Gromov-Hausdorff-Prokhorov convergence of vertex cut-trees of n-leaf Galton-Watson trees

Hui He* Matthias Winkel†

June 8, 2018

Abstract

In this paper we study the vertex cut-trees of Galton-Watson trees conditioned to have n leaves. This notion is a slight variation of Dieuleveut's vertex cut-tree of Galton-Watson trees conditioned to have n vertices. Our main result is a joint Gromov-Hausdorff-Prokhorov convergence in the finite variance case of the Galton-Watson tree and its vertex cut-tree to Bertoin and Miermont's joint distribution of the Brownian CRT and its cut-tree. The methods also apply to the infinite variance case, but the problem to strengthen Dieuleveut's and Bertoin and Miermont's Gromov-Prokhorov convergence to Gromov-Hausdorff-Prokhorov remains open for their models conditioned to have n vertices.

AMS 2010 subject classifications: Primary 60J80; Secondary 60J25, 60F17.

Keywords: Galton-Watson tree, stable tree, cut-tree, fragmentation at nodes, Invariance Principle, $\mathbb{R}$-tree, Continuum Random Tree, Gromov-Hausdorff-Prokhorov topology

1 Introduction

Consider a *rooted planar tree* $(\mathbf{t}, \rho)$. Specifically, $\mathbf{t}$ consists of a finite vertex set $V(\mathbf{t})$ including the *root* $\rho \in V(\mathbf{t})$, a set $E(\mathbf{t})$ of directed edges $u \to v$, one edge for each $u \in V(\mathbf{t}) \setminus \{\rho\}$ without creating cycles, and a planar order, which we describe below. We call v the *parent* of u and $k_v(\mathbf{t}) = \#\{w \in V(\mathbf{t})\colon w \to v\}$ the *number of children* or *degree* of $v \in V(\mathbf{t})$. A vertex $v \in V(\mathbf{t})$ with $k_v(\mathbf{t}) = 0$ is called a *leaf*. We denote by $\mathrm{Lf}(\mathbf{t}) = \{v \in V(\mathbf{t})\colon k_v(\mathbf{t}) = 0\}$ the set of leaves of $\mathbf{t}$, and by $\zeta(\mathbf{t}) = \#V(\mathbf{t})$ and $\lambda(\mathbf{t}) = \#\mathrm{Lf}(\mathbf{t})$ the numbers of vertices and leaves, respectively. Non-leaf vertices, including the root, if $\zeta(\mathbf{t}) \geq 2$, are called *branch points*. The set of branch points is $\mathrm{Br}(\mathbf{t}) = V(\mathbf{t}) \setminus \mathrm{Lf}(\mathbf{t})$. The *planar order* specifies for each $v \in \mathrm{Br}(\mathbf{t})$ a total order on the set of its $k_v(\mathbf{t})$ children. Unless otherwise stated, we will assume that $k_v(\mathbf{t}) \neq 1$ for all $v \in V(\mathbf{t})$.

- Let $n = \lambda(\mathbf{t})$. We use as *vertex splitting rule* the pruning rule of [3] and select a branch point at random, $v \in \mathrm{Br}(\mathbf{t})$ with probability $(k_v(\mathbf{t})-1)/(n-1)$. We fragment the vertex set into $k_v(\mathbf{t})+1$ connected components by removing the edges $w \to v$ from all the children w of the selected branch point v. The component of ρ now has v as a leaf, while the $k_v(\mathbf{t})$ other components are now rooted at the children of v. We apply the splitting rule independently and repeatedly until all components are singleton leaves. We define *our vertex cut-tree* $\mathrm{cut}^\circ_{\mathrm{HW}}(\mathbf{t})$ as the rooted planar tree taking as vertex set the set of components (subsets of $V(\mathbf{t})$) that ever exist, as edge relation the relation between each component and its fragments, as root the initial single component $(V(\mathbf{t}))$ that contains all vertices, and as planar order the order that has for the component split at v the component of v first and the other $k_v(\mathbf{t})$ components in the order their roots have in $\mathbf{t}$ as children of v.

This is illustrated in Figure 1. Our notion of a cut-tree appears to be new, but is closely related to other cut-trees that have been studied and indeed have motivated us for this work:

*Laboratory of Mathematics and Complex Systems, School of Mathematical Sciences, Beijing Normal University, Beijing 100875, P.R.China; email hehui@bnu.edu.cn

†Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK; email winkel@stats.ox.ac.uk

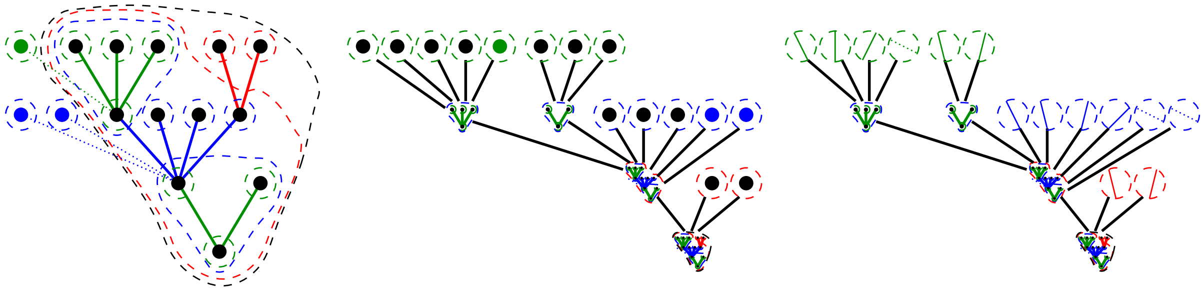

Figure 1: Illustration of $\mathbf{t}$, $\mathrm{cut}_{\mathrm{HW}}(\mathbf{t})$ and $\mathrm{cut}_{\mathrm{D}}(\widehat{\mathbf{t}})$ for $n=8$. Dotted lines capture components of $\mathrm{cut}_{\mathrm{HW}}(\mathbf{t})$. To see $\mathrm{cut}^{\circ}_{\mathrm{HW}}(\mathbf{t})$, omit the singleton components of $\mathrm{cut}_{\mathrm{HW}}(\mathbf{t})$ with coloured bullets.

- Let $n=\zeta(\mathbf{t})$. Meir and Moon [36] introduced an *edge splitting rule*, as follows. Select an edge uniformly at random. Remove the edge (as a singleton) and retain up to two further components (above/below). Pitman [43] and Bertoin [12] studied the forest of components in connection to additive coalescents and forest fires. Bertoin and Miermont [13] introduced the associated *edge cut-tree* $\mathrm{cut}_{\mathrm{BM}}(\mathbf{t})$. In the case of finite-variance Galton-Watson trees conditioned to have n vertices, they showed Gromov-Prokhorov (GP) convergence of tree and cut-tree to a pair $(\mathcal{T}_{\mathrm{Br}},\mathrm{cut}(\mathcal{T}_{\mathrm{Br}}))$ of Brownian Continuum Random Trees (CRTs).
- Let $n=\zeta(\mathbf{t})$. *Dieuleveut's [17] vertex splitting rule and vertex-cut tree* $\mathrm{cut}_{\mathrm{D}}(\mathbf{t})$ are, as follows. Select $v\in\mathrm{Br}(\mathbf{t})$ with probability $k_v(\mathbf{t})/(n-1)$. Fragment the edge set into up to $2k_v(\mathbf{t})+1$ components including all edges above the vertex as singletons. In the case of finite-variance Galton-Watson trees conditioned to have n vertices, Dieuleveut showed GP convergence of the tree and her cut-tree to the same pair $(\mathcal{T}_{\mathrm{Br}},\mathrm{cut}(\mathcal{T}_{\mathrm{Br}}))$. She also obtained an infinite-variance result with a pair of stable CRTs as limiting trees.
- Let $n=\zeta(\mathbf{t})$. Broutin and Wang [14] studied an *inhomogeneous vertex splitting rule and vertex cut-tree* $\mathrm{cut}_{p_n}(\mathbf{t})$ based on a distribution p_n on vertices, and applied this to Camarri and Pitman's [15] p_n-trees. They showed GP/Gromov-Hausdorff-Prokhorov (GHP) convergence of p_n-trees to Aldous and Pitman's inhomogeneous CRTs [10] implies the convergence of pairs of trees and cut-trees in the same mode of convergence. This does not include conditioned Galton-Watson trees beyond a result for uniform trees of [12].

Before the constructions of cut-trees, the evolution of the root component had received particular attention [1, 5, 12, 31, 36, 42]. In the cut-tree, this *pruning process* corresponds to a single spine. Pruning processes of Galton-Watson trees were studied by Aldous and Pitman [9] pruning at edges, and by Abraham et al. [3] pruning at vertices. Limit theorems for pruning processes were obtained in [29] in both cases. These are for forests of Galton-Watson trees. In the domain of attraction of the Brownian forest, this is the same (up to the conditioning on numbers of leaves or vertices) as the joint convergence of the tree and a spine of the cut-tree.

Let $\mathcal{G}$ be a Galton-Watson tree. In our vertex cut-tree model, conditioning on $\lambda(\mathcal{G})=n$, the splitting rule turns out to give some random number $k+1$ of conditioned Galton-Watson trees whose numbers of leaves add up to $n+1$. Hence, the cut-tree is almost a Markov branching tree in the sense of Haas and Miermont [28]. This property fails for all cut-trees of Galton-Watson trees conditioned on $\zeta(\mathcal{G})=n$, except for the edge cut-tree of a Poisson-Galton-Watson tree, which gives the uniform model studied in [9, 12, 43]. Informally, the root component is biased by the number of its leaves. While in general, GHP convergence appears to be much harder to prove than GP convergence (hence the weaker results in [13, 17]), we present here a way to apply the results of [28] and obtain the stronger mode of convergence. Kortchemski [34] obtained tight bounds between numbers of leaves and vertices, so there is scope to transfer asymptotic results for n-leaf trees to n-vertex trees. We leave this for future work.

One of the key ideas is not to focus on the number of leaves, but on $\overline{n}:=2n-1$. Then the "split" of n leaves into $n_1+\cdots+n_{k+1}=n+1$ means that $\overline{n}_1+\cdots+\overline{n}_{k+1}=2(n+1)-k-1\le\overline{n}$

for all $k \geq 2$. We will obtain a Markov branching cut-tree in terms of numbers $\overline{n} = 2n - 1$ associated with numbers n of leaves. For $k \geq 3$, there is loss of mass, so we proceed, as follows.

- Let $n = \lambda(\mathbf{t})$. We add $k-2$ singleton components to $\mathrm{cut}^{\circ}_{\mathrm{HW}}(\mathbf{t})$ for every split into $k+1$ components (summing to $2k-1 = \overline{k}$ components), $k \geq 2$. We modify our vertex cut-tree to include the additional singleton components. We denote this vertex cut-tree by $\mathrm{cut}_{\mathrm{HW}}(\mathbf{t})$.

Proposition 1.1. *Let $\mathcal{G}^{(n)}$ be an n-leaf Galton-Watson tree with offspring distribution ν. Then the vertex cut-tree $\mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)})$ is a Markov branching tree with splitting probabilities*

$$q_{\overline{n}}(\#\text{blocks} = \overline{k}) = \frac{k-1}{k+1}\nu_k \frac{n+1}{n-1}\frac{\mathbb{P}(S_{k+1} = n+1)}{\nu_0 \mathbb{P}(S_1 = n)}, \qquad k \geq 2,$$

where $S_k = X_1 + \cdots + X_k$ for independent $\mathrm{GW}(\nu)$-trees $\mathcal{G}_j$ with $X_j = \lambda(\mathcal{G}_j)$ leaves, $j \geq 1$; and given $\overline{k}$ blocks, the ranked block sizes are like the non-increasing rearrangement of $(\overline{X}_1, \ldots, \overline{X}_{k+1})$ conditionally given $X_1 + \cdots + X_{k+1} = n+1$, with an additional $k-2$ blocks of size 1 appended.

We provide a proof in Section 3.1. Now recall the following notation: let

- $\mathcal{G}^{(n)}$ be a Galton-Watson tree rooted at an ancestor and conditioned to have n leaves,
- $\mathrm{cut}^{\circ}_{\mathrm{HW}}(\mathcal{G}^{(n)})$ our vertex cut-tree of the beginning of this introduction, where in a tree with n leaves a branch point with k children is cut with probability $(k-1)/(n-1)$,
- and $\mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)})$ its modification as just above Proposition 1.1, i.e. $\mathrm{cut}^{\circ}_{\mathrm{HW}}(\mathcal{G}^{(n)})$ with $k-2$ singleton blocks added to the cut-tree when cutting a branch point with k children.

The goal is to show that suitably scaled, we get convergence to $(\mathcal{T}_{\mathrm{Br}}, \mathrm{cut}(\mathcal{T}_{\mathrm{Br}}))$, where $\mathcal{T}_{\mathrm{Br}}$ is the Brownian CRT and $\mathrm{cut}(\mathcal{T}_{\mathrm{Br}})$ is the Brownian cut-tree introduced by Bertoin and Miermont [13], see Section 2.3. We assume for simplicity that the offspring distribution ν satisfies $\nu_1 = 0$. This is no loss of generality since our conditioning does not affect single-child vertices. To pass from this special case to the case of a general offspring distribution, we can associate the offspring distribution conditioned not to produce a single child and represent the desired Galton-Watson tree with single-child vertices as the tree with the conditioned offspring distribution, but with edge lengths added that are independent geometrically distributed with success parameter $1 - \nu_1$. The impact on limiting distances of $\mathcal{G}^{(n)}$ is a constant factor of $1/(1-\nu_1)$. Cut-trees are unaffected.

