Invariance principles for spatial multitype Galton-Watson trees

Grégory Miermont*

November 7, 2007

Abstract. We prove that critical multitype Galton-Watson trees converge after rescaling to the Brownian continuum random tree, under the hypothesis that the offspring distribution is irreducible and has finite covariance matrices. Our study relies on an ancestral decomposition for marked multitype trees, and an induction on the number of types. We then couple the genealogical structure with a spatial motion, whose step distribution may depend on the structure of the tree in a local way, and show that the resulting discrete spatial trees converge once suitably rescaled to the Brownian snake, under some moment assumptions.

Key Words: Multitype Galton-Watson tree, discrete snake, invariance principle, Brownian tree, Brownian snake

Résumé. Nous montrons que les arbres de Galton-Watson multitypes, dont les lois de reproduction sont irréductibles et de matrices de covariance finies, admettent pour limite d'échelle l'arbre continu brownien. La clef de notre étude est une décomposition ancestrale pour les arbres multitypes marqués, et une méthode par récurrence sur le nombre de types. Nous couplons ensuite la structure généalogique avec des déplacements spaciaux, dont la loi de saut peut dépendre localement de la structure de l'arbre, et nous montrons que les arbres spatiaux obtenus convergent vers le serpent brownien, sous certaines hypothèses de moments.

Mots-clef: Arbre de Galton-Watson multitype, serpent discret, principe d'invariance, arbre brownien, serpent brownien.

M.S.C. 60J80, 60F17

*CNRS & Laboratoire de Mathématique, Équipe Probabilités, Statistique et Modélisation, Bât. 425, Université Paris-Sud, 91405 Orsay, France. Gregory.Miermont@math.u-psud.fr

1 Introduction and main results

1.1 Motivation

Multitype Galton-Watson (GW) processes arise as a natural generalization of usual GW processes, in which individuals are differentiated by *types* that determine their offspring distribution. They were first studied in 1947 by Kolmogorov and his coauthors. We refer to [7, Chapters 2 & 3] for a very nice introduction to these processes. It turns out that their analysis is considerably eased under an *irreducibility* assumption, namely, that every type has a positive probability to eventually 'lead' to all others. Under this hypothesis, one can use variants of the Perron-Frobenius theorem, which allow to quantify the asymptotic behavior of iterates of the *mean operator* of [7], and obtain qualitative and quantitative results on the GW process. Informally, the large-scale aspects of irreducible multitype GW processes are similar to that of monotype GW processes whose mean offspring distribution is the Perron eigenvalue of the mean operator. On the other hand, Janson [9] has shown that if the irreducibility assumption fails to hold, many different behaviors can occur.

The aim of the present paper is to investigate the ancestor trees and forests associated with irreducible GW processes, when the total number of types is finite. Under criticality hypotheses on the mean matrix, and a finiteness hypothesis on the covariance matrices of the offspring distributions, we show that the height process of these forests converges to a reflected Brownian motion, and hence behaves asymptotically in a similar way as monotype processes (Theorem 1). Similar results are proved for family trees conditioned on the number of their individuals, under extra exponential moments assumptions (Theorem 2). Although we do not focus on a continuum tree formalism here, this says roughly that under these hypotheses, multitype GW trees conditioned to have n individuals converge once suitably renormalized to the Brownian continuum random tree [1].

We also show that these ancestor trees and forests, when coupled with a spatial motion whose step distribution may locally depend on the tree structure, converge to the so-called Brownian snake [12], under mild extra hypotheses on the spatial motion step distribution (Theorem 3).

Applications of these results to random planar maps are discussed in the companion paper [14]. The approach of the present paper follows quite closely that of [13].

1.2 Multitype Galton-Watson processes

Let $K \in \mathbb{N} = \{1, 2, \dots\}$ be a positive integer. Write $[K] = \{1, 2, \dots, K\}$, and identify $\mathbb{Z}^{[K]}, \mathbb{R}^{[K]}$ with $\mathbb{Z}^K, \mathbb{R}^K$. Suppose given distributions $(\mu^{(i)}, i \in [K])$ on the space $\mathbb{Z}_+^K$ of integer-valued non-negative sequences of length K. We will often use the notation $\boldsymbol{\mu}$ as a shorthand for the $\mathbb{R}^K$-valued measure $(\mu^{(i)}, i \in [K])$.

A K-multitype GW process with offspring distributions $\boldsymbol{\mu}$ is a $\mathbb{Z}_+^K$-valued Markov process

$$\mathbf{Z}_n = (Z_n(i), i \in [K]), \qquad n \geq 0$$

such that the law of $\mathbf{Z}_{n+1}$ given $\mathbf{Z}_n = \mathbf{z} = (z_i, i \in [K])$ is the same as $\sum_{i\in[K]} \sum_{l=1}^{z_i} \mathbf{X}^{(i)}(l)$, where the vectors $\mathbf{X}^{(i)}(l), i \in [K], l \geq 1$ are all independent, and $\mathbf{X}^{(i)}(l)$ has law $\mu^{(i)}$ for all $l \geq 1$.

Otherwise said, this process can be considered as a model for population evolution where each individual is given a type in $[K]$, and where each type-i individual gives birth to a set of individuals with law $\mu^{(i)}$, this independently over individuals (although the different components of $\mathbf{X}^{(i)}$ may well be dependent).

We say that the process (or the measure $\boldsymbol{\mu}$) is *non-degenerate* if there exists at least one $i \in [K]$ so that $\mu^{(i)}(\{\mathbf{z} : \sum_j z_j \neq 1\}) > 0$. Failure of this last assumption entails that all particles a.s. give birth to exactly one particle, and the study of the process boils down to that of a Markov chain with values in $[K]$. All the processes that we consider here are assumed to be non-degenerate.

For $i, j \in [K]$, let

$$m_{ij} = \sum_{\mathbf{z}\in\mathbb{Z}_+^K} z_j \mu^{(i)}(\{\mathbf{z}\})$$

be the mean number of type-j offspring of an type-i individual. We let $M_{\boldsymbol{\mu}} = (m_{ij})_{i,j\in[K]}$ and call it the *mean matrix* of $\boldsymbol{\mu}$.

Definition 1 *The mean matrix (or the offspring distribution $\boldsymbol{\mu}$) is called* irreducible, *if for every $i, j \in [K]$, there is some $n \in \mathbb{N}$ so that $m_{ij}^{(n)} > 0$, where $m_{ij}^{(n)}$ is the ij-entry of the matrix $M_{\boldsymbol{\mu}}^n$.*

Notice that the n in the definition may depend on the choice of i, j, so that the definition is distinct of that of *aperiodicity*, namely that the above property holds jointly for every i, j, for some common n.

With the irreducibility assumption, the Perron-Frobenius Theorem recalled below (Proposition 3) ensures that the eigenvalue ϱ of $M_{\boldsymbol{\mu}}$ with maximal modulus is real, positive, and simple, and that a non-zero eigenvector of $M_{\boldsymbol{\mu}}$ with eigenvalue ϱ has only non-zero entries all having the same sign.

Proposition 1 ([2]) *Suppose that the process $(\mathbf{Z}_n, n \geq 0)$ is non-degenerate. Then it eventually becomes extinct a.s., whatever the starting value, if and only if $\varrho \leq 1$. The process (or the distribution $\boldsymbol{\mu}$) is called sub-critical if $\varrho < 1$, critical if $\varrho = 1$ and supercritical if $\varrho > 1$.*

It will be useful to introduce the generating functions $\boldsymbol{\varphi} = (\varphi^{(i)}, i \in [K])$ defined by

$$\varphi^{(i)}(\mathbf{s}) = \sum_{\mathbf{z}\in\mathbb{Z}_+^K} \mu^{(i)}(\{\mathbf{z}\})\mathbf{s}^{\mathbf{z}}$$

where $\mathbf{s} = (s_i, i \in [K]) \in [0, 1]^K$ and $\mathbf{s}^{\mathbf{z}} = \prod_{i\in[K]} s_i^{z_i}$. With these notations, we have $m_{ij} = \partial\varphi^{(i)}/\partial s_j(\mathbf{1})$, where $\mathbf{1}$ is the vector of $[0, 1]^K$ with all components equal to 1.

For $1 \leq i, j, k \leq K$, define

$$Q_{jk}^{(i)} = \frac{\partial^2 \varphi^{(i)}}{\partial s_j \partial s_k}(\mathbf{1}).$$

We say that $\boldsymbol{\mu}$ has *finite variance* if

$$Q_{jk}^{(i)} < \infty \qquad \text{for all} \qquad i, j, k \in [K]\,. \tag{1}$$

Under this assumption, for each i, $(Q_{jk}^{(i)}, 1 \leq j, k \leq K)$ is the Hessian matrix of the convex function $\varphi^{(i)}$ evaluated at $\mathbf{1}$, hence the matrix of a non-negative quadratic form on $\mathbb{R}^K$, which we call $Q^{(i)}(\mathbf{s}), \mathbf{s} \in \mathbb{R}^K$.

Assuming $M_{\boldsymbol{\mu}}$ irreducible and $\boldsymbol{\mu}$ non-degenerate, critical and with finite variance, we let $\mathbf{a}, \mathbf{b}$ be the left and right eigenvectors of $M_{\boldsymbol{\mu}}$ with eigenvalue 1, chosen so that $\mathbf{a} \cdot \mathbf{1} = 1$ and $\mathbf{a} \cdot \mathbf{b} = 1$, where $x \cdot y$ is the scalar product of the vectors $x, y \in \mathbb{R}^K$. Let

$$\sigma = \sqrt{\sum_{i=1}^{K} a_i Q^{(i)}(\mathbf{b})} = \sqrt{\mathbf{a} \cdot \mathbf{Q}(\mathbf{b})}, \tag{2}$$

where $\mathbf{Q}(\mathbf{s})$ is the K-dimensional vector $(Q^{(i)}(\mathbf{s}), 1 \leq i \leq K)$. This should be interpreted as the 'variance' of the offspring distribution of the multitype process, as it plays a role similar to the variance for monotype GW processes in the asymptotics of the survival probability, see [19].

Basic assumption (H). In the sequel, unless specified otherwise, we will exclusively be concerned with irreducible, non-degenerate, critical offspring distributions with finite variance. Notice that criticality implies finiteness of all coefficients of $M_{\boldsymbol{\mu}}$.

1.3 Multitype trees and forests

We now add a genealogical structure to the branching processes, by endowing it into a tree-valued random variable. For $n \geq 0$ let $\mathcal{U}$ be the infinite-regular tree

$$\mathcal{U} = \bigsqcup_{n \geq 0} \mathbb{N}^n,$$

where $\mathbb{N}^n$ is the set of words with n letters, and by convention $\mathbb{N}^0 = \{\varnothing\}$. For two words u, v, we let uv be their concatenation and $|u|, |v|$ their length (with the convention $|\varnothing| = 0$). If u is a word and $A \subseteq \mathcal{U}$ we let $uA = \{uv : v \in A\}$, and say that u is a prefix of v if $v \in u\mathcal{U}$, in which case we write $u \vdash v$. A *planar tree* is a finite subset $\mathbf{t}$ of $\mathcal{U}$ such that

- $\varnothing \in \mathbf{t}$, it is called the *root* of $\mathbf{t}$,
- for every $u \in \mathcal{U}$ and $i \in \mathbb{N}$, if $ui \in \mathbf{t}$ then $u \in \mathbf{t}$, and $uj \in \mathbf{t}$ for every $1 \leq j \leq i$.

We let $\mathcal{T}$ be the set of planar trees, which we simply refer to as *trees* in the sequel. For a tree $\mathbf{t} \in \mathcal{T}$ and $u \in \mathbf{t}$, the number $c_{\mathbf{t}}(u) = \max\{i \in \mathbb{Z}_+ : ui \in \mathbf{t}\}$, with the convention $u0 = u$, is the number of children of u. We say that v is an ancestor of u if $v \vdash u$. An element $u \in \mathbf{t}$ is called a *vertex* of $\mathbf{t}$, and the length $|u|$ of the word u is called the *height* of u in $\mathbf{t}$. The vertices of $\mathbf{t}$ with no children are called *leaves*. For $\mathbf{t}$ a planar tree and u a vertex of $\mathbf{t}$, we let $\mathbf{t}_u = \{v \in \mathcal{U} : uv \in \mathbf{t}\}$, and we call it the *fringe subtree* rooted at u (it is trivially checked that it is indeed a tree). The 'remaining part' $[\mathbf{t}]_u = \{u\} \cup (\mathbf{t} \setminus u\mathbf{t}_u)$ is called the *subtree of* $\mathbf{t}$ *pruned at* u. Any planar tree $\mathbf{t}$ is endowed with the linear order which is the restriction to $\mathbf{t}$ of the usual lexicographical order $\prec$ on $\mathcal{U}$ ($u \prec v$ if $u \vdash v$ or if $u = wu', v = wv'$ where $u'_1 < v'_1$). We call it the *depth-first order*.

In addition to trees, we will consider forests, which are defined as nonempty subsets of $\mathcal{U}$ of the form

$$\mathbf{f} = \bigcup_k k\mathbf{t}_{(k)},$$

where $(\mathbf{t}_{(k)})$ is a finite or infinite sequence of trees, which are called the *components* of $\mathbf{f}$. We let $\mathcal{F}$ be the set of forests. The quantity $c_{\mathbf{f}}(u)$ (number of children of $u \in \mathbf{f}$) is defined as for trees, and we let $c_{\mathbf{f}}(\varnothing) \in \mathbb{N} \sqcup \{\infty\}$ be the number of tree components of $\mathbf{f}$. We define $\mathbf{f}_u = \{v : uv \in \mathbf{f}\} \in \mathcal{T}$ if $u \in \mathbf{f}$, and $\mathbf{f}_u = \varnothing$ otherwise, so in particular $\mathbf{f}_k, 1 \le k \le c_{\mathbf{f}}(\varnothing)$ are the tree components of $\mathbf{f}$. Also, let $[\mathbf{f}]_u = \{u\} \cup \mathbf{f} \setminus u\mathbf{f}_u \in \mathcal{F}$. If $u \in \mathbf{f}$ is a vertex of the forest, we call $|u| - 1$ the height of u. Notice that the notion of height of a vertex relies on whether we are considering the vertex to belong to a tree or a forest, the reason being that we want the roots $1, 2, 3, \ldots$ (or *floor*) of the forest to be at height 0. There should be no ambiguity according to the context.

A K*-type planar tree*, or simply a multitype tree if the number K is clear from the context, is a pair $(\mathbf{t}, e_{\mathbf{t}})$ where $\mathbf{t} \in \mathcal{T}$ and $e_{\mathbf{t}} : \mathbf{t} \to [K]$. We let $\mathcal{T}^{(K)}$ be the set of K-type trees. For $u \in \mathbf{t}$, $e_{\mathbf{t}}(u)$ is called the type of u. If $(\mathbf{t}, e_{\mathbf{t}}) \in \mathcal{T}^{(K)}$ and for $u \in \mathbf{t}$, we let $(\mathbf{t}, e_{\mathbf{t}})_u \in \mathcal{T}^{(K)}$ be the pair $(\mathbf{t}_u, e_{\mathbf{t}_u})$ where $e_{\mathbf{t}_u}(v) = e_{\mathbf{t}}(uv)$. Similarly, $[(\mathbf{t}, e_{\mathbf{t}})]_u$ is the tree $[\mathbf{t}]_u$ marked by the restriction of $e_{\mathbf{t}}$ to $[\mathbf{t}]_u$. Similar definitions hold in a straightforward way for marked and K-type forests $(\mathbf{f}, e_{\mathbf{f}})$, whose set is denoted by $\mathcal{F}^{(K)}$.

In the sequel, we will often denote the marking functions $e_{\mathbf{t}}, e_{\mathbf{f}}$ by e when it is free of ambiguity, and will even denote elements of $\mathcal{T}^{(K)}, \mathcal{F}^{(K)}$ by $\mathbf{t}$ or $\mathbf{f}$, i.e. without explicitly mentioning e. It will be understood then that $\mathbf{t}_u, \mathbf{f}_u, \ldots$ are marked with the appropriate function. We let, for $i \in [K]$,

$$\mathcal{T}_i^{(K)} = \left\{\mathbf{t} \in \mathcal{T}^{(K)} : e(\varnothing) = i\right\},$$

and for $\mathbf{x} = (x_j)$ a finite or infinite sequence with terms in $[K]$,

$$\mathcal{F}_{\mathbf{x}}^{(K)} = \{\mathbf{f} \in \mathcal{F}^{(K)} : e(j) = x_j \quad \forall j\}.$$

For $\mathbf{t} \in \mathcal{T}^{(K)}$ and $i \in [K]$, we let $\mathbf{t}^{(i)} = \{u \in \mathbf{t} : e(u) = i\}$, and $\mathbf{f}^{(i)}$ is the corresponding notation for $\mathbf{f} \in \mathcal{F}^{(K)}$.

Let $\mathcal{W}_K = \bigsqcup_{n \ge 0} [K]^n$ be the set of finite, possibly empty $[K]$-valued sequences, and consider the natural projection $p : \mathcal{W}_K \to \mathbb{Z}_+^K$, where $p(\mathbf{w}) = (p_i(\mathbf{w}), i \in [K])$ and

$p_i(\mathbf{w}) = \#\{j : \mathbf{w}_j = i\}$ counts the number of elements of $\mathbf{w}$ equal to i. Notice that for every K-type tree $\mathbf{t}$, any $u \in \mathbf{t}$ determines a sequence $\mathbf{w_t}(u) = (e(ui), 1 \leq i \leq c_{\mathbf{t}}(u)) \in \mathcal{W}_K$ with length $|\mathbf{w_t}(u)| = c_{\mathbf{t}}(u)$. The vector $p(\mathbf{w_t}(u))$ counts the number of children of u of each type.

Finally, we will frequently have to count the number of ancestors of some vertex of a tree or a forest, that satisfy certain specific properties. For $\mathbf{t} \in \mathcal{T}^{(K)}$ and $u \in \mathbf{t}$, we let $\mathrm{Anc}^u_{\mathbf{t}}(\mathsf{P}(v))$ be the number of ancestors $v \vdash u$ that satisfy the property P. For instance, $\mathrm{Anc}^u_{\mathbf{t}}(e(v) = i, c_{\mathbf{t}}(v) = k)$ counts the number of ancestors of u with type i and k children.

1.4 Galton-Watson trees

Let $\boldsymbol{\zeta} = (\zeta^{(i)}, i \in [K])$ be a family of probability measures on the set $\mathcal{W}_K$. We call $\boldsymbol{\zeta}$ an *ordered* offspring distribution. It is said to be non-degenerate (resp. critical, resp. to have finite variance) if the family of measures $\boldsymbol{\mu} = p_*\boldsymbol{\zeta} := (p_*\zeta^{(i)}, i \in [K])$ is non-degenerate (resp. critical, resp. satisfies (1)), where $p_*\zeta^{(i)}$ is the push-forward of $\zeta^{(i)}$ by p.

For $i \in [K]$, we now construct a distribution on $\mathcal{T}^{(K)}_i$ such that

- different vertices have independent offspring, and
- type-j vertices have a set of children with types given by a sequence $\mathbf{w} \in \mathcal{W}_K$ with probability $\zeta^{(j)}(\mathbf{w})$.

To do this, let $(\mathbf{W}^i_u = (W^i_u(l), 1 \leq l \leq |\mathbf{W}^i_u|), 1 \leq i \leq K, u \in \mathcal{U})$ be a $[K] \times \mathcal{U}$-indexed family of independent random variables such that $\mathbf{W}^i_u$ has law $\zeta^{(i)}$. Then recursively, construct a subset $\mathbf{t} \subset \mathcal{U}$ together with a mark $e : \mathbf{t} \to [K]$ by letting $\varnothing \in \mathbf{t}, e(\varnothing) = i$, and if $u \in \mathbf{t}, e(u) = j$, then $ul \in \mathbf{t}$ if and only if $1 \leq l \leq |\mathbf{W}^j_u|$, and then $e(ul) = W^j_u(l)$.

It is straightforward to check that $\mathbf{t}$ has the properties of a planar tree, except that it might be infinite. Moreover, it is straightforward from the construction that the process

$$\mathbf{Z}_n(\mathbf{t}) = (\# \{u \in \mathbf{t} : |u| = n, e(u) = i\}, i \in [K]) , \qquad n \geq 0 ,$$

is a multitype GW process with offspring distribution $\boldsymbol{\mu} = p_*\boldsymbol{\zeta}$, and started from a single type-i individual. In particular, under the criticality assumption $\varrho = 1$, this process becomes extinct a.s., so that $\mathbf{t}$ is finite a.s. and hence is a tree a.s.. In this case, we let $P^{(i)}_{\boldsymbol{\mu}}$, or simply $P^{(i)}$, be the law of $\mathbf{t}$ on $\mathcal{T}^{(K)}_i$. The probability measure $P^{(i)}$ is entirely characterized by the formulae

$$P^{(i)}(T = \mathbf{t}) = \prod_{u \in \mathbf{t}} \zeta^{(e(u))}(\mathbf{w_t}(u)),$$

where $T : \mathcal{T}^{(K)} \to \mathcal{T}^{(K)}$ is the identity map and $\mathbf{t}$ ranges over finite K-type trees.

