathrm{rel}} := (1-\lambda_{n,2})^{-1}$. If we let X_n be a random variable with distribution equal to the stationary distribution of a reversible Markov chain on a finite state space $\mathcal{X}_n$ with transition probabilities $p_n(x_1, x_2)$, we get from [13] that

$$\tau_n^{\mathrm{rel}} = \sup_{f_n\colon \mathcal{X}_n\to\mathbb{R}} \frac{2\,\mathrm{Var} f_n(X_n)}{\sum_{x_1}\sum_{x_2} \mu_n(\{x_1\})p_n(x_1,x_2)\left(f_n(x_2) - f_n(x_1)\right)^2}, \tag{3.3}$$

which can be used effectively to obtain a lower bound for the relaxation time simply by constructing suitable functions $f_n\colon \mathcal{X}_n \to \mathbb{R}$. Specifically, one way of showing that τ_n^{rel} is asymptotically lower bounded by $g(n)$ is therefore to pick a suitable sequence of functions $(f_n)_{n\in\mathbb{N}}$ such that the denominator of (3.3) in all cases are bounded above by a constant, and such that $f_n(X_n)/\sqrt{g(n)} \xrightarrow{D} Y$ for some random variable Y.

In order to find an asymptotic upper bound for the relaxation time we will use the *distinguished paths method* that was successfully applied in [48] in the binary and exchangeable case. To this end, we will focus on the finite directed graph $\mathcal{G}_n = (\mathcal{X}_n, \mathcal{E}_n)$ where the edge set $\mathcal{E}_n$ is defined by letting $e_{x_1x_2} \in \mathcal{E}_n$ if and only if $p_n(x_1, x_2) > 0$.

Theorem 3.2.7 (Corollary 3 of [48])**.** *Let $\mathcal{X}_n' \subset \mathcal{X}_n$ and let $\Gamma_{x,y}$ be a random path in $\mathcal{G}_n$ between $x, y \in \mathcal{X}_n$ such that no edge appears in the path more than once. If it for all $x \in \mathcal{X}_n$ and $y \in \mathcal{X}_n'$ holds that $\Gamma_{x,y}$ is shorter than some constant $L_n \in \mathbb{N}$, then*

$$\tau_n^{\mathrm{rel}} \leq \frac{4L_n}{\mu_n(\mathcal{X}_n')} \max_{e\in\mathcal{E}_n} \frac{1}{Q_n(e)} \sum_{x\in\mathcal{X}_n} \sum_{y\in\mathcal{X}_n'} \mu_n(\{x\})\mu_n(\{y\})\mathrm{P}\left(e \in \Gamma_{x,y}\right), \tag{3.4}$$

where $Q_n(e) = \mu_n(\{x\})p_n(x, y)$ for $e = e_{xy} \in \mathcal{E}_n$.

The results above give us efficient ways of deducing the asymptotic behaviour of the relaxation time, and combining Theorem 12.3 and 12.4 of [35] makes it possible to deduce (less precise) bounds on the mixing time since

$$\left(\tau_n^{\mathrm{rel}} - 1\right) \log\left(\frac{1}{2\varepsilon}\right) \leq \tau_n^{\mathrm{mix}}(\varepsilon) \leq \log\left(\frac{1}{\varepsilon \min_{x\in\mathcal{X}_n} \mu_n(\{x\})}\right) \tau_n^{\mathrm{rel}} \tag{3.5}$$

for any *reversible, irreducible* and (for the lower bound) *aperiodic* Markov chain with stationary distribution μ_n.

The last result of this section links bounds on the relaxation time of a Markov chain to bounds on the relaxation time of a projection of the Markov chain. The result is natural and not difficult, but as we have been unable to locate it in the literature, we include a short proof here.

Proposition 3.2.8. *Let $(X_n(m))_{m\in\mathbb{N}_0}$ be a reversible Markov chain on a finite state space, $\mathcal{X}_n$, with transition kernel P_n. Let π_n be a surjective map $\pi_n\colon \mathcal{X}_n \to \mathcal{Y}_n$ for each $n \in \mathbb{N}$. If*

$(X_n(m))_{m\in\mathbb{N}_0}$ and π_n satisfy the Kemeny-Snell criterion, i.e. that for all $x_1, x_2 \in \mathcal{X}_n$ with $\pi_n(x_1) = \pi_n(x_2)$ it holds that

$$P_n\left(x_1, \pi_n^{-1}(\{y\})\right) = P_n\left(x_2, \pi_n^{-1}(\{y\})\right) \qquad \text{for any } y \in \mathcal{Y}_n,$$

then $(Y_n(m))_{m\in\mathbb{N}} := (\pi_n(X_n(m)))_{m\in\mathbb{N}_0}$ is a Markov chain. Let τ_n^{rel} and $\tau_n^{\pi,\mathrm{rel}}$ denote the relaxation times for $(X_n(m))_{m\in\mathbb{N}_0}$ and $(Y_n(m))_{m\in\mathbb{N}}$, respectively. Then $\tau_n^{\pi,\mathrm{rel}} \leq \tau_n^{\mathrm{rel}}$ for all $n \in \mathbb{N}$.

Proof. Fix $n \in \mathbb{N}$ and note that $\tau_n^{\pi,\mathrm{rel}}$ is well-defined as $(\pi_n(X_n(m)))_{m\in\mathbb{N}_0}$ is a Markov chain by assumption, and that reversibility follows from the fact that it is a projection of a reversible Markov chain. In the following we will abuse notation so that P_n also denotes the transition matrix of $(X_n(m))_{m\in\mathbb{N}_0}$ by P_n whilst the transition matrix of $(Y_n(m))_{m\in\mathbb{N}_0}$ is denoted by P_n^π. Let $\lambda \in \mathbb{R}$ be an eigenvalue of P_n^π. The corresponding eigenfunction $f_n^\pi \colon \mathcal{Y}_n \to \mathbb{R}$ satisfies that $P_n^\pi f_n^\pi = \lambda f_n^\pi$, i.e. where

$$\mathbb{E}_y\left(f_n^\pi(Y_n(1))\right) = \lambda f_n^\pi(y) \qquad \text{for all } y \in \mathcal{Y}_n$$

Now consider the function $f_n \colon \mathcal{X} \to \mathbb{R}$ given by $f_n(x) = f_n^\pi(\pi_n(x))$ for $x \in \mathcal{X}_n$. By assumption we have $\pi_n(X_n(0)) \overset{D}{=} Y_n(0)$ and so the Kemeny-Snell criterion yields

$$\mathbb{E}_x f_n(X_n(1)) = \mathbb{E}_x f_n^\pi(\pi_n(X_n(1))) = \mathbb{E}_{\pi_n(x)} f_n^\pi(Y_n(1)) = \lambda f_n^\pi(\pi(x)) = \lambda f_n(x)$$

for all $x \in \mathcal{X}$, and hence all eigenvalues of P_n^π are also eigenvalues of P_n. In particular the spectral gap of P_n^π is at least equal to the spectral gap of P_n, implying that

$$\tau_n^{\pi,\mathrm{rel}} = \frac{1}{1-\lambda_{n,2}^\pi} \leq \frac{1}{1-\lambda_{n,2}} = \tau_n^{\mathrm{rel}},$$

where $\lambda_{n,2}^\pi$ and $\lambda_{n,2}$ are the second largest eigenvalues of P_n^π and P_n, respectively. ∎

3.3 On the reversibility of down-up chains

The characterization in (3.3) and the bounds in (3.4) and (3.5) all depend on the Markov chain in question being reversible, and we are not aware of any obvious way to adapt the

above bounds to a non-reversible Markov chain.

Unfortunately, despite the exchangeability of the labels of the $(\alpha, 1-\alpha)$-growth process as discussed briefly following Definition 2.2.11, the (α, γ)-chain is non-reversible for any choice of parameters $0 \leq \gamma \leq \alpha \leq 1$. A proof of this can be found in Appendix A. In light of Proposition 3.2.8 we find it important to highlight this fact, since if we were able to analyze the (α, γ)-chain of Chapter 2, we would immediately have a host of bounds for the relaxation time of various interesting Markov chains such as the unlabelled version of the (α, γ)-chain on $\mathbb{T}^{\circ}_{[n]}$ and the decorated (α, γ)-chain on $\mathbb{T}^{\bullet n}_{[k]}$.

The label dynamics of both the unrooted and rooted exchangeable stable chains, as well as the unlabelled (α, γ)-chain, are much simpler, and the remainder of this section is dedicated to showing that they are all reversible Markov chains. We will start by focusing on the unlabelled (α, γ)-chain, see Definition 3.2.5. One is easily convinced that the stationary distribution of this chain is that of T°_n, i.e. the nth step of the unlabelled (α, γ)-growth process. Now recall the following key facts regarding this stationary distribution:

Proposition 3.3.1 (Sampling Consistency)**.** *Fix $n \geq 2$. Let $(T^{\circ}_m)_{m\in[n]}$ denote the first n steps of the unlabelled (α, γ)-growth process and let for each $T^{\circ\downarrow}_n$ denote T°_n after the down-step outlined in Definition 3.2.5. The following holds:*

(i) T°_n is sampling consistent for any choice of parameters $0 \leq \gamma \leq \alpha \leq 1$:

$$T^{\circ\downarrow}_n \overset{D}{=} T^{\circ}_{n-1}.$$

(ii) T°_n is strongly sampling consistent, i.e.

$$\left(T^{\circ}_n, T^{\circ\downarrow}_n\right) \overset{D}{=} \left(T^{\circ}_n, T^{\circ}_{n-1}\right)$$

if and only if $\gamma = 1 - \alpha$.

Proof. See Proposition 12 and 13 of [12]. ∎

It turns out, that strong sampling consistency of the stationary distribution implies reversibility of the down-up chain if started out in stationarity:

Theorem 3.3.2. *Let $(T_n^\circ(m))_{m\in\mathbb{N}_0}$ be the rooted unlabelled (α,γ)-down-up chain on $\mathbb{T}_n^\circ$, running in stationary with $T_n^\circ(0) \stackrel{D}{=} T_n^\circ$, where T_n° is the nth step of the unlabelled (α,γ)-growth model. $(T_n^\circ(m))_{m\in\mathbb{N}_0}$ is reversible if $\gamma = 1-\alpha$.*

Proof. Proposition 3.3.1 yields that the unlabelled (α,γ)-chain is strongly sampling consistent if and only if $\gamma = 1-\alpha$. Strong sampling consistency immediately implies that

$$\left(T_n^\circ(0), T_n^\circ(0)^\downarrow\right) \stackrel{D}{=} \left(T_n^\circ, T_{n-1}^\circ\right).$$

Additionally, from the construction of the down-up chain we get

$$\left(T_n^\circ, T_{n-1}^\circ\right) \stackrel{D}{=} \left(T_n^\circ(1), T_n^\circ(0)^\downarrow\right)$$

and hence $\left(T_n^\circ(0), T_n^\circ(0)^\downarrow\right) \stackrel{D}{=} \left(T_n^\circ(1), T_n^\circ(0)^\downarrow\right)$, and the Markov property after the down-step now ensures that

$$\left(T_n^\circ(0), T_n^\circ(0)^\downarrow, T_n^\circ(1)\right) \stackrel{D}{=} \left(T_n^\circ(1), T_n^\circ(0)^\downarrow, T_n^\circ(0)\right)$$

from which reversibility of the chain easily follows. ∎

The distributional equalities above give rise to the following question:

Question 3.3.3. Similar to above, let $(X_n(m))_{m\in\mathbb{N}_0}$ be a down-up chain taking values on a tree space such as $\widehat{\mathbb{T}}_{[n]}$, $\mathbb{T}_{[n]}$, $\mathbb{T}_{[k]}^{\bullet n}$ or $\mathbb{T}_n^\circ$, running in stationary with $X_n(0) \stackrel{D}{=} X_n$, where X_n is the nth step of the growth model governed by the up-step. If X_n is sampling consistent but **not** strongly sampling consistent, is the down-up chain then non-reversible?

As alluded to earlier in this section, we are actually able to circumvent the problem of non-reversibility in a different manner. In the case where $\gamma = 1-\alpha$ we can define a different version of the (α,γ)-chain that heavily relies on the exchangeability of the leaf labels.