Let us modify $\mathcal{G}^{(n)}$ to a

- random tree $\widehat{\mathcal{G}}^{(n)}$ in which every branchpoint of $\mathcal{G}^{(n)}$ with k children has $k-2$ more children added (to the left of the k children, say), who themselves have no offspring.

If $\mathcal{G}^{(n)}$ is binary, then $\widehat{\mathcal{G}}^{(n)} = \mathcal{G}^{(n)}$, with $2n-1$ vertices and $2n-2$ edges. In general, the effect of this modification is that the tree with previously n leaves but fewer than $2n-2$ edges receives $k-2$ new edges for any branch point of degree k, for all $k \geq 2$. We note an elementary lemma.

Lemma 1.2. *The random tree $\widehat{\mathcal{G}}^{(n)}$ has $2n-1$ vertices and $2n-2$ edges almost surely.*

The modification of adding $k-2$ edges to $\mathcal{G}^{(n)}$ is related to adding $k-2$ singleton components to $\mathrm{cut}^{\circ}_{\mathrm{HW}}(\mathcal{G}^{(n)})$ to form $\mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)})$, which we did in order to obtain a Markov branching tree without loss of mass in Proposition 1.1. In both cases the effect on the Gromov-Hausdorff (GH) distances of the trees is an elementary consequence of the definition (recalled in Section 2.2):

Lemma 1.3. *We have $d_{\mathrm{GH}}(\mathcal{G}^{(n)}, \widehat{\mathcal{G}}^{(n)}) \leq 1$ and $d_{\mathrm{GH}}(\mathrm{cut}^{\circ}_{\mathrm{HW}}(\mathcal{G}^{(n)}), \mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)})) \leq 1$.*

After scaling, as $n \to \infty$, the GH scaling limits will be identical, i.e. the scaled pair converges to the same limiting tree. Comparison in the GHP distance d_{GHP} is less straightforward.

Recall that Dieuleveut's vertex cut-tree $\mathrm{cut}_{\mathrm{D}}(\mathbf{t})$ selects each branch point with k children with probability proportional to k, while our vertex cut-tree $\mathrm{cut}_{\mathrm{HW}}(\mathbf{t})$ selects each branch point with k children with probability proportional to $k-1$. Now note that $\widehat{\mathcal{G}}^{(n)}$ has $2k-2 \geq 2$ children wherever $\mathcal{G}^{(n)}$ has $k \geq 2$ children, and in $\widehat{\mathcal{G}}^{(n)}$, Dieuleveut would select a branch point with $2k-2$ children with probability proportional to $2k-2 = 2(k-1)$. Hence, we can couple the constructions of $\mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)})$ and $\mathrm{cut}_{\mathrm{D}}(\widehat{\mathcal{G}}^{(n)})$. However, Dieuleveut proceeds slightly differently when building the cut-tree. The branch points of $\mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)})$ and $\mathrm{cut}_{\mathrm{D}}(\widehat{\mathcal{G}}^{(n)})$ can be taken the same, but the numbers of leaves at any particular branch point are typically different, while the total numbers of leaves are $2n-1$ and $2n-2$, respectively. See e.g. Figure 1.

Specifically, for $v \in \mathrm{Br}(\mathcal{G}^{(n)})$ with $k = k_v(\mathcal{G}^{(n)})$ children, our cut-tree $\mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)})$ always has $k+1$ main components, some of which may be singleton vertices, and $k-2$ more singleton components, giving $2k-1$ altogether. On the other hand, $\mathrm{cut}_{\mathrm{D}}(\widehat{\mathcal{G}}^{(n)})$ records components of the edge set, and depending on when the k edges are removed, they may or may not have subtrees above them. As an extreme example, suppose that all k initially had subtrees above them, and v is not the root. If this is the first split, there are $k+1$ components above and below, plus a further k singletons for the removed edges, $2k+1$ altogether. If, however, this is the last split, there are only the k singletons, all other "components" already being empty. In any case, this yields:

Proposition 1.4. *We have* $d_{\mathrm{GH}}(\mathrm{cut}_{\mathrm{D}}(\widehat{\mathcal{G}}^{(n)}), \mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)})) \leq 1$ *for a suitable coupling.*

Turning to d_{GHP}, the question arises what mass measures we place onto the cut-trees. Bertoin, Miermont and Dieuleveut actually consider trees with n edges ($n-1$ vertices) and obtain cut-trees with n leaves, so it is natural to put the uniform measure in leaves onto their cut-trees in their framework. In our framework, we equip $\mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)})$ with the uniform measure on its $2n-1 = \overline{n}$ leaves and $\mathrm{cut}_{\mathrm{D}}(\widehat{\mathcal{G}}^{(n)})$ with the uniform measure on its $2n-2$ leaves. We also equip $\mathcal{G}^{(n)}$ with the uniform measure on its n leaves and $\widehat{\mathcal{G}}^{(n)}$ with the uniform measure on its $2n-2$ edges. Our programme has three steps, here given for the finite variance case, for suitable c_n and c'_n, which will be discussed in Sections 2.1 and 3.1, respectively. We show

Step 1: $\mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)})/c'_n \to \mathcal{T}_{\mathrm{Br}}$ in GHP, using the Markov branching convergence criterion of [28], deduce $(\mathrm{cut}_{\mathrm{D}}(\widehat{\mathcal{G}}^{(n)})/c'_n, \mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)})/c'_n, \mathrm{cut}^{\circ}_{\mathrm{HW}}(\mathcal{G}^{(n)})/c'_n) \to (\mathcal{T}_{\mathrm{Br}}, \mathcal{T}_{\mathrm{Br}}, \mathcal{T}_{\mathrm{Br}})$ in GH^3;

Step 2: $\widehat{\mathcal{G}}^{(n)}/c_n \to \mathcal{T}_{\mathrm{Br}}$ in GHP, based on [37, 16], deduce $(\mathcal{G}^{(n)}/c_n, \widehat{\mathcal{G}}^{(n)}/c_n) \to (\mathcal{T}_{\mathrm{Br}}, \mathcal{T}_{\mathrm{Br}})$ in GHP^2;

Step 3: $(\widehat{\mathcal{G}}^{(n)}/c_n, \mathrm{cut}_{\mathrm{D}}(\widehat{\mathcal{G}}^{(n)})/c'_n) \to (\mathcal{T}_{\mathrm{Br}}, \mathrm{cut}(\mathcal{T}_{\mathrm{Br}}))$, in GP^2, adapting the arguments of [17].

Here GHP, GH^3, GHP^2 and GP^2 denote convergences in distribution on product spaces, where each component is equipped with the GHP, GH or GP topologies, as appropriate, see Section 2.2. We deduce that $\mathcal{T}_{\mathrm{Br}} \overset{d}{=} \mathrm{cut}(\mathcal{T}_{\mathrm{Br}})$, as was already shown in [13]. More importantly, we conclude:

Theorem 1.5. *With any finite-variance offspring distribution* $(\mathcal{G}^{(n)}/c_n, \widehat{\mathcal{G}}^{(n)}/c_n) \to (\mathcal{T}_{\mathrm{Br}}, \mathcal{T}_{\mathrm{Br}})$ *in* GHP^2 *in distribution, jointly with* $(\mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)})/c'_n, \mathrm{cut}_{\mathrm{D}}(\widehat{\mathcal{G}}^{(n)})/c'_n) \to (\mathrm{cut}(\mathcal{T}_{\mathrm{Br}}), \mathrm{cut}(\mathcal{T}_{\mathrm{Br}}))$ *in* GHP^2*, as* $n \to \infty$ *in* $\{n \geq 1 \colon \mathbb{P}(\lambda(\mathcal{G}) = n) > 0\}$ *for an associated Galton-Watson tree* $\mathcal{G}$*.*

Given the three steps, the remaining proof is mainly a standard argument via tightness and uniqueness of subsequential limit distributions, see Section 2.4, but also requires the following result, which is part of the folklore on the Brownian CRT $(\mathcal{T}_{\mathrm{Br}}, \mu_{\mathrm{Br}})$, but we were unable to locate it in the literature, so we quickly derive it from well-known results in Section 2.4.

Proposition 1.6. *The measured tree* $(\mathcal{T}_{\mathrm{Br}}, \mu_{\mathrm{Br}})$ *is a measurable function of the unmeasured* $\mathcal{T}_{\mathrm{Br}}$*.*

This proposition will also hold for stable trees, but the argument would be more involved, and since we do not need this here, we do not work out the details.

The structure of this paper is as follows. In Section 2, we note a local limit theorem for the number of leaves, recall the three relevant topologies GP, GH and GHP, we prove Proposition 1.6, and we deduce Theorem 1.5 from the three steps given above. In Section 3, we prove Proposition 1.1 and turn to the three main steps and hence complete the above programme in the finite variance case, and we indicate how corresponding results in the stable case can be approached. We also include the brief Appendix B summarising the use of different normalisations of the Brownian CRT in the literature.

2 Preliminaries

2.1 A local limit theorem for the number of leaves

Consider a critical offspring distribution ν in the domain of attraction of a stable distribution with index $\alpha \in (1,2]$. Specifically, suppose that for a random walk W with step distribution $\mathbb{P}(W_1 = n) = \nu_{n+1}$, $n \geq -1$,

$$\frac{W_n}{a_n} \xrightarrow[n\to\infty]{d} X_1, \tag{1}$$

where a_n is regularly varying with index α and $\mathbb{E}(\exp(-rX_1)) = \exp(r^\alpha)$. Then the classical local limit theorem holds for W, see Ibragimov and Linnik [30, Theorem 4.2.1], or Kortchemski [34, Theorem 1.10] for a statement:

$$\sup_{k\in\mathbb{Z}} \left| a_n \mathbb{P}(W_n = k) - p_1\left(\frac{k}{a_n}\right)\right| \to 0 \qquad \text{as } n \to \infty,$$

where p_1 is the continuous density of X_1, which is $p_1(x) = \frac{1}{2\sqrt{\pi}}\exp(-x^2/4)$, $x \in \mathbb{R}$, for $\alpha = 2$.

Consider the stopping times $K_0 = 0$ and $K_{n+1} = \inf\{k \geq K_n + 1\colon W_k - W_{k-1} = -1\}$ of down-moves and the time-changed process $\widetilde{W}_n = W_{K_n}$, $n \geq 0$, of values after down-moves. This can be viewed as a transformation on trees that in some sense removes all non-leaf branch points. See Rizzolo [47] for generalisations removing all branch points with multiplicities not in a set $A \subset \mathbb{N}$. Note that the original tree can be recovered from W, but not in general from $\widetilde{W}$. Effectively, some of the leaves of the tree encoded in W now act as branch points of the transformed tree encoded in $\widetilde{W}$ (replacing one or more removed branch points).

Lemma 2.1. *The increment distribution of $\widetilde{W}$ is in the domain of attraction of the same stable distribution as ν. Specifically,*

$$\frac{\widetilde{W}_n}{\widetilde{a}_n} \xrightarrow[n\to\infty]{d} X_1,$$

where $\widetilde{a}_n = a_n/\nu_0^{1/\alpha}$. If W_1 has finite variance σ^2, we can choose $a_n = \sigma\sqrt{n/2}$.

Proof. This is rather elementary: by definition, we can write $\widetilde{W}_1 = A_1 + \cdots + A_G - 1$, where $G \sim \operatorname{geom}(\nu_0)$ is independent of an independent and identically distributed sequence of up-moves A_n, $n \geq 1$, with $\mathbb{P}(A_n = j) = \nu_{j+1}/(1-\nu_0)$, $j \geq 0$. Here

$$\mathbb{E}\left[\exp\left(-r\widetilde{W}_1\right)\right] = e^r \frac{\nu_0}{1-(1-\nu_0)\mathbb{E}[e^{rA_1}]} = \frac{e^r\nu_0}{1-(\mathbb{E}[e^{-rW_1}]-\nu_0 e^r)}$$

By assumption,

$$\left(\mathbb{E}\left[\exp\left(-\frac{r}{a_n}W_1\right)\right]\right)^n \longrightarrow \exp(r^\alpha) \qquad \text{i.e.} \qquad n\left(\mathbb{E}\left[\exp\left(-\frac{r}{a_n}W_1\right)\right]-1\right) \longrightarrow r^\alpha.$$

Hence

$$n\left(\mathbb{E}\left[\exp\left(-\frac{r}{\widetilde{a}_n}\widetilde{W}_1\right)\right]-1\right) = \frac{n\left(1-\mathbb{E}\left[\exp\left(-\frac{r\nu_0^{1/\alpha}}{a_n}W_1\right)\right]\right)}{1-\mathbb{E}\left[\exp\left(-\frac{r\nu_0^{1/\alpha}}{a_n}W_1\right)\right]+\nu_0\exp\left(\frac{r\nu_0^{1/\alpha}}{a_n}\right)} \longrightarrow \frac{(r\nu_0^{1/\alpha})^\alpha}{\nu_0} = r^\alpha.$$

If $\sigma^2 < \infty$, then $a_n = \sigma\sqrt{n/2}$ is the central limit theorem with limiting variance 2. $\square$

Corollary 2.2. *Under the assumption* (1)*, the time-changed process $\widetilde{W}$ satisfies the local limit theorem*

$$\sup_{k\in\mathbb{Z}} \left| \widetilde{a}_n \mathbb{P}(\widetilde{W}_n = k) - p_1\left(\frac{k}{\widetilde{a}_n}\right)\right| \to 0 \qquad \textit{as } n \to \infty.$$

Now denote by S_j respectively S_j^V the random number of leaves respectively vertices in j independent Galton-Watson trees with offspring distribution ν. Following Haas and Miermont [28], we note the classical argument based on the observation that we can think of the steps of W as corresponding to vertices of the trees (e.g. exploring the trees in depth first order) adding each time the number of children minus one so that W_k is the number of unexplored vertices whose parent has been explored minus j while the jth tree is being explored. Then

$$S_j^V = n \quad \iff \quad W_n = -j \text{ and } W_m > -j,\ m < n,$$

which via the cyclic lemma (e.g. Feller [25, Lemma XII.6.1]) for the downward skip-free random walk W yields

$$\mathbb{P}(S_j^V = n) = \frac{j}{n}\mathbb{P}(W_n = -j).$$

The following result was noted in [47, Corollary 1] and has been implicit in Kortchemski [34].