Similarly, if $\mathbf{x} = (x_j, 1 \leq j \leq r) \in \mathcal{W}_K$, we define $P^{\mathbf{x}}$ as the image measure of $\bigotimes_{i=1}^r P^{(x_i)}$ by

$$(\mathbf{t}_{(1)}, \ldots, \mathbf{t}_{(k)}) \longmapsto \bigcup_{1 \leq k \leq r} k\mathbf{t}_{(k)},$$

i.e., it is the law that makes the identity map $F : \mathcal{F}^{(K)} \to \mathcal{F}^{(K)}$ the random forest whose tree components $F_i, 1 \leq j \leq r$ are independent with respective laws $P^{(x_i)}$. A similar definition holds for an infinite sequence $\mathbf{x} \in [K]^{\mathbb{N}}$. It will be convenient to use the notation $P^{\mathbf{x}}_r$ for $P^{(x_1,\dots,x_r)}$ when $\mathbf{x} \in [K]^{\mathbb{N}}$.

1.5 Convergence of height processes

For $\mathbf{t} \in \mathcal{T}$, we let $\varnothing = u_{\mathbf{t}}(0) \prec u_{\mathbf{t}}(1) \prec \ldots \prec u_{\mathbf{t}}(\#\mathbf{t}-1)$ be the list of vertices of $\mathbf{t}$ in depth-first order. When there is no ambiguity on $\mathbf{t}$, we simply denote them by $u(0), u(1), \ldots$. Let $(H^{\mathbf{t}}_n = |u(n)|, n \geq 0)$ be the *height process* of $\mathbf{t}$, with the convention that $|u(n)| = 0$ for $n \geq \#\mathbf{t}$. If $\mathbf{f} \in \mathcal{F}$, we similarly let $u_{\mathbf{f}}(0) \prec u_{\mathbf{f}}(1) \prec \ldots$, or simply $u(0), u(1), \ldots$ be the depth-first ordered list of its vertices ($u(0) = 1$), and define $(H^{\mathbf{f}}_n, n \geq 0)$ by $H^{\mathbf{f}}_n = (|u(n)| - 1)\mathbb{1}_{\{n \leq \#\mathbf{f}-1\}}$ (again, the convention differs because we want the floor $1, 2, \ldots$ of $\mathbf{f}$ to be at height 0). Also, for $n \geq 0$ and $\mathbf{f} \in \mathcal{F}$, let $\Upsilon^{\mathbf{f}}_n$ be the first letter of $u(n)$ with the convention that for $n \geq \#\mathbf{f}$, it equals the number of components of $\mathbf{f}$.

For $(\mathbf{t}, e) \in \mathcal{T}^{(K)}$ and $i \in [K]$, we let

$$\Lambda^{\mathbf{t}}_i(n) = \#\{0 \leq k \leq n : e(u(k)) = i\}$$

be the number of type-i individuals standing before the $n+1$-th individual in depth-first order. The quantity $\Lambda^{\mathbf{f}}_i$ is defined similarly for $(\mathbf{f}, e) \in \mathcal{F}^{(K)}$.

Theorem 1 *Let $\boldsymbol{\zeta}$ be an ordered offspring distribution such that $\boldsymbol{\mu} = p_*\boldsymbol{\zeta}$ satisfies* (H). *Recall the notations $\mathbf{a}, \mathbf{b}, \sigma$ around (2). Then*
(i) *Under $P^{\mathbf{x}}$, for some arbitrary $\mathbf{x} \in [K]^{\mathbb{N}}$, the following convergence in distribution holds for the Skorokhod topology on the space $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ of right-continuous functions with left limits:*

$$\left(\frac{H^F_{\lfloor ns \rfloor}}{\sqrt{n}}, s \geq 0\right) \xrightarrow[n\to\infty]{d} \left(\frac{2}{\sigma}|B_s|, s \geq 0\right),$$

where B is a standard one-dimensional Brownian motion.
(ii) *For every $i \in [K]$, if $\mathbf{i}$ is the constant sequence $(i, i, \ldots)$, then, under $P^{\mathbf{i}}$, the following convergence in distribution in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ holds jointly with that of* (i)*:*

$$\left(\frac{\Upsilon^F_{\lfloor ns \rfloor}}{\sqrt{n}}, s \geq 0\right) \xrightarrow[n\to\infty]{d} \left(\frac{\sigma}{b_i}L^0_s, s \geq 0\right),$$

where $(L^0_t, t \geq 0)$ is the local time of B at level 0, normalized as the density of the occupation measure of B at 0 before time t.
(iii) *Moreover, for any $\mathbf{x}$,*

$$\left(\frac{\Lambda^F_i(\lfloor ns \rfloor)}{n}, s \geq 0\right) \xrightarrow[n\to\infty]{P^{\mathbf{x}}} (a_i s, s \geq 0), \qquad i \in [K].$$

Here, the convergence is convergence in probability under $P^{\mathbf{x}}$, for the topology of uniform convergence over compact subsets of $\mathbb{R}_+$.

Note that Theorem 1 could be also stated purely within a tree formalism, without reference to height processes. In a rough way, Theorem 1 says that multitype GW forests converge once properly rescaled to a random forest, for a certain topology on the set of tree-like metric spaces. This limiting forest is made of tree components that are described by a Poisson process whose intensity measure is the σ-finite Brownian continuum tree measure Θ of [5].

Let us comment on this result. First, (i) says that the height process of a multitype forest, when properly rescaled, always looks the same as a reflected Brownian motion with some prescribed scale factor, whatever the roots are. Moreover, the scale factor depends only on $\boldsymbol{\mu} = p_*\boldsymbol{\zeta}$, meaning that the exact way in which each set of children is ordered does not affect the asymptotic distributional shape of the forest.

However, the value of $\mathbf{x}$ actually matters if one wants to get a closer look at the 'limiting forest'. Indeed, (ii) says that if one wants to extract every single tree of a forest grown from a type-i floor, one should proceed at a certain speed which does depend on i. In particular, in a general mixed floor as in (i), when taking a tree component in the limiting forest, one is in general unable to recover the rank of the tree it comes from in the discrete picture.

Last, (iii) implies that

$$n^{-1} d\Lambda_i^F(\lfloor ns \rfloor)\mathbb{1}_{\{s\geq 0\}} \xrightarrow[n\to\infty]{P^{\mathbf{x}}} a_i \,\mathrm{d}s \mathbb{1}_{\{s\geq 0\}}\,, \qquad 1 \leq i \leq K\,,$$

for the topology of vague convergence of measures, hence showing that $\mathbf{a}$ provides the asymptotic relative weights of different types, which are not influenced by the types of the roots. This is known as the *convergence of types theorem*, see [2], and it is proved by different methods in Proposition 6 below. Theorem 1 (iii) gives the extra information that all types are homogeneously distributed in the limiting tree.

We mention that the topology for the weak convergence of (i) and (ii) could simply be the the uniform topology over compact subsets, since all limits are continuous.

We also obtain as a corollary the following theorem of [19], in the more general case of irreducible mean matrix ([19] is in the aperiodic case).

Corollary 1 *Let* $\mathrm{ht}(\mathbf{t})$ *be the maximal height of a vertex in* $\mathbf{t}$*. Under the same assumptions, as* $n \to \infty$*, we have*

$$nP^{(i)}(\mathrm{ht}(T) \geq n) \underset{n\to\infty}{\longrightarrow} \frac{2b_i}{\mathbf{a}\cdot\mathbf{Q}(\mathbf{b})}.$$

Conditioned versions of Theorem 1 also hold. We say that $\boldsymbol{\mu}$ (or $\boldsymbol{\zeta}$) has small exponential moments if there exists $\varepsilon > 0$ such that

$$\sup_{i\in[K]} \sum_{\mathbf{z}\in\mathbb{Z}_+^K} \exp(\varepsilon|\mathbf{z}|_1)\mu^{(i)}(\mathbf{z}) = \sup_{i\in[K]} \sum_{\mathbf{w}\in\mathcal{W}_K} \exp(\varepsilon|\mathbf{w}|)\zeta^{(i)}(\mathbf{w}) < \infty\,, \qquad (3)$$

where $|\mathbf{z}|_1 = z_1 + \ldots + z_K$. In the following statement, as well as in all statements in the paper involving conditioned laws, we make the assumption that n goes to infinity along some subsequence, so that all the conditioning events that are considered have positive probabilities.

Theorem 2 *Assume that hypothesis* (H) *holds and that* $\boldsymbol{\mu}$ *has small exponential moments. Then for every* $i,j\in[K]$*, the following convergence in distribution holds on* $\mathbb{D}([0,1],\mathbb{R})$*:*

$$\left(\frac{H^T_{\lfloor \#Tt\rfloor}}{n^{1/2}},0\leq t\leq 1\right)\qquad \textit{under}\qquad P^{(i)}(\cdot|\#T^{(j)}=n)\xrightarrow[n\to\infty]{d}\left(\frac{2}{\sigma\sqrt{a_j}}B^{\mathrm{ex}}_t,0\leq t\leq 1\right),$$

where B^{ex} *is the standard Brownian excursion with duration* 1.

Moreover, $((\#T)^{-1}\Lambda^T_k(\lfloor \#Tt\rfloor),0\leq t\leq 1)$ *under* $P^{(i)}(\cdot|\#T^{(j)}=n)$ *converges in probability to* $(a_k t,0\leq t\leq 1)$ *for the uniform norm as* $n\to\infty$*, for every* $i,j,k\in[K]$.

The main ingredient used to prove Theorems 1 and 2 is the following result:

Proposition 2 (Corollary 2.5.1 in [4], Theorem 3.1 in [3]) *Theorems 1 and 2 are true in the case* $K=1$.

To be completely accurate, Theorem 2.5.1 in [4] gives the convergence in distribution of $n^{-1/2}(V^F_{\lfloor ns\rfloor},H^F_{\lfloor ns\rfloor},s\geq 0)$ to $(\sigma B_s,2(B_s-I_s)/\sigma,s\geq 0)$, where B is a standard Brownian motion with infimum process $I_s=\inf_{0\leq u\leq s}B_u$, and where V^F is the so-called Łukaciewicz walk of F (see the proof of Proposition 8 for the definition), whose only property we need at this point is that $\inf_{0\leq k\leq n}V^F_k=-\Upsilon^F_n$. This entails that $n^{-1/2}(\Upsilon^F_{\lfloor ns\rfloor},H^F_{\lfloor ns\rfloor},s\geq 0)$ converges in distribution to $(-\sigma I_s,2(B_s-I_s)/\sigma,s\geq 0)$, which by Lévy's theorem has same distribution as $(\sigma L^0_s,2|B_s|/\sigma,s\geq 0)$.

Also, Duquesne [3] makes the assumption that the offspring distribution is aperiodic to avoid conditioning events of zero probability, but the proofs still work by considering subsequences as we do. The conditioned version of Proposition 2 was first stated in Aldous [1], using the so-called *contour* process, rather than the height process, to encode the discrete trees. We stress that in [1, 3], the authors do not assume that the offspring distribution has small exponential moments, and we expect Theorem 2 to hold without this extra hypothesis. As a matter of fact, by making occasional changes in the proofs below, one can show that a sixth moment for $\boldsymbol{\mu}$ is sufficient, and we suspect that a second moment condition is enough.

We finally stress that Proposition 2 is in fact a particular case of Duquesne and Le Gall's results [4, 3], which deal with the case of offspring distributions belonging to other stable domains of attraction than the Gaussian one.

The idea of the proof of Theorems 1, 2 will be to use an inductive argument on K in order to apply the foregoing proposition. We rely strongly on an ancestral decomposition for multitype trees adapted from Kurtz *et al.* [11] and to be developed in Section 2.1, in which we also state some facts about monotype trees and the Perron-Frobenius theorem. The proof of Theorem 1 is then given in Sections 2.5 and 2.6.

1.6 Convergence of multitype snakes to the Brownian snake

Let us now couple the multitype branching process with a spatial motion. As in [13], we are interested in the case where the motion, which is parametrized by the vertices of the tree, has a step distribution around a given vertex that may depend locally on the tree, through the type and children of the vertex. We stress that our method could most likely be applied to other kinds of step distributions that depend on the structures of neighborhoods of the vertices which are not too large (i.e. have uniformly negligible scale compared to the height of the large trees).

Consider a family of probability distributions $\nu_{i,\mathbf{w}}$ respectively on $\mathbb{R}^{|\mathbf{w}|}$, and indexed by types $i \in [K]$ and $\mathbf{w} \in \mathcal{W}_K$. For $(\mathbf{t}, e)$ a multitype tree and every $u \in \mathbf{t}$ with type $e(u) = i$ and children vector $\mathbf{w}_\mathbf{t}(u) = \mathbf{w}$, we take a random variable $(Y_{uj}, 1 \leq j \leq |\mathbf{w}|)$ with law $\nu_{i,\mathbf{w}}$, independently over distinct vertices. Let $\nu_\mathbf{t}$ be the law of the random vector $(Y_u, u \in \mathbf{t})$ thus obtained, with the convention that $Y_\varnothing = 0$. We let

$$\mathbb{T}^{(K)} = \{(\mathbf{t}, e, (y_u, u \in \mathbf{t})) : (\mathbf{t}, e) \in \mathcal{T}^{(K)}, \mathbf{y} \in \mathbb{R}^\mathbf{t}\}$$

be the set of multitype trees with *spatial marks* on the vertices, and let $\mathbb{P}^{(i)}$ be the probability measure $\mathrm{d}P^{(i)}(\mathbf{t}, e) \otimes \mathrm{d}\nu_\mathbf{t}(\mathbf{y})$.

Similarly, for $(\mathbf{f}, e)$ a multitype forest we let $\nu_\mathbf{f}$ be the law of the random variable $(Y_u, u \in \mathbf{f})$, where $Y_n = 0$ for any $n \in \mathbb{N}$ and the random vectors $(Y_{ui} : 1 \leq i \leq c_\mathbf{f}(u))$ for $u \in \mathbf{f}$ are independent with respective laws $\nu_{e(u),\mathbf{w}_\mathbf{f}(u)}$. For $\mathbf{x} = (x_1, x_2, \ldots)$ a finite or infinite sequence of types we let $\mathbb{P}^\mathbf{x} = \mathrm{d}P^\mathbf{x}((\mathbf{f}, e)) \otimes \mathrm{d}\nu_\mathbf{f}(\mathbf{y})$, which is a probability distribution on the set

$$\mathbb{F}^{(K)} = \{(\mathbf{t}, e, (y_u, u \in \mathbf{t})) : (\mathbf{t}, e) \in \mathcal{T}^{(K)}, \mathbf{y} \in \mathbb{R}^\mathbf{t}\}.$$

Notice that in the definitions of $\mathbb{P}^{(i)}$ and $\mathbb{P}^\mathbf{x}$, only the measures $\nu_{i,\mathbf{w}}$ for $(i, \mathbf{w})$ such that $\zeta^{(i)}(\mathbf{w}) > 0$ matter. By convention, we let $\nu_{i,\mathbf{w}}$ be the Dirac mass at $0 \in \mathbb{R}^{|\mathbf{w}|}$ for all irrelevant indices. We say that the family $(\nu_{i,\mathbf{w}}, i \in [K], \mathbf{w} \in \mathcal{W}_K)$ is non-degenerate if $\nu_{i,\mathbf{w}}$ is not a Dirac mass for at least one of the relevant indices $(i, \mathbf{w})$, so that there is 'some randomness' in the spatial displacement. We say that $\boldsymbol{\nu} = (\nu_{i,\mathbf{w}}, i \in [K], \mathbf{w} \in \mathcal{W}_K)$ is centered if all distributions $\nu_{i,\mathbf{w}}$ are.

For $(\mathbf{t}, e, \mathbf{y}) \in \mathbb{T}^{(K)}$, define

$$S_u^{\mathbf{t},e,\mathbf{y}} = \sum_{v \vdash u} y_v,$$

and let $S_k^\mathbf{t}$ stand (a little improperly, but for lighter notations) for $S_{u(k)}^{\mathbf{t},e,\mathbf{y}}$, with the convention that it equals 0 for $k \geq \#\mathbf{t}$. A similar definition holds for $S^\mathbf{f}$ where $(\mathbf{f}, e, \mathbf{y}) \in \mathbb{F}^{(K)}$, where we use the convention $y_\varnothing = 0$. In the following statement and in the sequel, $|\cdot|_2$ is the Euclidean norm of the vector x.

Theorem 3 *Assume $\boldsymbol{\mu} = p_* \boldsymbol{\zeta}$ satisfies* (H) *and admits some exponential moments, and that $\boldsymbol{\nu}$ is non-degenerate and centered. Suppose also that every $\nu_{i,\mathbf{w}}$ admits a moment $\mathtt{M}_{i,\mathbf{w}} = \langle \nu_{i,\mathbf{w}}, |y|_2^{8+\xi} \rangle$, for some $\xi > 0$, such that*

$$\sup_{i \in [K]} \mathtt{M}_{i,\mathbf{w}} = O(|\mathbf{w}|^D) \tag{4}$$

for some $D > 0$. *Write*

$$\Sigma = \sqrt{\sum_{i\in[K]} a_i \sum_{\mathbf{w}\in\mathcal{W}_K} \zeta^{(i)}(\mathbf{w}) \sum_{j=1}^{|\mathbf{w}|} b_{w_j} \langle \nu_{i,\mathbf{w}}, y_j^2\rangle} \in (0,\infty)\,.$$

Then for any $\mathbf{x} \in [K]^{\mathbb{N}}$, *under* $\mathbb{P}^{\mathbf{x}}$, *the following convergence in distribution on* $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$ *holds jointly with that of* (i) *in Theorem 1:*

$$\left(\frac{1}{n^{1/4}} S^F_{\lfloor ns\rfloor}, s \geq 0\right) \xrightarrow[n\to\infty]{d} \left(\Sigma\sqrt{\frac{2}{\sigma}} R_s, s\geq 0\right),$$

where conditionally on the Brownian motion B *of* (i) *in Theorem 1,* R *is a Gaussian process with covariance*

$$\mathrm{Cov}\,(R_s, R_{s'}) = \inf_{s\wedge s'\leq u\leq s\vee s'} |B_u|.$$

If $\mathbf{x} = (i,i,\ldots)$ *for some* $i \in [K]$, *then the convergence holds jointly with that of* (ii) *in Theorem 1.*

The analogous statement for conditioned laws is:

Theorem 4 *Under the same hypotheses as Theorem 3, the following convergence in distribution on* $\mathbb{D}([0,1],\mathbb{R})$ *holds jointly with that of Theorem 2: for every* $i, j \in [K]$,

$$\left(\frac{S^T_{\lfloor \#Tt\rfloor}}{n^{1/4}}, 0\leq t\leq 1\right) \quad \text{under} \quad \mathbb{P}^{(i)}(\cdot|\#T^{(j)} = n) \xrightarrow[n\to\infty]{d} \left(\Sigma\sqrt{\frac{2}{\sigma\sqrt{a_j}}} R^{\mathrm{ex}}_t, 0\leq t\leq 1\right),$$

where conditionally on the Brownian excursion B^{ex} *of Theorem 2,* R^{ex} *is a Gaussian process with covariance*

$$\mathrm{Cov}\,(R^{\mathrm{ex}}_s, R^{\mathrm{ex}}_{s'}) = \inf_{s\wedge s'\leq u\leq s\vee s'} B^{\mathrm{ex}}_u\,.$$

Remark. By contrast with [10] and [13], our hypothesis on the spatial displacement is an $8+\xi$-moment assumption rather than a $4+\xi$-moment assumption. We believe such a weaker hypothesis to be sufficient, but were not able to prove it, essentially because we could not prove what we believe to be the best Hölder norm bounds in Proposition 8. See the remark after the latter's statement.

2 Proof of Theorem 1

2.1 Ancestral decomposition for multitype Galton-Watson trees and forests

Let $\boldsymbol{\zeta} = (\zeta^{(1)},\ldots,\zeta^{(K)})$ be a non-degenerate critical ordered offspring distribution. For $i \in [K]$, define the size-biased measure

$$\widehat{\zeta}^{(i)}(\mathbf{w}) = \frac{p(\mathbf{w})\cdot\mathbf{b}}{b_i}\zeta^{(i)}(\mathbf{w})\,, \qquad \mathbf{w}\in\mathcal{W}_K\,,$$

and notice that these are probability measures on $\mathcal{W}_K$ by the definition of $\mathbf{b}$, since they are pushed by p to the measure

$$\widehat{\mu}^{(i)}(\mathbf{z}) = \frac{\mathbf{z}\cdot\mathbf{b}}{b_i}\mu^{(i)}(\mathbf{z}), \qquad \mathbf{z}\in\mathbb{Z}_+^K,$$

and they do not charge the null sequence $\varnothing$. On some probability space $(\Omega, \mathcal{A}, P)$, let $((\mathbf{W}_u^i, \widehat{\mathbf{W}}_u^i, l_u^i, i \in [K]), u \in \mathcal{U})$ be a family of $\mathcal{U}$-indexed independent random vectors, such that $\mathbf{W}_u^i$ has law $\zeta^{(i)}$, $\widehat{\mathbf{W}}_u^i$ has law $\widehat{\zeta}^{(i)}$, and

$$P(l_u^i = k | \widehat{\mathbf{W}}_u^i) = \frac{b_{\widehat{W}_u^i(k)}}{p(\widehat{\mathbf{W}}_u^i)\cdot\mathbf{b}}.$$

Otherwise said, l_u^i is equal to k with probability proportional to b_{w_k} given $\widehat{\mathbf{W}}_u^i = \mathbf{w}$. Fix some $i \in [K]$. Recursively, we build a set $\widehat{\mathfrak{t}}$, a mark $\widehat{e} : \mathfrak{t} \to [K]$ and a sequence $(V_n, n \geq 0)$ by first letting $V_0 = \varnothing \in \widehat{\mathfrak{t}}$, $\widehat{e}(\varnothing) = i$, and given $V_0, \widehat{e}(V_0), \ldots, V_n, \widehat{e}(V_n)$ have been constructed with $\widehat{e}(V_n) = j$, we let $V_{n+1} = V_n l_{V_n}^j$ and $\widehat{e}(V_{n+1}) = \widehat{W}_{V_n}^j(l_{V_n}^j)$. Then, for $u \in \widehat{\mathfrak{t}}$ with $\widehat{e}(u) = j$,

- if $u = V_n$ for some n, then $ul \in \widehat{\mathfrak{t}}$ if and only if $1 \leq l \leq |\widehat{\mathbf{W}}_u^j|$. For such l, $\widehat{e}(ul) = \widehat{W}_u^j(l)$, and
- otherwise, $ul \in \widehat{\mathfrak{t}}$ if and only if $1 \leq l \leq |\mathbf{W}_u^j|$. For such l, $\widehat{e}(ul) = W_u^j(l)$.