It turns out that the exchangeability of the leaf labels, which in turn makes the swapping of labels in the down-step of the (α,γ)-chain superfluous, is enough to ensure reversibility of the chain:

Proposition 3.3.4. *The exchangeable stable chain is reversible.*

Proof. Note that it is enough to show that $(T_n^{\text{ex}}(0), T_n^{\text{ex}}(1)) \overset{D}{=} (T_n^e(1), T_n^e(0))$. Additionally, it is easy to check that the stationary distribution of the exchangeable stable chain is the nth step of the stable growth process, T_n. Letting G^e denote the growth rule governing the stable growth process, this directly implies that

$$\begin{aligned} & \mathrm{P}\left(T_n^{\text{ex}}(0) = \boldsymbol{r}, T_n^{\text{ex}}(0)^{\downarrow} = \boldsymbol{s}, T_n^e(1) = \boldsymbol{t}\right) = \frac{1}{n}\mathrm{P}\left(T_n = \boldsymbol{r}\right) G^e\left(\boldsymbol{s}, \{\boldsymbol{t}\}\right) \\ = \; & \frac{1}{n}\mathrm{P}\left(T_{n-1} = \boldsymbol{s}\right) G^e\left(\boldsymbol{s}, \{\boldsymbol{r}\}\right) G^e\left(\boldsymbol{s}, \{\boldsymbol{t}\}\right) \\ = \; & \frac{1}{n}\mathrm{P}\left(T_n = \boldsymbol{t}\right) G^e\left(\boldsymbol{s}, \{\boldsymbol{r}\}\right) = \mathrm{P}\left(T_n^{\text{ex}}(0) = \boldsymbol{t}, T_n^{\text{ex}}(0)^{\downarrow} = \boldsymbol{s}, T_n^e(1) = \boldsymbol{r}\right) \end{aligned}$$

for any combination of $\boldsymbol{r} \in \mathbb{T}_{[n]}, \boldsymbol{s} \in \mathbb{T}_{[n]\setminus\{i\}}$ for some $i \in [n]$, and $\boldsymbol{t} \in \mathbb{T}_{[n]}$, implying the distributional equality we wanted to obtain. ■

Comparing the (rooted) exchangeable stable chain to the unrooted version, the option to relocate the root in the unrooted exchangeable stable chain does not alter the argument above, except that there will be a $\frac{1}{n+1}$ appearing instead of $\frac{1}{n}$ due to the increase in choices of leaves to be deleted in the down-step. Hence we easily obtain the following corollary:

Corollary 3.3.5. *The unrooted exchangeable stable chain is reversible.*

3.4 Properties of the exchangeable stable chain

Having determined that the exchangeable stable chain is reversible, we now go on to discuss various aspects of the chain. Specifically, we will characterize transition probabilities of the chain and its stationary distribution. Both of these are useful to gain intuition about the behaviour of the chain, and in addition they are crucial in obtaining the bounds on the relaxation and mixing times of the chain.

Proposition 3.4.1. *The unique stationary distribution for the exchangeable stable chain is $\mu \sim T_n$, where T_n is the nth step of the $(\alpha, 1-\alpha)$-growth process, characterized by the following*

$$\mu(\{\boldsymbol{t}\}) = \mathrm{P}\left(T_n = \boldsymbol{t}\right) = \frac{(1-\alpha)^{l_{edge}^{\boldsymbol{t}}} \times \prod_{j=1}^{d_{max}^{\boldsymbol{t}}} \left((j+1)\alpha - 1\right)^{\sum_{j' \geq j} l_{j'}^{\boldsymbol{t}}}}{\prod_{j=1}^{n-1}(j-\alpha)} \tag{3.6}$$

for any $\boldsymbol{t} \in \mathbb{T}_n$, *where* $d^{\boldsymbol{t}}_{max} = \max_{v \in \mathrm{bp}(\boldsymbol{t})} \deg(v)$, *and* $d^{\boldsymbol{t}}_j = \#\{v \in \mathrm{bp}(\boldsymbol{t}) \mid \deg(v) = j\}$ *for each* $3 \leq j \leq d^{\boldsymbol{t}}_{max}$, *and* $l^{\boldsymbol{t}}_{edge} = n - 1 - \sum_{j=4}^{d^{\boldsymbol{t}}_{max}} (j-3) l^{\boldsymbol{t}}_j$. *Furthermore*

$$\min_{\boldsymbol{t} \in \mathbb{T}_n} \mu(\{\boldsymbol{t}\}) = \begin{cases} \frac{(1-\alpha)^{n-1}}{\prod_{j=1}^{n-1}(j-\alpha)} & \text{for } \frac{2}{3} \leq \alpha \leq 1, \\ \frac{(1-\alpha)^{1+\lceil (n-2)/2 \rceil}(2\alpha-1)^{\lfloor (n-2)/2 \rfloor}}{\prod_{j=1}^{n-1}(j-\alpha)} & \text{for } \frac{1}{2} \leq \alpha \leq \frac{2}{3}. \end{cases} \tag{3.7}$$

Proof. That the stationary distribution of the exchangeable stable down-up chain is that of the nth step of the stable growth process follows directly from the exchangeability of the leaf labels of the stationary distribution and the fact that the up-step of the down-up chain is governed by the stable growth process. To see that the stable probability mass function is characterized by (3.6) consider how the leaves have been inserted in the growth process, and note how the leaf labels of $\boldsymbol{t}$ uniquely identify if a leaf has been inserted into an edge or a branch point and, if the latter is the case, how many leaves with lower labels have been inserted into that branch point previously. A branch point of degree $j+3$ will have an associated weight of $(j+1)\alpha - 1$, for each $j \in \left[d^{\boldsymbol{t}}_{\max}\right]$ and $l^{\boldsymbol{t}}_{\mathrm{edge}}$ will be the number of leaves that have been inserted into an edge to form $\boldsymbol{t}$.

To see that the minimum is as described in (3.7) note that for $\frac{1}{2} \leq \alpha \leq \frac{2}{3}$ we have

$$j\alpha - 1 \geq 1 - \alpha \geq 2\alpha - 1 \qquad \text{for all } j \geq 3$$

whilst for $\frac{2}{3} \leq \alpha \leq 1$ we have $1 - \alpha \leq j\alpha - 1$ for all $j \geq 2$. Considering the insertion of the leaves in the stable growth rule, we thus construct a tree with the minimal probability by only doing edge insertions, in the case where $\frac{2}{3} \leq \alpha \leq 1$ and alternating between inserting into a binary branch point and into an edge to create a new binary branch point, in the case where $\frac{1}{2} \leq \alpha \leq \frac{2}{3}$. ■

The characterization in (3.7) is important because it appears in (3.5). To highlight this, let us do a brief interlude and focus on the binary down-up chain of Aldous from [7], i.e. the exchangeable stable chain on $\mathbb{T}^{\mathrm{bin}}_{[n]}$ with $\alpha = \frac{1}{2}$. In that paper, Aldous showed that there exist $c_1 > 0$ (not depending on n) and $n_1 \in \mathbb{N}$ such that $\tau^{\mathrm{rel}}_n \geq c_1 n^2$ for all $n \geq n_1$, and in [48] Schweinsberg showed that $\tau^{\mathrm{rel}}_n = O(n^2)$, i.e. that there exist $c_2 > 0$ (not depending on n) and $n_2 \in \mathbb{N}$ such that $\tau^{\mathrm{rel}}_n \leq c_2 n^2$ for all $n \geq n_2$. This clearly shows

that τ_n^{rel} is exactly of order n^2, but now consider the bounds in (3.5). The stationary distribution for the Aldous down-up chain, μ, is the uniform distribution on $\mathbb{T}_{[n]}^{\text{bin}}$ (and an explicit expression can be found by setting $\alpha = \frac{1}{2}$ in Proposition 3.4.1), and hence

$$1/\mu(\{\boldsymbol{t}\}) = (2n-3)(2n-5)\cdots 3 \approx 2^n n^{n-1} e^{-n+1}$$

where $\approx$ means that the ratio of the two sides are bounded away from 0 and infinity as n varies, by Stirling's formula. Hence there exist $c_3 > 0$ such that

$$\begin{aligned}(c_1 n^2 - 1)\log(2) \leq \tau_n^{\text{mix}} &\leq \left(\log(4) + \log\left(c_3 2^n n^{n-1} e^{-n}\right)\right) c_2 n^2 \\ &= \left(\log(4) + \log(c_3) + ((\log(2) - 1 + \log(n))n - \log(n))\right) c_2 n^2\end{aligned} \tag{3.8}$$

This seems to suggest that even in the binary case, the best mixing time bounds this method can achieve is that

$$c_1' n^2 \leq \tau_n^{\text{bin}} \leq c_2' \log(n) n^3 \tag{3.9}$$

for some constants $c_1', c_2' > 0$, when letting $n \to \infty$ in the sense where we for each $n \in \mathbb{N}$ have a Markov chain on $\mathbb{T}_{[n]}^{\text{bin}}$.

We will now give an explicit characterization of the transition probabilities of the exchangeable stable chain. This is something that is not feasible for the (α, γ)-chain for any other values of γ due to the label swapping and relabelling, and is therefore of interest in its own right. Additionally, it also enters the upper bound for the relaxation time given in Theorem 3.2.7.

The transition probabilities are interesting as there are sometimes multiple pathways for the transition between $\boldsymbol{s} \in \mathbb{T}_{[n]}$ and $\boldsymbol{t} \in \mathbb{T}_{[n]}$ that depend on the symmetries of either $\boldsymbol{s}$ or $\boldsymbol{t}$ or both. In particular, one needs to be particularly careful when a multifurcating branch point forms, or when two neighbouring leaves swap places in a transition. To make this precise, we will for a tree $\boldsymbol{t} \in \mathbb{T}_{[n]}$, say that all leaves that are only separated by binary branch points belong to the same *binary component* of $\boldsymbol{t}$. We need to capture two kinds of symmetries that require special attention. Firstly, we say that three leaves $i, j, k \in [n]$ are *at the top of a binary component* if there exist two binary branch points $u, v \in \text{bp}(\boldsymbol{t})$

such that v is the parent of u and, one of the leaves i, j or k, whilst u is the parent of the remaining two leaves (*the top pair*), and will encode this in a function:

$$\mathrm{top}(\boldsymbol{s}, i, j, k) = \begin{cases} 1 & \text{if leaves } i, j \text{ and } k \text{ are at the top} \\ & \text{of a binary component of } \boldsymbol{s}, \\ 0 & \text{otherwise.} \end{cases}$$

We will use the same notation to encode that just two leaves are the top of a binary component if they are the entire component, i.e.

$$\mathrm{top}(\boldsymbol{s}, i, j) = \begin{cases} 1 & \text{if leaves } i \text{ and } j \text{ are the only two leaves} \\ & \text{in the top of a binary component of } \boldsymbol{s}, \\ 0 & \text{otherwise,} \end{cases}$$

and we note that this only happens if $\boldsymbol{s} \in \mathbb{T}_{[2]}$ or if there exists branch points $v, w \in \boldsymbol{s}$ such that v is multifurcating and the parent of w, and such that the set of child vertices of w in $\boldsymbol{s}$, denoted by $\mathrm{ch}_{\boldsymbol{s}}(w)$, is exactly $\{i, j\}$.

Secondly, we will say that two leaves belonging to the same binary component are *adjacent* if there are exactly two binary branch points separating them and they do not both belong to the top of a binary component, and encode this into an *adjacency function* by saying that

$$\mathrm{adj}(\boldsymbol{s}, i, j) = \begin{cases} 1 & \text{if leaves } i \text{ and } j \text{ are adjacent} \\ & \text{in a binary component of } \boldsymbol{s}, \\ 0 & \text{otherwise.} \end{cases}$$

Although it might appear counter-intuitive, it is entirely on purpose, that we in the situation sketched above, say that the pairs $\{i, j\}$, $\{i, k\}$, and $\{j, k\}$, respectively, are at the top of a binary component but not adjacent. The reason for this is, that we wish the adjacency function to capture a specific kind of symmetry that is not easily nor quickly described using the parent-child relationship.

Proposition 3.4.2. *The exchangeable stable chain on $\mathbb{T}_{[n]}$ has the following transition probabilities with normalizing constant $n(n-1-\alpha)$:*

$$p^{\mathrm{ex}}(\boldsymbol{s},\boldsymbol{t}) \propto \begin{cases} 3(1-\alpha) & \text{if } \exists!\{i,j,k\}\subseteq[n]\colon \begin{array}{l}\mathrm{top}(\boldsymbol{t},i,j,k)=1 \text{ and either } \mathrm{top}(\boldsymbol{s},i,j,k)=1\\ \text{or } \ \exists! v\in \mathrm{bp}(\boldsymbol{s})\colon \ \mathrm{ch}_{\boldsymbol{s}}(v)=\{i,j,k\}\end{array}\\ 3(2\alpha-1) & \text{if } \exists!\{i,j,k\}\subseteq[n]\exists! v\in \mathrm{bp}(\boldsymbol{t})\colon \mathrm{top}(\boldsymbol{s},i,j,k)=1, \mathrm{ch}_{\boldsymbol{t}}(v)=\{i,j,k\}\\ 2\,(c_v\alpha-1) & \text{if } \exists!\{i,j\}\subseteq[n]\exists! v\in \mathrm{bp}(\boldsymbol{t})\colon \ \mathrm{top}(\boldsymbol{s},i,j)=1, \mathrm{ch}_{\boldsymbol{t}}(v)\supsetneq\{i,j\}\\ 2(1-\alpha) & \text{if } \exists!\{i,j\}\subseteq[n]\exists! v\in \mathrm{bp}(\boldsymbol{s})\colon \ \mathrm{ch}_{\boldsymbol{s}}(v)\supsetneq\{i,j\}, \mathrm{top}(\boldsymbol{t},i,j)=1\\ 2(1-\alpha) & \text{if } \exists!\{i,j\}\subseteq[n]\colon \ \mathrm{adj}(\boldsymbol{s},i,j)=\mathrm{adj}(\boldsymbol{t},i,j)=1\\ 2(2\alpha-1) & \text{if } \exists!\{i,j\}\subseteq[n]\exists!\{v,w\}\in \mathrm{bp}(\boldsymbol{t})\colon \mathrm{adj}(\boldsymbol{s},i,j)=1, \mathrm{ch}_{\boldsymbol{t}}(v)=\{i,j,w\}\\ 2(1-\alpha) & \text{if } \exists!\{i,j\}\subseteq[n]\exists!\{v,w\}\subseteq \mathrm{bp}(\boldsymbol{s})\colon \mathrm{ch}_{\boldsymbol{s}}(v)=\{i,j,w\}, \mathrm{adj}(\boldsymbol{t},i,j)=1\\ 1-\alpha & \text{if } \exists! i\in[n]\exists! e\in \mathrm{edge}(\mathring{\rho}(\boldsymbol{s},i))\colon \boldsymbol{t}=\varphi(\mathring{\rho}(\boldsymbol{s},i),e,i)\\ c_v\alpha-1 & \text{if } \exists! i\in[n]\exists! v\in \mathrm{bp}(\mathring{\rho}(\boldsymbol{s},i))\colon \boldsymbol{t}=\varphi(\mathring{\rho}(\boldsymbol{s},i),v,i), \end{cases} \tag{3.10}$$