Proposition 2.3. *We have* $\mathbb{P}(S_j = n) = \dfrac{j}{n}\mathbb{P}(\widetilde{W}_n = -j)$ *for all* $1 \leq j \leq n$.

Proof. Just note that $\widetilde{W}$ is also downward skip-free since it does not skip any down-moves of W. Each step now corresponds to a leaf and $-j$ is first reached when all leaves have been explored so that

$$S_j = n \quad \iff \quad \widetilde{W}_n = -j \text{ and } \widetilde{W}_m > -j, m < n,$$

and we conclude via the cyclic lemma for $\widetilde{W}$. □

Corollary 2.4. *We have* $\sup_{j\geq 1}\left| n\widetilde{a}_n \frac{1}{j}\mathbb{P}(S_j = n) - p_1\left(\frac{-j}{\widetilde{a}_n}\right)\right| \to 0$ *as* $n \to \infty$.

Recall that given a planar tree $\mathbf{t}$ with root ρ, we denote by $\zeta(\mathbf{t})$ and $\lambda(\mathbf{t})$ the total number of vertices and leaves of $\mathbf{t}$, respectively. For $v \in V(\mathbf{t})$ with $v = v_k \to v_{k-1} \to \cdots \to v_1 \to v_0 = \rho$, we say that v has *generation* $|v| = k$. Denote by $\zeta_k(\mathbf{t})$ and $\lambda_k(\mathbf{t})$ the number of vertices and leaves of $\mathbf{t}$ at generation k. Let $\mathbf{t}(k)$ be $\mathbf{t}$ restricted to generation at most k, i.e.

$$\mathbf{t}(k) = \{v \in \mathbf{t} : |v| \leq k\}.$$

Let $\mathcal{G}^{(n)}$ be a critical Galton-Watson tree conditioned to have n leaves, with offspring distribution ν, and $\widehat{\mathcal{G}}^{(n)}$ its modification with extra leaves as defined just before Lemma 1.2.

Lemma 2.5. *If the offspring distribution has finite variance, there exists a constant* $C > 0$ *such that*

$$\sup_{n\geq 1}\mathbb{E}\left[\zeta_k\left(\widehat{\mathcal{G}}^{(n)}\right)\right] \leq 2\sup_{n\geq 1}\mathbb{E}\left[\zeta_k\left(\mathcal{G}^{(n)}\right)\right] \leq Ck, \quad k \geq 1.$$

Proof. The first inequality is elementary. For the second inequality, we adapt Janson's idea of proving [31, Theorem 1.13]. Our proof will be divided into four subparts. We use $c, C, C_1, C_2, \ldots$ for constants independent of n and k.

Subpart 1. Let $\mathcal{G}$ be a Galton-Watson tree and $\mathcal{G}^\infty$ the so-called Kesten tree arising as local limit of $\mathcal{G}^{(n)}$ as $n \to \infty$; see Abraham and Delmas [2]. It is well-known [33, (1.15)] that for any tree $\mathbf{t}$

$$\mathbb{P}(\mathcal{G}(k) = \mathbf{t}(k)) = \zeta_k(\mathbf{t})\mathbb{P}(\mathcal{G}^\infty(k) = \mathbf{t}(k)).$$

Let $\mathbf{t}$ be a tree with $\zeta_k(\mathbf{t}) = m$. Define $N = n - \sum_{i\leq k-1}\lambda_i(\mathbf{t}(k))$. Then by conditioning on generation k and using Kortchemski [34, Theorem 3.1] and Proposition 2.3, we obtain

$$\begin{aligned}
\mathbb{P}(\mathcal{G}^{(n)}(k) = \mathbf{t}(k)) &= \frac{\mathbb{P}(\mathcal{G}(k) = \mathbf{t}(k), \lambda(\mathcal{G}) = n)}{\mathbb{P}(\lambda(\mathcal{G}) = n)} \\
&\leq C_1 n^{3/2}\mathbb{P}(\mathcal{G}(k) = \mathbf{t}(k))\mathbb{P}\left(S_m = N\right) \\
&= C_1 n^{3/2}\mathbb{P}(\mathcal{G}(k) = \mathbf{t}(k))\frac{m}{N}\mathbb{P}\left(\widetilde{W}_N = -m\right) \\
&\leq C_2 m\left(\frac{n}{N}\right)^{3/2} e^{-cm^2/N}\mathbb{P}(\mathcal{G}(k) = \mathbf{t}(k)) \\
&= C_2\left(\frac{n}{N}\right)^{3/2} e^{-cm^2/N}\mathbb{P}(\mathcal{G}^\infty(k) = \mathbf{t}(k)), \qquad (2)
\end{aligned}$$

where in the second inequality we use Lemma 2.1 above and [31, Lemma 2.1].

The argument in Subparts 2.–4. is very similar to the proof of [31, Theorem 1.13] with only slight modifications.

Subpart 2. For each $k \geq 1$, define

$$\Gamma_k = \left\{ \sum_{i \leq k-1} \lambda_i(\mathcal{G}^{(n)}(k)) \leq n/2 \right\} \qquad \text{and} \qquad \zeta_k^*(\mathcal{G}^{(n)}(k)) = \zeta_k(\mathcal{G}^{(n)}(k)) 1_{\Gamma_k}.$$

By (2), for any tree $\mathbf{t}$ with $\sum_{i \leq k-1} \lambda_i(\mathbf{t}(k)) \leq n/2$ and $\zeta_k(\mathbf{t}) > 0$, we have

$$\mathbb{P}(\mathcal{G}^{(n)}(k) = \mathbf{t}(k)) \leq C_3 \mathbb{P}(\mathcal{G}^{\infty}(k) = \mathbf{t}(k)),$$

which implies

$$\mathbb{P}(\zeta_k^*(\mathcal{G}^{(n)}) = i) \leq C_4 \mathbb{P}(\zeta_k(\mathcal{G}^{\infty}) = i), \qquad \text{for all } i \geq 1.$$

Thus

$$\mathbb{E}[\zeta_k^*(\mathcal{G}^{(n)})] = \mathbb{E}[\zeta_k(\mathcal{G}^{(n)}) 1_{\Gamma_k}] \leq C_4 \mathbb{E}[\zeta_k(\mathcal{G}^{\infty})] \leq C_5 k, \tag{3}$$

where the last inequality follows from [31, Lemma 2.3].

Subpart 3. On Γ_k^c, one can find a (random) integer $L \leq k$ such that

$$\sum_{i=1}^{L-1} \lambda_i(\mathcal{G}^{(n)}) \leq n/2 \qquad \text{and} \qquad \sum_{i=1}^{L} \lambda_i(\mathcal{G}^{(n)}) > n/2.$$

Thus on Γ_k^c,

$$\sum_{i=0}^{k} \zeta_i^*(\mathcal{G}^{(n)}) = \sum_{i=0}^{L} \zeta_i^*(\mathcal{G}^{(n)}) = \sum_{i=0}^{L} \zeta_i(\mathcal{G}^{(n)}) > \sum_{i=0}^{L} \lambda_i(\mathcal{G}^{(n)}) > n/2.$$

By the Markov inequality and (3),

$$\mathbb{P}(\Gamma_k^c) \leq \frac{2}{n} \mathbb{E}\left[\sum_{i=0}^{k} \zeta_i^*(\mathcal{G}^{(n)})\right] \leq \frac{C_6 k^2}{n}.$$

Hence, we obtain

$$\mathbb{E}\left[\zeta_k(\mathcal{G}^{(n)}) 1_{\Gamma_k^c} 1_{\{\zeta_k(\mathcal{G}^{(n)}) \leq \sqrt{n}\}}\right] \leq \sqrt{n}\mathbb{P}(\Gamma_k^c) \leq \sqrt{n\mathbb{P}(\Gamma_k^c)} \leq \sqrt{C_6} k. \tag{4}$$

Subpart 4. For any $\mathbf{t}$ with $\zeta_k(\mathbf{t}) \geq \sqrt{n}$, according to (2), we have

$$\mathbb{P}(\mathcal{G}^{(n)}(k) = \mathbf{t}(k)) \leq C_7 \left(\frac{n}{N}\right)^{3/2} e^{-cn/N} \mathbb{P}(\mathcal{G}^{\infty}(k) = \mathbf{t}(k)) \leq C_8 \mathbb{P}(\mathcal{G}^{\infty}(k) = \mathbf{t}(k)),$$

which, by reasoning similar as for (3), yields

$$\mathbb{E}[\zeta_k(\mathcal{G}^{(n)}) 1_{\{\zeta_k(\mathcal{G}^{(n)}) > \sqrt{n}\}}] \leq C_8 \mathbb{E}[\zeta_k(\mathcal{G}^{\infty})] \leq C_9 k. \tag{5}$$

Then the desired result follows from (3), (4) and (5). We have completed the proof. $\square$

2.2 GH, GP and GHP topologies

According to [23, 24, 26, 38] and references therein, we can define a Gromov-Hausdorff-Prokhorov (Gromov-Hausdorff or Gromov-Prokhorov) distance on the set of measure-preserving isometry classes of pointed measured compact metric spaces to turn the set (of equivalence classes modulo measure or modulo restriction to the support of the measure) into a Polish space.

Specifically, let (Z, d^Z) be a metric space. For Borel sets $A, B \subseteq Z$, set

$$d_{\mathrm{H}}^Z(A, B) = \inf\{\varepsilon > 0\colon A \subseteq B^\varepsilon \text{ and } B \subseteq A^\varepsilon\},$$

the Hausdorff distance between A and B, where $A^\varepsilon = \{x \in Z\colon \inf_{y\in A} d^Z(x, y) \leq \varepsilon\}$. Let $M_f(Z)$ be the set of all Borel probability measures on (Z, d^Z). For $\mu, \mu' \in M_f(Z)$, we define

$$d_{\mathrm{P}}^Z(\mu, \mu') = \inf\{\varepsilon > 0\colon \mu(A) \leq \mu'(A^\varepsilon) + \varepsilon \text{ and } \mu'(A) \leq \mu(A^\varepsilon) + \varepsilon \ \text{ for all closed } A \subseteq Z\},$$

the Prokhorov distance between μ and μ'.

A pointed measured metric space $\mathrm{T} = (T, d, \rho, \mu)$ is a metric space (T, d) with a distinguished element $\rho \in T$ and a Borel probability measure μ on (T, d). For two compact pointed measured metric spaces $\mathrm{T} = (T, d, \rho, \mu)$ and $\mathrm{T}' = (T', d', \rho', \mu')$, the Gromov-Hausdorff-Prokhorov distance is

$$d_{\mathrm{GHP}}(\mathrm{T}, \mathrm{T}') = \inf_{\Phi,\Phi',Z} \left(d_{\mathrm{H}}^Z(\Phi(T), \Phi'(T')) + d^Z(\Phi(\rho), \Phi'(\rho')) + d_{\mathrm{P}}^Z(\Phi_*\mu, \Phi'_*\mu')\right),$$

where the infimum is taken over all isometric embeddings $\Phi\colon T \hookrightarrow Z$ and $\Phi'\colon T' \hookrightarrow Z$ into some common Polish metric space (Z, d^Z) and $\Phi_*\mu$ is the measure μ transported by Φ. Similarly, we define Gromov-Hausdorff and Gromov-Prokhorov distances, respectively, as

$$d_{\mathrm{GH}}(\mathrm{T}, \mathrm{T}') = \inf_{\Phi,\Phi',Z} \left(d_{\mathrm{H}}^Z(\Phi(T), \Phi'(T')) + d^Z(\Phi(\rho), \Phi'(\rho'))\right),$$

$$d_{\mathrm{GP}}(\mathrm{T}, \mathrm{T}') = \inf_{\Phi,\Phi',Z} \left(d^Z(\Phi(\rho), \Phi'(\rho')) + d_{\mathrm{P}}^Z(\Phi_*\mu, \Phi'_*\mu')\right).$$

A compact metric space (T, d) is called a real tree if for any two $x, y \in T$, there is an isometry $f_{x,y}\colon [0, d(x, y)] \to T$ with $f_{x,y}(0) = x$ and $f_{x,y}(d(x, y)) = y$, and if for all injective $g\colon [0, 1] \to T$ with $g(0) = x$ and $g(1) = y$ we have $g([0, 1]) = f_{x,y}([0, d(x, y)])$. Every real tree (T, d) is naturally equipped with a (sigma-finite) length measure ℓ_T, for which $\ell_T(f_{x,y}([0, d(x, y)])) = d(x, y)$, $x, y \in T$. We refer to a pointed real tree (T, d, ρ) as a *rooted* real tree, to points $x \in T \setminus \{\rho\}$ for which $T \setminus \{x\}$ is connected, respectively, disconnected into three or more connected components, as *leaves*, respectively *branch points*.