The set $\widehat{\mathfrak{t}}$ thus obtained has the properties of a tree, except that it is infinite. More precisely, it consists of an infinite 'spine' $V_0, V_1, \ldots$ of distinguished vertices, which is interpreted as an infinite ancestral line along which individuals of type i have a $\widehat{\zeta}^{(i)}$-distributed offspring sequence, among which each element j is selected as a distinguished successor with probability proportional to b_j. Then, non-distinguished individuals have a regular GW descent with offspring distribution $\boldsymbol{\zeta}$. It is easy to prove that the spine is the unique infinite simple path of vertices in $\widehat{\mathfrak{t}}$ starting from $\varnothing$.

We let $\widehat{P}^{(i),h}$ be the law of $([\widehat{\mathfrak{t}}]_{V_h}, V_h)$, with the notation $[\mathfrak{t}]_u$ of Sect. 1.3. It is a distribution on the set $\{(\mathfrak{t}, e, v) : (\mathfrak{t}, e) \in \mathcal{T}^{(K)}, v \in \mathfrak{t}\}$ of pointed trees, on which we let (T, V) be the identity map. For any finite sequence $\mathbf{x}$ of types, let also $\widehat{P}^{\mathbf{x},j,h}$ be the law under which $(F_l, l \neq j), F_j$ are pairwise independent with respective laws $(P^{(x_l)}, l \neq j), \widehat{P}^{(x_j),h}$.

Notice that by construction, the types of the distinguished individuals $V_0, V_1, \ldots$ form a Markov chain, whose transition law is

$$\widehat{P}^{(i)}(e(V_{h+1}) = j' | e(V_h) = j) = \sum_{\mathbf{z}\in\mathbb{Z}_+^K} \frac{\widehat{\mu}^{(j)}(\mathbf{z}) b_{j'} z_{j'}}{\mathbf{b}\cdot\mathbf{z}} = \frac{b_{j'}}{b_j} m_{jj'}. \tag{5}$$

The stationary distribution of this Markov chain is easily checked to be the vector $(a_i b_i, 1 \leq i \leq K)$, because of the normalization $\mathbf{a}\cdot\mathbf{b} = 1$. This will be useful in the sequel.

Lemma 1 *For any finite sequence of types* $\mathbf{x} = (x_j, 1 \leq j \leq r)$*, and any non-negative functions* G_1, G_2*,*

$$E^{\mathbf{x}}\left[\sum_{v\in F} G_1(v,[F]_v)G_2(F_v)\right] = \sum_{j=1}^{r} b_{x_j} \sum_{h\geq 0} \widehat{E}^{\mathbf{x},j,h}\left[\frac{G_1(V,F)E^{(e(V))}\left[G_2(T)\right]}{b_{e(V)}}\right]. \tag{6}$$

Proof. Let $(\mathbf{f}, e_{\mathbf{f}})$ be a K-type forest, u a leaf of $\mathbf{f}$ (i.e. a vertex with no child) and $(\mathbf{t}, e_{\mathbf{t}})$ a K-type tree with $e_{\mathbf{t}}(\varnothing) = e_{\mathbf{f}}(u)$. Then it is enough to show the result for $G_1(u,\mathbf{f})G_2(\mathbf{t}) = \mathbb{1}_{\{u,\mathbf{f},\mathbf{t}\}}$, as one can then use linearity of the expectation. In this case, the left-hand side of (6) is equal to $P^{\mathbf{x}}(F = [\mathbf{f}, u, \mathbf{t}])$, where $[\mathbf{f}, u, \mathbf{t}]$ is the only forest $(\mathbf{f}', e') \in \mathcal{F}^{(K)}$ containing u with $[\mathbf{f}']_u = \mathbf{f}$ and $\mathbf{f}'_u = \mathbf{t}$. We let $S = \{v : v \vdash u, v \neq u\}$ and $A = \{v : v \notin S \cup \{u\}, \neg v \in S\}$, where $\neg v$ is the father of v, i.e. the word v with its last letter removed. We can redisplay

$$P^{\mathbf{x}}(F = [\mathbf{f}, u, \mathbf{t}]) = \prod_{v\in \mathbf{f}'} \zeta^{(e_{\mathbf{f}}(v))}(\mathbf{w}_{\mathbf{f}'}(v)),$$

as

$$\prod_{v\in\mathbf{t}} \zeta^{(e_{\mathbf{t}}(v))}(\mathbf{w}_{\mathbf{t}}(v)) \prod_{v\in\mathbf{f}, v\not\vdash u} \zeta^{(e_{\mathbf{f}}(v))}(\mathbf{w}_{\mathbf{f}}(v)) \prod_{v\in S} \zeta^{(e_{\mathbf{f}}(v))}(\mathbf{w}_{\mathbf{f}}(v)).$$

In this expression, one can factorize out the product of probabilities of the subtrees $\mathbf{f}'_u = \mathbf{t}$ and $\mathbf{f}'_j = \mathbf{f}_j$ for $j \in \{1, \ldots, r\} \setminus \{u_1\}$, u_1 the first letter of u, and $\mathbf{f}_v$ for $v \in A$. Letting $j = u_1$ and $u = ju'$, this shows

$$\begin{aligned}
&P^{\mathbf{x}}(F = [\mathbf{f}, u, \mathbf{t}]) \\
&= P^{(e_{\mathbf{f}}(u))}(T = \mathbf{t}) \prod_{l\neq j, 1\leq l\leq r} P^{(x_l)}(T = \mathbf{f}_l) \prod_{v\in A} P^{(e_{\mathbf{f}}(v))}(T = \mathbf{f}_v) \prod_{v\in S} \zeta^{(e_{\mathbf{f}}(v))}(\mathbf{w}_{\mathbf{f}}(v)).
\end{aligned}$$

We can also rewrite the last product as

$$\prod_{v\in S} \widehat{\zeta}^{(e_{\mathbf{f}}(v))}(\mathbf{w}_{\mathbf{f}}(v)) \frac{b_{e_{\mathbf{f}}(v)}}{p(\mathbf{w}_{\mathbf{f}}(v)) \cdot \mathbf{b}},$$

and we finally recognize

$$\begin{aligned}
&P^{\mathbf{x}}(F = [\mathbf{f}, u, \mathbf{t}]) \\
&= \frac{b_{x_j}}{b_{e_{\mathbf{f}}(u)}} \widehat{P}^{(x_j),|u|-1}((T,V) = (\mathbf{f}_j, u')) P^{(e_{\mathbf{f}}(u))}(T = \mathbf{t}) \left(\prod_{l\neq j, 1\leq l\leq r} P^{(x_l)}(T = \mathbf{f}_l)\right) \\
&= b_{x_j} \widehat{E}^{\mathbf{x},j,|u|}[\mathbb{1}_{\{u,\mathbf{f}\}}(V,F) E^{(e(V))}[\mathbb{1}_{\{\mathbf{t}\}}(T)]/b_{e(V)}],
\end{aligned}$$

which yields the result. $\square$

2.2 Around Perron-Frobenius' Theorem

Let us first recall the well-known Perron-Frobenius Theorem, which in this form can be found in [17].

Proposition 3 (Perron-Frobenius) *Let $M \in \mathfrak{M}_K(\mathbb{R}_+)$ be an irreducible matrix.*
(i) *The matrix M has a real eigenvalue ϱ with maximal modulus, which is positive, simple (i.e. it is a simple root of the characteristic polynômial of M), and every ϱ-eigenvector has only non-zero entries, all of the same sign.*
(ii) *Any eigenvector of M with non-negative entries is a ϱ-eigenvector, and hence has only positive entries.*

From this, we deduce the following useful

Lemma 2 (i) *Suppose $M \in \mathfrak{M}_K(\mathbb{R}_+)$ is irreducible. Then its spectral radius ϱ satisfies $\varrho > m_{ii}$, for all $1 \leq i \leq K$.*
(ii) *Suppose $K > 1$ and $\varrho = 1$. Then the matrix $\widetilde{M} \in \mathfrak{M}_{K-1}(\mathbb{R}_+)$ with entries*

$$\widetilde{m}_{ij} = m_{ij} + \frac{m_{iK}m_{Kj}}{1 - m_{KK}}, \qquad 1 \leq i, j \leq K - 1$$

is also irreducible with spectral radius 1.

Proof. (i) is immediate by writing $\sum_{j=1}^{K} m_{ij}u_j = \varrho u_i, 1 \leq i \leq K$, where $\mathbf{u}$ is a right ϱ-eigenvector of M with positive entries.

(ii) First, notice that the irreducibility of a matrix with nonnegative entries only depends on which entries are non-zero, and not on what their actual value is. Since $1 - m_{KK} > 0$ by (i), it is sufficient to show that the matrix with entries

$$\max(m_{ij}, m_{iK}m_{Kj}), \qquad 1 \leq i, j \leq K - 1 \tag{7}$$

is irreducible when M is. Assuming M irreducible and upon replacing m_{ij} by $\mathbb{1}_{\{m_{ij}>0\}}$, we may assume M is the adjacency matrix of a connected non-oriented graph on the vertices $1, \ldots, K$. But now, the matrix of (7) is the adjacency matrix of the graph on the vertices $1, \ldots, K-1$, where i is adjacent to j if and only if they are either adjacent or both adjacent to K in the initial graph. It is straightforward to see that this graph remains connected, so its adjacency matrix is irreducible.

It remains to show that $\widetilde{M}$ has spectral radius 1. First, it is immediate that if $\mathbf{a}, \mathbf{b}$ are left and right 1-eigenvectors of M, then $(a_1, \ldots, a_{K-1})$ and $(b_1, \ldots, b_{K-1})$ are left and right 1-eigenvectors of $\widetilde{M}$. Indeed, for $1 \leq j \leq K - 1$,

$$\sum_{i=1}^{K-1} a_i \widetilde{m}_{ij} = a_j - a_K m_{Kj} + a_K(1 - m_{KK})\frac{m_{Kj}}{1 - m_{KK}} = a_j, \text{ and } \sum_{j=1}^{K-1} b_j \widetilde{m}_{ij} = b_i$$

for $1 \leq i \leq K - 1$. This is enough to conclude by (ii) in Proposition 3. $\square$

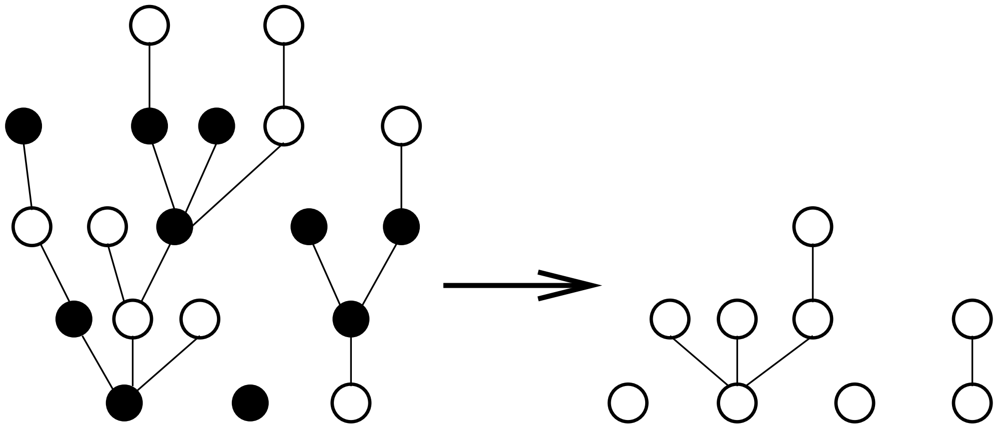

Figure 1: An illustration of the projection $\Pi^{(1)}$, with type-1 vertices represented in white, and type-2 vertices in black.

2.3 Reduction of trees

2.3.1 A projection on monotype trees

We describe a projection function $\Pi^{(i)}$ that goes from the set of K-type planar forests to the set of monotype planar forests, and which intuitively squeezes generations, keeping only the type-i individuals.

Precisely, if $(\mathbf{f}, e)$ is a K-type forest, we first let $v_1 \prec v_2 \prec \dots$ be the vertices of $\mathbf{f}^{(i)}$ such that all ancestors of v_k have type different from i, and listed in depth-first order. We consider a forest $\Pi^{(i)}(\mathbf{f}) = \mathbf{f}'$ with as many tree components as there are elements in $\{v_1, v_2, \dots\}$. Thus, we start with the set of roots $1, 2, \dots$ of $\mathbf{f}'$. Then recursively, for each $u \in \mathbf{f}'$, let $v_{u1}, \dots, v_{uk}$ be the vertices of $(v_u \mathbf{f}_{v_u}) \setminus \{v_u\}$ such that

- $e(v_{uj}) = i, 1 \leq j \leq k$,
- for every $1 \leq j \leq k$, if $v_{uj} = v_u l_1, \dots l_h$, then $e(ul_1 \dots l_r) \neq i$ for all $1 \leq r < h$, and
- $v_{u1}, \dots, v_{uk}$ are arranged in lexicographical order.

Then, we add the vertices $u1, \dots, uk$ to $\mathbf{f}'$, and continue iteratively. See Figure 1 for an example in the case $K = 2$. If $u \in \mathbf{f}'$ has children $u1, \dots, uk$, we let

$$N^{(i)}(u) = \# \left(\mathbf{f}_{v_u} \setminus \left(\bigcup_{j=1}^{k} \mathbf{f}_{v_{uj}} \right) \right) - 1 \,,$$

be the number of vertices that have been deleted between u and its children during the operation. We also let

$$N^{(i)}_{\text{floor}}(n) = \#\{v \in \mathbf{f}_n : e(w) \neq i \, \forall w \vdash v\} \,,$$

be the number of vertices of the n-th tree component of $\mathbf{f}$ that lie below the first layer of type-i vertices.

If $(\mathbf{t}, e) \in \mathcal{T}^{(K)}$, we may apply the map $\Pi^{(i)}$ to the forest $1\mathbf{t}$, and get as a result a forest $\mathbf{f}'$. We denote this forest by $\Pi^{(i)}(\mathbf{t})$ with a slight abuse of notation. Notice that $\Pi^{(i)}(\mathbf{t})$ has one tree component if $e(\varnothing) = i$, in which particular case we denote it by $\mathbf{f}'_1 = \Pi^{(i)}(\mathbf{t})$, with further abuse of notations.

If $\boldsymbol{\mu}$ is an unordered offspring distribution and $i, j \in [K]$, we let $\overline{\mu}_j^{(i)}$ be the distribution of $c_{\Pi^{(i)}(T)}(\varnothing)$ under $P^{(j)}$ with the above conventions: if $i \neq j$ then this counts the number of components of the reduced forest, while if $i = j$ this is the number of children of the root of the reduced tree.

Proposition 4 (i) *Let $\mathbf{x} \in [K]^{\mathbb{N}}$. Under the law $P^{\mathbf{x}}$, the forest $\Pi^{(i)}(F)$ is a (monotype) GW forest with offspring distribution $\overline{\mu}_i^{(i)}$, whose mean and variance are equal to*

$$\sum_{z \geq 0} z \overline{\mu}_i^{(i)}(\{z\}) = 1 \qquad \textit{and} \qquad \mathrm{Var}\,(\overline{\mu}_i^{(i)}) = \frac{\sigma^2}{a_i b_i^2}.$$

In particular, it is critical with finite variance.
(ii) *For each i and still under $E^{\mathbf{x}}$, the random variables $(N^{(i)}(u(n)), n \geq 0), (N_{\text{floor}}^{(i)}(n), n \geq 1)$ are all independent, and the variables $(N^{(i)}(u(n)), n \geq 0)$ are i.i.d. The number of vertices which are deleted in the operation $\Pi^{(i)}$ between two generations has mean*

$$E^{\mathbf{x}}\left[N^{(i)}(u(0))\right] = \frac{1}{a_i} - 1\,,$$

and finite variance. Similarly, the random variables $(N_{\text{floor}}^{(i)}(n), n \geq 1)$ have finite (x_n-dependent) variance, as well as the laws $\overline{\mu}_i^{(j)}$ for $i, j \in [K]$.
(iii) *More generally, if $\boldsymbol{\mu}$ admits a finite p-th moment with $p \in \mathbb{N}$ (resp. admits some exponential moments), then so do $\overline{\mu}_i^{(j)}$ and the variables $N^{(i)}(u(n)), N_{\text{floor}}^{(i)}(n+1), n \geq 0$ under $P^{\mathbf{x}}$.*

The GW property of $\Pi^{(i)}(F)$ under $E^{\mathbf{x}}$ is easy to obtain from Jagers' theorem on stopping lines [8], each subtree rooted at a vertex of type i being a copy of the whole tree. This also gives the independence statement in (ii). A detailed proof of these intuitive statements would be cumbersome, so that we leave the details to the interested reader, whom we refer to [8].

The rest of the proof of this proposition will be done by removing types one by one, and using an induction argument.

2.3.2 From K to $K-1$ types

In this section, we will suppose that type K is deleted, keeping in mind that the general case is similar. If $\mathbf{f} \in \mathcal{F}^{(K)}$, we now let $v_1 \prec v_2 \prec \ldots$ be the ordered list of vertices of $\mathbf{f}$ such that $e(v_i) \neq K$ and $e(v) = K$ for every $v \vdash v_i$. Recursively, given $v_u \in \mathbf{f}$ has been constructed, we let $v_{u1} \prec \ldots \prec v_{uk}$ be the descendents of v_u such that $e(v) = K$ for every $v_u \vdash v \vdash v_{uj}$, $v \notin \{v_u, v_{uj}\}$, while $e(v_{uj}) \neq K$. We then let $\widetilde{\Pi}(\mathbf{f})$ be the set of $u \in \mathcal{U}$ such that v_u has been defined by our recursive construction, and naturally associate a type $\widetilde{e}(u) = e(v_u)$ with $\widetilde{\Pi}(\mathbf{f})$. If $u \in \widetilde{\Pi}(\mathbf{f})$, we let $\widetilde{N}(u)$ be the number of vertices of type K that have been deleted in this construction between v_u and $v_{u1}, \ldots, v_{uk}$, namely,

$$\widetilde{N}(u) = \#\left(\mathbf{f}_{v_u} \setminus \bigcup_{i=1}^{k} \mathbf{f}_{v_{ui}}\right) - 1\,,$$

with the above notations. We also let $\widetilde{N}_{\text{floor}}(n)$ be the number of vertices $v \in \mathbf{f}_n^{(K)}$ that have only ancestors of type K.

Lemma 3 *Let $\mathbf{x} \in [K]^{\mathbb{N}}$. Then, under $P^{\mathbf{x}}$:*

(i) *for any $1 \leq i \leq K-1$, the forest $\widetilde{\Pi}(F)$ is a non-degenerate, irreducible, critical $K-1$-type GW forest. The (unordered) offspring distribution $\widetilde{\boldsymbol{\mu}} = (\widetilde{\mu}^{(i)}, i \in [K-1])$ has generating functions*

$$\widetilde{\varphi}^{(i)}(\mathbf{s}) = \varphi^{(i)}(\mathbf{s}, \widetilde{\varphi}^{(K)}(\mathbf{s}))\,, \tag{8}$$

for $1 \leq i \leq K-1$ and $\mathbf{s} \in [0,1]^{K-1}$, where $\widetilde{\varphi}^{(K)}$ is implicitly defined by

$$\widetilde{\varphi}^{(K)}(\mathbf{s}) = \varphi^{(K)}(\mathbf{s}, \widetilde{\varphi}^{(K)}(\mathbf{s}))\,. \tag{9}$$

(ii) *the $K-1$ sequences $(\widetilde{N}(u^{(i)}(n)), n \geq 0), i \in [K-1]$ are independent and formed of i.i.d. elements, where $u^{(i)}(0) \prec u^{(i)}(1) \prec \ldots$ is the ordered list of elements of F with type i. Their generating functions $\widetilde{\psi}^{(i)}$ respectively satisfy*

$$\widetilde{\psi}^{(i)}(s) = \varphi^{(i)}(1, \ldots, 1, \widetilde{\psi}^{(K)}(s))\,, \tag{10}$$

where $\widetilde{\psi}^{(K)}$ is implicitly defined by

$$\widetilde{\psi}^{(K)}(s) = s\varphi^{(K)}(1, \ldots, 1, \widetilde{\psi}^{(K)}(s))\,. \tag{11}$$

The random variables $\widetilde{N}_{\text{floor}}(n), n \geq 1$ are independent as well.