where c_v denotes the number of children of $v\in\mathrm{bp}(\boldsymbol{s})$, for each $\boldsymbol{t}\neq\boldsymbol{s}$ where there exists $i\in[n]$ such that $\boldsymbol{t}=\varphi(\mathring{\rho}(\boldsymbol{s},i),x,i)$ for some $x\in\mathrm{ins}(\mathring{\rho}(\boldsymbol{s},i))$. Additionally

$$p^{\mathrm{ex}}(\boldsymbol{s},\boldsymbol{s}) \propto (l^{\boldsymbol{s}}_{3,1}+2l^{\boldsymbol{s}}_{3,2})(1-\alpha)+\sum_{j=4}^{d^{\boldsymbol{s}}_{max}}\sum_{m=1}^{j-1} m l^{\boldsymbol{s}}_{j,m}\,((j-2)\alpha-1) \tag{3.11}$$

where $d^{\boldsymbol{s}}_{max}=\max_{v\in\mathrm{bp}(\boldsymbol{s})}\deg(v)$ and

$$l^{\boldsymbol{s}}_{j,m}=\#\left\{v\in\mathrm{bp}(\boldsymbol{s})\mid \deg(v)=j \text{ and } m=\max\{\#A\mid A\subseteq[n], \underset{\boldsymbol{s}}{\mathrm{ch}}(v)\supseteq A\}\right\}$$

is the number of vertices of $\boldsymbol{s}$ of degree j with m children that are leaves, for each $3\leq j\leq d_{max}$ and $m\leq j-1$.

Remark 3.4.3. One could consider a (generalized) version of the exchangeable stable chain, where the up-step is governed by the general (α,γ)-growth rule, see Definition 2.2.10, rather than the $(\alpha,1-\alpha)$-growth rule. Whilst other results presented in this paper would not generalize to this setting, the characterization in Proposition 3.4.2 does, and one simply needs to take care to distinguish between insertions into internal and external edges.

Proof of Proposition 3.4.2. Firstly note that, as opposed to a down-up chain with relabelling, we are almost always able to see exactly which leaf has moved during one step of the chain from $\boldsymbol{s}$ to $\boldsymbol{t}$ if $\boldsymbol{s} \neq \boldsymbol{t}$. Irrespective of the leaf that has been moved, there is a probability of $1/n$ of moving that particular leaf, and the total weight of the tree that the leaf is being inserted into by performing the up-step of the chain is $n-1-\alpha$, giving rise to the normalization constant. In all cases where we can see which label has moved, we can also see exactly where it was inserted, and hence the probability of the corresponding transition is simply $\frac{w_x}{n(n-1-\alpha)}$ where w_x is the weight of the insertable part that the leaf was inserted into in the up-step. This immediately explains the last two lines of (3.10).

We will now argue that the remaining cases are an exhaustive list of the cases where the same transition from $\boldsymbol{s}$ to $\boldsymbol{t}$ can result from several different down-steps followed by the appropriate up-step, whenever $\boldsymbol{s} \neq \boldsymbol{t}$, and that the probabilities are as claimed.

Let us start with the very general observation, that in order for us to not be able to see which label that was moved in a transition the leaf cannot move very far. In most cases this means that the leaf can only be relocated to one of the edges or branch points adjacent to its parent branch point, with the caveat that in the event where two branch points "merge" after the deletion of a leaf, this will extend to edges adjacent to the branch point further away from the leaf. This is illustrated in Figure 3.1.

Case 1: Transitions involving the top of a binary component. Lines $1-4$ of (3.10) all have in common that they involve the top of a binary component.

The first two cases involve three leaves, say $\{i,j,k\}$ such that $\boldsymbol{t}$ is obtained by relocating one of these leaves to a different location (encapsulated by requiring that $\boldsymbol{s} \neq \boldsymbol{t}$), and such that they constitute the top of a binary component (either in $\boldsymbol{s}$, in $\boldsymbol{t}$, or in both).

When we form a new top of a binary component, there are only two scenarios where we cannot see which label has moved; either they also constitute the top of a binary component in $\boldsymbol{s}$, i.e. such that $\mathrm{top}(\boldsymbol{s},i,j,k) = 1$, but where the top pair is different compared to $\boldsymbol{t}$, or they all have the same parent, say $v \in \mathrm{bp}(\boldsymbol{s})$, such that v has no other children, i.e. $\mathrm{ch}_{\boldsymbol{s}}(v) = \{i,j,k\}$. Now in either of these two scenarios, if we delete one of the three leaves, say k, we are left with a binary branch point, w, with a parent edge e and two leaf edges e_i and e_j. If we reinsert k anywhere in the tree but w, e, e_i or e_j we will be able to see that exactly leaf k has been moved. But if we reinsert into e_i we could obtain the same

tree by deleting leaf i and inserting it into e_k, or by deleting leaf j and inserting into e. As insertions into edges have weight $1-\alpha$ this shows line 1 of (3.10).

For line 2, we follow the argument above where we started out with the top of a binary component, but rather than insert the deleted leaf into one of the remaining edges, we now insert it into the branch point w. This is symmetric in a very obvious way, and since the insertion weight to a branch point with two children is $2\alpha-1$ we obtain line 2 of (3.10).

The last two cases involve just two leaves constituting the top of a binary component by themselves, which is simpler to handle than the above. For $\boldsymbol{s} \in \mathbb{T}_{[2]}$ there is nothing to prove, and for $\boldsymbol{s} \in \mathbb{T}_{[n]}$ for $n > 2$ we do a similar argument to above: If $\mathrm{top}(\boldsymbol{s}, i, j) = 1$, the only way we can relocate i or j and not see which one has moved, is if we insert it to the multifurcating branch point, w, which is the grandparent of both i and j in $\boldsymbol{s}$, which will have insertion weight $c\alpha - 1$ where c is the number of children of w in $\boldsymbol{s}$.

In order to form a new binary component during a transition, and not be able to identify the label which has moved, we require the leaves i and j to have the same multifurcating parent in $\boldsymbol{s}$ and then deleting one of i and j and inserting it into the other leaf edge, thus forming a new binary component. This proves lines 3 and 4 of (3.10).

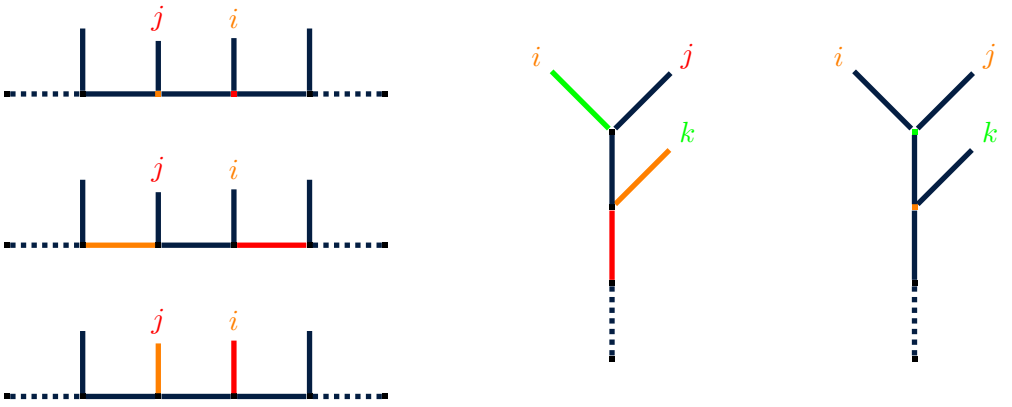


Figure 3.1: Illustration of transitions that require special attention. Three examples of two adjacent leaves, i and j, in a binary componenet are depicted in the left column. And the two right most trees depict leaves i, j and k forming the top of a binary component. In each of the cases above, we obtain the same tree by deleting leaf i and inserting it into the orange insertable part, by deleting j and inserting it into the red insertable part (in the rightmost tree, j should be inserted into the orange insertable part as well) or by deleting k, where applicable, and inserting it into the green insertable part. In (3.10), from bottom to top in the left column, these scenarios correspond to lines 4,5, and 6, respectively. The middle example correspond to one of the cases in line 1, whilst the rightmost example illustrates line 2. The remaining scenarios are similar in style.

Case 2: Transitions involving adjacent leaves in a binary component. Lines $5-7$ of (3.10) all involve adjacent leaves in binary components.

We first note that the requirement for two leaves, i and j, to be adjacent in $\boldsymbol{s}$ implies that there are exactly three branch points v_i, v_j and w that (if, say, i is closer to the root than j) satisfy that $\mathrm{ch}_{\boldsymbol{s}}(v_i) = \{i, v_j\}$ and $\mathrm{ch}_{\boldsymbol{s}}(v_j) = \{j, w\}$ and that $e_{v_i\uparrow}, e_{v_iv_j}$ and e_{v_jw} are all internal edges of $\boldsymbol{s}$ where $e_{v_i\uparrow}$ is the parent edge of v_i. The only way we can move one of two leaves that are adjacent, without being able to detect which one, is if they "switch places" thus ending up being adjacent again (by deleting leaf i and inserting it into e_{v_jw}, or by deleting j and inserting it into $e_{v_i\uparrow}$, as defined above), or if one is inserted onto the other's leaf edge, whereas in all other cases we would be able to see which leaf was moved. This directly proves lines 5 and 6 of (3.10).

The only remaining case, is when we have a multifurcating branch point, $v \in \mathrm{bp}(\boldsymbol{s})$, of exactly degree 4 where i and j are the only two children that are leaves (note that this is the only way we can end go from i and j having the same parent to them being adjacent and hence this is exactly the event that there exists another branch point $w \in \mathrm{bp}(\boldsymbol{s})$ such that $\mathrm{ch}_{\boldsymbol{s}}(v) = \{i, j, w\}$ and $\mathrm{adj}\{\boldsymbol{t}, i, j\} = 1$). The non-leaf edges that are connected to v are by definition internal edges, and given which of the leaves i and j we are required to insert the leaf into exactly one of these two edges to obtain $\boldsymbol{t}$. This gives rise to the weight of $2(1-\alpha)$ in line 7 of (3.10). ■

The factors of 2 and 3 that arise in the special cases outlined in Proposition 3.4.2 are not a consequence of the chain being multifurcating, but rather of inherent properties of tree symmetries and some of them are hence also special cases in the binary chain proposed by Aldous in [7], which are unaccounted for in that paper. These special cases will not affect the arguments in later sections, just as the special cases for the binary chain do not affect the relaxation or mixing time arguments in [7] and [48]. However, the special cases are not taken into account in the characterization of the transition probabilities of the binary chain in [7] and so it is worthwhile to explicitly supply these in that case.

For each $n \in \mathbb{N}$, recall that the subset of $\mathbb{T}_{[n]}$ consisting of binary trees is denoted by $\mathbb{T}^{\mathrm{bin}}_{[n]}$.

Corollary 3.4.4 (Binary Exchangeable Chain)**.** *The binary exchangeable stable on* $\mathbb{T}^{\text{bin}}_{[n]}$ *has transition probabilities*

$$p^{\text{bin}}(\boldsymbol{s},\boldsymbol{t}) \propto \begin{cases} 3 & \text{if } \exists!\{i,j,k\} \subseteq [n] \colon \operatorname{top}(\boldsymbol{s},i,j,k) = \operatorname{top}(\boldsymbol{t},i,j,k) = 1 \\ 2 & \text{if } \exists!\{i,j\} \subseteq [n] \colon \operatorname{adj}(\boldsymbol{s},i,j) = \operatorname{adj}(\boldsymbol{t},i,j) = 1 \\ 2 & \text{if } \exists!\{i,j\} \subseteq [n] \exists! v \in \operatorname{bp}(\boldsymbol{s}) \colon \operatorname{ch}_{\boldsymbol{s}}(v) = \{i,j\}, \operatorname{adj}(\boldsymbol{t},i,j) = 1 \\ 2 & \text{if } \exists!\{i,j\} \subseteq [n] \exists! v \in \operatorname{bp}(\boldsymbol{t}) \colon \operatorname{adj}(\boldsymbol{s},i,j) = 1, \operatorname{ch}_{\boldsymbol{t}}(v) = \{i,j\} \\ 1 & \text{if } \exists! i \in [n] \exists! e \in \operatorname{edge}(\rho(\boldsymbol{s},i)) \colon \boldsymbol{t} = \varphi(\rho(\boldsymbol{s},i),e,i) \end{cases} \tag{3.12}$$

and the normalization constant is $n(2n-3)$*.*

Remark 3.4.5. Aldous specifies the transition probabilities of the unrooted exchangeable stable binary chain. It is easy to check that the cases and weights for this chain should be the same as the ones outlined in (3.12). Going from the rooted to the unrooted case, the only difference is that the normalization constant changes to $(n+1)(2n-3)$ due to the option of picking the leaf labelled by 0 for relocation.