For any rooted real tree (T, d, ρ), we define the height $\mathrm{ht}(T) = \max\{d(\rho, x), x \in T\}$. For any $x \in T$, we define the subtree $T_x = \{y \in T\colon x \in f_{\rho,y}([0, d(\rho, y)])\}$ above x. For $\varepsilon > 0$, we define Neveu's [39] notion of ε-erasure of T as $R_\varepsilon(T) = \{\rho\} \cup \{x \in T\colon \mathrm{ht}(T_x) \geq \varepsilon\}$. Then $R_\varepsilon(T)$ is a rooted real tree with finitely many leaves and branch points; see also [23, 40, 41].

Examples of pointed measured compact real trees are obtained from continuous functions $h\colon [0, 1] \to [0, \infty)$ with $h(0) = h(1) = 1$. For $s, t \in [0, 1]$, let $d_h(s, t) = h(t) + h(s) - 2 \inf\{h(r), \min(s, t) \leq r \leq \max(s, t)\}$ and $s \sim_h t$ iff $d_h(s, t) = 0$. Then the quotient space $T_h = [0, 1]/\sim_h$ is a compact real tree when equipped with the quotient metric, again denoted by d_h. We further equip (T_h, d_h) with the root $\rho_h = [0]_{\sim_h}$ and the measure μ_h obtained as the push-forward of Lebesgue measure on $[0, 1]$ under the quotient map. The function h is called the *height function* of $(T_h, d_h, \rho_h, \mu_h)$. See e.g. [20, Section 2].

2.3 Bertoin and Miermont's Brownian cut-tree

A Brownian Continuum Random Tree (CRT) is a random pointed measured compact metric space introduced by Aldous [6]. One construction is to take $h = 2B^{\mathrm{ex}}$ as height function, for a normalised excursion B^{ex} of linear Brownian motion.

Let $(\mathcal{T}_{\mathrm{Br}}, \mu_{\mathrm{Br}})$ be a Brownian CRT. Conditionally on $\mathcal{T}_{\mathrm{Br}}$, let $\sum_{i\in I} \delta_{(t_i,x_i)}(dt, dx)$ be a Poisson point measure on $[0, \infty) \times \mathcal{T}_{\mathrm{Br}}$ with intensity $dt \times d\ell_{\mathrm{Br}}$, where ℓ_{Br} is the length measure on $\mathcal{T}_{\mathrm{Br}}$.

Denote by $\mathcal{T}_{\mathrm{Br}}(t)$ the "forest" obtained by removing points $\{x_i : i \in I, t_i \le t\}$ that are marked before t. For any $x \in \mathcal{T}_{\mathrm{Br}}$, let $\mathcal{T}_{\mathrm{Br}}(x,t)$ be the connected component of $\mathcal{T}_{\mathrm{Br}}(t)$ that contains x with the convention that $\mathcal{T}_{\mathrm{Br}}(x,t) = \emptyset$ if $x \notin \mathcal{T}_{\mathrm{Br}}(t)$. Define $\mu_{\mathrm{Br}}(x,t) = \mu_{\mathrm{Br}}(\mathcal{T}_{\mathrm{Br}}(x,t))$. We further define a function δ from $(\mathcal{T}_{\mathrm{Br}} \cup \{0\})^2$ into $[0,+\infty]$ such that $\delta(0,0) = 0$ and

$$\delta(0,x) = \delta(x,0) = \int_0^\infty \mu_{\mathrm{Br}}(x,t)dt \qquad \text{and} \qquad \delta(x,y) = \int_{t(x,y)}^\infty (\mu_{\mathrm{Br}}(x,t) + \mu_{\mathrm{Br}}(y,t))\, dt,$$

where $t(x,y) = \inf\{t \ge 0 : \mathcal{T}_{\mathrm{Br}}(x,t) \ne \mathcal{T}_{\mathrm{Br}}(y,t)\}$.

Let $\xi_0 = 0$ and $(\xi_i, i \in \mathbb{N})$ be an i.i.d. sequence distributed as μ_{Br}. For all $k \ge 1$, let $\mathcal{R}_k$ be the random real tree spanned by $\{\xi_0, \xi_1, \ldots, \xi_k\}$ with metric δ. Then $\mathrm{cut}(\mathcal{T}_{\mathrm{Br}})$ is defined as

$$\mathrm{cut}(\mathcal{T}_{\mathrm{Br}}) = \overline{\bigcup_{k\ge 1} \mathcal{R}_k}\,,$$

the completion of the metric space $(\bigcup_{k\ge1} \mathcal{R}_k, \delta)$. Then $(\mathrm{cut}(\mathcal{T}_{\mathrm{Br}}), \delta, 0)$, equipped with the limiting empirical measure of $(\xi_i, i \in \mathbb{N})$, is again a Brownian CRT; see Bertoin and Miermont [13] for details of how to make this construction precise.

2.4 Deduction of Theorem 1.5 from the statements of Steps 1–3.

Since the proof of Theorem 1.5 requires Proposition 1.6, we prove the proposition first.

Proof of Proposition 1.6. First consider $H = 2B$ for a Brownian motion B. For $\varepsilon > 0$, we follow [44, Section 7.6] and define alternating up- and down-crossing times as $D_0^{(\varepsilon)} = 0$ and, for $m \ge 0$,

$$\begin{aligned} U_{m+1}^{(\varepsilon)} &= \inf\{t \ge D_m^{(\varepsilon)} : H(t) - \min\{H(s), D_m^{(\varepsilon)} \le s \le t\} = \varepsilon\}, \\ D_{m+1}^{(\varepsilon)} &= \inf\{t \ge U_{m+1}^{(\varepsilon)} : H(t) - \max\{H(s), U_{m+1}^{(\varepsilon)} \le s \le t\} = -\varepsilon\}. \end{aligned}$$

Then $D_m^{(\varepsilon)}$ is precisely ε below a previous local maximum of H for all $m \ge 1$. Let $X_m^{(\varepsilon)} = H(D_m^{(\varepsilon)})$ and $Y_m^{(\varepsilon)} = \min\{H(s) : D_m^{(\varepsilon)} \le s \le D_{m+1}^{(\varepsilon)}\}$, $m \ge 0$.

The excursions above the minimum of H are scaled copies of $2B^{\mathrm{ex}}$ and hence encode scaled Brownian CRTs. The subtrees spanned by $D_m^{(\varepsilon)}$, $m \ge 1$, are ε-erasures of the Brownian CRTs with leaves at heights $X_m^{(\varepsilon)} - \min\{Y_k^{(\varepsilon)}, 0 \le k \le m-1\}$, $m \ge 1$, and roots and branch points at heights $Y_m^{(\varepsilon)}$, $m \ge 0$. Consider the function $H^{(\varepsilon)}$, which is piecewise linear at alternating slopes of $\pm 2/\varepsilon$ interpolating the alternating walk $X_0^{(\varepsilon)}, Y_0^{(\varepsilon)}, X_1^{(\varepsilon)}, Y_1^{(\varepsilon)}, \ldots$. By [44, Corollary 7.17], we have $H^{(\varepsilon)} \to H$ locally uniformly and almost surely, as $\varepsilon \downarrow 0$.

Our aim is to deduce that the ε-erasure $R_\varepsilon(\mathcal{T}_{\mathrm{Br}})$ equipped with a scaled length measure $\mu_\varepsilon = \varepsilon \ell_{\mathrm{Br}}|_{R_\varepsilon(\mathcal{T}_{\mathrm{Br}})}$ converges to $(\mathcal{T}_{\mathrm{Br}}, \mu_{\mathrm{Br}})$ in GHP. Measurability (actually GH-GHP-continuity) of $(\varepsilon, T) \mapsto (T, \varepsilon \ell_T|_{R_\varepsilon(T)})$ is formally established in Lemma A.1.

The convergence $H^{(\varepsilon)} \to H$ includes the first excursion of height greater than $r > 0$, jointly with the excursion length, so that convergence holds under the Brownian Itô excursion measure n_{Br} conditioned on excursions of height greater than r, for all $r > 0$. See [46, Chapter XII]. By disintegration and the scaling property of n_{Br} (e.g. [32, Theorem 22.15]), this convergence also holds under the distribution of $2B^{\mathrm{ex}}$, which is the normalised excursion measure $n_{\mathrm{Br}}(\,\cdot\,|\,\zeta = 1)$, where $\zeta(h) = \inf\{t \ge 0 : h(t) = 0\}$, for continuous $h : [0,1] \to [0,\infty)$.

For any continuous $h : [0,1] \to [0,\infty)$, let $h^{(\varepsilon)}$ be constructed from h as $H^{(\varepsilon)}$ was constructed from H. Then $T_{h^{(\varepsilon)}}$ is isometric to $R_\varepsilon(T)_h)$ and, with this isometry, the quotient map $\sim_{h^{(\varepsilon)}}$ pushes forward Lebesgue measure onto ε times the length measure of $R_\varepsilon(T_h)$. Uniform convergence jointly with excursion lengths implies GHP convergence of encoded trees equipped with the push-forward of Lebesgue measure (see e.g. [4]). This completes the proof. $\square$

We noted in the introduction that while Proposition 1.6 will also hold for stable trees, the argument will be more involved and beyond the scope of this paper, since we focus on the

Brownian case here. While ε-erasure of stable trees has been studied in [21], this paper does not construct the mass measure from the length measure. [19] study height functions, but "Poisson sampling" instead of ε-erasure. For Poisson sampling, their results yield the analogous almost sure and locally uniform convergence of contour functions. While [21] have shown that ε-erasure and Poisson sampling yield the same marginal distribution, the joint distributions are not the same, and hence we only obtain convergence in distribution. But this is not good enough here. To study ε-erasure directly and get almost sure convergence in GHP back to the stable tree, [22] may help, where a reconstruction procedure demonstrates how subtrees (which contain all the mass) are attached to the ε-erased tree in order to get the stable tree back.

Proof of Theorem 1.5. From the three steps listed in the introduction (and completed in Section 3), we have marginal convergence in GH or GHP for each of the four components of $(\mathcal{G}^{(n)}/c_n, \widehat{\mathcal{G}}^{(n)}/c_n, \mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)}/c_n', \mathrm{cut}_{\mathrm{D}}(\widehat{\mathcal{G}}^{(n)}/c_n'))$. As GH-tightness implies GHP-tightness (see Miermont [38, Proposition 8]), the joint laws are GHP^4-tight. Take any subsequence along which we have convergence in distribution in GHP^4. By Skorokhod's representation theorem, we may assume that convergence holds almost surely, to a vector $((\mathcal{T}_1, \mu_1), \ldots, (\mathcal{T}_4, \mu_4))$ of measured limiting trees.

As GHP^2-convergence implies GP^2-convergence, we get from Step 3 that $((\mathcal{T}_2, \mu_2), (\mathcal{T}_4, \mu_4)) \sim ((\mathcal{T}_{\mathrm{Br}}, \mu_{\mathrm{Br}}), (\mathrm{cut}(\mathcal{T}_{\mathrm{Br}}), \mu_{\mathrm{cut}}))$, by uniqueness of GP^2-limits. By Step 2, we obtain $(\mathcal{T}_1, \mu_1) = (\mathcal{T}_2, \mu_2)$ a.s.. By Step 1, we obtain $\mathcal{T}_3 = \mathcal{T}_4$ a.s.. Finally, $(\mathcal{T}_{\mathrm{Br}}, \mu_{\mathrm{Br}})$ is a measurable function of $\mathcal{T}_{\mathrm{Br}}$, by Proposition 1.6, and therefore, both $(\mathcal{T}_3, \mu_3)$ by GP convergence in Step 3 and $(\mathcal{T}_4, \mu_4)$ by GHP convergence in Step 1 are this measureable function of $\mathcal{T}_3 = \mathcal{T}_4$ a.s.. This completely specifies the joint distribution of $((\mathcal{T}_1, \mu_1), \ldots, (\mathcal{T}_4, \mu_4))$, which furthermore does not depend on the chosen subsequence. Therefore, joint convergence in distribution holds with the limiting distribution thus identified. $\square$