(iii) *any integer or exponential moment conditions on $\boldsymbol{\mu}$ is also satisfied by the laws $\widetilde{\boldsymbol{\mu}}$ and the random variables $\widetilde{N}(u^{(i)}(n)), \widetilde{N}_{\text{floor}}(n+1), n \geq 0, i \in [K-1]$.*

Proof. (i) Again, the GW property follows from the construction. On some probability space, let $\widetilde{\mathbf{X}}^{(j)}$ have same distribution as the $\mathbb{Z}_+^{K-1}$-valued random vector of children of a type-j vertex of $\widetilde{\Pi}(F)$ under $P^{\mathbf{x}}$. Then, by separating the offspring of this vertex with types equal and different from K, we obtain the identity in law

$$\widetilde{\mathbf{X}}^{(j)} \stackrel{d}{=} (X_1^{(j)}, \ldots, X_{K-1}^{(j)}) + \sum_{l=1}^{X_K^{(j)}} \widetilde{\mathbf{X}}^{(K)}(l)\,,$$

where $\mathbf{X}^{(i)}$ has distribution $\mu^{(i)}$ and is independent of $(\widetilde{\mathbf{X}}^{(K)}(l), l \geq 1)$, which are independent with same distribution, and a vector $\widetilde{\mathbf{X}}^{(K)}$ with this distribution must satisfy

$$\widetilde{\mathbf{X}}^{(K)} \stackrel{d}{=} (X_1^{(K)}, \dots, X_{K-1}^{(K)}) + \sum_{l=1}^{X_K^{(K)}} \widetilde{\mathbf{X}}^{(K)}(l)\,,$$

with similar notations. These two expressions immediately translate as (8) and (9). Now let

$$\widetilde{m}_{ij} = \frac{\partial \widetilde{\varphi}^{(i)}}{\partial s_j}(\mathbf{1})\,, \qquad 1 \leq i, j \leq K-1,$$

so that $\widetilde{M} = (\widetilde{m}_{ij})_{1 \leq i,j \leq K}$ is the mean matrix associated with the $K-1$-type GW forest $\widetilde{\Pi}(F)$ under $P^{\mathbf{x}}$. Differentiating (8) and (9) and letting $\mathbf{s}$ increase to $(1, \dots, 1)$ gives

$$\widetilde{m}_{ij} = m_{ij} + \frac{m_{iK} m_{Kj}}{1 - m_{KK}},$$

for $1 \leq i, j \leq K-1$. Hence, $\widetilde{M}$ is defined as in Lemma 2, so it is irreducible with spectral radius 1. Moreover, the K-type GW process associated with $\widetilde{\Pi}(F)$ under $P^{\mathbf{x}}$ has to be non-degenerate, because it dies in finite time a.s..

(ii) The independence statement is again a consequence of Jager's theorem, and Formulas (10) and (11) are obtained by similar distributional equations arguments as above, namely,

$$\widetilde{N}(u^{(i)}(n)) \stackrel{d}{=} \sum_{l=1}^{X_K^{(i)}} \widetilde{N}_l^{(K)}\,,$$

where $\mathbf{X}^{(i)}$ has law $\mu^{(i)}$ and $\widetilde{N}_l^{(K)}, l \geq 1$ are i.i.d. random elements independent of $\mathbf{X}^{(i)}$, that satisfy

$$\widetilde{N}_1^{(K)} \stackrel{d}{=} 1 + \sum_{l=1}^{X_K^{(K)}} \widetilde{N}_l^{(K)}\,.$$

On the other hand, $\widetilde{N}_{\text{floor}}(n)$ is either 0 or equal in distribution to $\widetilde{N}_1^{(K)} + 1$ with the same notation, according to whether $x_n \neq K$ or $x_n = K$.

(iii) is obtained by differentiating equations (8), (9), (10) and (11) p times, while the assertion on small exponential moments is obtained by applying the implicit function theorem to the implicit functions $\widetilde{\varphi}^{(K)}, \widetilde{\psi}^{(K)}$. Details are left as an exercise to the reader. □

Notice that this provides an alternative way of showing that the spectral radius of $\widetilde{M}$ is ≤ 1, since the GW process has to be (sub)-critical in order to become extinct a.s. Recall that $\mathbf{a}, \mathbf{b}$ are the left and right 1-eigenvectors of M with $\mathbf{a} \cdot \mathbf{1} = 1 = \mathbf{a} \cdot \mathbf{b}$. In view of the proof of Lemma 2, the left and right 1-eigenvectors $\widetilde{\mathbf{a}}, \widetilde{\mathbf{b}}$ of $\widetilde{M}$ satisfying $\widetilde{\mathbf{a}} \cdot \mathbf{1} = 1 = \widetilde{\mathbf{a}} \cdot \widetilde{\mathbf{b}}$ are given by

$$\widetilde{\mathbf{a}} = \frac{1}{1-a_K}(a_1, \dots, a_{K-1}), \qquad \widetilde{\mathbf{b}} = \frac{1-a_K}{1-a_K b_K}(b_1, \dots, b_{K-1}).$$

We are now ready to give the

Proof of Proposition 4. (i) We prove this by induction on K, in the case $i = 1$, without losing generality. The case $K = 1$ is obvious, since in this case $M = (1)$, $Q^{(1)} = (\sigma^2)$ is indeed the variance of the offspring distribution and $\Pi^{(1)}$ is the identity. According to Lemma 3, under $P^{\mathbf{x}}$, it is licit to do the K-to $K-1$-type operation $\widetilde{\Pi}$, without changing the hypothesis that the GW processes under consideration are nondegenerate, irreducible and critical. This immediately gives the result on the mean of the offspring distribution of $\Pi^{(1)}(F)$ by induction. The only statement that remains to be proved is the formula for the variance of its offspring distribution. Using again (8) and (9), straightforward (but tedious) computations show that, letting

$$\widetilde{Q}^{(i)}_{jk} = \frac{\partial^2 \widetilde{\varphi}^{(i)}}{\partial s_j \partial s_k}(\mathbf{1}), \qquad 1 \le i, j, k \le K$$

be the quadratic forms associated with the offspring distributions of $\widetilde{\Pi}(F)$ under $P^{\mathbf{x}}$,

$$\begin{aligned} \widetilde{Q}^{(i)}_{jk} &= Q^{(i)}_{jk} + \frac{m_{Kk} Q^{(i)}_{jK} + m_{Kj} Q^{(i)}_{kK}}{1 - m_{KK}} + \frac{m_{Kj} m_{Kk}}{(1-m_{KK})^2} Q^{(i)}_{KK} \\ &\quad + \frac{m_{iK}}{1-m_{KK}} \left(Q^{(K)}_{jk} + \frac{m_{Kk} Q^{(K)}_{jK} + m_{Kj} Q^{(K)}_{kK}}{1-m_{KK}} + \frac{m_{Kj} m_{Kk}}{(1-m_{KK})^2} Q^{(K)}_{KK} \right). \end{aligned}$$

It is then easy to check that

$$\widetilde{\mathbf{a}} \cdot \widetilde{\mathbf{Q}}(\widetilde{\mathbf{b}}) = \frac{1-a_K}{(1-a_K b_K)^2} \mathbf{a} \cdot \mathbf{Q}(\mathbf{b}).$$

Using the induction hypothesis, we obtain $\widetilde{\mathbf{a}} \cdot \widetilde{\mathbf{Q}}(\widetilde{\mathbf{b}}) = \widetilde{a}_1 \widetilde{b}_1^2 \mathrm{Var}\,(\overline{\mu}_1^{(1)})$, so that

$$a_1 b_1^2 \frac{1-a_K}{(1-a_K b_K)^2} \mathrm{Var}\,(\overline{\mu}_1^{(1)}) = \frac{1-a_K}{(1-a_K b_K)^2} \mathbf{a} \cdot \mathbf{Q}(\mathbf{b}),$$

giving the result.

(ii) We again prove this in the case $i = 1$. When $K = 1$ there is nothing to prove. If $K = 2$, one checks that the number of type-2 vertices trapped between two 1-type generations of $\Pi^{(1)}(T)$ under $P^{(1)}$ has mean $m_{12}/(1-m_{22})$ and finite variance (resp. some exponential moment if $\boldsymbol{\mu}$ has some), by differentiating (10) and (11) once. Then, one obtains by direct computations that this is $a_2/a_1 = (1-a_1)/a_1$.

So suppose $K \geq 3$. The idea is to apply the projection operation $\widetilde{\Pi}$, $K-2$ times, removing types $K, K-1, \ldots, 3$ one after the other. When this is performed, a two-type tree $\widetilde{\Pi}^{\circ k}(F)$ is obtained, and the number of type 2 vertices that have only the root as type 1 ancestor is precisely the number of type 2 individuals that are trapped between two generations of $\Pi^{(1)}(F)$. By a direct inductive argument using Lemma 3 and the discussion after its proof, the mean matrix of this contracted two-type tree $\widetilde{\Pi}^{\circ k}(F)$ has (a_1, a_2) as left 1-eigenvector. In view of the $K = 2$ case above, the mean number of deleted type 2 vertices in a generation is thus a_2/a_1. By symmetry, the average number of type $i \in \{2, \ldots, K\}$ vertices deleted in a generation of $\Pi^{(1)}(F)$ is a_i/a_1. The average total number of deleted vertices is thus $(a_2 + \ldots + a_K)/a_1 = (1-a_1)/a_1$, as claimed.

Finally, (iii) is obtained by applying point (iii) in Lemma 3 in a similar induction argument. □

2.4 Two exponential bounds

Let $\boldsymbol{\zeta}$ be an ordered offspring distribution with $p_*\boldsymbol{\zeta} = \boldsymbol{\mu}$ satisfying (H). The following lemma allows to control the height and number of components in a GW forest.

Lemma 4 *There exist two constants $0 < C, C' < \infty$ depending only on $\boldsymbol{\mu}$, such that for every $n \in \mathbb{N}$, $\mathbf{x} \in [K]^{\mathbb{N}}$ and $\eta > 0$,*

$$P^{\mathbf{x}}\left(\max_{0\leq k\leq n} |u(n)| \geq n^{1/2+2\eta}\right) \leq C\exp(-C'\,n^{\eta}),$$

and

$$P^{\mathbf{x}}\left(\Upsilon_n^F \geq n^{1/2+2\eta}\right) \leq C\exp(-C' n^{\eta}),$$

Proof. In the case $K = 1$, this is a straightforward consequence of [13, Lemma 13]. In the general case, it suffices to note that under $P^{\mathbf{x}}$, independently of $\mathbf{x}$,

$$\max_{0\leq k\leq n} |u_F(k)| \leq \sum_{i\in[K]} \max_{0\leq k\leq n} |u_{\Pi^{(i)}(F)}(k)|\,,$$

and

$$\Upsilon_n^F \leq \sum_{i\in[K]} \Upsilon_n^{\Pi^{(i)}(F)}\,.$$

We may then apply the $K = 1$ case to each of the forests $\Pi^{(i)}(F), i \in [K]$, which under $P^{\mathbf{x}}$ are critical non-degenerate monotype GW forests by Proposition 4. □

2.5 Convergence of types

Note. From now on, we will make a frequent use of exponential bounds for real sequences $(x_n, n \geq 0)$, namely that $|x_n| \leq \exp(-n^{\varepsilon})$ for some $\varepsilon > 0$ and large enough n. To simplify notations and avoid referring to changing ε's, we write $x_n = \mathrm{oe}(n)$ in this case.

A natural way to proceed to prove Theorem 1 is now to use the known results of the monotype forest $\Pi^{(1)}(F)$, and to try and pull them back to the projected multitype tree. To do this, we must take care of two kinds of loss of information: the number of vertices with type $\neq 1$ of F that stand between two consecutive 1-type vertices seen in $\Pi^{(1)}(F)$ ('time' information), and the number of vertices of F that actually stand between a type-1 vertex of $\Pi^{(1)}(F)$ and one of its sons ('height' information).

For $u \in \mathbf{f} \in \mathcal{F}^{(K)}$, let $\mathrm{Anc}^u_{\mathbf{f}}(i) := \mathrm{Anc}^u_{\mathbf{f}}(e(v) = i)$ be the number of ancestors u such that $e(v) = i$.

Proposition 5 *Under* (H), *for every* $\gamma > 0$ *and* $\mathbf{x} \in [K]^{\mathbb{N}}$,

$$\max_{i\in[K]} P^{\mathbf{x}}\left(\max_{0\leq k\leq n}\left|H^F_k - \frac{\mathrm{Anc}^{u(k)}_F(i)}{a_i b_i}\right| > n^{1/4+\gamma}\right) = \mathrm{oe}(n)\,. \tag{12}$$

For this, we need the following moderate deviations estimate for Markov chains.

Lemma 5 *On some probability space* $(\Omega, \mathcal{A}, P)$, *let* $(X_n, n \geq 0)$ *be an irreducible Markov chain taking values in a finite set* S. *Let* π *be its stationary distribution, and* $\xi_n = n^{-1}\sum_{k=0}^{n-1}\delta_{X_k}$ *be its empirical distribution at time* n. *Then, for any* $f : S \to \mathbb{R}$ *and* $\gamma > 0$, *there exists* $N(f,\gamma) > 0$ *for every* $n \geq N(f,\gamma)$,

$$\max_{0\leq k\leq n} P(k|\xi_k(f) - \pi(f)| \geq n^{1/2+\gamma}) \leq \exp(-n^{\gamma})\,.$$

Proof. If the Markov chain is also aperiodic, then according to Wu [20, Theorem 2.1 (a)], for every f there exists a constant C depending on f such that for every large enough k,

$$P(k|\xi_k(f) - \pi(f)| \geq k^{1/2+\gamma}) \leq \exp(-Ck^{2\gamma})\,.$$

If the chain has period $d > 1$, then the same result is easily obtained by partitioning the state space into the d periodic classes with equal π-masses and considering the d shifted chains $(X_{nd+r}, n \geq 0)$ for $0 \leq r \leq d-1$, which under $P(\cdot|X_0 = x)$ are aperiodic for every $x \in S$.

Next, notice that for large enough n, $\max_{0\leq k\leq n^{1/2+\gamma/2}} P(k|\xi_k(f) - \pi(f)| \geq n^{1/2+\gamma}) = 0$ since $|\xi(f) - \pi(f)| \leq 2\|f\|_\infty$. Thus, we have, for large enough n,

$$\begin{aligned}\max_{0\leq k\leq n} P(k|\xi_k(f) - \pi(f)| \geq n^{1/2+\gamma}) &\leq \max_{n^{1/2+\gamma/2}\leq k\leq n} P(k|\xi_k(f) - \pi(f)| \geq k^{1/2+\gamma}) \\ &\leq \exp(-Cn^{\gamma+\gamma^2})\,,\end{aligned}$$

entailing the result. □

Proof of Proposition 5. Suppose $i = 1$ with no loss of generality. By Lemma 4, the probability that either $\Upsilon^F_n > n^{1/2+\gamma}$ or $\max_{0\leq k\leq n}|u(k)| > n^{1/2+\gamma}$ is an $\mathrm{oe}(n)$, so we can restrict ourselves to the complementary event. Thus, it suffices to bound the quantity

$$\begin{aligned}&P^{\mathbf{x}}\left(\mathbb{1}_{\{\Upsilon^F_n\leq n^{1/2+\gamma}, \max_{0\leq k\leq n}|u(k)|\leq n^{1/2+\gamma}\}}\max_{0\leq k\leq n}|a_1b_1|u(k)| - \mathrm{Anc}^{u(k)}_F(1)| > n^{1/4+\gamma}\right) \\ &\qquad\qquad \leq P^{\mathbf{x}}_{\lfloor n^{1/2+\gamma}\rfloor}\left(\max_{u\in F, |u|\leq n^{1/2+\gamma}}|a_1b_1|u| - \mathrm{Anc}^u_F(1)| > n^{1/4+\gamma}\right)\end{aligned}$$

for large n. By bounding the max by a sum over the same set, and then making use of the ancestral decomposition (Lemma 1), this is less than

$$\begin{aligned} &E^{\mathbf{x}}_{\lfloor n^{1/2+\gamma}\rfloor}\left[\sum_{u\in F}\mathbb{1}_{\{|u|\le n^{1/2+\gamma}\}}\mathbb{1}_{\left\{|a_1b_1|u|-\mathrm{Anc}^u_F(1)|>n^{1/4+\gamma}\right\}}\right] \\ &\le\ C\sum_{j=1}^{\lfloor n^{1/2+\gamma}\rfloor}\sum_{h=0}^{\lfloor n^{1/2+\gamma}\rfloor}\widehat{P}^{(x_j),h}\left(h|h^{-1}\mathrm{Anc}^V_T(1)-a_1b_1|>n^{1/4+\gamma}\right) \\ &\le\ Cn^{1+2\gamma}\max_{i\in[K]}\max_{0\le h\le n^{1/2+\gamma}}\widehat{P}^{(i),h}\left(h|h^{-1}\mathrm{Anc}^V_T(1)-a_1b_1|>n^{1/4+\gamma}\right)\,, \end{aligned}$$

where $C=\max_i b_i/\min_i b_i>0$. Recall that under $\widehat{P}^{(i),h}$, the sequence $(e(V_0),\dots,e(V_h)=e(V))$ is a Markov chain in $[K]$ started at i with step transition $p_{j,j'}=m_{jj'}b_{j'}/b_j$, and which admits $\mathbf{ab}:=(a_1b_1,\dots,a_Kb_K)$ as invariant probability. Notice that $h^{-1}\mathrm{Anc}^V_T(1)$ is the empirical measure of $\{1\}$ for this Markov chain. The result is now a straightforward consequence of Lemma 5. □

Next, recall the notation $\Lambda^{\mathbf{f}}_i(k)$ and let also $G^{\mathbf{f}}_i(k)=\mathrm{Card}\,\{u\prec u^{(i)}(k)\}$, where $u^{(i)}(0),u^{(i)}(1),\dots$ is the list of type-i vertices of $\mathbf{f}$, arranged in depth-first order. A similar notation holds for trees instead of forests, and we adopt the convention $G^{\mathbf{t}}_i(\#\mathbf{t}^{(i)})=\#\mathbf{t}$.

Proposition 6 (i) *For any $\mathbf{x}\in[K]^{\mathbb{N}}$, under $P^{\mathbf{x}}$, as $n\to\infty$, $(\Lambda^F_i(\lfloor ns\rfloor)/n,s\ge 0)$ converges in probability to $s\mapsto a_is$, for the topology of uniform convergence over compact sets.*

(ii) *Moreover, if $\boldsymbol{\mu}$ admits small exponential moments, it holds that for every $\gamma>0$ and $\mathbf{x}\in[K]^{\mathbb{N}}$,*

$$P^{\mathbf{x}}\left(|G^F_i(n)-a_i^{-1}n|>n^{1/2+\gamma}\right)=\mathrm{oe}(n)\,. \tag{13}$$

Proof. With the notations of Section 2.3.1, for $\mathbf{f}\in\mathcal{F}^{(K)}$, let $N(k):=N^{(i)}(u(k))$ be the number of descendents v of $u^{(i)}(k)$ such that the types of vertices in $\{w: u^{(i)}(k)\vdash w\vdash v, w\neq u^{(i)}(k)\}$ are all $\neq i$, and $N'(k)$ is the similar quantity, but counting only the vertices w that come before $u^{(i)}(n)$ in depth-first order. Then

$$G^{\mathbf{f}}_i(n)=\sum_{k=0}^{n-1}(1+N(k))+\underbrace{\sum_{k=0}^{n-1}(N'(k)-N(k))\mathbb{1}_{\{u^{(i)}(k)\vdash u^{(i)}(n)\}}}_{:=R_1(n)}+\underbrace{\sum_{k=1}^{\Upsilon^{\mathbf{f}}_n}N^{(i)}_{\text{floor}}(k)}_{:=R_2(n)}\,, \tag{14}$$

Now, we estimate the probability that $R_1(n)+R_2(n)$ is large, and by Lemma 4, for any fixed $0<\delta<1/2$, we may restrict ourselves to the event that the number of ancestors of type 1 of $u^{(i)}(n)$ is $\le n^{1/2+\delta}$ and that the tree containing $u^{(i)}(n)$ has rank $\le n^{1/2+\delta}$, up to losing an $\mathrm{oe}(n)$ term. Then under this event, the probability of $\{R_2(n)>n^{1-\delta}\}$ is less than

$$P^{\mathbf{x}}\left(\sum_{k=1}^{n^{1/2+\delta}}N^{(i)}_{\text{floor}}(k)>n^{1-\delta}\right)\,.$$

Now under $P^{\mathbf{x}}$, the $N^{(i)}_{\text{floor}}(k)$'s are independent, with respective laws that of Card $\{v : i \notin \{e(w) : w \vdash v\}\}$ under $P^{(x_k)}$. By (ii) in Proposition 4 and Chebychev's inequality, this goes to 0 as $n \to \infty$.