3.5 Lower bound for the relaxation time

In this section we will generalize Aldous's argument for the lower bound, which in the binary exchangeable can be found in Section 2 of [7]. Specifically, we will prove that the relaxation time — and hence the mixing time — of the rooted and unrooted exchangeable stable chain on $\mathbb{T}_{[n]}$ as well as the unlabelled stable chain on $\mathbb{T}^{\circ}_{[n]}$ is at least of order n^2. We will do so by using the characterization of the relaxation time detailed in (3.3). In [7] Aldous selected a sequence of functions that for each tree would output the number of leaves in the largest subtree of a branch point close to the centre of the tree — a so called centroid. We will use the same kind of functions here, but we provide a slightly different proof in order to be able to treat multifurcating trees efficiently.

Theorem 3.5.1. *For each* $n \geq 2$*, let* $\tau_n^{\text{ex, rel}}$ *be the relaxation time of the rooted exchangeable stable chain on* $\mathbb{T}_{[n]}$*. Then there exists* $c > 0$ *such that* $\tau_n^{\text{rel}} \geq cn^2$ *for all* $n \geq 2$*.*

Proof. Fix $n \in \mathbb{N}$ and consider $\boldsymbol{t} \in \mathbb{T}_{[n]}$ and a branch point $v \in \operatorname{bp}(\boldsymbol{t})$ with c_v children. Let $A := \{i_1, \ldots, i_{c_v}\}$ denote the set of least labels in the c_v subtrees of $\boldsymbol{t}$ rooted at v,

and consider the projection of $\boldsymbol{t}$ to the decorated tree with tree shape spanned by the root and A, $(\boldsymbol{s}, \mathbf{y}_v) := \pi_A^{\bullet n}(\boldsymbol{t})$ where $\boldsymbol{s} = \rho(\boldsymbol{t}, [n] \setminus A)$. Now, for each $n \in \mathbb{N}$, define the function $f_n \colon \mathbb{T}_{[n]} \to \mathbb{R}$ to be the smallest of the maximal subtree sizes of all the branch points of $\boldsymbol{t}$, i.e.

$$f_n(\boldsymbol{t}) = \min \{\max \mathbf{y}_v \mid v \in \mathrm{bp}(\boldsymbol{t})\}. \tag{3.13}$$

To avoid confusion, and since we are working with both rooted and unrooted trees in this chapter, we reiterate that for for each $\boldsymbol{t} \in \mathbb{T}_{[n]}$ and $v \in \mathrm{bp}(\boldsymbol{t})$ there is exactly one element of the vector $\mathbf{y}_v$ that count the number of leaves found subtrees attached to the ancestral line between v and the root (thus not including the root).

Now, similar to the binary case of Aldous,

$$|f_n(\boldsymbol{s}) - f_n(\boldsymbol{t})| \leq 1 \qquad \text{if} \qquad \mathrm{P}\left(T_n^{\mathrm{ex}}(1) = \boldsymbol{t} \mid T_n^{\mathrm{ex}}(0) = \boldsymbol{s}\right) > 0.$$

By 3.3 this immediately implies that

$$\tau_n^{u,\mathrm{rel}} \geq 2\mathrm{Var} f_n(T_n)$$

where T_n is the nth step of the stable growth growth process (the $(\alpha, 1-\alpha)$-growth process). It is not necessary to calculate $\mathrm{Var} f_n(T_n)$ precisely, we simply need to argue that is at least of order n^2.

We shall therefore utilize that we have much better control over the masses of the subtrees of a branch point if we can fix a branch point based on exactly which leaves it is separating in the tree. So consider $\left(\boldsymbol{s}_2, (Y_n^0, Y_n^1, Y_n^2, Y_n^{bp})\right) := \pi_{[2]}^{\bullet n}(T_n)$ where $Y_{bp} = n - \sum_{i=0}^2 Y_n^i$ and $\boldsymbol{s}_2$ is the unique element of $\mathbb{T}_{[2]}$. By standard urn scheme arguments we have that

$$\frac{1}{n}\left(Y_n^0, Y_n^1, Y_n^2, Y_n^{bp}\right) \xrightarrow{D} \mathrm{Dir}\left(1-\alpha, 1-\alpha, 1-\alpha, 2\alpha - 1\right) \tag{3.14}$$

for $n \to \infty$. Additionally, observe that

$$B_n := \left\{\max\{Y_n^0, Y_n^1, Y_n^2\} < \frac{n}{2}, \sum_{i=0}^{2} Y_n^i > \frac{3}{4}n\right\} \supseteq \left\{f_n(T_n) = \max\{Y_n^0, Y_n^1, Y_n^2\}\right\} \tag{3.15}$$

and that $\lim_{n\to\infty} \mathrm{P}(B_n) > 0$ due to 3.14. Now, observe that $1_{B_n} f_n(T_n) \in \left(\frac{n}{4}, \frac{n}{2}\right)$, and consider the two sequences of events:

$$B_n^- = B_n \cap \left\{ \frac{n}{4} < f_n(T_n) < \frac{5}{16} n \right\} \qquad \text{and} \qquad B_n^+ = B_n \cap \left\{ \frac{7}{16} n < f_n(T_n) < \frac{n}{2} \right\}$$

Then let $c = \frac{1}{16^2} \inf_{n\geq 32} \min\{\mathrm{P}(B_n^-), \mathrm{P}(B_n^+)\}$, and note that we know have

$$\mathrm{Var} f_n(T_n) \geq \frac{n^2}{16^2} \left(\mathrm{P}\left(B_n^+\right) 1_{E f_n(T_n) \leq \frac{3}{8} n} + \mathrm{P}\left(B_n^-\right) 1_{E f_n(T_n) > \frac{3}{8} n} \right) \geq c n^2 \tag{3.16}$$

finishing our proof. ■

Corollary 3.5.2. *For each $n \geq 2$, let $\tau_n^{\circ,\mathrm{rel}}$ and $\tau_n^{\mathrm{un,\ rel}}$ be the relaxation times of unlabelled stable chain and the unrooted exchangeable stable chain, respectively. There exists $c > 0$ such that $\min\{\tau_n^{\circ,\mathrm{rel}}, \tau_n^{rel}\} \geq c_1 n^2$ for all $n \geq 2$.*

Proof. This follows almost immediately from Theorem 3.5.1 as there are close to no differences in argument between the unrooted and rooted case, other than counting the root (meaning that for each $\boldsymbol{t} \in \mathbb{T}_{[n]}$ and $v \in \mathrm{bp}(\boldsymbol{t}), \sum \mathbf{y}_v = n + 1$) as we will still have $|f_n(\boldsymbol{s}) - f_n(\boldsymbol{t})| \leq 1$ if $\mathrm{P}\left(T_n^{\mathrm{un}}(1) = \boldsymbol{t} \mid T_n^{\mathrm{un}}(0) = \boldsymbol{s}\right) > 0$. For the unlabelled case, one would simply need to sample leaf labels as well and otherwise do the above argument. ■

3.6 Upper bound for the relaxation time

We finally turn our attention to the upper bound for the relaxation time of the down-up chains that we are investigating. In contrast to the discussion of the lower bound, we find fixing a root makes a rather big difference in Schweinsberg's distinguished paths argument from [48]. As a consequence we will start out with proving that the relaxation time of the unrooted exchangeable stable chain has relaxation time of exactly order n^2, and then discuss how the rooted exchangeable stable chain is different and why — using the same method — we obtain an upper bound of order $n^{3-\alpha}$ rather than n^2.

We begin this section with a scaling result for stable trees.

Lemma 3.6.1. *Fix $n \in \mathbb{N}$ and let T_n denote the nth step of the stable growth process. There exists a constant $c > 0$, not depending on n, such that*

$$\mathrm{P}\left(\mathrm{height}(T_n) \leq cn^{1-\alpha}\right) \geq \frac{1}{2}. \tag{3.17}$$

for all $n \in \mathbb{N}$.

Remark 3.6.2. There is an unfortunate clash in parametrization between the stable tree introduced in [37] and the stable tree as a subclass of the (α, γ)-trees from [12]. The translation is that the stable tree parametrized by $\theta \in (1, 2]$ in [37] corresponds to the stable tree parametrized by $\alpha = \frac{1}{\theta}$, $\alpha \in (1/2, 1]$, in [12]. We will continue to use the parametrization from the (α, γ)-trees.

Proof of Lemma 3.6.1. Corollary 24 of [32] shows that

$$\frac{T_n^\circ}{\alpha n^{1-\alpha}} \xrightarrow{P} \mathcal{T}_{stable-\alpha} \qquad \text{as } n \to \infty \tag{3.18}$$

in the Gromov-Hausdorff metric, and that

$$\frac{\mathrm{height}(T_n)}{\alpha n^{1-\alpha}} \xrightarrow{P} H \tag{3.19}$$

where H, the height of the α-stable tree, is a positive and almost-surely finite random variable. Hence in particular

$$\mathrm{P}\left(\mathrm{height}(T_n) \leq \alpha K n^{1-\alpha}\right) \to \mathrm{P}\left(H \leq K\right)$$

for any $K > 0$. For fixed $\varepsilon > 0$, choose K such that $\mathrm{P}\left(H \leq K\right) \geq \frac{1}{2} + \varepsilon$, and conclude the existence of $n_0 \in \mathbb{N}$ such that

$$\left|\mathrm{P}\left(\mathrm{height}(T_n) \leq \alpha K n^{1-\alpha}\right) - \mathrm{P}\left(H \leq K\right)\right| \leq \varepsilon$$

for all $n \geq n_0$. Hence by increasing c if necessary, this also holds for $1 \leq n \leq n_0$, proving the assertion. ■

Lemmas 3.6.1 provides some of the entities that appear in Theorem 3.2.7 if we choose $\mathcal{X}'$ to be the collection of trees with height less than $cn^{1-\alpha}$. In order to get any further we now need to elucidate the construction of the random path $\Gamma_{\boldsymbol{s},\boldsymbol{t}}$ between two trees $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{T}_{[n]}$ in the set-up outlined in Section 3.2.2. We will start out by doing so in the unrooted case, and then amend it slightly for the rooted case.

Let σ be a uniform permutation of $[n]_0 := \{0\} \cup [n] = \{0, \dots, n\}$, where we for brevity will use the notation that $\sigma(A) = \{\sigma(i) \mid i \in A\}$ for a set $A \subseteq [n]$. Next, note how for any subset $A \subset [n]$ and $\boldsymbol{t} \in \mathbb{T}_{[n]}$, we can talk about the subtree $\boldsymbol{t}^A := \pi_A(\boldsymbol{t}) \in \mathbb{T}_{[A]}$, i.e. the reduced subtree labelled by the elements of A and the root in $\boldsymbol{t}$, where all vertices of degree 2 have been contracted away. If no contraction other than possibly ones that will turn a string of edges into the root edge is necessary in order for $\boldsymbol{t}^A$ to be an element of $\mathbb{T}_A$ — no vertices having degree 2, that is — we say that $\boldsymbol{t}^A$ is a *fringe subtree* of $\boldsymbol{t}$.