3 Proof of the statements of Steps 1–3.

3.1 Step 1: GHP convergence of vertex cut-trees as Markov branching trees

Proof of Proposition 1.1. Denote by $\mathbb{T}$ the set of (combinatorial) rooted planar trees. Let $\mathcal{G}$ be a $\mathbb{T}$-valued Galton-Watson tree, and denote by $X = \lambda(\mathcal{G})$ the number of leaves of $\mathcal{G}$. First note that for all trees $\mathbf{t} \in \mathbb{T}$ with n leaves and root ρ, we have

$$\mathbb{P}(\mathcal{G} = \mathbf{t} \,|\, \lambda(\mathcal{G}) = n) = \mathbb{P}(\mathcal{G} = \mathbf{t} \,|\, X = n) = \frac{\mathbb{P}(\mathcal{G} = \mathbf{t})}{\mathbb{P}(X = n)} = \frac{1}{\mathbb{P}(X = n)} \prod_{v \in V(\mathbf{t})} \nu_{k_v(\mathbf{t})},$$

where $V(\mathbf{t})$ denotes the set of vertices of $\mathbf{t}$ and $k_v(\mathbf{t})$ the degree (number of subtrees of vertex $v \in V(\mathbf{t})$, not counting the component containing ρ). For any branch point $v \in \mathrm{Br}(\mathbf{t})$, splitting $\mathbf{t}$ into $\mathbf{t}_1, \ldots, \mathbf{t}_{k+1}$ by removing the edges $w \to v$ for all children w of v, where $\mathbf{t}_1$ is the component containing ρ and v, and $\mathbf{t}_2, \ldots, \mathbf{t}_{k+1}$ are the components of each of the children of v, in planar order, we obtain

$$\prod_{j=1}^{k+1} \mathbb{P}(\mathcal{G} = \mathbf{t}_j) = \frac{\nu_0}{\nu_k} \mathbb{P}(\mathcal{G} = \mathbf{t}).$$

Note that if we also record the new leaf $v \in \mathrm{Lf}(\mathbf{t}_1)$, we can uniquely reconstruct $(\mathbf{t}, v)$ from $(\mathbf{t}_1, \ldots, \mathbf{t}_{k+1}, v)$. Hence, the probability that the first cut is at a branch point with k children is

$$\begin{aligned} q_{\overline{n}}(\#\text{blocks} = \overline{k}) \quad &:= \sum_{\mathbf{t} \in \mathbb{T} \colon \lambda(\mathbf{t}) = n} \; \sum_{v \in \mathrm{Br}(\mathbf{t}) \colon k_v(\mathbf{t}) = k} \mathbb{P}(\mathcal{G} = \mathbf{t} \,|\, \lambda(\mathcal{G}) = n) \, \frac{k-1}{n-1} \\ &= \sum_{\substack{\mathbf{t}_1, \ldots, \mathbf{t}_{k+1} \in \mathbb{T} \colon \\ \lambda(\mathbf{t}_1) + \cdots + \lambda(\mathbf{t}_{k+1}) = n+1}} \; \sum_{v \in \mathrm{Lf}(\mathbf{t}_1)} \frac{1}{\mathbb{P}(X = n)} \, \frac{k-1}{n-1} \, \frac{\nu_k}{\nu_0} \prod_{j=1}^{k+1} \mathbb{P}(\mathcal{G} = \mathbf{t}_j). \end{aligned}$$

By symmetry, this value is exactly the same if the second sum is taken over $v \in \mathrm{Lf}(\mathbf{t}_i)$ for any $i = 1, \ldots, k+1$. Hence, summing over i and dividing by $k+1$, the second sum captures all $n+1$ leaves leaving the first sum to sum over all $(k+1)$-tuples of trees with total $n+1$ leaves, so that

$$q_{\overline{n}}(\#\text{blocks} = \overline{k}) = \frac{(n+1)(k-1)\nu_k \mathbb{P}(X_1 + \cdots + X_{k+1} = n+1)}{(k+1)\mathbb{P}(X = n)(n-1)\nu_0},$$

for independent X_j, $1 \leq j \leq k+1$, with the same distribution as $X = \lambda(\mathcal{G})$, as required. The joint distribution of the $k+1$ non-trivial and $k-2$ trivial components follows by a refinement of the above argument: denote by $\mathcal{S}_1$ the root component and by $\mathcal{S}_2, \ldots, \mathcal{S}_{k+1}$ the subtrees, then the argument yields a probability to see a Galton-Watson tree $\mathcal{G}^{(n)}$ with n leaves split into $\mathcal{S}_1 = \mathbf{s}_1$ and $\mathcal{S}_2 = \mathbf{s}_2, \ldots, \mathcal{S}_{k+1} = \mathbf{s}_{k+1}$ of

$$\frac{1}{\mathbb{P}(X = n)} \frac{k-1}{n-1} \frac{\nu_k}{\nu_0} \lambda(\mathbf{s}_1)\mathbb{P}(\mathcal{G} = \mathbf{s}_1) \prod_{j=2}^{k+1} \mathbb{P}(\mathcal{G} = \mathbf{s}_j),$$

and a simple combinatorial argument to handle equal block sizes yields the probability that the ranked split of $n+1$ is $(\lambda(\mathcal{S}_1), \ldots, \lambda(\mathcal{S}_{k+1}))^{\downarrow} = (m_1, \ldots, m_{k+1})$ as

$$\frac{1}{\mathbb{P}(X = n)} \frac{k-1}{n-1} \frac{\nu_k}{\nu_0} \frac{(n+1)k!}{\prod_{1 \leq \ell \leq n} r_\ell!} \prod_{j=1}^{k+1} \mathbb{P}(X = m_j),$$

where $r_\ell = \#\{1 \leq j \leq k+1 \colon m_j = \ell\}$ is the number of block sizes equal to ℓ. Hence, the conditional probability to see a split into $\mathcal{S}_1 = \mathbf{s}_1, \ldots, \mathcal{S}_{k+1} = \mathbf{s}_{k+1}$ given a ranked split of $(m_1, \ldots, m_{k+1})$ is

$$\frac{\lambda(\mathbf{s}_1)}{n+1} \frac{\prod_{1 \leq \ell \leq n} r_\ell!}{k!} \prod_{j=1}^{k+1} \mathbb{P}(\mathcal{G} = \mathbf{s}_j \,|\, X = \lambda(\mathbf{s}_j)).$$

The Markov branching property follows if we can show that conditionally given the ranked split $(m_1, \ldots, m_{k+1})$, the multiset of trees $\{\{\mathcal{S}_1, \ldots, \mathcal{S}_{k+1}\}\}$ has the same distribution as the multiset of $k+1$ independent trees with respective distribution $\mathbb{P}(\mathcal{G} = \cdot \,|\, X = m_j)$, $1 \leq j \leq k+1$. First suppose that the trees $\mathbf{t}_1, \ldots, \mathbf{t}_{k+1}$ are distinct. Then the probability that the multiset of trees $\{\{\mathcal{S}_1, \ldots, \mathcal{S}_{k+1}\}\}$ equals $\{\mathbf{t}_1, \ldots, \mathbf{t}_{k+1}\}$ is the sum over all $\mathbf{s}_1, \ldots, \mathbf{s}_{k+1}$ that are permutations of $\mathbf{t}_1, \ldots, \mathbf{t}_{k+1}$. In particular, $\mathbf{s}_1$ can be any $\mathbf{t}_i$, giving different factors $\lambda(\mathbf{t}_i)$, and there are $k!$ equally likely ways to match the others:

$$k! \sum_{i=1}^{k+1} \frac{\lambda(\mathbf{t}_i)}{n+1} \frac{\prod_{1 \leq \ell \leq n} r_\ell!}{k!} \prod_{j=1}^{k+1} \mathbb{P}(\mathcal{G} = \mathbf{s}_j \,|\, X = \lambda(\mathbf{s}_j)) = \prod_{1 \leq \ell \leq n} r_\ell! \prod_{j=1}^{k+1} \mathbb{P}(\mathcal{G} = \mathbf{s}_j \,|\, X = \lambda(\mathbf{s}_j)).$$

When some of the trees $\mathbf{t}_1, \ldots, \mathbf{t}_{k+1}$ are equal, there is duplication in some of the matchings of $\mathbf{s}_1, \ldots, \mathbf{s}_{k+1}$ and $\mathbf{t}_1, \ldots, \mathbf{t}_{k+1}$, and we lose some factors from $\prod_{1 \leq \ell \leq n} r_\ell!$. In each case, we get the probability that the multiset of independent conditioned Galton-Watson trees equals the multiset of $\mathbf{t}_1, \ldots, \mathbf{t}_{k+1}$, as required. $\square$

Proposition 3.1. *Suppose $\alpha = 2$ and the offspring variance σ^2 is finite. Let $(T_n, n \geq 1)$ be a family of Markov branching trees with splitting rule as given in Proposition 1.1, so that T_n is the genealogical tree of a fragmentation process starting from an initial block of size $\overline{n} = 2n - 1$, equipped with the uniform measure on the $\overline{n}$ leaves of T_n. Then*

$$\frac{T_n}{\sqrt{n}} \to \frac{\sqrt{\nu_0}}{\sigma} \mathcal{T}_{\mathrm{Br}}, \qquad \textit{in distribution in GHP},$$

where $\mathcal{T}_{\mathrm{Br}}$ is a Brownian Continuum Random Tree equipped with its usual mass measure.

Proof. Like Rizzolo [47] who applied the arguments of [28, Section 5.1] for his results on trees with numbers of vertices in a given set of degrees, we only present the part of the argument that differs from their's in some details and thereby reveals the constants in the limiting expression. Let $l_1([0,\infty))$ be the space of nonnegative summable sequences with sum bounded by 1 equipped with the l_1-norm, and $f\colon l_1([0,\infty))\to[0,\infty)$ bounded continuous. Set $g(x)=(1-\max x)f(x)$. Then numerous applications of the local limit theorem (Corollary 2.4) yield that for all $\eta>0$ and $\eta'<\eta$ small enough there is $n_0\ge 1$ and $\varepsilon>0$ such that for all $n\ge n_0$, $1\le k\le\varepsilon\sqrt{n}$ and $m=(m_1,\dots,m_{\overline{k}})$ with $n^{1/8}\le m_1\le(1-\eta)n$ and $(1-\eta')n\le m_1+m_2\le n$ and $\overline{m}_1+\cdots+\overline{m}_{\overline{k}}=\overline{n}$

$$\begin{aligned}
\frac{(k-1)\nu_k}{\nu_0}(1-\eta) \le{}& \quad q_{\overline{n}}(\#\text{blocks}=\overline{k}) &&\le \frac{(k-1)\nu_k}{\nu_0}(1+\eta)\\
\left(g\left(\frac{(\overline{m}_1,\overline{n-m_1},1,0,\dots)}{\overline{n}}\right)-\eta\right)^+ \le{}& \quad g\left(\tfrac{\overline{m}}{\overline{n}}\right) &&\le g\left(\frac{(\overline{m}_1,\overline{n-m_1},1,0,\dots)}{\overline{n}}\right)+\eta\\
\frac{k+1}{(n+1)^{3/2}}\frac{\sqrt{\nu_0}}{\sqrt{2\pi\sigma^2}}(1-\eta) \le{}& \quad \mathbb{P}(\tau_{k+1}=n+1) &&\le \frac{k+1}{(n+1)^{3/2}}\frac{\sqrt{\nu_0}}{\sqrt{2\pi\sigma^2}}(1+\eta)\\
\frac{1}{m_1^{3/2}}\frac{1}{m_2^{3/2}}\frac{\nu_0}{2\pi\sigma^2}(1-\eta)^2 \le{}& \mathbb{P}(X_1=m_1)\mathbb{P}(X_2=m_2) &&\le \frac{1}{m_1^{3/2}}\frac{1}{m_2^{3/2}}\frac{\nu_0}{2\pi\sigma^2}(1+\eta)^2\\
1-\eta \le{}& \quad \mathbb{P}(\tau_{\varepsilon\sqrt{n}}\le\eta' n) &&\le 1\\
\left(1-\frac{\eta'}{\eta}\right)(n-m_1) \le{}& \quad m_2 &&\le n-m_1.
\end{aligned}$$

These estimates allow us to check the criterion of [28, Display (3)] in terms of normalised splitting rules, which take the following form in our context:

$$\overline{q}^*_{\overline{n}}(g):=\sum_{k\ge 1}\ \sum_{m:=(m_1,\dots,m_{k+1},1,\dots,1)\colon\,\overline{m}_1+\cdots+\overline{m}_{\overline{k}}=\overline{n}} q_{\overline{n}}(m)g\left(\frac{\overline{m}}{\overline{n}}\right).$$

Specifically, by taking lim sup and lim inf as $n\to\infty$ and then the limit as $\eta\to 0$, under which contributions outside the above ranges of k, m_1 and m_2 vanish, we see that

$$\begin{aligned}
\sqrt{n}\overline{q}^*_{\overline{n}}(g)\quad\sim\quad & \sqrt{n}\sum_{k\ge1}q_{\overline{n}}(\#\text{blocks}=k)\sum_m g\left(\frac{\overline{m}}{\overline{n}}\right)\frac{(k+1)\overline{m}_1}{\overline{n}}\frac{k\overline{m}_2}{\overline{n}-\overline{m}_1}\frac{\mathbb{P}(X_1=m_1)\mathbb{P}(X_2=m_2)}{\mathbb{P}(\tau_{k+1}=n+1)}\\
&\qquad \mathbb{P}(X_3^*=m_3,\dots,X^*_{\overline{k}}=m_{\overline{k}}|X_1^*=m_1,X_2^*=m_2,\tau_{k+1}=n+1)\\
\longrightarrow\quad & \frac{1}{\sqrt{2\pi\sigma^2\nu_0}}\sum_{k\ge1}(k-1)k\nu_k\int_0^1 g(x,1-x,0,\dots)\frac{1}{x^{1/2}(1-x)^{3/2}}dx,
\end{aligned}$$

where the first line only fails to be an equality because $X^*=(X_1^*,\dots,X^*_{\overline{k}})$ is a size-biased rearrangement of $(X_1,\dots,X_{k+1},1,\dots,1)$, so the exact expressions in the negligible cases where $m_1=1$ or $m_2=1$ are different. Since $\sum_{k\ge1}(k-1)k\nu_k=\sigma^2$, we conclude by the convergence theorem of Haas and Miermont, [28, Theorem 1]. In particular we see from the multiplicative constant of the limiting measure that the limiting tree has (ranked) dislocation measure

$$\frac{\sigma}{\sqrt{2\pi\nu_0}}\left(\frac{1}{x^{1/2}(1-x)^{3/2}}+\frac{1}{(1-x)^{1/2}x^{3/2}}\right)1_{[1/2,1)}(x)dx=\frac{\sigma}{2\sqrt{\pi\nu_0}}\nu_B(dx),$$

which is associated with $\sqrt{\nu_0}\sigma^{-1}\mathcal{T}_{\mathrm{Br}}$; see Appendix B for a discussion of normalisations of the Brownian CRT and its dislocation measure. □