On the other hand, notice that $|R_1(n)| \leq \sum_{k=0}^{n-1} N(k) \mathbb{1}_{\{u^{(i)}(k) \vdash u^{(i)}(n)\}}$. Let $r_n = \sup\{k : P^{\mathbf{x}}(N(0) > k) > n^{-1}\}$. Since $N(0)$ has finite variance under $P^{\mathbf{x}}$ by (ii) in Proposition 4, it holds that $P^{\mathbf{x}}(N(0) > t) = o(t^{-2})$ as $t \to \infty$, so that $r_n = o(n^{1/2})$, as otherwise $r_{\phi(n)} \geq c\phi(n)^{1/2}$ for some extraction ϕ, so that

$$\phi(n)^{-1} < P^{\mathbf{x}}(N(0) > r_{\phi(n)}) \leq P^{\mathbf{x}}(N(0) > c\phi(n)^{1/2}) = o(\phi(n)^{-1}),$$

a contradiction. Therefore, for any $r'_n = o(n^{1/2})$ such that $r_n = o(r'_n)$, we obtain that

$$P^{\mathbf{x}}\left(\max_{0 \leq k \leq n} N(k) > r'_n\right) \leq 1 - (1 - P(N(0) > r'_n))^n \underset{n \to \infty}{\longrightarrow} 0.$$

On the other hand, we know from (i) in Proposition 4 and Proposition 2 that

$$n^{-1/2} \max_{0 \leq k \leq n} \mathrm{Anc}_F^{u(k)}(i) \leq n^{-1/2} \max_{0 \leq k \leq n} H_k^{\Pi^{(i)}(F)},$$

which converges in distribution as $n \to \infty$ to the supremum of a properly scaled Brownian excursion. Consequently, noticing that $R_1(n)$ is a sum involving $\mathrm{Anc}_F^{u^{(i)}(n)}(i)$ terms, we obtain that for every $\varepsilon > 0$, there exists $C > 0$ such that

$$P^{\mathbf{x}}\left(n^{-1} R_1(n) \leq C n^{-1} n^{1/2} r'_n\right) > 1 - \varepsilon$$

for every n large, whence $R_1(n) = o(n)$ in probability.

These estimates, when combined with (14) and the law of large numbers, entail that $G_i^F(n)/n$ converges in probability to the mean of $1 + N(0)$ under $P^{(i)}$, which by (ii) in Proposition 4 is $1/a_i$. Therefore, $G_i^F(\lfloor ns \rfloor)/n \to a_i^{-1} s$ in probability for every rational s and we claim that the convergence holds for the uniform topology over compact subsets of $\mathbb{R}$. To see this, one can use Skorokhod's representation theorem and assume that the convergence of $G_i^F(\lfloor ns \rfloor)$ is almost-sure for every rational s, and then apply a standard monotonicity, continuity and compactness argument. It is then elementary to conclude that the right-continuous inverse function $(\Lambda_i^F(\lfloor ns \rfloor)/n, s \geq 0)$ converges in probability to $s \mapsto a_i s$ for the uniform topology over compact sets.

Part (ii) of the statement is obtained along closely related lines, by first noting that this time, for $0 < \delta < \gamma/2$, and using similar notations as above,

$$\begin{aligned} P^{\mathbf{x}}\left(R_2(n) > n^{1/2+2\delta}\right) &\leq P^{\mathbf{x}}\left(\sum_{k=1}^{n^{1/2+\delta}} N^{(i)}_{\text{floor}}(k) > n^{1/2+2\delta}\right) + \mathrm{oe}(n) \\ &\leq \exp(A n^{1/2+\delta} - \varepsilon n^{1/2+2\delta}) + \mathrm{oe}(n) = \mathrm{oe}(n) \end{aligned}$$

for some $A, \varepsilon > 0$. Then, we have $P^{\mathbf{x}}\left(\max_{0 \leq k \leq n} N(k) > n^{\delta}\right) = \mathrm{oe}(n)$, so that

$$P^{\mathbf{x}}\left(R_1(n) > n^{1/2+2\delta}\right) \leq P^{\mathbf{x}}\left(\max_{0 \leq k \leq n} |u(k)| \geq n^{1/2+\delta}\right) + P^{\mathbf{x}}\left(\max_{0 \leq k \leq n} N(k) > n^{\delta}\right) = \mathrm{oe}(n).$$

Finally, the estimate

$$P^{\mathbf{x}}\left(\left|\sum_{k=0}^{n-1}(1+N(k))-a_i^{-1}n\right|\geq n^{1/2+\gamma}\right)=\mathrm{oe}(n)$$

is a standard moderate deviations estimate for random variables admitting small exponential moments, see [15, Theorem 2.6]. This is enough to conclude. $\square$

Notice that the previous statement immediately implies point (iii) in the statement of Theorem 1.

2.6 Proof of (i) and (ii) in Theorem 1

For any $s\geq 0$, we have

$$\left|H^F_{\lfloor ns\rfloor}-\frac{H^{\Pi^{(i)}(F)}_{\Lambda_i^F(\lfloor ns\rfloor)-1}}{a_ib_i}\right|\leq\left|u(\lfloor ns\rfloor)-\frac{\mathrm{Anc}_F^{u(\lfloor ns\rfloor)}(i)}{a_ib_i}\right|+\frac{\left|H^{\Pi^{(i)}(F)}_{\Lambda_i^F(\lfloor ns\rfloor)-1}-\mathrm{Anc}_F^{u(\lfloor ns\rfloor)}(i)\right|}{a_ib_i}\,.$$

By Proposition 5, we obtain that, for every $A>0$,

$$\sup_{0\leq s\leq A}n^{-1/2}\left|u(\lfloor ns\rfloor)-\frac{\mathrm{Anc}_F^{u(\lfloor ns\rfloor)}(i)}{a_ib_i}\right|\to 0$$

in probability as $n\to\infty$. On the other hand, we claim that

$$\left|H^{\Pi^{(i)}(F)}_{\Lambda_i^F(\lfloor ns\rfloor)-1}-\mathrm{Anc}_F^{u(\lfloor ns\rfloor)}(i)\right|\leq\left|H^{\Pi^{(i)}(F)}_{\Lambda_i^F(\lfloor ns\rfloor)-1}-H^{\Pi^{(i)}(F)}_{\Lambda_i^F(\lfloor ns\rfloor)}\right|+1\,. \tag{15}$$

Indeed, if $u^{(i)}(\lfloor ns\rfloor)$ is an ancestor of $u(\lfloor ns\rfloor)$, then the left-hand side is zero and there is nothing to prove. Else, the left-hand side equals the number of ancestors of type i of $u^{(i)}(\lfloor ns\rfloor)$ which are not ancestors of $u(\lfloor ns\rfloor)$, and so $H^{\Pi^{(i)}(F)}_{\Lambda_i^F(\lfloor ns\rfloor)-1}-\mathrm{Anc}_F^{u(\lfloor ns\rfloor)}(i)\geq 0$. On the other hand, the strict ancestors of $u^{(i)}(\lfloor ns\rfloor+1)$ that are not ancestors of type i of $u(\lfloor ns\rfloor)$, cannot be themselves of type i by definition (otherwise, such an ancestor would come after $u(\lfloor ns\rfloor)$ and before $u^{(i)}(\lfloor ns\rfloor+1)$ in depth-first order). Hence, $H^{\Pi^{(i)}(F)}_{\Lambda_i^F(\lfloor ns\rfloor)}-\mathrm{Anc}_F^{u(\lfloor ns\rfloor)}(i)\leq 1$, so that

$$0\leq H^{\Pi^{(i)}(F)}_{\Lambda_i^F(\lfloor ns\rfloor)-1}-\mathrm{Anc}_F^{u(\lfloor ns\rfloor)}(i)\leq H^{\Pi^{(i)}(F)}_{\Lambda_i^F(\lfloor ns\rfloor)-1}-H^{\Pi^{(i)}(F)}_{\Lambda_i^F(\lfloor ns\rfloor)}+1\,,$$

and the claimed inequality follows.

Under $P^{\mathbf{x}}$, the forest $\Pi^{(i)}(F)$ is a single-type GW forest whose offspring distribution has finite variance by Proposition 4, so that by Proposition 2,

$$n^{-1/2}\max_{0\leq k\leq n}|H^{\Pi^{(i)}(F)}_{k-1}-H^{\Pi^{(i)}(F)}_{k}|\xrightarrow[n\to\infty]{P^{\mathbf{x}}}0\,,$$

and it follows that under $P^{\mathbf{x}}$,

$$\left(n^{-1/2}\left(H^F_{\lfloor ns\rfloor}-(a_ib_i)^{-1}H^{\Pi^{(i)}(F)}_{\Lambda^F_i(\lfloor ns\rfloor)}\right), s\geq 0\right)\xrightarrow[n\to\infty]{P^{\mathbf{x}}} 0 \tag{16}$$

for the topology of uniform convergence over compact sets.

Using Propositions 6, 4 and 2, and composing $s\mapsto H^{\Pi^{(i)}(F)}_{\lfloor ns\rfloor}$ with $s\mapsto \Lambda^F_i(\lfloor ns\rfloor)/n$, we now obtain that $(n^{-1/2}H^{\Pi^{(i)}(F)}_{\Lambda^F_i(\lfloor ns\rfloor)}, s\geq 0)$ converges in distribution to $(2\overline{\sigma}_i^{-1}|B_{a_is}|, s\geq 0)$ where $\overline{\sigma}_i^2=\mathrm{Var}\,(\overline{\mu}^{(i)}_i)$, which is also equal in law to $(2a_i^{1/2}\overline{\sigma}_i^{-1}|B_s|, s\geq 0)$. One way of seeing this is to use Skorokhod's representation theorem to exhibit a probability space where the convergences of Propositions 2 and 6 hold a.s. rather than in distribution. Point (i) of the theorem is now proved by using (16).

Let us prove (ii). By definition, since all the roots are of type i, $u(\lfloor ns\rfloor)$ and the last node with type i before $u(\lfloor ns\rfloor)$ in depth-first order belong to the same tree. Therefore, the label of the tree of F containing $u(\lfloor ns\rfloor)$ is always the same as the label of the tree of $\Pi^{(i)}(F)$ containing the $\Lambda^F_i(\lfloor ns\rfloor)$-th node. This implies that $\Upsilon^{\Pi^{(i)}(F)}_{\Lambda^F_i(\lfloor ns\rfloor)}=\Upsilon^F_{\lfloor ns\rfloor}$. Now, the result is a plain consequence of Proposition 2 and of similar arguments as above. $\square$

Proof of Corollary 1. From (ii) in Theorem 1, we obtain that $(n^{-1}H^F_{n^2s}, 0\leq s\leq \tau_n)$ converges in distribution to $2\sigma^{-1}(|B_s|, 0\leq s\leq \mathfrak{T}_{b_i\sigma^{-1}})$, where τ_n is the first hitting time of n by Υ^F and $\mathfrak{T}_x$ is the first hitting time of x by L^0.

Now,

$$P^{\mathbf{i}}\left(\forall 1\leq k\leq n, \mathrm{ht}(F_k)<n\right)\underset{n\to\infty}{\longrightarrow} P(2\sigma^{-1}|B_s|\leq 1, 0\leq s\leq \mathfrak{T}_{b_i\sigma^{-1}})\,,$$

which can be rewritten as

$$\left(1-P^{(i)}(\mathrm{ht}(T)\geq n)\right)^n\underset{n\to\infty}{\longrightarrow}\exp\left(-\frac{b_i}{\sigma}N\left(\frac{2}{\sigma}\sup \mathrm{e}\geq 1\right)\right),$$

where $N(\mathrm{de})$ is the Ito excursion measure of the standard Brownian motion (see e.g. [16, Chapter XII] for definitions and the results recalled below), and where we have used the Ito decomposition of a Brownian motion into a Poisson process of excursions in the local time scale. Taking logarithms and using $N(\sup \mathrm{e}\geq x)=1/x$, gives the result. $\square$

Let us also mention that similar arguments, following the same lines as in [4, Proposition 2.5.2], actually show the more general result:

Corollary 2 *For every $a>0$ the probability measures $P^{(i)}(n^{-1}H^T\in\cdot\,|\mathrm{ht}(T)\geq an)$ converge in distribution as $n\to\infty$ towards $N(2\sigma^{-1}\mathrm{e}\in\cdot\,|2\sigma^{-1}\sup\mathrm{e}\geq a)$.*

2.7 Conditioned results: Theorem 2

Our main tool for conditioning is the following estimate for the size of GW trees.

Lemma 6 *Let $\boldsymbol{\mu}$ be a critical non-degenerate offspring distribution with finite variance. Then for every $i, j \in [K]$, one has*

$$n^{3/2} P^{(i)}(\# T^{(j)} = n) \underset{n\to\infty}{\longrightarrow} C_{ij}\,,$$

where if necessary the limit is taken along a subsequence for which the probability on the left-hand side is non-zero, and for some constant $C_{ij} > 0$.

Proof. This is very similar to Lemma 14 in [13]. If $i = j$, then using the fact that the reduced tree of Section 2.3.1 is a monotype GW tree, the result is a well-known fact. To treat the general case we elaborate slightly on the proof.

Let $r \in \mathbb{N}$ be fixed, and recall $\mathbf{j} = (j, j, \ldots)$ and the notation at the very end of Section 1.4. Let $\tilde{\mu} := \overline{\mu}_j^{(j)}$ be the offspring distribution of the GW tree $\Pi^{(j)}(T)$ under $P^{(j)}$, and on some probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$, let $(W_n, \geq 0)$ be a random walk with step distribution $\tilde{\mu}(\cdot + 1)$ on $\{-1, 0, 1, 2, \ldots\}$. Under $P_r^{\mathbf{j}}$, the forest $\Pi^{(j)}(F)$ is a monotype GW forest with offspring distribution $\tilde{\mu}$ and r tree components. It is then well-known that

$$P_r^{\mathbf{j}}(\#\Pi^{(j)}(F) = n) = P_r^{\mathbf{j}}(\# F^{(j)} = n) = \frac{r}{n} \tilde{P}(W_n = -r)\,.$$

By the local limit theorem in the lattice case [6, Theorem XV.5.3], $n^{1/2}\tilde{P}(W_n = -r) \to C$ as $n \to \infty$ for some $C > 0$, and for every r such that the probabilities under consideration are > 0. Moreover, there is a common uniform bound for all the terms as r varies along the admissible values.

Let $p(r; i, j)$ be the probability that there are r tree components in $\Pi^{(j)}(T)$ under $P^{(i)}$. Notice that the probability distribution $(p(r; i, j), r \in \mathbb{N})$ has finite expectation (its generating function is $\overline{\varphi}_i^{(j)}$ with the notations of Section 2.3.1), so that $\sum_r r p(r; i, j) < \infty$. Then

$$P^{(i)}(\# T^{(j)} = n) = \sum_{r \geq 1} p(r; i, j) P_r^{\mathbf{j}}(\#\Pi^{(j)}(F) = n)\,,$$

and an application of the previous paragraph and dominated convergence (using the fact that $n^{3/2} P_r^{\mathbf{j}}(\#\Pi^{(j)}(F) = n)$ is uniformly bounded) gives that

$$n^{3/2} P^{(i)}(\# T^{(j)} = n) \to C \sum_{r \geq 1} r p(r; i, j)\,, \tag{17}$$

which is the wanted result. □

Lemma 7 *The respective laws of the number of tree components of $\Pi^{(j)}(T)$ under the probability distributions $P^{(i)}(\cdot | \# T^{(j)} = n)$ converge weakly as $n \to \infty$.*

Proof. We use the notations of the previous proof, as well as the expression (17) of the constant C_{ij}. Observe that the $P^{(i)}$-probability that $\Pi^{(j)}(T)$ has r components given it has n individuals is

$$P^{(i)}(c_{\Pi^{(j)}(T)}(\varnothing) = r | \# T^{(j)} = n) = \frac{p(r; i, j) P_r^{\mathbf{j}}(\# F^{(j)} = n)}{P^{(i)}(\# T^{(j)} = n)} \underset{n\to\infty}{\longrightarrow} \frac{r p(r; i, j)}{\sum_{r'} r' p(r'; i, j)}\,,$$

and this does define a probability distribution. □

The following modification of Proposition 2 for forests with a fixed number of trees also holds:

Lemma 8 *In the case $K = 1$, assume* (H) *and take $r \in \mathbb{N}$. Let $P_r := P_r^{\mathbf{1}}$ be the law of a monotype GW forest with r tree components and offspring distribution $\mu := \mu^{(1)}$. Then the process $(n^{-1/2}H^F_{\lfloor ns\rfloor}, 0 \le s \le 1)$ under $P_r(\cdot|\#F = n)$ converges in distribution to $2\sigma^{-1}B^{\mathrm{ex}}$ as $n \to \infty$.*

Proof. Under $P_r(\cdot)$, the random variables $\#F_i, 1 \le i \le r$ are independent and distributed as the total size of a GW tree with offspring distribution μ. Simple estimates using Lemma 6 in the case $K = 1$ imply that $\liminf_{n\to\infty} n^{3/2}P_r(\#F = n) > 0$ (at least along a subsequence for which this quantity is non-zero) and for every $\varepsilon > 0$,

$$n^2 P_r\left(\exists i \neq j : \#F_i \wedge \#F_j \ge \varepsilon n, \#F = n\right) = O(1)\,,$$

so that $\max_{1\le i\le r} \#F_i/n \to 1$ in probability under $P_r(\cdot|\#F = n)$. Thus, when taking r independent copies of GW trees with offspring distribution μ, and conditioning their sum to be n, only one of the trees has a size of order n, while the others $r-1$ trees have $o(n)$ size, hence have maximal height $o(n^{1/2})$ according to Proposition 2. The result is easily deduced from these considerations, and we leave the details to the reader. □

Proof of Theorem 2. The proof starts like the one of Theorem 1. For $0 \le s \le 1$, write

$$\left| H^T_{\lfloor \#Ts\rfloor} - \frac{H^{\Pi^{(j)}(T)}_{\Lambda^T_j(\lfloor \#Ts\rfloor)}}{a_jb_j} \right| \le \left| u(\lfloor \#Ts\rfloor) - \frac{\mathrm{Anc}_T^{u(\lfloor \#Ts\rfloor)}(j)}{a_jb_j} \right| + R_n(s)\,, \tag{18}$$

where $|R_n(s)| \le (a_jb_j)^{-1}(2\max_{0\le k\le n}|H^{\Pi^{(j)}(T)}_{k-1} - H^{\Pi^{(j)}(T)}_k| + 1)$.

We start by showing the convergence of the processes $(n^{-1/2}H^{\Pi^{(j)}(T)}_{\Lambda^T_j(\lfloor \#Ts\rfloor)}/a_jb_j, 0 \le s \le 1)$ under $P^{(i)}(\cdot|\#T^{(j)} = n)$. Since $\Pi^{(j)}(T)$ under $P^{(i)}$ is a GW forest, and by first conditioning on $c_{\Pi^{(j)}(T)}(\varnothing) = r$, we obtain using Lemmas 7 and 8 that $(n^{-1/2}H^{\Pi^{(j)}(T)}_{\lfloor ns\rfloor}, 0 \le s \le 1)$ under $P^{(i)}(\cdot|\#T^{(j)} = n)$ converges in distribution to $2\overline{\sigma}_j^{-1}B^{\mathrm{ex}}$, where $\overline{\sigma}_j^2 = \mathrm{Var}\,(\overline{\mu}_j^{(j)})$, with definitions from Proposition 4.

We now show that $(n^{-1}\Lambda^T_j(\lfloor \#Ts\rfloor), 0 \le s \le 1)$ converges in probability to the identity on $[0,1]$, under the conditioned measures. Using Lemma 6 and (ii) in Proposition 6, one obtains that for some $C > 0$, for every $s \in [0,1]$,

$$\begin{aligned} &P^{\mathbf{i}}\left(\left|G^F_j(\lfloor ns\rfloor) - a_j^{-1}sn\right| \ge n^{1/2+\gamma}\middle|\#F_1^{(j)} = n\right) \\ &\le \; Cn^{3/2}P^{\mathbf{i}}\left(\left|G^F_j(\lfloor ns\rfloor) - a_j^{-1}sn\right| \ge n^{1/2+\gamma}\right) = \mathrm{oe}(n) \end{aligned} \tag{19}$$

Since F_1 under $P^{\mathbf{i}}$ has same distribution as T under $P^{(i)}$, we thus obtain that for $0 \le s \le 1$,

$$P^{(i)}(|G^T_j(\lfloor ns\rfloor) - a_j^{-1}sn| \ge n^{1/2+\gamma}|\#T^{(j)} = n) = \mathrm{oe}(n)\,, \tag{20}$$

which shows that $n^{-1}G_j^T(\lfloor ns \rfloor)$ under $P^{(i)}(\cdot|\#T^{(j)} = n)$ converges to $a_j^{-1}s$ for every $s \in [0,1]$ in probability, and thus, by the same reasoning as in Section 2.6, we obtain that $(n^{-1}G_j^T(\lfloor ns \rfloor), 0 \le s \le 1)$ converges in probability to $(a_j^{-1}s, 0 \le s \le 1)$ for the uniform topology In particular, for $s = 1$ we obtain that $n^{-1}\#T$ under $P^{(i)}(\cdot|\#T^{(j)} = n)$ converges to a_j^{-1} in probability, where we recall that we adopted the convention $G_j^T(n) = \#T$.

Now, $(n^{-1}\Lambda_j^T(\lfloor \#Ts \rfloor), 0 \le s \le 1)$ is the right-continuous inverse of $(\#T^{-1}G_j^T(\lfloor ns \rfloor), 0 \le s \le 1)$, and as such, it converges to the identity of $[0,1]$ in probability for the uniform topology.

It remains to show that the two terms on the right-hand side of (18) are $o(n^{1/2})$ in probability, uniformly in $s \in [0,1]$. First, notice that letting $\Upsilon^j := c_{\Pi^{(j)}(T)}(\varnothing)$ be the number of tree components of $\Pi^{(j)}(T)$, then the law of $\Pi^{(j)}(T)$ under $P^{(i)}(\cdot|\Upsilon^j = r)$ is the same as that of $\Pi^{(j)}(T)$ under $P_r^{\mathbf{j}}$, i.e. is that of a monotype GW forest with r tree components. Using Lemma 8, one concludes that $P^{(i)}(\sup_{0\le s\le 1} n^{-1/2}|R_n(s)| \ge \varepsilon|\#T^{(j)} = n, \Upsilon^j = r)$ converges to 0 for any $\varepsilon > 0$. Using Lemma 6, we know that the laws of Υ^j under $P^{(i)}(\cdot|\#T^{(j)} = n)$ are tight as n varies, so that we get $P^{(i)}(\sup_{0\le s\le 1} n^{-1/2}|R_n(s)| \ge \varepsilon|\#T^{(j)} = n) \to 0$ as well.