We will now hook on to the very precise notation we have already introduced for the insertion of leaves into and deletion of leaves from a tree in Chapter 2. Fix a tree $\boldsymbol{s}$, and recall that $\mathring{\rho}$ denotes the deletion operation (see Definition 2.2.8) such that $\rho(\boldsymbol{s}, i)$ is the tree with leaf $i \in [n]$ deleted (and the parent branch point contracted away if necessary) from $\boldsymbol{s}$, and that we write $\pi_A(\boldsymbol{s}) := \mathring{\rho}(\boldsymbol{s}, [n] \setminus A)$ if we wish to delete all leaves labelled by elements of $[n]$ that are not in A. For the insertion operation, we have $\varphi(\boldsymbol{s}, x, i)$ (see Definition 2.2.7) being the tree where we have inserted the leaf i into the insertable part $x \in \mathrm{ins}(\boldsymbol{s})$ of a tree $\boldsymbol{s} \in \mathbb{T}_{[n]\setminus\{i\}}$. In the case of the unrooted chain, we will allow for the root to be relocated in one transition of the chain, and without a common reference point it is beneficial to explicitly incorporate the root into our projection notation in this case, i.e. write $\pi_{\{0\}\cup A}(\boldsymbol{t})$ when projecting $\boldsymbol{t} \in \mathbb{T}_{[n]}$ to the reduced subtree spanned by the leaves labelled by $\{0\} \cup A$, for a set $A \subseteq [n]$. Now, let $\boldsymbol{s}_0 = \boldsymbol{s}$ and construct a collection of trees $\boldsymbol{s}_1, \dots, \boldsymbol{s}_n = \boldsymbol{t}$ successively by for each $k \in \{1, \dots, n\}$ relocating the leaf labelled by $\sigma(k)$ such that we construct a fringe subtree within $\boldsymbol{s}_k$ labelled by $\sigma(1), \dots, \sigma(k)$ which is exactly the reduced subtree of $\boldsymbol{t}$ spanned by those leaves, see Figure 3.2 for an illustration of the beginning of such a path. Put differently, this sequence incrementally reconstructs $\boldsymbol{t}$ in a fringe subtree within $\boldsymbol{s}$, by relocating leaves in the order assigned by σ. In more

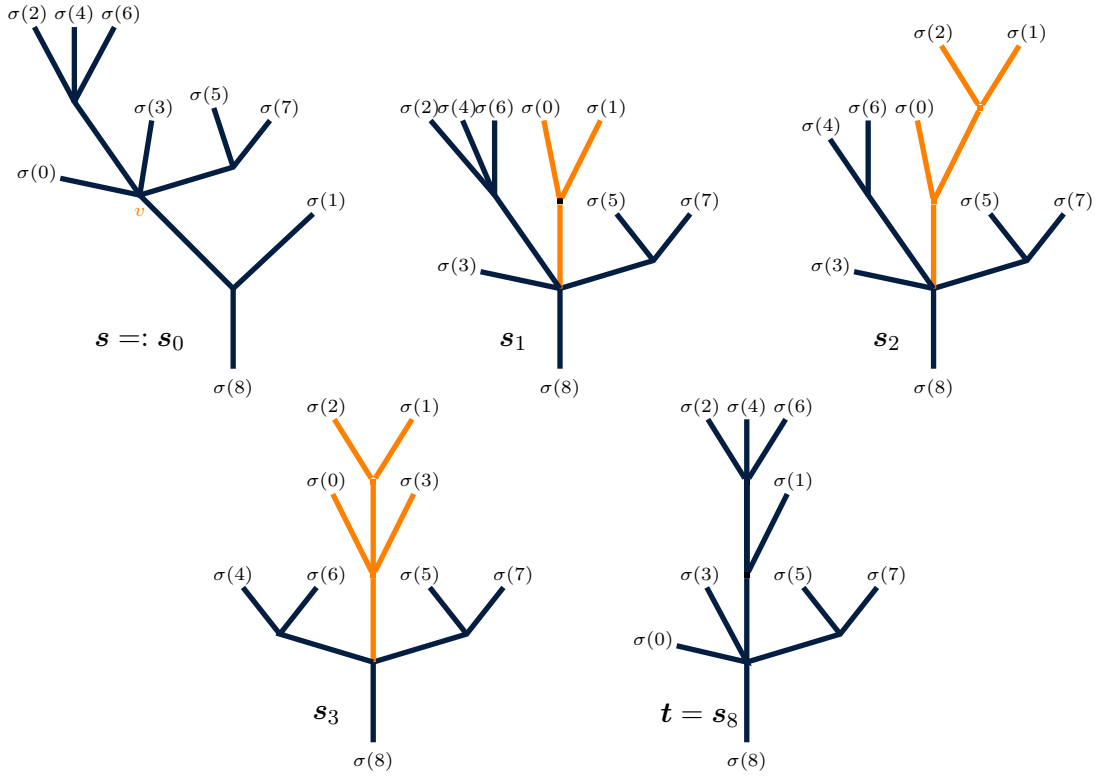


Figure 3.2: Constructing $\boldsymbol{t}$ *from* $\boldsymbol{s}$ *for a fixed permutation* σ *of* $[8]$. *Starting from* $\boldsymbol{s}_0 = \boldsymbol{t}$, $\boldsymbol{s}_1, \boldsymbol{s}_2$, *and* $\boldsymbol{s}_3$ *are the three trees that one encounter in the first three steps of* $\Gamma^{\mathrm{un}}_{\boldsymbol{s},\boldsymbol{t}}$.

mathematical terms, we set

$$\boldsymbol{s}_k = \varphi\left(\rho\left(\boldsymbol{s}_{k-1}, \sigma(k)\right), x_k, \sigma(k)\right)$$

for each $k \in [n]$, where x_k is the unique insertable part of $\mathring{\rho}\left(\boldsymbol{s}_{k-1}, \sigma(k)\right)$ such that $\pi_{\sigma([k]_0)}\left(\boldsymbol{s}_k\right)$ is a fringe subtree of $\boldsymbol{s}_k$ and for $k \in [n-1]$ we have

$$\pi_{\sigma([k]_0 \cup \{n\})}\left(\boldsymbol{s}_k\right) = \pi_{\sigma([k]_0 \cup \{n\})}(\boldsymbol{t}),$$

where we include $\{n\}$ to obtain an unrooted tree where we can identify the leaf that marks the "beginning" of the fringe subtree of $\boldsymbol{s}_k$ that we are constructing.

Recalling that $\left(T^{\mathrm{un}}_n(m)\right)_{m \in \mathbb{N}_0}$ denotes the unrooted exchangeable stable chain, consider the graph $\left(\mathbb{T}_{[n]}, \mathcal{E}^{\mathrm{un}}_n\right)$ where

$$\mathcal{E}^{\mathrm{un}}_n = \left\{ e_{\boldsymbol{st}} \mid \exists \boldsymbol{s}, \boldsymbol{t} \in \mathbb{T}_{[n]} \colon \mathrm{P}(T^{\mathrm{un}}_n(1) = \boldsymbol{t} \mid T^{\mathrm{un}}_n(0) = \boldsymbol{s}) > 0 \right\},$$

and observe that $\{e_{\boldsymbol{s}_{k-1}\boldsymbol{s}_k} \mid k \in [n]\} \subset \mathcal{E}_n^{\mathrm{un}}$. Hence

$$\tilde{\Gamma}_{\boldsymbol{s},\boldsymbol{t}}^{\mathrm{un}} := \left(e_{\boldsymbol{s}_0\boldsymbol{s}_1}, \ldots, e_{\boldsymbol{s}_{n-1}\boldsymbol{s}_n}\right)$$

is a path from $\boldsymbol{s}$ to $\boldsymbol{t}$ in the graph $\left(\mathbb{T}_{[n]}, \mathcal{E}_n^{\mathrm{un}}\right)$, and if we from $\tilde{\Gamma}_{\boldsymbol{s},\boldsymbol{t}}^{\mathrm{un}}$ remove all edges $e_{\boldsymbol{s}\boldsymbol{s}'}$ where $\boldsymbol{s} = \boldsymbol{s}'$ we traverse each edge at most once and the resulting path, $\Gamma_{\boldsymbol{s},\boldsymbol{t}}^{\mathrm{un}}$, has length at most n. We will say that e is in $\Gamma_{\boldsymbol{s},\boldsymbol{t}}^{\mathrm{un}}$ *at step* k if $e = e_{\boldsymbol{s}_{k-1}\boldsymbol{s}_k}$ from the above construction, and write $e = \Gamma_{\boldsymbol{s},\boldsymbol{t}}^{\mathrm{un},k}$.

Remark 3.6.3. Although we have technically not introduced $\mathring{\rho}$ and φ in the cases involving the relocation of the root (i.e. the leaf labelled by 0) it should be completely clear that this works in exactly the same way as for any other leaf.

Remark 3.6.4. The construction of the random path, $\Gamma_{\boldsymbol{s},\boldsymbol{t}}^{\mathrm{ex}}$, in the case of the (rooted) exchangeable stable chain is almost identical, except that the random permutation is a permutation on $[n]$ rather than $[n]_0$, that we consider the path in $\left(\mathbb{T}_{[n]}, \mathcal{E}_n^{\mathrm{ex}}\right)$ where

$$\mathcal{E}_n^{\mathrm{ex}} = \left\{e_{\boldsymbol{s}\boldsymbol{t}} \mid \exists \boldsymbol{s}, \boldsymbol{t} \in \mathbb{T}_{[n]} \colon \mathrm{P}(T_n^{\mathrm{ex}}(1) = \boldsymbol{t} \mid T_n^{\mathrm{ex}}(0) = \boldsymbol{s}) > 0\right\}.$$

and start by deleting the leaf labelled by $\sigma(2)$ and insert it onto the leaf labelled by $\sigma(1)$ as opposed to relocating the leaf labelled by $\sigma(1)$ to the leaf labelled by $\sigma(0)$.

Lemma 3.6.5. *Let* $\mathbb{T}'_{[n]} := \left\{\boldsymbol{t} \in \mathbb{T}_{[n]} \mid \mathrm{height}(\boldsymbol{t}) \leq An^{1-\alpha}\right\}$ *for some* $A > 0$ *and define* $K(e) := \left\{k \in [n] \;\middle|\; \exists \boldsymbol{s} \in \mathbb{T}_{[n]}, \boldsymbol{t} \in \mathbb{T}'_{[n]} \colon \mathrm{P}(e = \Gamma_{\boldsymbol{s},\boldsymbol{t}}^{\mathrm{un},k}) > 0\right\}$ *for each* $e \in \mathcal{E}_n^{\mathrm{un}}$. *Then* $\#K(e) \leq 2An^{1-\alpha}$ *for all* $e \in \mathcal{E}_n^{\mathrm{un}}$.

Proof. Fix $\boldsymbol{s}', \boldsymbol{t}' \in \mathbb{T}_{[n]}$ such that $\boldsymbol{s}' \neq \boldsymbol{t}'$, and consider the edge $e = e_{\boldsymbol{s}',\boldsymbol{t}'} \in \mathcal{E}_n^{\mathrm{un}}$. We note that apart from the inconsequential and very exceptional circumstance where $k = \lfloor n/2 \rfloor$ and there are two adjacent leaves i and j in $\boldsymbol{s}'$ that swap places to form $\boldsymbol{t}'$, we will due to the definition of $\Gamma_{\boldsymbol{s},\boldsymbol{t}}^{\mathrm{un}}$ always be able to see from $\boldsymbol{s}', \boldsymbol{t}'$ and k exactly which leaf was moved.

Disregarding this oddity, say that we can obtain $\boldsymbol{t}'$ from $\boldsymbol{t}$ be relocating leaf i, and let x denote the insertable part of $\mathring{\rho}(\boldsymbol{s}', i)$ such that $\boldsymbol{t}' = \varphi(\mathring{\rho}(\boldsymbol{s}', i), x, i)$. Given i and x we can construct a path Υ in $\boldsymbol{s}'$ starting at leaf i and ending at the insertable part x of $\boldsymbol{s}'$, and to avoid ambiguity we will say that the length of Υ is the number of edges of $\boldsymbol{s}'$ in Υ.

Now, suppose that $e = e_{\boldsymbol{s}'\boldsymbol{t}'} = \Gamma^{\mathrm{un},k}_{\boldsymbol{s},\boldsymbol{t}}$ for some $k \in [n]$ and some $\boldsymbol{s} \in \mathbb{T}_{[n]}, \boldsymbol{t} \in \mathbb{T}'_{[n]}$. Then, by construction of $\Gamma^{\mathrm{un}}_{\boldsymbol{s},\boldsymbol{t}}$, we have $\boldsymbol{s}_k = \boldsymbol{s}'$ and $\boldsymbol{s}_{k+1} = \boldsymbol{t}'$. Furthermore, $\boldsymbol{s}_k$ contains a fringe subtree, S, labelled by $\sigma([k]_0)$ that contains x, immediately implying that the root of S, call it $r \in \mathrm{bp}(\boldsymbol{s}_k)$, is in Υ. Since $\boldsymbol{t} \in \mathbb{T}'_{[n]}$ the part of Υ that starts at r and ends at x is shorter than $2An^{1-\alpha}$ (if the leaf labelled by 0 is already in the fringe subtree, we only know that the distance from r to 0 and from 0 to x are both bounded by $An^{1-\alpha}$). Now, r has at least two — and possibly several more — subtrees, but only one of them will contain x, and so there are at most $2An^{1-\alpha}$ subtrees of $\boldsymbol{s}' = \boldsymbol{s}_k$ that could be S. And since S has exactly k leaves as well as the root this implies that there are only $2An^{1-\alpha}$ possible values for k, i.e. that $\#K(e) \leq 2An^{1-\alpha}$. ■

The construction of the random path $\Gamma^{\mathrm{un}}_{\boldsymbol{s},\boldsymbol{t}}$ between two trees $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{T}_{[n]}$ means that we obtain a lot of information regarding the two trees $\boldsymbol{s}', \boldsymbol{t}'$ if we know that $e_{\boldsymbol{s}'\boldsymbol{t}'}$ is in $\Gamma^{\mathrm{un}}_{\boldsymbol{s},\boldsymbol{t}}$ at step k for some $k \in [n]$:

Lemma 3.6.6. *Let $T = (T_k)_{k\in[n]}$ and $T' = (T'_k)_{k\in[n]}$ be the first n steps of two independent $(\alpha, 1-\alpha)$-growth processes. Let Σ be a uniform random permutation of $\{0, \dots, n\}$, independent of T and T', and let $\Gamma^{\mathrm{un}}_{\boldsymbol{s},\boldsymbol{t}}$ be the random path induced by Σ as defined prior to Remark 3.6.4 for each $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{T}_{[n]}$. Then for each $\boldsymbol{s}', \boldsymbol{t}' \in \mathbb{T}_{[n]}$ we have for all $k \in [n]$ where $\mathrm{P}(\Gamma^{\mathrm{un},k}_{T_n,T'_n} = e_{\boldsymbol{s}',\boldsymbol{t}'}) > 0$ that*

$$\begin{aligned}\mathrm{P}\left(e_{\boldsymbol{s}'\boldsymbol{t}'} = \Gamma^{\mathrm{un},k}_{T_n,T'_n}\right) = \mathrm{P}\Big(&\left\{T'_{k+1} = \pi_{\sigma([k]_0\cup\{n\})}(\boldsymbol{t}')^{\mathrm{re}}\right\} \cap \left\{T_{n-k+1} = \pi_{\sigma([n]_0\setminus[k-1])}(\boldsymbol{s}')^{\mathrm{re}}\right\} \\ &\cap \{\Sigma([k-1]_0) = \sigma([k-1]_0)\} \\ &\cap \{\Sigma(k) = \sigma(k)\}\Big)\end{aligned}$$

where σ is some permutation of $[n]_0$ and $\pi_A(\boldsymbol{r})^{\mathrm{re}}$ is the subtree of $\boldsymbol{r} \in \mathbb{T}_{[n]}$ spanned by the leaves labelled by the elements of $A \subseteq [n]_0$ and subsequently relabelled by the increasing bijection $A \mapsto [\#A - 1]_0$.