This identifies $c'_n=\sqrt{\nu_0}\sqrt{n}/\sigma$. Note that Step 3 of our programme therefore is to show, for Galton-Watson trees $\mathcal{G}^{(n)}$ with n leaves, joint GP convergence in distribution of

$$\left(\frac{\sqrt{\nu_0}}{\sqrt{n}}\sigma\mathcal{G}^{(n)},\frac{1}{\sqrt{\nu_0}\sqrt{n}}\sigma\mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)})\right)\to(\mathcal{T}_{\mathrm{Br}},\mathrm{cut}(\mathcal{T}_{\mathrm{Br}})).$$

For Galton-Watson trees $\mathcal{G}_V^{(n)}$ with n vertices, Bertoin, Miermont and Dieuleveut showed

$$\left(\frac{1}{\sqrt{n}}\sigma\mathcal{G}_V^{(n)}, \frac{1}{\sqrt{n}}\frac{1}{\sigma}\mathrm{cut}_{\mathrm{BM}}(\mathcal{G}_V^{(n)})\right) \to (\mathcal{T}_{\mathrm{Br}}, \mathrm{cut}(\mathcal{T}_{\mathrm{Br}})),$$
$$\left(\frac{1}{\sqrt{n}}\sigma\mathcal{G}_V^{(n)}, \frac{1}{\sqrt{n}}\left(\sigma+\frac{1}{\sigma}\right)\mathrm{cut}_{\mathrm{D}}(\mathcal{G}_V^{(n)})\right) \to (\mathcal{T}_{\mathrm{Br}}, \mathrm{cut}(\mathcal{T}_{\mathrm{Br}})).$$

We understand the appearance of $\sqrt{\nu_0}$, which is simply due to the different conditioning: $\mathcal{G}^{(n)}$ is conditioned to have n leaves, while $\mathcal{G}_V^{(n)}$ is conditioned to have n vertices. Dieuleveut gave a heuristic interpretation of her factor $\sigma + 1/\sigma$, comparing to just $1/\sigma$ for Bertoin and Miermont by referring to the fact that the number of edges removed in a vertex fragmentation is $k \geq 2$, so she gets $\sum_k k\nu_k \times k = \sigma^2 + 1$ as the speed-up compared to Bertoin and Miermont. Here, the first k in the sum reflects the fact that a branch point with k children is selected with probability proportional to k. This is what we have changed. Therefore, the average number of edges we remove is smaller when we drop the rate to being proportional to $k-1$, and we get $\sum_k (k-1)\nu_k \times k = \sigma^2$ as the speed-up compared to Bertoin and Miermont.

We deduce $(\mathrm{cut}_{\mathrm{D}}(\widehat{\mathcal{G}}^{(n)})/c'_n, \mathrm{cut}_{\mathrm{HW}}(\mathcal{G}^{(n)})/c'_n, \mathrm{cut}^{\circ}_{\mathrm{HW}}(\mathcal{G}^{(n)})/c'_n) \to (\mathcal{T}_{\mathrm{Br}}, \mathcal{T}_{\mathrm{Br}}, \mathcal{T}_{\mathrm{Br}})$ in GH^3 by Lemma 1.3 and Proposition 1.4 and since GHP convergence implies GH convergence, completing Step 1 for finite variance offspring distribution. Step 1 for offspring distributions in the domain of an infinite variance stable distribution is beyond the scope of this paper. The interested reader is referred to [28, Section 5.2], where Haas and Miermont establish the invariance principle for infinite-variance Galton-Watson trees using their convergence criterion. Their arguments would need to be adapted to cut-trees with splitting rule given in Proposition 1.1.

3.2 Step 2: Coding function convergence of modified Galton-Watson trees

Given a rooted planar tree $\mathbf{t}$, recall that $\zeta(\mathbf{t})$ and $\lambda(\mathbf{t})$ denote the total number of vertices and leaves of $\mathbf{t}$, respectively. Define the Lukasiewicz path, contour function and height function, denoted by $\mathcal{X}(\mathbf{t}), \mathcal{C}(\mathbf{t}), \mathcal{H}(\mathbf{t})$, as follows. To define $\mathcal{C}(\mathbf{t})$, consider a particle that visits the tree in planar order, starting from the root and moving continuously at unit speed up and down the edges of unit length, for each branch point exploring the subtrees in the (left to right) planar order. Then for $s \in [0, 2\zeta(\mathbf{t})]$, let $\mathcal{C}_s(\mathbf{t})$ be the distance of the particle to the root at time s. To define $\mathcal{X}(\mathbf{t})$ and $\mathcal{H}(\mathbf{t})$, let $\{v_j(\mathbf{t})\colon j = 0, 1, \ldots, \zeta(\mathbf{t}) - 1\}$ be the vertices of $\mathbf{t}$ in the order encountered by $\mathcal{C}(\mathbf{t})$, without duplication. The height function $\mathcal{H}(\mathbf{t})$ is defined by letting $\mathcal{H}_j(\mathbf{t})$ be the generation or *height* $|v_j(\mathbf{t})|$ of vertex $v_j(\mathbf{t})$. The Lukasiewicz path is defined by $\mathcal{X}_0(\mathbf{t}) = 0$ and
$$\mathcal{X}_{j+1}(\mathbf{t}) = \mathcal{X}_j(\mathbf{t}) + k_{v_j(\mathbf{t})}(\mathbf{t}) - 1, \quad j = 0, \ldots, \zeta(\mathbf{t}) - 1,$$
where $k_{v_j(\mathbf{t})}(\mathbf{t})$ is the number of children of $v_j(\mathbf{t})$ in $\mathbf{t}$. Further denote by
$$\Lambda_0(\mathbf{t}) = 0, \qquad \Lambda_k(\mathbf{t}) = \#\{j \leq k\colon \mathcal{X}_j(\mathbf{t}) - \mathcal{X}_{j-1}(\mathbf{t}) = -1\}, \quad 1 \leq k \leq \zeta(\mathbf{t}),$$
the leaf counting process of $\mathbf{t}$.

Let $\mathcal{G}^{(n)}$ be a critical Galton-Watson tree with n leaves. We recall from [34, Theorem 8.1] and [35, Theorem 3.3] the invariance principle for Galton-Watson trees in terms of coding functions, expressed as a joint convergence on the Skorokhod space $D[0,1]$ of càdlàg functions on $[0,1]$.

Proposition 3.2. *In the setting of Section 2.1, we have*
$$\sup_{0\leq t\leq 1}\left|\frac{\Lambda_{[\zeta(\mathcal{G}^{(n)})t]}(\mathcal{G}^{(n)})}{n} - t\right| \tag{6}$$
together with
$$\left(\frac{1}{a_{\zeta(\mathcal{G}^{(n)})}}\mathcal{X}_{[\zeta(\mathcal{G}^{(n)})t]}(\mathcal{G}^{(n)}), \frac{a_{\zeta(\mathcal{G}^{(n)})}}{\zeta(\mathcal{G}^{(n)})}\mathcal{C}_{2\zeta(\mathcal{G}^{(n)})t}(\mathcal{G}^{(n)}), \frac{a_{\zeta(\mathcal{G}^{(n)})}}{\zeta(\mathcal{G}^{(n)})}\mathcal{H}_{[\zeta(\mathcal{G}^{(n)})t]}(\mathcal{G}^{(n)})\right)_{0\leq t\leq 1} \tag{7}$$
converge in distribution in $[0,1]\times(D[0,1])^3$*, as* $n \to \infty$*, to* $(0, X, H, H)$*, where* X *is a normalised stable excursion and* H *is Duquesne and Le Gall's [19] stable height function. If* $a_n = \sigma\sqrt{n/2}$*, then* $H = X = \sqrt{2}B^{\mathrm{ex}}$ *is a multiple of the normalized excursion* B^{ex} *of linear Brownian motion.*

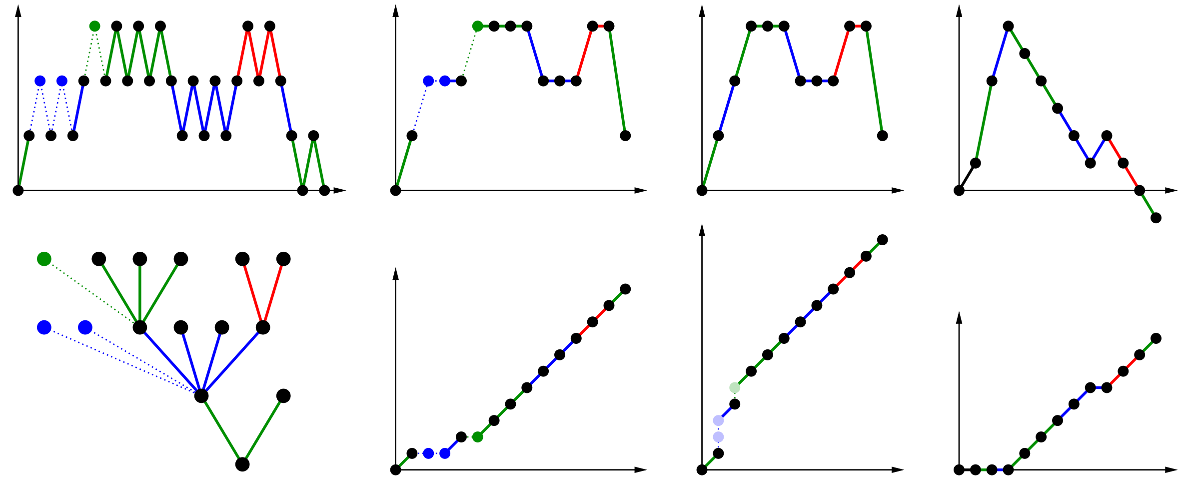

Figure 2: Illustration for some $\mathcal{G}^{(n)}$ with $n = 8$. Top row: $\mathcal{C}(\widehat{\mathcal{G}}^{(n)})$, $\mathcal{H}(\widehat{\mathcal{G}}^{(n)})$, $\mathcal{H}(\mathcal{G}^{(n)})$, $\mathcal{X}(\mathcal{G}^{(n)})$, bottom row: $\widehat{\mathcal{G}}^{(n)}$, φ_n and ψ_n and $\Lambda(\mathcal{G}^{(n)})$. Coloured lines only illustrate how functions relate.

Recall that $\widehat{\mathcal{G}}^{(n)}$ is the modified tree associated with $\mathcal{G}^{(n)}$ as introduced just before Lemma 1.2. Let us be precise and extend the planar order of $\mathcal{G}^{(n)}$ to $\widehat{\mathcal{G}}^{(n)}$ by placing all extra children to the left. Following ideas of Miermont [37] and de Raphélis [16], we introduce the following notation. Let $\widehat{u}(i) = v_i(\widehat{\mathcal{G}}^{(n)})$, $i = 0, 1, \ldots, \zeta(\widehat{\mathcal{G}}^{(n)}) - 1$, and $u(j) = v_j(\mathcal{G}^{(n)})$, $j = 0, 1, \ldots, \zeta(\mathcal{G}^{(n)}) - 1\}$, be the planar enumerations of $V(\widehat{\mathcal{G}}^{(n)})$ and $V(\mathcal{G}^{(n)})$, respectively. For $0 \le i \le \zeta(\widehat{\mathcal{G}}^{(n)}) - 1$, define

$$\varphi_n(i) = j \qquad \text{if} \qquad \begin{array}{l} u(j) = \widehat{u}(i) \in \mathcal{G}^{(n)} \\ \text{or } u(j) \text{ is the parent of } \widehat{u}(i) \notin \mathcal{G}^{(n)}. \end{array} \tag{8}$$

This means that $\varphi_n(i)$ is the index of the corresponding vertex in $\mathcal{G}^{(n)} \subseteq \widehat{\mathcal{G}}^{(n)}$, or if $\widehat{u}(i)$ is an extra child, $\varphi_n(i)$ is the index of its parent, which will be in $\mathcal{G}^{(n)}$ as no extra children have offspring. For $0 \le j \le \zeta(\mathcal{G}^{(n)}) - 1$, define $\psi_n(j) = \#\{v \in \widehat{\mathcal{G}}^{(n)} : v \prec u(j)\}$, where $v = \widehat{u}(k) \prec u(j) = \widehat{u}(\varphi_n(i))$ if and only if $k < j$, i.e. v has strictly smaller index in $\widehat{\mathcal{G}}^{(n)}$ than $u(j)$. Then $\varphi_n(\psi_n(j)) = j$, but $\psi_n(\varphi_n(i)) = i$ only if $\widehat{u}(i) \in \mathcal{G}^{(n)}$.