Finally, by applying (20) for $s = 1$, we obtain that $P^{(i)}(\#T > An|\#T^{(j)} = n) = \mathrm{oe}(n)$ for any $A > a_j^{-1}$. Combining that with Proposition 5, gives for such A and some $C > 0$:

$$
\begin{aligned}
&P^{(i)}\left(\max_{0\le k\le \#T}\left|u(k) - \frac{\mathrm{Anc}_T^{u(k)}(j)}{a_j b_j}\right| \ge n^{3/8} \Big| \#T^{(j)} = n\right) \\
&\le \; Cn^{3/2}P^{\mathbf{i}}\left(\max_{0\le k\le An}\left|u(k) - \frac{\mathrm{Anc}_F^{u(k)}(j)}{a_j b_j}\right| \ge n^{3/8}\right) + \mathrm{oe}(n) = \mathrm{oe}(n)\,,
\end{aligned}
$$

hence the result. □

Remark. In the companion paper [14], a similar statement as Theorem 4 was needed, with the law $P^{(i)}(\cdot|\#T^{(j)} = n)$ replaced by $P^{(i,i)}(\cdot|\#F^{(j)} = n)$, i.e. by a forest with two trees conditioned by their total number of vertices of type j. The proof of such a statement should be clear from the previous methods: by applying the transformation $\Pi^{(j)}$ to this forest, one obtains a monotype Galton-Watson forest with a random number of roots that is tight as n varies, and conditioned to have n vertices. By conditioning on its number of roots and applying Lemma 8, this implies that none but one of these trees have more that $o(n)$ vertices. Hence in the initial forest with two components, only one of the components has more than $o(n)$ vertices, and the result is now a consequence of Theorem 4, which will be proved in the next section.

3 Proof of Theorem 3

The key technical results needed to prove Theorem 3 are, like in [13], a control on the frequencies of branching events in GW trees, which will allow to prove the convergence of finite-dimensional marginals of the snake, and a bound on a Hölder norm-like quantity for the rescaled height process, which will imply the tightness.

3.1 Exponential control of branching events

For $u \in \mathbf{f} \in \mathcal{F}^{(K)}$, let $\mathrm{Anc}^u_{\mathbf{f}}(i,\mathbf{w},l,h) := \mathrm{Anc}^u_{\mathbf{f}}(e(v)=i, \mathbf{w}_{\mathbf{f}}(v)=\mathbf{w}, u \in vl\mathbf{f}_{vl}, |v| \geq |u|-h)$, i.e. the number of ancestors v of u with type i, with children's types $\mathbf{w}$, such that u is the descendent of the l-th child of v, and that are at distance at most h from u. In the sequel, when dealing with quantities of the form $\mathrm{Anc}^u_{\mathbf{f}}(\mathsf{P}_v)$, and if the last argument is h, we will understand that we consider only those ancestors v of u such that $|v| \geq |u| - h$.

Lemma 9 *Assume* $\boldsymbol{\mu}$ *satisfies* (H) *and has small exponential moments. Then for every* $\mathbf{x} \in [K]^{\mathbb{N}}$ *and* $\gamma > 0$*,*

$$P^{\mathbf{x}}\left(\max_{0\leq k\leq n}\ \max_{n^\gamma \leq h \leq |u(k)|, \mathbf{w}\in\mathcal{W}_K, 1\leq l\leq|\mathbf{w}|} \frac{|\mathrm{Anc}^{u(k)}_F(i,\mathbf{w},l,h) - h a_i b_{w_l}\zeta^{(i)}(\mathbf{w})|}{h^{1/2+\gamma}} \geq 1\right) = \mathrm{oe}(n)\,.$$

We first state an intermediate lemma. Recall the construction of the size-biased infinite tree $\widehat{\mathfrak{t}}$ of Sect. 2.1, and the spinal path $\varnothing = V_0, V_1, \ldots$. We assume $\widehat{\mathfrak{t}}$ to be constructed on some probability space $(\Omega, \mathcal{A}, P)$. If $u \in \widehat{\mathfrak{t}}$, we let $\mathbf{w}_{\widehat{\mathfrak{t}}}(u) \in \mathcal{W}_K$ be the ordered sequence of its children's types, as for finite trees.

Lemma 10 (i) *The sequence* $((e(V_n), e(V_{n+1})), n \geq 0)$ *is a Markov chain with step transition* $p_{(i,j),(i',j')} = \delta_{ji'} b_{j'} m_{jj'}/b_j$ *and equilibrium measure* $\pi_{(i,j)} = a_i b_j m_{ij}$*.*
(ii) *Conditionally on* $(e(V_k), k \geq 0)$*, the variables* $(\mathbf{w}_{\widehat{\mathfrak{t}}}(V_n), V_{n+1}(n+1), n \geq 0)$ *are independent, (here* $V_{n+1}(n+1)$ *is the last letter of* V_{n+1}*, so that* $V_{n+1} = V_n V_{n+1}(n+1)$*), with law defined by*

$$P\left((\mathbf{w}_{\widehat{\mathfrak{t}}}(V_n), V_{n+1}(n+1)) = (\mathbf{w}, l) \,|\, (e(V_k), k \geq 0)\right) = \frac{\zeta^{(e(V_n))}(\mathbf{w})}{m_{e(V_n)e(V_{n+1})}}\,,$$

for every $\mathbf{w} \in \mathcal{W}_K$ *and* $1 \leq l \leq |\mathbf{w}|$ *such that* $w_l = e(V_{n+1})$*.*

Proof. The first statement is immediate to check since we already know that $(e(V_n), n \geq 0)$ is Markov with step transition $p_{jj'} = b_{j'} m_{jj'}/b_j$ and stationary distribution $\mathbf{ab}$. The conditional independence of $\mathbf{w}_{\widehat{\mathfrak{t}}}(V_n), V_{n+1}(n+1), n \geq 0$ is easy from the construction of $\widehat{\mathfrak{t}}$, and we have

$$P((\mathbf{w}_{\widehat{\mathfrak{t}}}(V_n), V_{n+1}(n+1)) = (\mathbf{w}, l) | e(V_k), k \geq 0) = \frac{P((\mathbf{w}_{\widehat{\mathfrak{t}}}(V_n), V_{n+1}(n+1)) = (\mathbf{w}, l) | e(V_n))}{P(e(V_{n+1}) = w_l | e(V_n))}\,,$$

which amounts to the desired result. □

Proof of Lemma 9. Fix $\gamma > 0$ and choose $0 < \eta < \gamma^2$. First, we claim that

$$P^{\mathbf{x}}\left(\max_{0\leq k\leq n} c_F(u(k)) \geq n^\eta\right) = \mathrm{oe}(n)\,.$$

Indeed, under $P^{\mathbf{x}}$, the sequences $(\mathbf{w}_F(u^{(i)}(k)), k \geq 0)$, for $i \in [K]$, are independent i.i.d. sequences with respective common distribution $\zeta^{(i)}$, as follows from the Markov branching property of [8] and the fact that, when exploring the forest in depth-first order, no information on the set of children of the vertex explored at step n can be obtained before this step. Hence, the fact that $\boldsymbol{\mu}$ admits small exponential moments gives that

$$\begin{aligned} P^{\mathbf{x}}\left(\max_{i\in[K]}\max_{0\leq k\leq n} c_F(u^{(i)}(k)) > n^\eta\right) &\leq Kn\max_{i\in[K]}\zeta^{(i)}(|\mathbf{w}| > n^\eta) \\ &= Kn\max_{i\in[K]}\mu^{(i)}(|\mathbf{z}|_1 > n^\eta)\,, \end{aligned}$$

which by (3) and Markov's inequality is an $\mathrm{oe}(n)$. Since

$$\max_{0\leq k\leq n} c_{\mathbf{f}}(u(k)) \leq \max_{i\in[K]}\max_{0\leq k\leq n} c_{\mathbf{f}}(u^{(i)}(k))\,,$$

this gives the claim. Considering this, and using Lemma 4, we may restrict ourselves to showing that

$$P^{\mathbf{x}}_{\lfloor n^{1/2+\gamma}\rfloor}\left(\max_{u\in F,|u|\leq n^{1/2+\gamma}}\max_{(h,\mathbf{w},l)\in B_{n,|u|}}\frac{|\mathrm{Anc}^u_F(i,\mathbf{w},l,h)-ha_ib_{w_l}\zeta^{(i)}(\mathbf{w})|}{h^{1/2+\gamma}}\geq 1\right) = \mathrm{oe}(n)\,, \quad (21)$$

where $B_{n,h'} = \{(h,\mathbf{w},l) \in \mathbb{Z}_+ \times \mathcal{W}_K \times \mathbb{Z}_+ : n^\gamma \leq h \leq h', 1 \leq l \leq |\mathbf{w}| \leq n^\eta\}$. By using Lemma 1 in a similar way as in the proof of Proposition 5, the probability in the left-hand side of (21) is bounded above by

$$C\sum_{r=1}^{\lfloor n^{1/2+\gamma}\rfloor}\sum_{h'=\lfloor n^\gamma\rfloor}^{\lfloor n^{1/2+\gamma}\rfloor}\widehat{P}^{(x_r),h'}\left(\max_{(h,\mathbf{w},l)\in B_{n,h'}}\frac{|\mathrm{Anc}^V_T(i,\mathbf{w},l,h)-ha_ib_{w_l}\zeta^{(i)}(\mathbf{w})|}{h^{1/2+\gamma}}\geq 1\right)\,, \quad (22)$$

where $C = \max_j b_j/\min_j b_j$. Writing $Q_{i,\mathbf{w},l,h}$ for the quotient appearing in the probability, we have

$$\widehat{P}^{(x_r),h'}\left(\max_{(h,\mathbf{w},l)\in B_{n,h'}} Q_{i,\mathbf{w},l,h}\geq 1\right) \leq n^{1/2+\gamma+\eta}K^{n^\eta}\max_{(h,\mathbf{w},l)\in B_{n,h'}}\widehat{P}^{(x_r),h'}\left(Q_{i,\mathbf{w},l,h}\geq 1\right)\,, \quad (23)$$

where $n^{1/2+\gamma+\eta}K^{n^\eta}$ bounds the cardinality of $B_{n,h'}$ with $h' \leq n^{1/2+\gamma}$. Let $\mathrm{Anc}^u_{\mathbf{t}}(i,j,h) = \mathrm{Anc}^u_{\mathbf{t}}(e(v)=i, e(vu_{|v|+1})=j, |v|\geq|u|-h)$ be the number of strict ancestors v of $u \in \mathbf{t} \in \mathcal{T}^{(K)}$ such that $e(v)=i$, $|v|\geq|u|-h$ and $e(v')=j$ whenever v' is the child of v with $v\vdash v'\vdash u$, so that according to (ii) in Lemma 10, under $\widehat{P}^{(x_r),h'}$ and given $\mathrm{Anc}^V_T(i,j,h)=k$, the random variable $\mathrm{Anc}^V_T(i,\mathbf{w},l,h)$ is binomial with parameters $(k, \zeta^{(i)}(\mathbf{w})/m_{ij})$, for every $\mathbf{w}, l$ with $w_l = j$.

By Lemmas 10, (i) and 5, we have for every $n^\gamma \leq h \leq h'$,

$$\begin{aligned} \max_{i,j\in[K]}\widehat{P}^{(x_r),h'}(|\mathrm{Anc}^V_T(i,j,h)-ha_ib_jm_{ij}|\geq h^{1/2+\gamma}) &\leq \exp(-h^\gamma) \\ &\leq \exp(-n^{\gamma^2})\,, \quad (24) \end{aligned}$$

for every large enough n. In particular, for some $c > 0$, $\inf_{i,j\in[K]} \mathrm{Anc}_T^V(i,j,h) \geq ch$ with $\widehat{P}^{(x_r),h'}$-probability $\leq \exp(-n^{\gamma^2})$ for every $n^\gamma \leq h \leq h'$ and large enough n. Moreover,

$$\begin{aligned} &\widehat{P}^{(x_r),h'}(Q_{i,\mathbf{w},l,h} \geq 1) \\ &\leq \widehat{P}^{(x_r),h'}\left(\left|\mathrm{Anc}_T^V(i,\mathbf{w},l,h) - \frac{\zeta^{(i)}(\mathbf{w})\mathrm{Anc}_T^V(i,w_l,h)}{m_{ij}}\right| \geq \frac{h^{1/2+\gamma}}{2}\right) \\ &\quad + \widehat{P}^{(x_r),h'}\left(|\mathrm{Anc}_T^V(i,w_l,h) - ha_j b_{w_l} m_{ij}|\zeta^{(i)}(\mathbf{w})/m_{ij} \geq \frac{h^{1/2+\gamma}}{2}\right). \end{aligned}$$

Since $\zeta^{(i)}(\mathbf{w}) \leq 1$ and using (24), the second probability is $\leq \exp(-n^{\gamma^2})$ for large n, as long as $(h,\mathbf{w},l) \in B_{n,h'}$ for $h' \geq n^\gamma$.

Also, according to Hoeffding's inequality for binomial distributions, it holds that

$$\begin{aligned} \widehat{P}^{(x_r),h'}\left(\frac{|\mathrm{Anc}_T^V(i,\mathbf{w},l,h) - k\zeta^{(i)}(\mathbf{w})/m_{ij}|}{h^{1/2+\gamma}} \geq 1 \,\middle|\, \mathrm{Anc}_T^V(i,j,h) = k\right) \\ \leq 2\exp(-2h^{1+2\gamma}/k), \end{aligned}$$

which for $h \geq n^\gamma$ and $k \leq ch$ is less than $2\exp(-2n^{2\gamma^2}/c)$. This is enough to conclude that for every $n^\gamma \leq h \leq h' \leq n^{1/2+\gamma}$ and large enough n,

$$\max_{(h,\mathbf{w},l)\in B_{n,h'}} \widehat{P}^{(x_r),h'}(Q_{i,\mathbf{w},l,h} \geq 1) \leq \exp(-n^{\gamma^2}).$$

Combining with (23) and then (22), gives the result (notice that by our choice of η, we have $\gamma^2 > \eta$ so that the quantity on the right-hand side of (23) is an $\mathrm{oe}(n)$). □

We also have the following controls

Lemma 11 *Under the same hypotheses as Lemma 9,*
(i) *For every $\Gamma, \eta > 0$ there exists $C > 0$ so that for every large n,*

$$P^{\mathbf{x}}\left(\max_{0\leq k\leq n} c_F(u(k)) \geq C\log n\right) \leq n^{-\Gamma}.$$

(ii) *For every $c > 0$ there exists $\gamma \in (0,1)$ so that*

$$P^{\mathbf{x}}\left(\max_{0\leq k\leq n}\max_{n^\eta\leq h\leq|u(k)|} h^{-\gamma}\mathrm{Anc}_F^{u(k)}(c_F(v) \geq c\log n, h) \geq 1\right) = \mathrm{oe}(n).$$

(iii) *For every $M > 0$, there is a constant $C_M > 0$ such that*

$$P^{\mathbf{x}}\left(\max_{0\leq k\leq n}\max_{n^\eta\leq h\leq|u(k)|}\sup_{m\geq 1} m^M h^{-1}\mathrm{Anc}_F^{u(k)}(c_F(v) = m, h) \geq C_M\right) = \mathrm{op}(n).$$

Here $x_n = \mathrm{op}(n)$ means that for every $\Gamma > 0$, $x_n \leq n^{-\Gamma}$ for every large n.

Proof. Point (i) follows similar lines as the beginning of the proof of Lemma 9. We bound the probability of interest by

$$Kn \max_{i\in[K]} \mu^{(i)}(|\mathbf{z}|_1 \geq C\log n) \leq Kn^{1-C\varepsilon} \max_{i\in[K]} \mu^{(i)}(\exp(\varepsilon|\mathbf{z}|_1))$$

for some appropriate $\varepsilon > 0$, and choose C so that $1 - C\varepsilon < -\Gamma$.

For (ii), we see that, using Lemma 4, it suffices to bound the probability

$$P^{\mathbf{x}}_{\lfloor n^{1/2+\eta}\rfloor}\left(\max_{u\in F, |u|\leq n^{1/2+\eta}} \max_{n^\eta\leq h\leq |u|} h^{-\gamma}\mathrm{Anc}^u_F(c_F(v) \geq c\log n, h) \geq 1\right).$$

Using Lemma 1, this is bounded up to some fixed multiplicative constant by

$$\sum_{r=1}^{\lfloor n^{1/2+\eta}\rfloor} \sum_{h'=\lfloor n^\eta\rfloor}^{\lfloor n^{1/2+\eta}\rfloor} \max_{n^\eta\leq h\leq h'} \widehat{P}^{(x_r),h'}\left(\mathrm{Anc}^V_T(c_T(v)\geq c\log n, h) \geq h^\gamma\right).$$

By conditioning on the types of the ancestors of V under $\widehat{P}^{(x_r),h'}$, we may again use (ii) in Lemma 10 to argue that the number of ancestors v of V such that $c_F(v) \geq c\log n$ and $|v| \geq |V| - h$ is a sum of independent Bernoulli random variables with parameters of the form

$$\frac{\zeta^{(i)}(|\mathbf{w}| \geq c\log n)}{m_{ij}} \leq n^{-c\varepsilon}\frac{\max_{i\in[K]}\zeta^{(i)}(\exp(\varepsilon|\mathbf{w}|))}{\min_{ij:m_{ij}\neq 0} m_{ij}},$$

so that it is stochastically bounded by a sum of h i.i.d. Bernoulli random variables with parameter $h^{-c\varepsilon/(1/2+\eta)} \leq h^{-c\varepsilon}$ as soon as $0 < \eta < 1/2$. Now if $X_1, \ldots, X_n$ are i.i.d. Bernoulli random variables with common parameter p on some probability space $(\Omega, \mathcal{A}, P)$, $P(\sum_{i=1}^n X_i \geq x) \leq \exp(-\lambda x + np(\exp(\lambda) - 1))$ for every $\lambda > 0$, by Markov's inequality. Therefore, by choosing λ small enough, for some $C > 0$, and by unconditioning on the sequence of types in the end,

$$\widehat{P}^{(x_r),h'}\left(\mathrm{Anc}^V_T(c_T(v) \geq c\log n, h) \geq 2h^{1-c\varepsilon}\right) \leq \exp(-Ch^{1-c\varepsilon}),$$

hence the result by choosing $\gamma \in (1 - c\varepsilon, 1)$, provided $n^\eta \leq h$ and n is large enough.

Point (iii) is obtained following similar lines. We are even going to prove the result with $\mathrm{Anc}^{u(k)}_F(c_F(v) = m, h)$ replaced by $\mathrm{Anc}^{u(k)}_F(c_F(v) = m, h)$. Fix $\Gamma > 0$, then the probability under consideration is bounded up to a multiplicative constant by

$$\sum_{r=1}^{\lfloor n^{1/2+\eta}\rfloor} \sum_{h'=\lfloor n^\eta\rfloor}^{\lfloor n^{1/2+\eta}\rfloor} \widehat{P}^{(x_r),h'}\left(\max_{1\leq m\leq C\log n} \max_{n^\gamma\leq h\leq h'} m^M h^{-1}\mathrm{Anc}^V_T(c_T(v) \geq m, h) \geq C_M\right) + o(n^{-\Gamma}),$$

for some $C > 0$, where we used point (i) of the lemma to bound the maximal degree of a vertex in the tree, and then Lemma 1. The first term is in turn bounded by

$$n^{1+2\eta} \max_{i\in[K]} \max_{n^\eta\leq h\leq h'\leq n^{1/2+\eta}} \max_{1\leq m\leq C\log n} \widehat{P}^{(i),h'}(m^M h^{-1}\mathrm{Anc}^V_T(c_T(v) \geq m, h) \geq C_M m^{-M} h).$$

Using again the large deviation inequality as above, we obtain that the probability on the right hand-side is bounded by $\exp(-\lambda h(C_M m^{-M} - C' \exp(-\xi m)))$ for some $\lambda > 0$ and where C', ξ are such that

$$\max_{i\in[K]} \zeta^{(i)}(|\mathbf{w}| \geq x) \leq 2^{-1} C' \exp(-\xi x)\,.$$

By choosing C_M large enough, we obtain that

$$\exp(-\lambda h(C_M m^{-M} - C' \exp(-\xi m))) \leq \exp(-\lambda' h m^{-M})$$

for some $\lambda' > 0$ and every $m \geq 1$. Since we consider only the terms $m \leq C \log n$ and $h \geq n^\eta$, we get that $hm^{-M} \geq n^\eta (C \log n)^{-M}$ which is $\geq n^{\eta/2}$ for large n, and this allows to conclude. □

3.2 Finite-dimensional marginals

We are now able to show the convergence of finite-dimensional marginals for the spatial process S. Notice the slightly loosened hypotheses when compared with Theorem 3.