Remark 3.6.7. This result holds in the rooted case as well by carefully updating it to account for the differences between $\Gamma^{\mathrm{un}}_{T_n,T'_n}$ and $\Gamma^{\mathrm{ex}}_{T_n,T'_n}$ and the restriction of Σ to be a

uniform random permutation on $[n]$ rather than $[n]_0$, thus obtaining

$$\mathrm{P}\left(e_{\boldsymbol{s}'\boldsymbol{t}'} = \Gamma^{\mathrm{ex},k}_{T_n,T'_n}\right) = \mathrm{P}\Big(\left\{T'_{k+1} = \pi_{\sigma([k+1])}(\boldsymbol{t}')^{\mathrm{re}}\right\} \cap \left\{T_{n-k+1} = \pi_{\sigma([n]\setminus\{2,\dots,k\})}(\boldsymbol{s}')^{\mathrm{re}}\right\} \\ \cap \{\Sigma([k]) = \sigma([k]), \Sigma(k+1) = \sigma(k+1)\}\Big)$$

for all $1 \leq k \leq n-1$ and $\boldsymbol{s}', \boldsymbol{t}' \in \mathbb{T}_{[n]}$.

Proof of Lemma 3.6.6. First consider $\boldsymbol{s}, \boldsymbol{t}, \boldsymbol{s}', \boldsymbol{t}' \in \mathbb{T}_{[n]}$ and $k \in [n]$ such that $e_{\boldsymbol{s}'\boldsymbol{t}'} = \Gamma^{\mathrm{un},k}_{\boldsymbol{s},\boldsymbol{t}}$. Just as in the proof of Lemma 3.6.5 we note that we given $\boldsymbol{s}', \boldsymbol{t}' \in \mathbb{T}_{[k]}$ and $k \in [n]$ almost always can identify which label was moved to construct $\boldsymbol{t}'$ from $\boldsymbol{s}'$ in the kth step of $\Gamma^{\mathrm{un}}_{\boldsymbol{s},\boldsymbol{t}}$, as well as which labels have been moved in the previous $k-1$ steps (the inconsequential case where we are not able to see which label has moved will be dealt with at the end of the proof). Let $i \in [n]$ denote the label of the leaf moved in step k and let $B \subseteq [n]_0$ denote the set of leaf labels that has been moved prior to step k. Construct a permutation $\sigma\colon [n]_0 \to [n]_0$ by saying that σ is increasing on the sets $[k-1]_0$ and $[n]_0 \setminus [k]_0$ such that $\sigma([k-1]_0) = B$, $\sigma(k) = i$, and $\sigma([n]_0 \setminus [k]_0) = [n]_0 \setminus (B \cup \{i\})$.

Now, by definition of $e_{\boldsymbol{s}'\boldsymbol{t}'}$ being in $\Gamma^{\mathrm{un}}_{\boldsymbol{s},\boldsymbol{t}}$ at step k there are

- a unique fringe subtree, S, of $\boldsymbol{t}'$ labelled by the elements of $\sigma([k]_0)$ which is equal to $\pi_{\sigma([k]_0)}(\boldsymbol{t})$, rooted at a vertex in $\boldsymbol{t}'$ which we, when considering S as an unrooted tree by itself, will label $\sigma(n)$, so that we can write $S = \pi_{\sigma([k]_0\cup\{n\})}(\boldsymbol{t}')$, and

- a unique fringe subtree, R, of $\boldsymbol{s}'$ labelled by the elements of $\sigma([n] \setminus [k-1])$ is equal to $\pi_{\sigma([n]\setminus[k-1])}(\boldsymbol{s})$, rooted at a vertex in $\boldsymbol{s}'$ which we, when considering S as an unrooted tree by itself, will label $\sigma(0)$. We can therefore write $R = \pi_{\sigma([n]_0\setminus[k-1])}(\boldsymbol{s}')$.

The first of these is obvious from the construction of $\Gamma^{\mathrm{un}}_{\boldsymbol{s},\boldsymbol{t}}$. The second is true as at step k, the leaves labelled by $\sigma([n] \setminus [k-1])$ have not been moved.

We can reconstruct both $\boldsymbol{s}'$ and $\boldsymbol{t}'$ by deleting $\sigma([k])$ from either S (to obtain $\boldsymbol{s}'$) or R (to obtain $\boldsymbol{t}'$), respectively, and attach R to S by replacing the leaf labelled by $\sigma(n)$ in S and the leaf labelled by $\sigma(0)$ in R with an edge between their two parent branch points, respectively. It is easily seen that the corresponding edge between $\boldsymbol{s}'$ and $\boldsymbol{t}'$ in $\mathcal{E}^{\mathrm{un}}_n$ will indeed be in the kth step of $\Gamma^{\mathrm{un}}_{\boldsymbol{s},\boldsymbol{t}}$.

This shows that the event $\{e_{\boldsymbol{s}'\boldsymbol{t}'} = \Gamma^{\mathrm{ex},k}_{T_n,T'_n}\}$ is equivalent to the event composed of the following three events:

- $\Sigma([k-1]_0) = \sigma([k-1]_0), \Sigma(k) = \sigma(i)$,
- $\pi_{\sigma([n]_0\setminus[k-1])}(T_n) = \pi_{\sigma([k]_0)}(\boldsymbol{s}')$, and
- $\pi_{\sigma([k]_0\cup\{n\})}(T'_n) = \pi_{\sigma([k]_0\cup\{n\})}(\boldsymbol{t}')$.

Using that $\pi_{A\cup\{0\}}(T_n)^{\mathrm{re}} \overset{D}{=} T_{\#A}$ for any set $A \subseteq [n]$ now yields the result.

Returning to the odd case where we cannot identify which leaf is moved in the kth step based on $\boldsymbol{s}', \boldsymbol{t}' \in \mathbb{T}_{[k]}$ and $k \in [n]$, we observe — as in the proof of Lemma 3.6.5 — that a requirement for this to occur is that $k = \lfloor \frac{n}{2} \rfloor$ and R and S being of the same size, and that can only ever be two possible options for i. With this in mind, simply observe that the expression in this case is completely symmetric in k and $n-k$ and so the probability is the same no matter which i pick. ■

Equipped with Lemmas 3.6.1, 3.6.5, and 3.6.6 we are now ready to prove that the relaxation time of the unrooted exchangeable stable chain is at most of order n^2.

Theorem 3.6.8. *There exists $c_2 > 0$ such that $\tau_n^{\mathrm{un,\,rel}} \leq c_2 n^2$ for all $n \geq 3$.*

Proof. Translating to our current setting, the upper bound for the relaxation time characterized in Theorem 3.2.7 is

$$\tau_n^{\mathrm{un,\,rel}} \leq \frac{4n}{\mathrm{P}\left(T_n \in \mathbb{T}'_{[n]}\right)} \max_{e_{\boldsymbol{s}'\boldsymbol{t}'} \in \mathcal{E}_n^{\mathrm{un}}} \left[\frac{1}{\mathrm{P}\left(T_n = \boldsymbol{s}'\right) p^{\mathrm{un}}(\boldsymbol{s}', \boldsymbol{t}')} \times \sum_{\boldsymbol{s} \in \mathbb{T}_{[n]}} \sum_{\boldsymbol{t} \in \mathbb{T}'_{[n]}} \mathrm{P}\left(T_n = \boldsymbol{s}\right) \mathrm{P}\left(T_n = \boldsymbol{t}\right) \mathrm{P}\left(e_{\boldsymbol{s}'\boldsymbol{t}'} \in \Gamma^{\mathrm{un}}_{\boldsymbol{s},\boldsymbol{t}}\right) \right] \tag{3.20}$$

where p^{un} denotes the transition kernel for the unrooted exchangeable stable chain, for any choice of $\mathbb{T}'_{[n]} \subseteq \mathbb{T}_{[n]}$, as $\Gamma^{\mathrm{un}}_{\boldsymbol{s},\boldsymbol{t}}$ satisfies that an edge of $\mathcal{E}_n^{\mathrm{un}}$ can appear at most once and that its maximal length is at most n for any $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{T}_{[n]}$. Now, letting $(K(e))_{e\in\mathcal{E}_n^{\mathrm{un}}}$ and $\mathbb{T}'_{[n]}$ be defined as in Lemma 3.6.5 (with $A = c$ from Lemma 3.6.1), we can deploy both of the

upper bounds from those lemmas to obtain

$$\begin{aligned}
\tau_n^{\text{un, rel}} \leq & 8n \max_{e_{\boldsymbol{s}'\boldsymbol{t}'} \in \mathcal{E}_n^{\text{un}}} \left[\frac{1}{\mathrm{P}\left(T_n = \boldsymbol{s}'\right) p^{\text{un}}(\boldsymbol{s}', \boldsymbol{t}')} \right. \\
& \left. \times \sum_{\boldsymbol{s} \in \mathbb{T}_{[n]}} \sum_{\boldsymbol{t} \in \mathbb{T}'_{[n]}} \mathrm{P}\left(T_n = \boldsymbol{s}\right) \mathrm{P}\left(T_n = \boldsymbol{t}\right) \left(\sum_{k \in K(e_{\boldsymbol{s}'\boldsymbol{t}'})} \mathrm{P}\left(e_{\boldsymbol{s}'\boldsymbol{t}'} = \Gamma_{\boldsymbol{s},\boldsymbol{t}}^{\text{un},k} \right) \right) \right] \\
\leq & 16An^{2-\alpha} \max_{e_{\boldsymbol{s}'\boldsymbol{t}'} \in \mathcal{E}_n^{\text{un}}} \max_{k \in K(e_{\boldsymbol{s}'\boldsymbol{t}'})} \left[\frac{1}{\mathrm{P}\left(T_n = \boldsymbol{s}'\right) p^{\text{un}}(\boldsymbol{s}', \boldsymbol{t}')} \right. \\
& \left. \times \sum_{\boldsymbol{s} \in \mathbb{T}_{[n]}} \sum_{\boldsymbol{t} \in \mathbb{T}_{[n]}} \mathrm{P}\left(T_n = \boldsymbol{s}\right) \mathrm{P}\left(T_n = \boldsymbol{t}\right) \mathrm{P}\left(e_{\boldsymbol{s}'\boldsymbol{t}'} = \Gamma_{\boldsymbol{s},\boldsymbol{t}}^{\text{un},k} \right) \right] \\
= & 16An^{2-\alpha} \max_{e_{\boldsymbol{s}'\boldsymbol{t}'} \in \mathcal{E}_n^{\text{un}}} \max_{k \in K(e_{\boldsymbol{s}'\boldsymbol{t}'})} \frac{\mathrm{P}\left(e_{\boldsymbol{s}'\boldsymbol{t}'} = \Gamma_{T_n, T'_n}^{\text{un},k} \right)}{\mathrm{P}\left(T_n = \boldsymbol{s}'\right) p^{\text{un}}(\boldsymbol{s}', \boldsymbol{t}')}
\end{aligned} \tag{3.21}$$

where $T = (T_n)_{n \in \mathbb{N}}$ and $T' = (T'_n)_{n \in \mathbb{N}}$ are i.i.d. $(\alpha, 1-\alpha)$-growth processes.

Now, let us focus on the ratio in (3.21) and use Lemma 3.6.6 and the independence of T, T' and Σ to write

$$\begin{aligned}
\frac{\mathrm{P}\left(e_{\boldsymbol{s}'\boldsymbol{t}'} = \Gamma_{T_n, T'_n}^{\text{un},k} \right)}{\mathrm{P}\left(T_n = \boldsymbol{s}'\right) p^{\text{un}}(\boldsymbol{s}', \boldsymbol{t}')} = & \frac{\mathrm{P}(T_{k+1} = \pi_{\sigma([k]_0 \cup \{n\})}(\boldsymbol{t}')^{\text{re}}) \mathrm{P}\left(T_{n-k+1} = \pi_{\sigma([n]_0 \setminus [k-1])}(\boldsymbol{s}')^{\text{re}} \right)}{\mathrm{P}\left(T_n = \boldsymbol{s}'\right) p^{\text{un}}(\boldsymbol{s}', \boldsymbol{t}')} \\
& \times \mathrm{P}\left(\Sigma([k-1]_0) = \sigma([k-1]_0), \Sigma(k) = \sigma(k) \right).
\end{aligned} \tag{3.22}$$

Now note that we have characterized the probabilities related to the $(\alpha, 1-\alpha)$-growth process explicitly in Proposition 3.4.1, and let us discuss the weights that appear in the numerator of (3.6). Comparing $\mathrm{P}(T_n = \boldsymbol{s}')$ to $\mathrm{P}(T_n = \boldsymbol{t}')$ the two expressions can differ by at most one weight, so we let $w_{\boldsymbol{t}'}$ denote the weight that appears in the latter but not in the former (define it to be 1 otherwise). We have also explicitly characterized the transition probability $p_n^{\text{un}}(\boldsymbol{s}', \boldsymbol{t}')$ in Proposition 3.4.2, and we note that the weight appearing in the numerator of the transition probability, call it $w_{\boldsymbol{s}',\boldsymbol{t}'}$, is either exactly the $w_{\boldsymbol{t}'}$, or a factor of at most 3 of that.