Hence, the function $\left(\frac{1}{\zeta(\widehat{\mathcal{G}}^{(n)})}\psi_n([\zeta(\mathcal{G}^{(n)})t]), 0 \le t \le 1\right)$ can be regarded as the right inverse of $\left(\frac{1}{\zeta(\mathcal{G}^{(n)})}\varphi_n([\zeta(\widehat{\mathcal{G}}^{(n)})t]), 0 \le t \le 1\right)$, in the approximate sense that their composition is the step function with steps $1/\zeta(\mathcal{G}^{(n)})$ at times $j/\zeta(\mathcal{G}^{(n)})$, $1 \le j \le \zeta(\mathcal{G}^{(n)})$. As $\zeta(\mathcal{G}^{(n)}) \ge \lambda(\mathcal{G}^{(n)}) = n$ will tend to infinity, this composition will approach the identity on $[0, 1]$ uniformly in $t \in [0, 1]$.

Proposition 3.3. *In the setting of the previous theorem, with* $\widetilde{a}_n = a_n/\nu_0^{1/\alpha}$*, we also have joint convergence in distribution in* $[0,1]^2 \times (D[0,1])^3$ *of*

$$\sup_{0\le t\le 1}\left|\frac{\Lambda_{[\zeta(\mathcal{G}^{(n)})t]}(\mathcal{G}^{(n)})}{n} - t\right|, \qquad \sup_{0\le t\le 1}\left|\frac{\varphi_n([\zeta(\widehat{\mathcal{G}}^{(n)})t])}{\zeta(\mathcal{G}^{(n)})} - t\right|,$$

together with

$$\left(\frac{1}{\widetilde{a}_n}\mathcal{X}_{[\zeta(\mathcal{G}^{(n)})t]}(\mathcal{G}^{(n)}), \frac{\widetilde{a}_n}{n}\mathcal{C}_{2\zeta(\mathcal{G}^{(n)})t}(\mathcal{G}^{(n)}), \frac{\widetilde{a}_n}{n}\mathcal{H}_{[\zeta(\widehat{\mathcal{G}}^{(n)})t]}(\widehat{\mathcal{G}}^{(n)})\right)_{0\le t\le 1} \tag{9}$$

to $(0, 0, X, H, H)$.

The proof will be based on the following lemma.

Lemma 3.4. *We have*

$$\sup_{0\le t\le 1}\left|\frac{1}{\zeta(\mathcal{G}^{(n)})}\varphi_n([\zeta(\widehat{\mathcal{G}}^{(n)})t]) - t\right| \to 0, \quad \textit{in probability, as } n \to \infty.$$

Proof. According to the definition of ψ_n, and by the convention that extra children are placed to the left of other children (and hence enumerated first), we have for any $\ell < \zeta(\mathcal{G}^{(n)})$,

$$\psi_n(\ell+1) = \Lambda_{\ell+1}(\mathcal{G}^{(n)}) + \sum_{j\leq \ell} 1_{\{u(j)\in \mathrm{Br}(\mathcal{G}^{(n)})\}} + \sum_{j\leq \ell} 1_{\{u(j)\in \mathrm{Br}(\mathcal{G}^{(n)})\}}(k_{u(j)}(\mathcal{G}^{(n)}) - 2).$$

Meanwhile, by definition of the Lukasiewicz path, we have

$$\sum_{j\leq \ell} 1_{\{u(j)\in \mathrm{Br}(\mathcal{G}^{(n)})\}}(k_{u(j)}(\mathcal{G}^{(n)}) - 1) - \Lambda_\ell(\mathcal{G}^{(n)}) = \mathcal{X}_\ell(\mathcal{G}^{(n)}).$$

Thus,

$$\psi_n(\ell+1) = \Lambda_{\ell+1}(\mathcal{G}^{(n)}) + \Lambda_\ell(\mathcal{G}^{(n)}) + \mathcal{X}_\ell(\mathcal{G}^{(n)}).$$

Since $\zeta(\widehat{\mathcal{G}}^{(n)}) = 2n-1$ and $\widetilde{a}_n \sim o(n)$, one can immediately see from Proposition 3.2 that $\mathcal{X}_\ell(\mathcal{G}^{(n)})$ is asymptotically negligible when scaling by $\zeta(\widehat{\mathcal{G}}^{(n)}$ and

$$\sup_{0\leq t\leq 1}\left|\frac{1}{\zeta(\widehat{\mathcal{G}}^{(n)})}\psi_n([\zeta(\mathcal{G}^{(n)})t]) - t\right| \longrightarrow 0, \quad \text{in probability.}$$

By definition of φ_n and ψ_n, one sees $\varphi_n(\psi_n(k)) = k$. So for fixed $t \in [0,1]$, as $n \to \infty$,

$$\frac{1}{\zeta(\mathcal{G}^{(n)})}\varphi_n([\zeta(\widehat{\mathcal{G}}^{(n)})t]) \longrightarrow t, \quad \text{in probability.}$$

Since $t \mapsto \varphi_n([\zeta(\widehat{\mathcal{G}}^{(n)})t])$ is non-decreasing for each $n \geq 1$, Dini's Theorem yields

$$\left(\frac{1}{\zeta(\mathcal{G}^{(n)})}\varphi_n([\zeta(\widehat{\mathcal{G}}^{(n)})t]), \quad 0\leq t\leq 1\right) \longrightarrow (t,\, 0\leq t\leq 1) \quad \text{in distribution.}$$

And hence the desired result holds since the identity function is deterministic and continuous. □

Remark 3.5. The tree $\widehat{\mathcal{G}}^{(n)}$ can be regarded as a 2-type Galton-Watson tree. The analogue of Lemma 3.4 was obtained by Miermont [37] for irreducible and non-degenerate multi-type Galton-Watson trees under a "small exponential moment" condition; see Lemma 6 and the proof of Theorem 2 there.

Proof of Proposition 3.3. We see that

$$\left|\mathcal{H}_{\varphi_n([\zeta(\widehat{\mathcal{G}}^{(n)})t])}(\mathcal{G}^{(n)}) - \mathcal{H}_{[\zeta(\widehat{\mathcal{G}}^{(n)})t]}(\widehat{\mathcal{G}}^{(n)})\right| \leq 1.$$

Thus with Lemma 3.4 and Proposition 3.2, we obtain as $n \to \infty$,

$$\left(\frac{a_{\zeta(\mathcal{G}^{(n)})}}{\zeta(\mathcal{G}^{(n)})}\mathcal{H}_{\varphi_n(\zeta(\widehat{\mathcal{G}}^{(n)})t)}(\mathcal{G}^{(n)})\right)_{0\leq t\leq 1} \xrightarrow{d} H.$$

Meanwhile, by [34, Lemma 2.7], we have

$$\frac{\zeta(\mathcal{G}^{(n)})}{n} \longrightarrow \frac{1}{\nu_0},$$

in distribution and hence in probability. Then a standard argument based on the Skorokhod representation theorem establishes the desired result. □

Uniform convergence of either height functions or contour functions with measures being pushed forward by the quotient maps implies GHP convergence, cf. [45, Lemma 1]. Hence, Proposition 3.3 together with (6) completes Step 2 with $c_n = n/(\sqrt{2}\widetilde{a}_n)$, not just in the finite-variance case with $c_n = \sqrt{n}/(\sigma\sqrt{\nu_0})$ by Lemma 2.1, but also for offspring distributions in the stable domain of attraction. In fact, the convergence of Lukasiewicz paths of $\widehat{\mathcal{G}}^{(n)}$ can be proved similarly.

3.3 Step 3: Joint GP convergence of the modified tree and its cut-tree

In the sequel, we mainly have the case of a finite-variance modified Galton-Watson tree in mind, but we include the stable case, where the argument is the same. From here, we follow Dieuleveut [17, Section 4] closely (and [17, Section 2] for the stable case, which contains some of the details also needed for the finite variance case). Let $\zeta_n = \zeta(\mathcal{G}^{(n)})$ and $\widehat{\zeta}_n = \zeta(\widehat{\mathcal{G}}^{(n)})$. Also write $(X^{(n)}, H^{(n)}, C^{(n)})$ for suitably scaled Lukasiewicz path $\mathcal{X}(\mathcal{G}^{(n)})$, height function $\mathcal{H}(\mathcal{G}^{(n)})$ and contour function $\mathcal{C}(\mathcal{G}^{(n)})$, $n \geq 1$, which converge to the corresponding triplet (X, H, H) associated with a stable tree $\mathcal{T}$ (including the Brownian CRT, in which case $X = H = \sqrt{2}B^{\mathrm{ex}}$).

Lemma 3.6 (cf. [17] Lemmas 2.4, 4.2)**.** *If $(H^{(n)}, C^{(n)}, X^{(n)}) \to (H, H, X)$ in distribution in $(D[0,1])^3$, then $(H^{(n)}, X^{(n)}, \widetilde{X}^{(n)}) \to (H, X, \widetilde{X})$ in distribution in $(D[0,1])^3$, where $\widetilde{X}^{(n)}$ and $\widetilde{X}$ are Lukasiewicz paths with all orders of children reversed.*

Proof. Dieuleveut's argument only uses the identical distribution of reversed quantities $(\widetilde{X}, \widetilde{C})$, the fact that $\widetilde{X}$ is a measurable function of the jump sizes and jump times of X to identify the limit in the stable case, and the symmetry $\widetilde{C}_t^{(n)} = C_{1-t}^{(n)}$ and continuity of H to identify the limit in the case of a Brownian limit. Hence, her argument also establishes this analogous result. □

Lemma 3.7 (cf. [17] Lemmas 2.7, 2.8, 4.3, 4.4)**.** *Let $(X^{(n)}, H^{(n)}, \widetilde{X}^{(n)}, U^{(n)}) \to (X, H, \widetilde{X}, U)$ almost surely, for some $U^{(n)} = (U_i^{(n)}, i \geq 1)$ and $U = (U_i, i \geq 1)$ with $U_i^{(n)} \in \{\frac{j}{\zeta_n}, 1 \leq j \leq \zeta_n\}$, and i.i.d. $U_i \sim \mathrm{Unif}(0,1)$ independent of $(X, H, \widetilde{X})$. Then we also have the following limits.*

- *The shape of the subtree $\mathcal{R}^{(n)}(k)$ of $\mathcal{G}^{(n)}$ spanned by $0, U_1^{(n)}, \ldots, U_k^{(n)}$ is constant a.s. for n large enough, equal to the shape $R(k)$ of the subtree $\mathcal{R}(k)$ of $\mathcal{T}$ spanned by $0, U_1, \ldots, U_k$.*

- *For every edge $e = (v \to v') \in E(R(k))$, denote by $e_+^{(n)}(k), e_-^{(n)}(k) \in V(\mathcal{R}^{(n)}(k))$ the vertices corresponding to $v = e^+(k)$ and $v' = e^-(k)$, and by $V_e^{(n)}(k)$ the set of vertices between $e_+^{(n)}(k)$ and $e_-^{(n)}(k)$. Then the rescaled lengths of the edge converge a.s.:*

$$\frac{\widetilde{a}_n}{n}\left(1 + \#V_e^{(n)}(k)\right) = H^{(n)}_{b_n(e_+^{(n)}(k))} - H^{(n)}_{b_n(e_-^{(n)}(k))} \to H_{b(e_+(k))} - H_{b(e_-(k))},$$

  where $b_n(w)$ is the first time of $H^{(n)}$ corresponding to $w \in V(\mathcal{G}^{(n)})$, similarly $b(w)$, $w \in \mathcal{T}$.

- *For every branch point $v \in \mathrm{Br}(R(k))$, rescaled numbers of children converge a.s., i.e.*

$$\frac{1}{\widetilde{a}_n} k_v(\mathcal{G}^{(n)}) \sim \frac{1}{\widetilde{a}_n}\left(k_v(\mathcal{G}^{(n)}) - 1\right) = \Delta X^{(n)}_{b_n(v)} \to \Delta X_{b(v)},$$

  which vanishes in the finite-variance case.

- *For every edge $e \in E(R(k))$, sums of rescaled numbers of children converge a.s., as follows:*

$$\frac{1}{\widetilde{a}_n} \sum_{v \in V_e^{(n)}(k)} \left(k_v(\mathcal{G}^{(n)}) - 1\right) \to (X_{b(e^+)} + \widetilde{X}_{\widetilde{b}(e^+)}) - (X_{b(e^-)} + \widetilde{X}_{\widetilde{b}(e^-)}) - \Delta X_{b(e^-)},$$

  which in the finite-variance case simplifies to $H_{b(e^+)} - H_{b(e^-)}$. If we replace $(k_v(\mathcal{T}_n) - 1)$ by $k_v(\mathcal{T}_n)$, we get the same limit in the stable case, while in the finite-variance case, $n/a_n^2 = 1/\sigma^2$ and we obtain a limit $(1 + 1/\sigma^2)(H_{b(e^+)} - H_{b(e^-)})$ instead.

Proof. Dieuleveut's arguments are entirely deterministic, just requiring the limiting random variables to avoid certain degeneracies a.s. In our context, take independent $(U_i, i \geq 1)$ and use Skorokhod's representation theorem to have the convergences of Proposition 3.3 and Lemma 3.6 jointly and almost surely. □

To apply Lemma 3.7, we now use U_i to sample a uniform edge in $\widehat{\mathcal{G}}^{(n)}$ and take as $U_i^{(n)}$ the corresponding time of $H^{(n)}$, i.e. $U_i^{(n)} = \varphi_n([(2n-2)U_i]+1)/\zeta(\mathcal{G}^{(n)})$, which is not independent of $\mathcal{G}^{(n)}$, but since φ_n converges uniformly to the identity on $[0,1]$, the almost sure convergence needed to apply Lemma 3.7 holds, with limit U_i independent of the limiting coding functions.