Proposition 7 *Assume that $\boldsymbol{\zeta}$ satisfies* (H) *and admits small exponential moments. Also assume that the spatial displacement laws $\nu_{i,\mathbf{w}}$ are non-degenerate, centered and have a finite variance $\langle \nu_{i,\mathbf{w}}, |y|_2^2\rangle$, satisfying*

$$\max_{i\in[K]} \langle \nu_{i,\mathbf{w}}, |y|_2^2\rangle = O(|\mathbf{w}|^D)\,, \tag{25}$$

for some $D > 0$. Then, for any $\mathbf{x} \in [K]^{\mathbb{N}}$, jointly with the convergence (i) *in Theorem 1, for any $0 \leq s_1 < s_2 < \ldots < s_k$, it holds that under $P^{\mathbf{x}}$,*

$$\left(\frac{S^F_{\lfloor ns_j\rfloor}}{n^{1/4}}, 1 \leq j \leq k\right) \xrightarrow[n\to\infty]{d} \left(R_{s_j}, 1 \leq j \leq k\right)\,,$$

where R has the law described in Theorem 3. If $\mathbf{x} = \mathbf{i}$, it also holds jointly with (ii) *in Theorem 1.*

Proof. Since the proof is very similar to that of [13, Proposition 28] and the general case does not involve more sophisticated tools besides notational annoyances, we give the detailed argument only for $k = 2$.

Making use of Skorokhod's representation theorem, we will assume that the discrete snakes we are considering are defined on a probability space $(\Omega, \mathcal{A}, P)$ so that following holds. First, this space supports a sequence of processes $H^n, \Lambda_i^n, i \in [K]$ with same distribution as $n^{-1/2} H^F_{\lfloor n\cdot\rfloor}, \Lambda_i^F(\lfloor n\cdot\rfloor), i \in [K]$ under $P^{\mathbf{x}}$, such that H^n converges almost-surely to a process B which is $2\sigma^{-1}$ times a standard Brownian motion, while Λ_i^n/a_i converges almost-surely to the identity function. Otherwise said, the convergence of Theorem 1 is almost-sure. Second, for every $n \geq 1$, the processes $(n^{1/2} H^n_{k/n}, \Lambda_i^n(k/n), i \in [K], k \geq 0)$

are the height and type-counting processes of a unique multitype random forest $(F^n, e^n) \in \mathcal{F}^{(K)}$, and we assume that the discrete snakes are defined by using a family of random variables $(Y^n_u, u \in F^n)$ that are supported on $(\Omega, \mathcal{A}, P)$. That is, given F^n, the vectors $(Y^n_{ul}, 1 \le l \le c_{F^n}(u)), u \in F^n$ are independent with respective distributions $\nu_{e^n(u),\mathbf{w}_{F^n}(u)}$, and we let

$$S^n_k = \sum_{v \vdash u_{F^n}(k)} Y^n_v \,, \qquad k \ge 0$$

be the associated discrete snake.

Let $s < t$ be given. For simplicity write $u = u_{F^n}(\lfloor ns \rfloor), u' = u_{F^n}(\lfloor nt \rfloor)$ and $\check{u}$ the most recent common ancestor of u, u'. The vector (S^n_{ns}, S^n_{nt}) is the image by the application $(x, y, z) \mapsto (x + z, y + z)$ of the vector $(\mathsf{L}^n_1, \mathsf{R}^n_1, \mathsf{C}^n)$ where

$$\mathsf{L}^n_1 = \sum_{\check{u} \vdash v \vdash u, v \neq \check{u}} Y^n_v \,, \quad \mathsf{R}^n_1 = \sum_{\check{u} \vdash v \vdash u', v \neq \check{u}} Y^n_v \,, \quad \mathsf{C}^n = \sum_{v \vdash \check{u}} Y^n_v \,.$$

We aim at proving that given B, $n^{-1/4}(\mathsf{L}^n_1, \mathsf{R}^n_1, \mathsf{C}^n)$ converges in distribution to a triple of independent Gaussian variables with respective variances $\Sigma^2 B_s, \Sigma^2 B_t, \Sigma^2 \check{B}_{s,t}$, where $\check{B}_{s,t} = \inf_{s \le u \le t} B_u$.

First note that conditionally on F^n, $\mathsf{L}^n_1, \mathsf{R}^n_1, \mathsf{C}^n$ are almost independent as they are sums involving terms which are all independent but for one: if l, l' are such that $u \in \check{u}lF^n_{\check{u}l}$ and $u' \in \check{u}l'F^n_{\check{u}l'}$, i.e. u and u' are in the l-th and l'-th subtree pending from $\check{u}$, then $Y^n_{\check{u}l}$ and $Y^n_{\check{u}l'}$ may be dependent. However, letting $\mathsf{L}^n_1 = Y^n_{\check{u}l} + \mathsf{L}^n_2$ and $\mathsf{R}^n_1 = Y^n_{\check{u}l'} + \mathsf{R}^n_2$, the vector $(\mathsf{L}^n_2, \mathsf{R}^n_2, \mathsf{C}^n)$ has three independent components. It is thus sufficient to prove that $n^{-1/4} \max_l Y^n_{\check{u}l}$ and $n^{-1/4} Y^n_{\check{u}}$ both converge to 0 as $n \to \infty$, and to show the individual convergences of $n^{-1/4}\mathsf{L}^n, n^{-1/4}\mathsf{R}^n, n^{-1/4}\mathsf{C}^n$ to three Gaussian variables with the correct variances, where L^n and R^n are defined as L^n_1 and R^n_1, but allowing the term $v = \check{u}$ in the sum.

Using Lemma 11 for $\Gamma > 1$ and applying the Borel-Cantelli Lemma, we have that almost-surely, for n large enough, $c_{F^n}(u(k)) \le C \log n$ for every $k \le nt$ and for some $C > 0$. Conditionally on F^n and on this event, the expectation of $\max_l (Y^n_{\check{u}l})^2$ is bounded above by $(C \log n)^D$ by our assumption on moments, from which it follows that $n^{-1/4} \max_l Y^n_{\check{u}l}$ does converge to 0. The reasoning is similar for estimating $n^{-1/4} Y^n_{\check{u}}$.

Let us now deal with L^n, the case of R^n being similar. The sum defining L^n is a sum along a part of the ancestral line of u, whose height is approximately $\sqrt{n}(B_s - \check{B}_{s,t})$. Note that a.s., $B_s - \check{B}_{s,t} > 0$ by standard properties of Brownian motion, so that the sum in question has an order of $\sqrt{n}$ terms. Write

$$\mathsf{L}^n = \sum_{\check{u} \vdash v \vdash u} Y^n_v \mathbb{1}_{\{c_{F^n}(u) \le c\}} + \sum_{\check{u} \vdash v \vdash u} Y^n_v \mathbb{1}_{\{c_{F^n}(u) > c\}} \,.$$

We call these two sums L^n_c and $\tilde{\mathsf{L}}^n_c$.

We first argue that

$$\lim_{c \to \infty} \limsup_{n \to \infty} P(n^{-1/4} |\tilde{\mathsf{L}}^n_c| > \varepsilon | B) = 0 \,. \tag{26}$$

Letting $h = |u| - |\check{u}|$, we noticed that given B, h is of the order $n^{1/2}$. In particular, using the Borel-Cantelli Lemma and (iii) in Lemma 11 for $M = D + 2$, for some constant C_D, and every $m \geq 1$, a.s. $\mathrm{Anc}^u_{F^n}(c_{F^n}(v) = m, h) \leq C_D m^{-D-2} h$, for every large enough n. Then by Chebychev's Inequality and conditional independence of the terms,

$$\begin{aligned}
P(n^{-1/4}|\tilde{\mathsf{L}}^n_c| > \varepsilon | B, F^n) &\leq n^{-1/2}\varepsilon^{-2} \sum_{i\in[K]} \sum_{\mathbf{w}\in\mathcal{W}_K, |\mathbf{w}|\geq c} \sum_{1\leq l\leq |\mathbf{w}|} \mathrm{Anc}^u_{F^n}(i,\mathbf{w},l,h)\langle \nu_{i,\mathbf{w}}, y_l^2\rangle \\
&\leq Cn^{-1/2}\varepsilon^{-2} \sum_{m\geq c} m^D \mathrm{Anc}^u_{F^n}(c_{F^n}(v) = m, h) \\
&\leq C_D \varepsilon^{-2} n^{-1/2} h \sum_{m\geq c} m^{-2},
\end{aligned}$$

at least for (ω-dependent) large enough n. Notice that this upper bound does not depend on F^n. Since $n^{-1/2}h$ converges to $B_s - \check{B}_{s,t}$, taking the conditional expectation given B and applying the reverse Fatou Lemma, we obtain

$$\begin{aligned}
\limsup_{n\to\infty} P(n^{-1/4}|\tilde{\mathsf{L}}^n_c| > \varepsilon | B) &\leq E[\limsup_{n\to\infty} P((n^{-1/4}|\tilde{\mathsf{L}}^n_c| > \varepsilon | B, F^n)|B] \\
&\leq C_D\varepsilon^{-2}(B_s - \check{B}_{s,t}) \sum_{m\geq c} m^{-2},
\end{aligned}$$

which goes to 0 as $c \to \infty$.

On the other hand, fixing $\eta, \varepsilon > 0$, we may apply Lemma 9 and Borel-Cantelli to obtain that a.s., for large enough n, and every $0 \leq k \leq nt, n^\eta \leq h \leq |u(k)|, \mathbf{w} \in \mathcal{W}_K$ with $|\mathbf{w}| \leq c$ and $1 \leq l \leq |\mathbf{w}|$,

$$|\mathrm{Anc}^{u(k)}_{F^n}(i,\mathbf{w},l,h) - h a_i b_{w_l} \zeta^{(i)}(\mathbf{w})| \leq h^{1/2+\eta} \leq \varepsilon h a_i b_{w_l} \zeta^{(i)}(\mathbf{w}). \tag{27}$$

Now, we can redisplay L^n_c as the finite sum

$$\sum_{i\in[K]} \sum_{\mathbf{w}\in\mathcal{W}_K, |\mathbf{w}|\leq c} \sum_{l=1}^{|\mathbf{w}|} \sum_{\check{u}\vdash v\vdash u} Y^n_v \mathbb{1}_{\{e(v)=i, \mathbf{w}_{F^n}(v)=\mathbf{w}, u\in vlF^n_{vl}\}},$$

and given F^n, the last summation has $\mathrm{Anc}^u_{F^n}(i,\mathbf{w},l,h)$ i.i.d. terms (for $h = |u| - |\check{u}|$) with variance $\langle \nu_{i,\mathbf{w}}, y_l^2\rangle$. Using (27), we can apply the the central limit theorem conditionally on B, F^n, and obtain that $n^{-1/4}\mathsf{L}^n_c$ converges in distribution to a Gaussian variable with variance $(B_s - \check{B}_{s,t})\Sigma_c^2$ where

$$\Sigma_c^2 = \sum_{i\in[K]} a_i \sum_{\mathbf{w}\in\mathcal{W}_K, |\mathbf{w}|\leq c} \zeta^{(i)}(\mathbf{w}) \sum_{l=1}^{|\mathbf{w}|} b_{w_l} \langle \nu_{i,\mathbf{w}}, y_l^2\rangle.$$

As $c \to \infty$, this converges to the constant Σ^2 of the statement of the theorem, which implies the result when combined with (26).

To complete the proof, it remains to show the convergence of $n^{-1/4}\mathsf{C}^n$. The argument is essentially the same, where the height h under consideration is of the order $\sqrt{n}\check{B}_{s,t}$ rather than $\sqrt{n}(B_s - \check{B}_{s,t})$. The argument above is unchanged in the case $\check{B}_{s,t} > 0$, but is not valid anymore if $\check{B}_{s,t} = 0$. In that case, we claim that the result is in fact trivial as we have $|\check{u}| = 0$, so that $\mathsf{C}^n = 0$. Indeed, assume for a moment that $\mathbf{x} = \mathbf{i}$ for some $i \in [K]$. By Theorem 1, and plugging a new random element in our use of Skorokhod's representation Theorem, we may assume that $(n^{-1/2}\Upsilon^{F^n}_{\lfloor ns \rfloor}, s \geq 0)$ converges a.s. to a multiple of the local time L^0 of B at level 0. Since $\check{B}_{s,t} = 0$ and $B_s B_t > 0$, there is an increase time of L^0 between times s and t. Therefore, the function Υ^{F^n} increases as well between these times, at least for large enough n. This means that u and u' are in two different tree components of F^n, and thus $\check{u} = \varnothing$.

The case where $\mathbf{x}$ is any element of $[K]^{\mathbb{N}}$ is a slight elaboration of the preceding argument, which we briefly sketch. If $\check{B}_{s,t} = 0$ but $|\check{u}| > 0$ for infinitely many n, this means that for these values of n, the vertices u, u' belong to the same tree component of F^n, whose root is of type i, say. Now, $n^{-1/2}H^{\Pi^{(i)}(F^n)}_{\lfloor n\cdot \rfloor}$ converges to a multiple of the Brownian motion $B_{\cdot/a_i}$ by the proof of Theorem 1, while $n^{-1/2}\Upsilon^{\Pi^{(i)}(F^n)}_{\lfloor n\cdot \rfloor}$ converges to a multiple of the local time of the latter. Necessarily, there is a time of increase of this local time between the times $a_i s$ and $a_i t$, hence the function $n^{-1/2}\Upsilon^{\Pi^{(i)}(F^n)}_{\lfloor n\cdot \rfloor}$ increases also during that time interval, whose length corresponds asymptotically to the fraction of vertices of type i that appear between u and u' in depth-first order. This is a contradiction with the fact that u and u' belong to the same tree component of F^n. $\square$

3.3 Hölder norm bounds

Proposition 8 *Let $\boldsymbol{\mu}$ satisfy the basic assumption* (H), *and admit small exponential moments. For every $A > 0$ and $\alpha \in (0, 1/4)$, for every $\varepsilon > 0$, there exists $C > 0$ such that for every $\mathbf{x} \in [K]^{\mathbb{N}}$,*

$$\sup_{n\in\mathbb{N}} P^{\mathbf{x}}\left(\sup_{s\neq t\in[0,A], ns, nt\in\mathbb{Z}_+} \frac{|H^F_{ns} - H^F_{nt}|}{\sqrt{n}|s-t|^{\alpha}} > C\right) \leq \varepsilon\,.$$

Remark. As will be shown below, in the particular case $K = 1$, we are able to prove the same assertion with $\alpha < 1/2$ rather than $\alpha < 1/4$. In [13], we could also obtain the result for $\alpha < 1/2$ because of the particular nature of the multitype trees we were considering, i.e. alternating types, corresponding to an antidiagonal mean matrix. In the general case, our method of approximation by the monotype case does not seem to be fine enough to obtain the best estimate.

Proof. We first prove the result in the case $K = 1$, and let $\mu := \mu^{(1)}$, $P := P^{\mathbf{1}}$. Our proof is partly inspired from that of [4, Theorem 1.4.4], of which we can interpret Proposition 8 to be a discrete counterpart.

Recall e.g. from [4] that if we let

$$V_n^{\mathbf{f}} = \sum_{k=1}^{n} (c_{\mathbf{f}}(u(k)) - 1), \qquad n \geq 0$$

be the *Lukaciewicz* walk associated with the forest $\mathbf{f}$, then the height process of $\mathbf{f}$ is given by

$$H_n^{\mathbf{f}} = \# \left\{ k \in \{0, 1, \ldots, n-1\} : V_k^{\mathbf{f}} = \min_{k \leq l \leq n} V_l^{\mathbf{f}} \right\}. \tag{28}$$

Under P, V^F is a random walk on $\mathbb{Z}$ with centered step distribution $\mu(\cdot + 1)$ on $\{-1, 0, 1, 2, \ldots\}$.

Now, suppose $0 \leq s < t \leq A$ are such that $ns, nt \in \mathbb{Z}_+$. Write $\lambda(x) = \max\{l \in [0, ns] : V_l^F \leq x\}$. Using (28), we have

$$\begin{aligned} H_{nt}^F - H_{ns}^F &= \# \left\{ k \in [ns, nt) : V_k^F = \min_{k \leq l \leq nt} V_l^F \right\} \\ &\quad - \# \left\{ \lambda \left(\min_{ns \leq l \leq nt} V_l^F \right) < k < ns : V_l^F = \min_{k \leq l \leq ns} V_l^F \right\}, \end{aligned} \tag{29}$$

and the rest of the proof will consist in estimating the moments of the two terms above, which correspond to the lengths of the branches of F from $u(ns), u(nt)$ down to their most recent common ancestor. By the time reversal property for walks,

$$(\widehat{V}_k^{(n)} = V_n^F - V_{n-k}^F, 0 \leq k \leq n) \stackrel{d}{=} (V_k^F, 0 \leq k \leq n),$$

the first term in (29) is equal in distribution, under P, to

$$W_{n(t-s)} := \# \left\{ 1 \leq k \leq n(t-s) : V_k^F = \max_{0 \leq l \leq k} V_l^F \right\},$$

the number of (weak) records of V^F before time $n(t-s)$.

Let $M_n = \max_{0 \leq k \leq n} V_k^F$. Let $\tau_0 = 0$, and $\tau_i, i \geq 1$ be the i-th record time, i.e. the i-th time $\tau \geq 1$ such that $V_\tau^F = M_\tau$. Then it is easy and well-known that $(\tau_i - \tau_{i-1}, i \geq 1)$ are a sequence of i.i.d. random variables. Moreover, since V^F is centered and its increments have finite second moment under P, it is a consequence of the proof of [6, XII,7 Theorem 1a] and the discussion before that the Laplace exponent $\phi(s) = -\log E[\exp(-s\tau_1)] \sim C s^{1/2}$ as $s \to 0$ for some $C > 0$ (Feller considers the case of strict ladder epochs, but the treatment of weak ones is similar). Now, for any $p > 1$, and integer u,

$$\begin{aligned} E[W_u^p] &= p \int_0^\infty x^{p-1} P(W_u \geq x) \, \mathrm{d}x \\ &= p \int_0^\infty x^{p-1} P\left(\sum_{i=1}^{x} (\tau_i - \tau_{i-1}) \leq u \right) \mathrm{d}x \\ &\leq pe \int_0^\infty x^{p-1} E\left[\exp\left(-\sum_{i=1}^{x} \frac{\tau_i - \tau_{i-1}}{u} \right) \right] \mathrm{d}x \leq C' \phi(u^{-1})^{-p} \leq C'' u^{p/2}, \end{aligned}$$

for some $C', C'' > 0$. Since ns, nt are distinct integers, we showed that $E[W^p_{n(t-s)}] \leq C_1 n^{p/2}|s-t|^{p/2}$ uniformly in such n, s, t, where $C_1 = C_1(\mu, p) > 0$.

Let us now handle the second term in (29). Using time-reversal, we see that this equals

$$\#\left\{n(t-s) < k < nt \wedge \kappa\left(\max_{1\leq l\leq n(t-s)} V^F_l\right) : V^F_k = \max_{n(t-s)\leq l\leq k} V^F_l\right\}$$

in distribution, where $\kappa(x) = \min\{k \geq n(t-s) : V^F_k \geq x\}$ (with the convention $\min\varnothing = \infty$). By using Markov's property at time $n(t-s)$, this has same distribution as $W_{ns\wedge\kappa(\widetilde{M}_{n(t-s)}-\widetilde{V}_{n(t-s)})-1}$, where W is defined as above, while $\widetilde{V}$ is an independent copy of V^F with maximum process $\widetilde{M}$. By monotonicity this is less than $W_{\kappa(\widetilde{M}_{n(t-s)}-\widetilde{V}_{n(t-s)})}$. Let us prove that $E[W^p_{\kappa(x)}] \leq Cx^p$ for every $x \geq 0$, for some $C > 0$. To this end, notice that

$$M_{n(t-s)} = \sum_{i=1}^{W_{n(t-s)}} (V^F_{\tau_i} - V^F_{\tau_{i-1}}), \tag{30}$$

and it is a classical result of fluctuation theory that the variables $V^F_{\tau_i} - V^F_{\tau_{i-1}}$ are independent with common distribution $P(V^F_{\tau_1} = i) = \mu([i+1, \infty)), i \geq 0$, so their mean is $\sigma^2/2$, where σ^2 is the variance of μ, and notice that these variables have small exponential moments. Now, the usual large deviations theorem shows that for some $a, N > 0$ and for every $n \geq N$,

$$P\left(\frac{\sum_{i=1}^n (V^F_{\tau_i} - V^F_{\tau_{i-1}})}{n} < \frac{\sigma^2}{4}\right) \leq \exp(-an). \tag{31}$$

Now, using (30) in the second equality,

$$\begin{aligned}
E[W^p_{\kappa(x)}] &= p\int_0^\infty u^{p-1} P(W_{\kappa(x)} > u)\mathrm{d}u \\
&= p\int_0^\infty u^{p-1} P\left(\sum_{i=1}^{\lceil u\rceil}(V^F_{\tau_i} - V^F_{\tau_{i-1}}) < x\right)\mathrm{d}u \\
&= px^p\int_0^\infty v^{p-1}\, P\left(\sum_{i=1}^{\lceil xv\rceil}(V^F_{\tau_i} - V^F_{\tau_{i-1}}) < \frac{xv}{v}\right)\mathrm{d}v \\
&\leq px^p\left(\int_0^{4\sigma^{-2}} v^{p-1}\mathrm{d}v + \int_{4\sigma^{-2}}^\infty v^{p-1} P\left(\sum_{i=1}^{\lceil xv\rceil}(V^F_{\tau_i} - V^F_{\tau_{i-1}}) < \frac{\sigma^2 xv}{4}\right)\mathrm{d}v\right).
\end{aligned}$$

Now, as soon as x is large enough, i.e. $4x\sigma^{-2} \geq N$, where N is defined before (31), the probability in the second integral is bounded by $\exp(-axv) \leq \exp(-v)$ if we further ask $x > a^{-1}$. Thus the wanted bound on $E[W^p_{\kappa(x)}]$. By the independence of $\widetilde{V}$ and V^F, we conclude that

$$\begin{aligned} E[(W_{\kappa(\widetilde{M}_{n(t-s)}-\widetilde{V}_{n(t-s)})})^p] &\leq CE[(M_{n(t-s)}-V^F_{n(t-s)})^p] \\ &\leq 2^{p-1}C\left(1+\left(\frac{p}{p-1}\right)^p\right)E[(V^F_{n(t-s)})^p], \end{aligned} \tag{32}$$

where we used Doob's inequality $E[M^p_{n(t-s)}] \leq (p/(p-1))^p E[(V^F_{n(t-s)})^p]$, since V^F is centered. Now we use the following consequence of Rosenthal's inequality [15, Theorem 2.10]: if $X_1,\ldots,X_n$ are independent centered random variables (not necessarily identically distributed) defined on some probability space $(\tilde{\Omega},\tilde{\mathcal{A}},\tilde{P})$, then for every $p\geq 2$ there exists $C(p)$ such that

$$\tilde{E}[|X_1+\ldots+X_n|^p] \leq C(p)n^{p/2-1}\sum_{i=1}^n \tilde{E}[|X_i|^p]. \tag{33}$$

This shows that $E[(V^F_{n(t-s)})^p] \leq C'(p)n^{p/2}|s-t|^{p/2}$ for some $C'(p)>0$, for every s,t such that $ns,nt\in\mathbb{Z}_+$, and therefore the same kind of upper bound holds for the quantity in (32).