Furthermore, observe how there are a combined number of n weights appearing in the expression for $\mathrm{P}\left(T_{k+1} = \pi_{\sigma([k]_0 \cup \{n\})}(\boldsymbol{t}')^{\text{re}} \right) \mathrm{P}\left(T_{n-k+1} = \pi_{\sigma([n]_0 \setminus [k-1])}(\boldsymbol{s}')^{\text{re}} \right)$ of which the $n-k$ appearing in $\mathrm{P}\left(T_{n-k+1} = \pi_{\sigma([n]_0 \setminus [k-1])}(\boldsymbol{s}')^{\text{re}} \right)$ can be found amongst the $n-1$ weights in the expression for $\mathrm{P}\left(T_n = \boldsymbol{s}'\right)$. There is at most one of the k weights of $\mathrm{P}(T_{k+1} = \pi_{\sigma([k]_0 \cup \{n\})}(\boldsymbol{t}')^{\text{re}})$ that cannot be found in a similar way. Call the extra weight

$w^0_{\boldsymbol{t}'}$, and observe that this is either exactly $w_{\boldsymbol{t}'}$ or at most $w_{\boldsymbol{s}',\boldsymbol{t}'}$ and in either case we have $w^0_{\boldsymbol{t}'} \leq w_{\boldsymbol{s}',\boldsymbol{t}'}$.

Lastly, note that $\prod_{i=1}^{j}(i-\alpha) = \frac{\Gamma(j+1-\alpha)}{\Gamma(1-\alpha)}$ for any $j \in \mathbb{N}$.

Combine the above to obtain

$$\begin{aligned}
&\frac{\mathrm{P}(T_{k+1} = \pi_{\sigma([k]_0\cup\{n\})}(\boldsymbol{t}')^{\mathrm{re}})\mathrm{P}\left(T_{n-k+1} = \pi_{\sigma([n]_0\setminus[k-1])}(\boldsymbol{s}')^{\mathrm{re}}\right)}{\mathrm{P}\left(T_n = \boldsymbol{s}'\right)p^{\mathrm{un}}(\boldsymbol{s}',\boldsymbol{t}')} \\
\leq\ & \frac{\frac{\Gamma(n-\alpha)}{\Gamma(1-\alpha)}(n+1)(n-1-\alpha)}{\frac{\Gamma(k+1-\alpha)}{\Gamma(1-\alpha)}\frac{\Gamma(n-k+1-\alpha)}{\Gamma(1-\alpha)}} \leq \Gamma(1-\alpha)\frac{\Gamma(n-\alpha)n(n+1)}{\Gamma(k+1-\alpha)\Gamma(n-k+1-\alpha)}.
\end{aligned}$$

Plugging this into (3.22), combined with the following inequality due to Wendel

$$\frac{x}{(x+s)^{1-s}} \leq \frac{\Gamma(x+s)}{\Gamma(x)} \leq x^s \qquad \text{for all } 0 < s < 1 \text{ and } x > 0$$

yields that

$$\begin{aligned}
\frac{\mathrm{P}\left(e_{\boldsymbol{s}'\boldsymbol{t}'} = \Gamma^{\mathrm{un},k}_{T_n,T'_n}\right)}{\mathrm{P}\left(T_n = \boldsymbol{s}'\right)p^{\mathrm{un}}(\boldsymbol{s}',\boldsymbol{t}')} &\leq \Gamma(1-\alpha)\frac{\Gamma(n-\alpha)n(n+1)}{\Gamma(k+1-\alpha)\Gamma(n-k+1-\alpha)}\frac{k!(n-k+1)!}{(n+1)!(n-k+1)} \\
&= \Gamma(1-\alpha)\frac{\Gamma(n-\alpha)}{\Gamma(n)}\frac{\Gamma(k+1)}{\Gamma(k+1-\alpha)}\frac{\Gamma(n-k+1)}{\Gamma(n-k+1-\alpha)} \\
&\leq c\Gamma(1-\alpha)n^{-\alpha}k^{\alpha}(n-k)^{\alpha},
\end{aligned}$$

for some constant $c > 0$. This expression is maximized at $k = \frac{n}{2}$ and so

$$\frac{\mathrm{P}\left(e_{\boldsymbol{s}'\boldsymbol{t}'} = \Gamma^{\mathrm{un},k}_{T_n,T'_n}\right)}{\mathrm{P}\left(T_n = \boldsymbol{s}'\right)p^{\mathrm{un}}(\boldsymbol{s}',\boldsymbol{t}')} \leq \frac{c\Gamma(1-\alpha)}{2^{2\alpha}}n^{-\alpha}n^{2\alpha} = \frac{c\Gamma(1-\alpha)}{2^{2\alpha}}n^{\alpha}$$

for all $e_{\boldsymbol{s}'\boldsymbol{t}'} \in \mathcal{E}^{\mathrm{un}}_n$, and combining this with (3.21) yields the result. ■

Theorem 3.1.1 now follows by combining Corollary 3.5.2 and Theorem 3.6.8.

Before proving Theorem 3.1.2 we will now highlight a few key points that turn out to be crucial in making the above proof work. Fundamentally, allowing for the root to be relocated during a transition of the down-up chain means that we can allow the random permutation Σ to also assign a random label in $\{0,\ldots,n\}$ to the root, which means that the chance of seeing a particular permutation is a factor of n smaller. Comparing this with Lemma 3.6.6 it means that the denominator of the probability of seeing a particular permu-

tation is $(n+1)! = \Gamma(n+2)$ which is *larger* than the denominators of $\mathrm{P}(T_n = \boldsymbol{t}')p^{\mathrm{un}}(\boldsymbol{s}', \boldsymbol{t}')$ which is $\Gamma(n-\alpha)(n+1)(n-1-\alpha) \approx \Gamma(n+2-\alpha)$.

In the rooted case we do not have a possible transition that only relocates the root, and so we cannot allow the random path in $\mathcal{E}_n^{\mathrm{ex}}$ to do so, limiting our options for the random permutation. This will mean that the denominator of the probability of seeing a particular permutation is now only of order $\Gamma(n+1-\alpha)$ even though everything else stays the same.

Proof of Theorem 3.1.2. Just as in the proof of Theorem 3.6.8 we can obtain the analogue of (3.21), i.e. that

$$\tau_n^{\mathrm{ex,\ rel}} \leq 16An^{2-\alpha} \max_{e_{\boldsymbol{s}'\boldsymbol{t}'} \in \mathcal{E}_n^{\mathrm{ex}}} \max_{k \in K(e_{\boldsymbol{s}'\boldsymbol{t}'})} \frac{\mathrm{P}\left(e_{\boldsymbol{s}'\boldsymbol{t}'} = \Gamma^{\mathrm{ex},k}_{T_n, T_n'}\right)}{\mathrm{P}\left(T_n = \boldsymbol{s}'\right) p^{\mathrm{ex}}(\boldsymbol{s}', \boldsymbol{t}')}$$

but we now only obtain that

$$\begin{aligned}
\frac{\mathrm{P}\left(e_{\boldsymbol{s}'\boldsymbol{t}'} = \Gamma^{\mathrm{ex},k}_{T_n, T_n'}\right)}{\mathrm{P}\left(T_n = \boldsymbol{s}'\right) p^{\mathrm{ex}}(\boldsymbol{s}', \boldsymbol{t}')} &\leq \Gamma(1-\alpha) \frac{\Gamma(n-\alpha)n(n-1)}{\Gamma(k+1-\alpha)\Gamma(n-k+1-\alpha)} \frac{k!(n-k)!}{n!(n-k)} \\
&\leq \Gamma(1-\alpha) \frac{\Gamma(n-\alpha)}{\Gamma(n-1)} \frac{\Gamma(k+1)}{\Gamma(k+1-\alpha)} \frac{\Gamma(n-k)}{\Gamma(n-k+1-\alpha)} \\
&\leq c\Gamma(1-\alpha) n^{1-\alpha} k^{\alpha} (n-k)^{-1+\alpha}
\end{aligned}$$

for some constant $c > 0$. Crucially, we no longer have a symmetric expression in k and $n-k$ to maximize, and the maximum is attained at $k = n-1$ rather than at $\frac{n}{2}$ as in the proof of Theorem 3.6.8. Hence we only get that

$$\tau_n^{\mathrm{ex,\ rel}} \leq 16Ac\Gamma(1-\alpha)n^{3-\alpha}$$

finishing the proof. ■

Remark 3.6.9. We tried several variations of the above argument for the rooted exchangeable stable case by looking at different ways to construct the random path $\Gamma^{\mathrm{ex}}_{\boldsymbol{s},\boldsymbol{t}}$ in $\mathcal{E}_n^{\mathrm{ex}}$ between $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{T}_{[n]}$. One approach was to try and incrementally reconstruct $\boldsymbol{t}$ in a fringe subtree at the root edge of $\boldsymbol{s}$. Whilst this lead to a slighter longer path (of length n rather than length $n-1$) it does not change the fundamental deficit that is created by not allowing the random permutation to assign another value to the root. We have tried

two different ways of constructing these fringe subtrees near the root in $\boldsymbol{s}$; one where we just from knowing at which step we are in the path can see exactly which subtree we are constructing, and one where we retain the uncertainty as in the proof of Theorems 3.1.1 and 3.1.2. In the former we end up picking up additional factors of n up elsewhere, and obtain a weaker upper bound of $n^{2+\alpha}$, whilst we an upper bound of $n^{3-\alpha}$ as in Theorem 3.1.2.

It seems that the way forward in trying to resolve the relaxation time related part of Conjecture 3.1.4 is to find a way to allow the random permutation to assign a different label to the root, but simultaneously ensure that the root is not actually moved, as this would create a path not consisting of edges in $\mathcal{E}_n$.

Corollary 3.6.10. *The relaxation time of the unlabelled $(\alpha, 1-\alpha)$-chain on $\mathbb{T}_{[n]}$ is at least of order n^2 and at most of order $n^{3-\alpha}$.*

Proof. This follows by combining Theorem 3.1.2, Proposition 3.2.8 and Corollary 3.5.2. ■

We finish this chapter of by combining all of the above bounds for the various chains to give us bounds for the mixing times of the chains.

Proof of 3.1.3. This follows immediately by combining Theorems 3.1.1, 3.1.2, and Corollary 3.6.10 with (3.5) and (3.7). ■

Appendix A

Non-reversibility of the (α, γ)-chain

This final appendix serves a single purpose — to bridge the gap between Chapters 2 and 3 by providing a proof that all the down-up chains with label swapping and relabelling analyzed in Chapter 2 are non-reversible and hence cannot be approached using the methods we deploy in Chapter 3.

Recall therefor the definition of the transition kernel of the (α, γ)-*chain on* $\mathbb{T}_{[n]}$, see Definition 2.1.1, denoted by $(T_n(m))_{m\in\mathbb{N}_0}$:

(i) Conditional on $T_n(m) = \boldsymbol{t}$, select leaf $I = i$ uniformly at random.

(ii) Let c_v denote the number of children of v, the parent of i in $\boldsymbol{t}$.

- If $c_v = 2$, let $\tilde{I} = \max\{i, a, b\}$ where

$$a = \min\left\{\begin{matrix}\text{leaf labels in the first spinal bush}\\ \text{on the ancestral line from leaf } i\end{matrix}\right\},$$
$$b = \min\left\{\begin{matrix}\text{leaf labels in the second spinal bush}\\ \text{on the ancestral line from leaf } i\end{matrix}\right\}.$$

- If $c_v > 2$, let $i_1, \ldots, i_{c_v}$ denote the smallest leaf labels in the c_v subtrees of v, enumerated in increasing order, and say that $i = i_j$ for some $j \in [c_v]$. Define the

conditional distribution of $\tilde{I}$ by

$$\mathrm{P}\left(\tilde{I}=i_{j'}\;\middle|\;T_n(m)=\boldsymbol{t},I=i\right)=\begin{cases}\frac{\alpha-\gamma}{(c_v-1)\alpha-\gamma} & \text{if } j\leq 2, j'=3\\ \frac{\alpha}{(c_v-1)\alpha-\gamma} & \text{if } j\leq 2, 3<j'\leq c_v\\ \frac{\alpha}{(c_v-1)\alpha-\gamma} & \text{if } 2<j<j'\leq c_v\\ \frac{(c_v-1-j)\alpha-\gamma}{(c_v-1)\alpha-\gamma} & \text{if } 2<j=j'\leq c_v.\end{cases}$$

(iii) Swap i and $\tilde{I}$, delete $\tilde{I}$.

(iv) Relabel $\{\tilde{I}+1,\ldots,n\}$ by $\{\tilde{I},\ldots,n-1\}$ using the increasing bijection.

(v) Perform an up-step according to the (α,γ)-growth process.

Lemma A.1. *The (α,γ)-chain is not reversible for any $0\leq\gamma\leq\alpha\leq 1$.*

Remark A.2. In passing, we note that although the above statement concerns the (α,γ)-chain of Definition 2.1.1, the proof is easily amended to cover the semi-planar (α,γ)-chain on $\widehat{\mathbb{T}}_{[n]}$ of Definition 2.4.3, as well as the decorated (α,γ)-chain of Definition 2.6.3, for any $3\leq k\leq n$.

Proof. Let $(T_n(m))_{m\in\mathbb{N}_0}$ denote the (α,γ)-chain and recall that its stationary distribution is that of T_n, the nth step of the (α,γ)-growth process. We will prove the assertion by finding a pair of trees $\boldsymbol{t},\boldsymbol{t}'\in\mathbb{T}_{[n]}$ such that $\mathrm{P}\left(T_n(1)=\boldsymbol{t}'\;\middle|\;T_n(0)=\boldsymbol{t}\right)>0$, but where

$$\begin{aligned}&\mathrm{P}\left(T_n=\boldsymbol{t}\right)\mathrm{P}\left(T_n(1)=\boldsymbol{t}'\;\middle|\;T_n(0)=\boldsymbol{t}\right)\\ \neq\;&\mathrm{P}\left(T_n=\boldsymbol{t}'\right)\mathrm{P}\left(T_n(1)=\boldsymbol{t}\;\middle|\;T_n(0)=\boldsymbol{t}'\right)\end{aligned}$$

which immediately implies non-reversibility of the (α,γ)-chain.

These pairs are not difficult to find as they arise whenever a small label chosen for deletion in the down-step, and swapped to a location "far away" in the tree. Specifically, we can take $\boldsymbol{t}$ and $\boldsymbol{t}'$ as in Figure A.1 where we observe that $\mathrm{P}\left(T_n=\boldsymbol{t}'\right)=\frac{1-\alpha}{\gamma}\mathrm{P}\left(T_n=\boldsymbol{t}\right)$. The only option to transition from $\boldsymbol{t}$ to $\boldsymbol{t}'$ is to

(i) Select leaf $i=1$ for deletion in the down-step.

(ii) Carry out the local search resulting resulting in $\tilde{\imath} = \max\{i, a, b\} = \max\{1, 2, 0\} = 2$.

(iii) Swap leaf $i = 1$ and $\tilde{\imath} = 2$, and delete leaf 2.

(iv) Relabel $\{3, \dots, n\}$ by $\{2, \dots, n-1\}$ using the increasing bijection.

(v) Insert leaf n into the root edge.

Hence

$$\mathrm{P}\left(T_n(1) = \boldsymbol{t}' \;\middle|\; T_n(0) = \boldsymbol{t}\right) = \frac{1}{n}\frac{\gamma}{n-1-\alpha}$$

but it is not possible to relocate leaf 1 back to the root edge in $\boldsymbol{t}'$ in a single step of the chain as selecting $i = 1$ leads to a swap with $\tilde{\imath} = 3$ and a selection of $i \neq 1$ never leads to a displacement of 1, and so

$$\mathrm{P}\left(T_n(1) = \boldsymbol{t} \;\middle|\; T_n(0) = \boldsymbol{t}'\right) = 0. \qquad \blacksquare$$

Faced with the above conundrum, we were left with two options: either we find ways of proving that the mixing time is of a certain order for non-reversible Markov chains, or we focus on variations of the chain that are reversible. We constrained ourselves to the latter scenario in two related ways. When studying the counter-example it is clear that the main problem of non-reversibility has to do with the label swapping — the counter-example in itself does not dictate that a version of the chain where we focus solely on the tree shapes would be non-reversible.

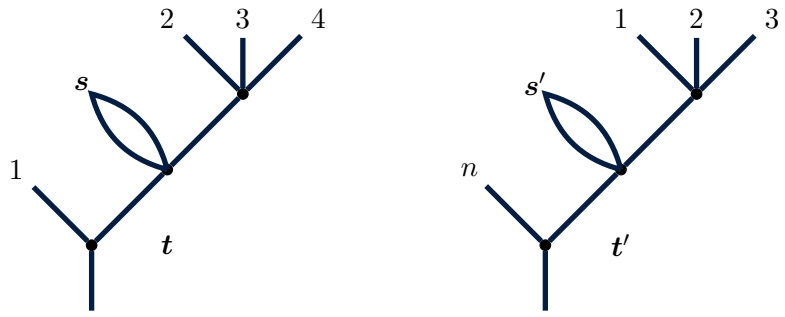


Figure A.1: Illustration of pairs of trees demonstrating non-reversibility of the (α, γ)-chain. Here, $\boldsymbol{s}' \in \mathbb{T}_{[n]\setminus\{1,2,3,n\}}$ is simply $\boldsymbol{s}$ relabelled by the increasing bijection $[n] \setminus [4] \mapsto [n-1] \setminus [3]$ and we can take any $\boldsymbol{s} \in \mathbb{T}_{[n]\setminus[4]}, n \geq 5$. In the (α, γ) we can transition from $\boldsymbol{t}$ to $\boldsymbol{t}'$ but not the converse.

Bibliography

[1] D. Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991.

[2] D. Aldous. Probability distributions on cladograms. In *Random discrete structures (Minneapolis, MN, 1993)*, volume 76 of *IMA Vol. Math. Appl.*, pages 1–18. Springer, New York, 1996.

[3] D. Aldous and J. A. Fill. Reversible Markov chains and random walks on graphs. `https://www.stat.berkeley.edu/~aldous/RWG/book.pdf`, 2002.

[4] D. J. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, 1985.

[5] D. J. Aldous. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discrete Math.*, 3(4):450–465, 1990.

[6] D. J. Aldous. Wright-Fisher diffusions with negative mutation rate! `http://www.stat.berkeley.edu/~aldous/Research/OP/fw.html`, 1999.

[7] D. J. Aldous. Mixing time for a Markov chain on cladograms. *Combin. Probab. Comput.*, 9(3):191–204, 2000.

[8] V. Anantharam and P. Tsoucas. A proof of the Markov chain tree theorem. *Statistics & Probability Letters*, 8(2):189–192, 1989.

[9] D. M. Blei, T. L. Griffiths, and M. I. Jordan. The nested Chinese restaurant process and Bayesian nonparametric inference of topic hierarchies. *J. ACM*, 57(2):Art. 7, 30, 2010.

[10] A. Borodin and G. Olshanski. Infinite-dimensional diffusions as limits of random walks on partitions. *Probab. Theory Related Fields*, 144(1-2):281–318, 2009.

[11] A. Broder. Generating random spanning trees. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, SFCS '89, page 442–447, USA, 1989. IEEE Computer Society.

[12] B. Chen, D. Ford, and M. Winkel. A new family of Markov branching trees: the alpha-gamma model. *Electron. J. Probab.*, 14:400 – 430, 2009.

[13] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991.

[14] R. Dong, C. Goldschmidt, and J. B. Martin. Coagulation-fragmentation duality, Poisson-Dirichlet distributions and random recursive trees. *Ann. Appl. Probab.*, 16 (4):1733–1750, 2006.

[15] T. Duquesne and J.-F. Le Gall. Random trees, Lévy processes and spatial branching processes. *Astérisque*, 281:vi+147.

[16] S. N. Ethier and T. G. Kurtz. *Markov processes: characterization and convergence.* Wiley series in Probability and Statistics. Wiley-Interscience, Hoboken, N.J., 2005.

[17] S. N. Evans and A. Winter. Subtree prune and regraft: a reversible real tree-valued Markov process. *Ann. Probab.*, 34(3):918–961, 2006.

[18] S. N. Evans, J. Pitman, and A. Winter. Rayleigh processes, real trees, and root growth with re-grafting. *Probab. Theory Related Fields*, 134(1):81–126, 2006.

[19] J. Felsenstein. *Inferring phylogenies.* Sinauer Associates, 2003.

[20] D. J. Ford. *Probabilities on cladograms: Introduction to the alpha model.* PhD thesis, Stanford University, 2006.

[21] N. Forman, S. Pal, D. Rizzolo, and M. Winkel. Projections of the Aldous chain on binary trees: Intertwining and consistency. arXiv:1802.00862, 2018.

[22] N. Forman, S. Pal, D. Rizzolo, and M. Winkel. Interval partition evolutions with emigration related to the Aldous diffusion. arXiv:1804.01205, 2018.

[23] N. Forman, S. Pal, D. Rizzolo, and M. Winkel. Aldous diffusion I: a projective system of continuum k-tree evolutions. arXiv:1809.07756, 2018.

[24] N. Forman, S. Pal, D. Rizzolo, and M. Winkel. Projections of the Aldous chain on binary trees: intertwining and consistency. *Random Structures & Algorithms*, 57(3): 745–769, 2020.

[25] N. Forman, S. Pal, D. Rizzolo, and M. Winkel. Diffusions on a space of interval partitions: construction from marked Lévy processes. *Electron. J. Probab.*, 25:1–46, 2020.

[26] N. Forman, S. Pal, D. Rizzolo, and M. Winkel. Diffusions on a space of interval partitions: Poisson–Dirichlet stationary distributions. *Ann. Probab.*, 49(2):793–831, 2021.

[27] J. Fulman. Stein's method and Plancherel measure of the symmetric group. *Trans. Amer. Math. Soc.*, 357(2):555–570, 2005.

[28] J. Fulman. Commutation relations and Markov chains. *Probab. Theory Related Fields*, 144(1-2):99–136, 2009.

[29] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman & Hall, London, 1996.

[30] C. Goldschmidt, B. Haas, and D. Sénizergues. Stable graphs: distributions and line-breaking construction. arXiv:1811.06940, 2020.

[31] B. Haas and G. Miermont. Scaling limits of Markov branching trees with applications to Galton-Watson and random unordered trees. *Ann. Probab.*, 40(6):2589–2666, 2012.

[32] B. Haas, G. Miermont, J. Pitman, and M. Winkel. Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *Ann. Probab.*, 36 (5):1790–1837, 2008.

[33] J. G. Kemeny and J. L. Snell. *Finite Markov chains*. The University Series in Undergraduate Mathematics. D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto-London-New York, 1960.

[34] I. Kortchemski. Invariance principles for Galton-Watson trees conditioned on the number of leaves. *Stochastic Process. Appl.*, 122(9):3126–3172, 2012.

[35] D. A. Levin and Y. Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017.

[36] W. Löhr, L. Mytnik, and A. Winter. The Aldous chain on cladograms in the diffusion limit. *Ann. Probab.*, 48(5):2565–2590, 2020.

[37] P. Marchal. A note on the fragmentation of a stable tree. In *Fifth Colloquium on Mathematics and Computer Science*, Discrete Math. Theor. Comput. Sci. Proc., AI, pages 489–499. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2008.

[38] G. Miermont. Tessellations of random maps of arbitrary genus. *Ann. Sci. Éc. Norm. Supér. (4)*, 42(5):725–781, 2009.

[39] J. Nussbaumer and A. Winter. The algebraic α-Ford tree under evolution. arXiv:2006.09316, 2020.

[40] S. Pal. Wright-Fisher diffusion with negative mutation rates. *Ann. Probab.*, 41(2): 503–526, 2013.

[41] L. A. Petrov. A two-parameter family of infinite-dimensional diffusions on the Kingman simplex. *Funktsional. Anal. i Prilozhen.*, 43(4):45–66, 2009.

[42] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006.

[43] J. Pitman and M. Winkel. Regenerative tree growth: binary self-similar continuum random trees and Poisson-Dirichlet compositions. *Ann. Probab.*, 37(5):1999–2041, 2009.

[44] J.-L. Rémy. Un procédé itératif de dénombrement d'arbres binaires et son application à leur génération aléatoire. *RAIRO Inform. Théor.*, 19(2):179–195, 1985.

[45] D. J. Rogers. *Up-down ordered Chinese Restaurant Processes: representations and asymptotics*. PhD thesis, University of Oxford, 2020.

[46] L. C. G. Rogers and J. W. Pitman. Markov functions. *Ann. Probab.*, 9(4):573–582, 1981.

[47] N. Saitou and M. Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4):406–425, 07 1987.

[48] J. Schweinsberg. An $O(n^2)$ bound for the relaxation time of a Markov chain on cladograms. *Random Structures & Algorithms*, 20(1):59–70, 2002.

[49] C. Semple and M. Steel. *Phylogenetics*, volume 24 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2003.

[50] R. Sokal and C. Michener. A statistical method for evaluating systematic relationships. *University of Kansas science bulletin*, 38:1409–1438, 1958.

[51] F. Sørensen. A down-up chain with persistent labels on multifurcating trees. arXiv:1909.02584, 2020.

[52] D. L. Swofford. Phylogeny reconstruction. *Molecular systematics*, pages 411–501, 1990.

[53] C. Whidden and I. Matsen, Frederick A. Quantifying MCMC Exploration of Phylogenetic Tree Space. *Systematic Biology*, 64(3):472–491, 01 2015.

[54] I. J. Wilson, M. E. Weale, and D. J. Balding. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J. Roy. Statist. Soc. Ser. A*, 166(2):155–201, 2003.