Putting things together, we have obtained that for every $p\geq 2$ and some $C_2 = C_2(\mu,p)>0$,

$$\sup_{n\geq 1}\sup_{s,t\geq 0, ns,nt\in\mathbb{Z}_+} E\left[\left|\frac{H^F_{ns}-H^F_{nt}}{\sqrt{n}}\right|^p\right] \leq C_2|s-t|^{p/2}.$$

Let now $H^F_{\{s\}}, s\geq 0$ be defined by linear interpolation between linear abscissa. Then, it is elementary that

$$\sup_{n\geq 1}\sup_{s\neq t\geq 0} E\left[\left|\frac{H^F_{\{ns\}}-H^F_{\{nt\}}}{\sqrt{n}}\right|^p\right] \leq C_3|s-t|^{p/2},$$

for some $C_3>0$. The uniform estimate in Kolmogorov's criterion [18, Theorem 3.4.16] finally entails the result, for $K=1$.

The general case is obtained by using the contraction function $\Pi^{(1)}$. We have, for $0\leq s,t\leq A$ and $ns,nt\in\mathbb{Z}_+$,

$$|H^F_{ns}-H^F_{nt}| \leq \frac{|H^{\Pi^{(1)}(F)}_{\Lambda^F_1(ns)-1}-H^{\Pi^{(1)}(F)}_{\Lambda^F_1(nt)-1}|}{a_1b_1} + 2\max_{0\leq k\leq An}\left|H^F_k - \frac{H^{\Pi^{(1)}(F)}_{\Lambda^F_1(k)-1}}{a_1b_1}\right|.$$

On the one hand, by the case $K=1$ and Proposition 4, we have with high $P^{\mathbf{x}}$-probability, uniformly in $n\in\mathbb{N}$,

$$\frac{|H^{\Pi^{(1)}(F)}_{\Lambda^F_1(ns)-1}-H^{\Pi^{(1)}(F)}_{\Lambda^F_1(nt)-1}|}{\sqrt{n}|s-t|^\alpha} \leq C\frac{|n^{-1}\Lambda^F_1(ns)-n^{-1}\Lambda^F_1(nt)|^\alpha}{|s-t|^\alpha} \leq C,$$

since Λ^F is a counting function. On the other hand, using the same inequalities as around (15),

$$\max_{0\le k\le An}\left|H_k^F - \frac{H^{\Pi^{(1)}(F)}_{\Lambda_1^F(k)-1}}{a_1 b_1}\right| \le \max_{0\le k\le An}\left|H_k^F - \frac{\mathrm{Anc}_F(1,u(k))}{a_1b_1}\right| + \max_{0\le k\le An}\frac{|H^{\Pi^{(1)}(F)}_{k-1} - H^{\Pi^{(1)}(F)}_{k}|+1}{a_1b_1}.$$

According to Proposition 5, the first term on the right-hand side is bounded above by $n^{1/4+\gamma}$ with high probability for large n, where we chose $0<\gamma<1/4-\alpha$, so that $n^{1/4+\gamma}\le\sqrt{n}|s-t|^\alpha$ for every large n (recall that $ns, nt\in\mathbb{Z}_+$ so that $|s-t|\ge n^{-1}$). As for the second term, by [13, Lemma 21] (while the statement is on conditioned trees, the first part of its proof yields the result on forests), it holds that it is $o(n^\gamma)$ in $P^{\mathbf{x}}$-probability for any $\gamma>0$, which gives the wanted bound. □

3.4 Tightness

This section is devoted to the proof of last building block needed to prove Theorem 3, namely

Proposition 9 *Under the hypotheses of Theorem 3, for every $\mathbf{x}\in[K]^{\mathbb{N}}$, the laws of the processes $((n^{-1/4}S^F_{\lfloor ns\rfloor}, s\ge 0), n\ge 1)$ under $\mathbb{P}^{\mathbf{x}}$ are tight in the Skorokhod space $\mathbb{D}(\mathbb{R}+,\mathbb{R})$.*

Proof. In this proof, $C_1,\dots,C_{10}$ will be denoting strictly positive constants. Our first task is to obtain an upper bound for expectations of the form $\mathbb{E}^{\mathbf{x}}[|S^F_{nt}-S^F_{ns}|^p]$. To this end, we first choose ξ, D so that the moment condition (4) holds, and write $p=8+\xi$. Also, fix $0<\alpha<1/4$ so that $\alpha(8+\xi)>2$, $0<\eta<1/4$, and $c>0$ such that $c<\eta/(4\log K)$. According to Proposition 3.3, we may choose $C_1>0$ so that if

$$A_n=\left\{\max_{0\le s\ne t\le A: ns,nt\in\mathbb{Z}_+}\frac{|H^F_{ns}-H^F_{nt}|}{n^{1/2}|s-t|^\alpha}\le C_1\right\},$$

then $P^{\mathbf{x}}(A_n)\ge 1-\varepsilon/2$ for every $n\ge1$. We let B_n be the intersection of A_n with the three events $\{\max_{0\le k\le An} c_F(u(k))\le C\log n\}$,

$$\left\{\max_{0\le k\le An}\max_{n^\eta\le h\le|u(k)|}\max_{\mathbf{w}\in\mathcal{W}_K,1\le l\le|\mathbf{w}|}\max_{i\in[K]}\frac{|\mathrm{Anc}^{u(k)}_F(i,\mathbf{w},l,h)-ha_ib_{w_l}\zeta^{(i)}(\mathbf{w})|}{h^{1/2+\eta}}\le 1\right\}$$

and

$$\left\{\max_{0\le k\le An}\max_{n^\eta\le h\le|u(k)|}\mathrm{Anc}^{u(k)}_F(c_F(v)\ge c\log n, h)\le h^\gamma\right\},$$

where $C>0, 0<\gamma<1$ are chosen so that $P^{\mathbf{x}}(B_n)\ge 1-\varepsilon$ for every n sufficiently large, which is possible according to Lemmas 9 and 11. We take $n\in\mathbb{N}$, choose $0\le s\ne t\le A$

be such that $k = ns \in \mathbb{Z}_+, k' = nt \in \mathbb{Z}_+$, and write $u = u(k), u' = u(k')$. Then, by definition, we have

$$S_k^F - S_{k'}^F = \sum_{v \vdash u, |v| > |\breve{u}|} Y_v - \sum_{v \vdash u', |v| > |\breve{u}|} Y_v \,,$$

whenever $\breve{u}$ is the most recent common ancestor to u and u'. Assume $u = \breve{u}rw, u' = \breve{u}r'w'$ for some $r \neq r' \in \mathbb{N}$ and $w, w' \in \mathcal{U}$. This allows to redisplay the previous expression as

$$\begin{aligned} S_k^F - S_{k'}^F &= (Y_{\breve{u}r} - Y_{\breve{u}r'}) \\ &\quad + \sum_{\mathbf{w} \in \mathcal{W}_K} \sum_{1 \le l \le |\mathbf{w}|} \sum_{i \in [K]} \sum_{v \vdash u, v \neq u, v \in F^{(i)}} Y_{vl} \mathbb{1}_{\{|v| > |\breve{u}|, \mathbf{w}_F(v) = \mathbf{w}, vl \vdash u\}} \\ &\quad - \sum_{\mathbf{w} \in \mathcal{W}_K} \sum_{1 \le l \le |\mathbf{w}|} \sum_{i \in [K]} \sum_{v \vdash u', v \neq u', v \in F^{(i)}} Y_{vl} \mathbb{1}_{\{|v| > |\breve{u}|, \mathbf{w}_F(v) = \mathbf{w}, vl \vdash u'\}} \,. \end{aligned}$$

By construction, under $\mathbb{E}^{\mathbf{x}}$, all the terms of this sum are independent of each other conditionally on F, except possibly for $Y_{\breve{u}r}$ and $Y_{\breve{u}r'}$. Let $h = |u| - |\breve{u}| - 1$ and $h' = |u'| - |\breve{u}| - 1$, and let $R(k, k') = |u| + |u'| - 2|\breve{u}| = h + h' + 2$ be the number of random variables of the form Y_v that are involved in the expression $S_k^F - S_{k'}^F$. Using (33) for $p = 8 + \xi$ gives,

$$\begin{aligned} &\mathbb{E}^{\mathbf{x}}\left[\left|S_k^F - S_{k'}^F\right|^p \mid F\right] && (34) \\ &\le C_2 R(k,k')^{p/2-1} \times \begin{pmatrix} \mathbb{E}^{\mathbf{x}}[|Y_{\breve{u}r} - Y_{\breve{u}r'}|^p \mid F] \\ + \sum_{\mathbf{w} \in \mathcal{W}_K} \sum_{l=1}^{|\mathbf{w}|} \sum_{i \in [K]} \mathrm{Anc}_F^u(i, \mathbf{w}, l, h) \langle \nu_{i,\mathbf{w}}, |y_l|^p \rangle \\ + \sum_{\mathbf{w} \in \mathcal{W}_K} \sum_{l=1}^{|\mathbf{w}|} \sum_{i \in [K]} \mathrm{Anc}_F^{u'}(i, \mathbf{w}, l, h') \langle \nu_{i,\mathbf{w}}, |y_l|^p \rangle \end{pmatrix} \\ &\le C_3 R(k,k')^{p/2-1} \times \begin{pmatrix} c_F(\breve{u})^D \\ + \sum_{m \ge 1} m^D \sum_{\mathbf{w} \in \mathcal{W}_K, |\mathbf{w}| = m} \sum_{l=1}^m \sum_{i \in [K]} \mathrm{Anc}_F^u(i, \mathbf{w}, l, h) \\ + \sum_{m \ge 1} m^D \sum_{\mathbf{w} \in \mathcal{W}_K, |\mathbf{w}| = m} \sum_{l=1}^m \sum_{i \in [K]} \mathrm{Anc}_F^{u'}(i, \mathbf{w}, l, h') \end{pmatrix} \end{aligned}$$

Now, on the event B_n, we have $c_F(\breve{u})^D \le (C \log n)^D \le n^{1/2}|s-t|^\alpha$, since $|s - t| \ge 1/n$.

It remains to bound the two above sums, by symmetry it suffices to deal with the first one. In the case where $h \le n^\eta$ and on the event B_n, notice that

$$\sum_{m \ge 1} m^D \sum_{\mathbf{w} \in \mathcal{W}_K, |\mathbf{w}| = m} \sum_{l=1}^m \sum_{i \in [K]} \mathrm{Anc}_F^u(i, \mathbf{w}, l, h) = \sum_{1 \le m \le C \log n} m^D \mathrm{Anc}_F^u(c_F(v) = m, h) \,,$$

which is less than $(C \log n)^D h \le (C \log n)^D n^\eta$, and this in turn is less than $n^{1/2}|s-t|^\alpha \ge n^{1/2-\alpha}$ since $\eta < 1/4 < 1/2 - \alpha$.

Assume now that $h \ge n^\eta$. Still on B_n, it holds that $\mathrm{Anc}_F^u(i, \mathbf{w}, l, h) \le h a_i b_{w_l} \zeta^{(i)}(\mathbf{w}) + h^{1/2+\eta}$. We split the sum under consideration into

$$\sum_{1 \le m \le c \log n} m^D \sum_{\mathbf{w} \in \mathcal{W}_K, |\mathbf{w}| = m} \sum_{l=1}^m \sum_{i \in [K]} \mathrm{Anc}_F^u(i, \mathbf{w}, l, h) + \sum_{c \log n < m \le C \log n} m^D \mathrm{Anc}_F^u(c_F(v) = m, h)$$

The second term is bounded by $(C\log n)^D h^\gamma$ by definition of B_n and since $h \geq n^\eta$, hence by $C_4 h$. The first term is bounded above by

$$C_5 h \max_{i\in[K]} \sum_{\mathbf{w}\in\mathcal{W}_K} |\mathbf{w}|^{D+1}\zeta^{(i)}(\mathbf{w}) + C_5 h^{1/2+\eta}(C\log n)^{D+1}\#\{\mathbf{w}\in\mathcal{W}_K : |\mathbf{w}|\leq c\log n\}\,,$$

and since $\#\{\mathbf{w}\in\mathcal{W}_K : |\mathbf{w}|\leq c\log n\} \leq K^{c\log n} = n^{c\log K} \leq h^{c\log K/\xi}$ and by our choice of c, the whole is bounded by $C_6 h \leq C_6 R(k,k')$.

As in [13, Proposition 27], we argue that the number $R(k,k')$, which equals $|u|+|u'|-2|\check{u}| = h+h'+2$, satisfies $R(k,k') \leq C_7 n^{1/2}|s-t|^\alpha$ on B_n, for all choices of n,s,t. Hence by inspection of all the cases discussed above, we have on B_n,

$$\mathbb{E}^{\mathbf{x}}[|S_k^F - S_{k'}^F|^p|F] \leq C_8(n^{1/2}|t-s|^\alpha)^{p/2}\,,$$

so that

$$\mathbb{E}^{\mathbf{x}}\left[\left|\frac{S_{ns}^F - S_{nt}^F}{n^{1/4}}\right|^p \Big| B_n\right] \leq C_9|t-s|^{\alpha p/2}\,.$$

As in the previous proof, if we let $(S_{\{ns\}}^F, s\geq 0)$ be the linearly interpolated version of $(S_{\lfloor ns\rfloor}^F, s\geq 0)$ between abscissa points of the form $k/n, k\geq 0$, then it is elementary that a similar bound holds up to taking a larger C_9, this time for all $0\leq s\neq t\leq A$. Also, by our choice of α, we have $\alpha p/2 > 1$. Hence, an application of Kolmogorov's criterion to $(n^{-1/4}S_{\{ns\}}^F, s\geq 0)$ gives that for every $A,\varepsilon>0$, there exists $C_{10},\beta>0$ such that for every $n\geq 1$,

$$\mathbb{P}^{\mathbf{x}}\left(\max_{0\leq t\neq s\leq A}\frac{|S_{\{nt\}}^F - S_{\{ns\}}^F|}{n^{1/4}|t-s|^\beta} \geq C_{10}\Big| B_n\right) \leq \varepsilon\,, \tag{35}$$

and since $P^{\mathbf{x}}(B_n)\geq 1-\varepsilon$, we may as well forget the conditioning on B_n.

This is enough to conclude that the laws of the continuous processes $(n^{-1/4}S_{\{ns\}}^F, s\geq 0)$ under $\mathbb{P}^{\mathbf{x}}$ form a tight sequence, and the result will follow from the fact that these processes are respectively uniformly close to $(n^{-1/4}S_{\lfloor ns\rfloor}^F, s\geq 0)$ over compact intervals. This immediately comes from $\max_{1\leq k\leq An}|Y_{u(k)}^F| = o(n^{1/4})$ in probability, which itself can be inferred from the fact that with high probability, no vertex $u(k)$ with $0\leq k\leq An$ has more than $C\log n$ children, and the moment control (4). Hence the result. $\square$

3.5 Conditioned results: Theorem 4

Obtaining Theorem 4 from Theorem 2 now consists in reproducing faithfully the proofs of Propositions 7 and 9, so we only sketch the plan of the proof. The important results that are needed are generalizations to conditioned measures of Lemmas 9 and 11, where $P^{\mathbf{x}}$ must be replaced by $P^{(i)}(\cdot|\#T^{(j)}=n)$. This is straightforward from the latter lemmas and Lemma 6, which we use as we did around (19). This is enough to obtain the exact analog of Proposition 7 for the conditioned probabilities $P^{(i)}(\cdot|\#T^{(j)}=n)$ (it is even simpler as the issue encountered at the very end of the proof of that proposition disappears).

The analog of the tightness statement (Proposition 9) is then a consequence of the following version of Proposition 8 for conditioned measures.

Proposition 10 *Assume that* $\boldsymbol{\mu}$ *satisfies* (H) *and admits small exponential moments. For every* $i, j \in [K]$*, for every* $\alpha \in (0, 1/4)$ *and* $\varepsilon > 0$*, there exists* $C > 0$ *such that*

$$\sup_{n\in\mathbb{N}} P^{(i)}\left(\sup_{s\neq t\in[0,1],\#Ts,\#Tt\in\mathbb{Z}_+} \frac{|H^T_{\lfloor\#Ts\rfloor} - H^T_{\lfloor\#Tt\rfloor}|}{\sqrt{n}|s-t|^\alpha} > C\Big|\#T^{(j)} = n\right) \leq \varepsilon\,.$$

Proof. We rest on [13, Theorem 24], which is essentially the monotype result ($K = 1$), with the extra freedom that we consider the conditioned law $P^{\mathbf{1}}_r(\cdot|\#F = n)$ of a forest with r components. Indeed, when applying the mapping $\Pi^{(j)}$ to T under $P^{(i)}(\cdot|\#T^{(j)} = n)$, one obtains such a conditioned forest with a random number of roots, although the laws of these random numbers form a tight sequence by Lemma 7. Hence, up to conditioning, we can assume that this number of roots is fixed and apply the monotype result. The conclusion is then the exact analog of the last lines of the proof of Proposition 8. Details are left to the interested reader. □

References

[1] D. J. Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.

[2] K. B. Athreya and P. E. Ney. *Branching processes.* Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.

[3] T. Duquesne. A limit theorem for the contour process of conditioned Galton-Watson trees. *Ann. Probab.*, 31(2):996–1027, 2003.

[4] T. Duquesne and J.-F. Le Gall. Random trees, Lévy processes and spatial branching processes. *Astérisque*, 281:vi+147, 2002.

[5] T. Duquesne and J.-F. Le Gall. Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields*, 131(4):553–603, 2005.

[6] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[7] T. E. Harris. *The theory of branching processes.* Dover Phoenix Editions. Dover Publications Inc., Mineola, NY, 2002. Corrected reprint of the 1963 original [Springer, Berlin; MR0163361 (29 #664)].

[8] P. Jagers. General branching processes as Markov fields. *Stochastic Process. Appl.*, 32(2):183–212, 1989.

[9] S. Janson. Limit theorems for triangular urn schemes. *Probab. Theory Relat. Fields*, 134(3):417–452, 2005.

[10] S. Janson and J.-F. Marckert. Convergence of discrete snakes. *J. Theor. Probab.*, 18(3):615–645, 2005.

[11] T. Kurtz, R. Lyons, R. Pemantle, and Y. Peres. A conceptual proof of the Kesten-Stigum theorem for multi-type branching processes. In *Classical and modern branching processes (Minneapolis, MN, 1994)*, volume 84 of *IMA Vol. Math. Appl.*, pages 181–185. Springer, New York, 1997.

[12] J.-F. Le Gall. *Spatial branching processes, random snakes and partial differential equations.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 1999.

[13] J.-F. Marckert and G. Miermont. Invariance principles for random bipartite planar maps. *Ann. Probab.*, 35(5):1642–1705, 2007.

[14] G. Miermont. An invariance principle for random planar maps. In *Fourth Colloquium on Mathematics and Computer Sciences CMCS'06*, Discrete Math. Theor. Comput. Sci. Proc., AG, pages 39–58 (electronic). Nancy, 2006.

[15] V. V. Petrov. *Limit theorems of probability theory*, volume 4 of *Oxford Studies in Probability.* The Clarendon Press Oxford University Press, New York, 1995. Sequences of independent random variables, Oxford Science Publications.

[16] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, Berlin, third edition, 1999.

[17] E. Seneta. *Nonnegative matrices and Markov chains.* Springer Series in Statistics. Springer-Verlag, New York, second edition, 1981.

[18] D. W. Stroock. *Probability theory, an analytic view.* Cambridge University Press, Cambridge, 1993.

[19] V. A. Vatutin and E. E. Dyakonova. The survival probability of a critical multitype Galton-Watson branching process. In *Proceedings of the Seminar on Stability Problems for Stochastic Models, Part II (Nałęczow, 1999)*, volume 106, pages 2752–2759, 2001.

[20] L. M. Wu. Moderate deviations of dependent random variables related to CLT. *Ann. Probab.*, 23(1):420–445, 1995.