Proposition 3.8 (cf. [17] Propositions 2.5 and 4.1). *Consider edge samples $\xi_n(i)$ in $\widehat{\mathcal{G}}^{(n)}$ and the continuous-time Dieuleveut vertex fragmentation of $\widehat{\mathcal{G}}^{(n)}$ that removes the edges above vertex $v \in \mathrm{Br}(\widehat{\mathcal{G}}^{(n)})$ at rate $k_v(\widehat{\mathcal{G}}^{(n)})/2\widetilde{a}_n$. Define mass processes $(\mu_{n,\xi_n(i)}(t))_{t\geq 0}$ capturing the evolution of the proportion of leaves in the component containing $\xi_n(i)$, $i \geq 1$, and separation times $\tau_n(i,j)$ of $\xi_n(i)$ and $\xi_n(j)$, $i,j \geq 1$. Then in $\mathrm{GP}\times[0,\infty)^{\mathbb{N}} \times (D[0,1])^{\mathbb{N}}$, in distribution, as $n \to \infty$,*

$$\left(\frac{\widetilde{a}_n}{n}\widehat{\mathcal{G}}^{(n)}, (\tau_n(i,j))_{i,j\geq 1}, (\mu_{n,\xi_n(i)}(t))_{t\geq 0, i\geq 1}\right) \to \left(\mathcal{T}, (c^{-1}\tau(i,j))_{i,j\geq 1}, (\mu_{\xi(i)}(ct))_{t\geq 0, i\geq 1}\right),$$

where $c = 1$ in the stable case and/or when rates are proportional to $k-1$, while it is $c = 1+1/\sigma^2$ only in the finite variance case when rates are proportional to k.

Proof. Dieuleveut's arguments work since we can still sample $\xi_n(i)$ from $U_i^{(n)}$ in $[0,1]$, and the remaining arguments only depend on tree convergences and rate convergences (up to a factor of c), both of which we have, from Proposition 3.3, Lemma 3.6 and Lemma 3.7. □

The convergences achieved so far imply the convergence of certain modified distances for the discrete cut-trees. These modified distances resemble the Brownian cut-tree distances and take the following form. We enumerate the $2n-2$ edges of $\widehat{\mathcal{G}}^{(n)}$ by $1, \ldots, 2n-2$ and define for $i,j \in \{1,\ldots,2n-2\}$

$$\delta'_n(0,i) = \int_0^\infty \mu_{n,i}(t)dt \quad \text{and} \quad \delta'(i,j) = \int_{t_n(i,j)}^\infty (\mu_{n,i}(t) + \mu_{n,j}(t))dt,$$

where $t_n(i,j)$ is the most recent time when edges i and j were in the same component in the continuous-time vertex fragmentation of $\widehat{\mathcal{G}}^{(n)}$.

Lemma 3.9 (cf. [17] Lemma 2.1). *For all $i,j \in \{1,\ldots,2n-2\}$, we have*

$$\mathbb{E}\left[\left|\frac{\widetilde{a}_n}{n-1}\delta_n(i,j) - \delta'_n(i,j)\right|^2\right] \leq \frac{\widetilde{a}_n}{n-1}\mathbb{E}\left[\delta'_n(0,i) + \delta'_n(0,j)\right].$$

Proof. Dieuleveut works conditionally given the tree, so the argument applies to the tree $\widehat{\mathcal{G}}^{(n)}$ with $2n-2$ edges and the rates $k_v(\widehat{\mathcal{G}}^{(n)})/2\widetilde{a}_n$ that specify the continuous-time cutting. □

Lemma 3.10 (cf. [17] Lemma 4.5). *Assume $\nu_1 = 0$ and finite variance σ^2. Let ξ_n be uniform on $\{1,\ldots,2n-2\}$. Then*

$$\lim_{\ell\to\infty}\sup_{n\geq 1}\mathbb{E}\left[\int_{2^\ell}^\infty \mu_{n,\xi_n}(t)dt\right] = 0 \qquad \textit{and} \qquad \mathbb{E}[\delta'_n(0,\xi_n)] \leq C_0$$

for some $C_0 \in (0,\infty)$.

Proof. We use the same ideas from [13, Corollary 2] and [17, Lemma 4.5] to prove the result. We focus on where the arguments differ. As Dieuleveut pointed out, there is a coupling between vertex-fragmentation and edge-fragmentation by a deterministic procedure. So we directly follow the argument in [13] by considering uniform edge-cutting on $\widehat{\mathcal{G}}^{(n)}$. Recall that $V(\mathbf{t})$ is the set of vertices of $\mathbf{t}$. For a vertex $u \in V(\widehat{\mathcal{G}}^{(n)})$, let e_u be the edge pointing down from u towards the root, and for an edge e of $\widehat{\mathcal{G}}^{(n)}$, let $v(e)$ be the vertex such that $e_{v(e)} = e$. Then given $\widehat{\mathcal{G}}^{(n)}$, $v(\xi_n)$ is uniform in $V^*(\widehat{\mathcal{G}}^{(n)}) = V(\widehat{\mathcal{G}}^{(n)}) \setminus \{\rho\}$. Following Bertoin and Miermont's argument, we obtain

$$\begin{aligned}
\mathbb{E}[n\mu_{n,\xi_n}(t)] &\leq e^{-t/\sqrt{n}} + \mathbb{E}\left[\sum_{u\in V^*(\widehat{\mathcal{G}}^{(n)})\setminus\{v(\xi_n)\}} e^{-d(u,v(\xi_n))t/\sqrt{n}}\right] \\
&= e^{-t/\sqrt{n}} + \frac{1}{2n-2}\mathbb{E}\left[\sum_{u,v\in V^*(\widehat{\mathcal{G}}^{(n)}),\, u\neq v} e^{-d(u,v)t/\sqrt{n}}\right] \\
&\leq e^{-t/\sqrt{n}} + \frac{4}{2n-2}\mathbb{E}\left[\sum_{u,v\in V^*(\mathcal{G}^{(n)}),\, u\neq v} e^{-d(u,v)t/\sqrt{n}}\right], \qquad (10)
\end{aligned}$$

where the last inequality follows from the following observation: for each vertex $v \in V(\mathcal{G}^{(n)})$ with $k_v \geq 2$ children, say $v_1, \ldots, v_{k_v}$, there are $k_v - 2$ further children in $V(\widehat{\mathcal{G}}^{(n)}) \setminus V(\mathcal{G}^{(n)})$, say $v'_1, \ldots, v'_{k_v-2}$. Then for $u \in V(\widehat{\mathcal{G}}^{(n)})$, we have $d(u, v_i) + 2 = d(u, v'_i)$ if v_i is an ancestor of u; and $d(u, v_i) = d(u, v'_i)$ otherwise. If we replace $d(u, v'_i)$ by $d(u, v_i)$, then each v_i would be counted at most twice. We can similarly reduce the sum over u and gain another factor 2.

Denote by GW_* the sigma-finite measure on the space of pointed trees such that

$$GW_*(\overline{\mathbf{t}}, v) = \mathbb{P}(\overline{\mathcal{G}} = \overline{\mathbf{t}}),$$

where $\overline{\mathcal{G}}$ is the planted version of $\mathcal{G}$, with an edge and vertex added below the root, $\overline{\mathbf{t}}$ denotes a generic planted planar tree and $v \in V(\mathbf{t})$; see Sections 1.2 and 4 in [13]. Then notice that the set of pointed trees $(\overline{\mathbf{t}}, v)$ with exactly n leaves has GW_*-measure equal to $\mathbb{E}[\zeta(\mathcal{G})1_{\{\lambda(\mathcal{G})=n\}}] \in (0, \infty)$. So the conditional law $GW_*(\,\cdot \mid \lambda(\mathbf{t}) = n)$ on the space of pointed tree with n leaves is well defined and is the same to the distribution of $(\overline{\mathcal{G}}^{(n)}, \eta)$, where η is a uniformly chosen vertex in $V(\mathcal{G}^{(n)})$. We also note that if $\nu_1 = 0$, then $\#V(\mathcal{G}^{(n)}) = \zeta(\mathcal{G}^{(n)}) \leq 2n$. Thus one can deduce that

$$\begin{aligned}
\mathbb{E}[n\mu_{n,\xi_n}(t)] &\leq e^{-t/\sqrt{n}} + \frac{2}{n-1}\mathbb{E}\left[\sum_{u,v\in V^*(\mathcal{G}^{(n)}),\, u\neq v} e^{-d(u,v)t/\sqrt{n}}\right] \\
&\leq e^{-t/\sqrt{n}} + \frac{4n}{n-1}GW_*\left[\sum_{u\in V(\mathbf{t})\setminus\{v\}} e^{-d(u,v)t/\sqrt{n}} \;\middle|\; \lambda(\mathbf{t}) = n\right] \\
&\leq e^{-t/\sqrt{n}} + \frac{8n}{n-1}\sum_{k\geq 1} e^{-kt/\sqrt{n}}\mathbb{E}[\zeta_k(\mathcal{G}^{(n)})],
\end{aligned}$$

where the last inequality follows from the same argument as [13] by replacing $\#V(\mathbf{t})$ with $\lambda(\mathbf{t})$. Using Lemma 2.5, we obtain

$$\mathbb{E}(\mu_{n,\xi_n}(t)) \leq \frac{e^{-t/\sqrt{n}}}{n} + \frac{4C}{n}\sum_{k\geq 1} ke^{-kt/\sqrt{n}} \leq \frac{C' \exp(-t/\sqrt{n})}{n(1-\exp(-t/\sqrt{n}))^2}.$$

Then it is easy to see that

$$\lim_{l\to\infty} \sup_{n\geq 1} \mathbb{E}\int_{2^l}^{\infty} \mu_{n,\xi_n}(t)dt = 0 \quad \text{and} \quad \sup_{n\geq 1} \mathbb{E}(\delta'_n(\xi_n, 0)) = \sup_{n\geq 1}\int_0^{\infty} \mathbb{E}(\mu_{n,\xi_n}(t))dt < \infty.$$

This completes the proof. $\square$

Recall that $\widehat{\mathcal{G}}^{(n)}$ is the modified Galton-Watson tree conditioned to have n leaves, where the modification is the addition of $k-2$ extra leaves attached to branch points with k children, for every $k \geq 3$, for every branch point. This tree has $2n-2$ edges and is equipped with the uniform measure on those $2n-2$ edges. Recall further that $\mathrm{cut}_{\mathrm{D}}(\widehat{\mathcal{G}}^{(n)})$ denotes the Dieuleveut cut-tree of $\widehat{\mathcal{G}}^{(n)}$. This tree has $2n-2$ leaves and is equipped with the uniform measure on those $2n-2$ leaves, which we enumerate $1, \ldots, 2n-2$. Let $c_n = \sqrt{n}/(\sigma\sqrt{\nu_0})$ and $c'_n = \sqrt{\nu_0}\sqrt{n}/\sigma$.

Theorem 3.11 (cf. [17] Theorem 1.4)**.** *If the offspring distribution ν has finite variance σ^2, then we have* $\left(\frac{1}{c_n}\widehat{\mathcal{G}}^{(n)}, \frac{1}{c'_n}\mathrm{cut}_{\mathrm{D}}(\widehat{\mathcal{G}}^{(n)})\right) \longrightarrow (\mathcal{T}_{\mathrm{Br}}, \mathrm{cut}(\mathcal{T}_{\mathrm{Br}}))$ *in distribution in* GP^2*, as* $n \to \infty$.

Proof. With Proposition 3.8 and Lemmas 3.9 and 3.10, we have provided all ingredients for Dieuleveut's proof to apply to $\widehat{\mathcal{G}}^{(n)}$. $\square$

It should be possible to approach Lemmas 2.5 and 3.10 in the stable case and hence complete Step 3 also in the stable case, at least under some technical assumptions on the tail of the offspring distribution. Dieuleveut's corresponding arguments for her vertex cut-trees in [17, Section 2.3] are by far the most technical part of her paper, spread over 14 pages, and we do not see any new insights from adapting them, hence we do not pursue this here.

A Measurable construction of ε-erased length measures

Recall from Section 2.2 the definitions of the GH- and GHP-spaces of equivalence classes of compact real trees, which we denote by $\mathbb{T}$ and $\mathbb{T}_{\mathrm{w}}$. We further recall the notion of ε-erasure and the definition of the length measure ℓ_T of a real tree (T, d). Specifically, for a rooted real tree (T, d, ρ), we denote by $\ell_T|_{R_\varepsilon(T)} := \ell_T(\,\cdot\, \cap R_\varepsilon(T))$ the length measure of T restricted to the ε-erased subtree $R_\varepsilon(T) \subset T$, as a finite measure on T. Recall from [38] that there are alternative representations of the GH- and GHP-metrics, as follows:

$$d_{\mathrm{GH}}((T,d,\rho),(T',d',\rho')) = \frac{1}{2} \inf_{\mathcal{R}\in\mathcal{C}_c(T,T')\colon (\rho,\rho')\in\mathcal{R}} \mathrm{dis}(\mathcal{R}),$$

where $\mathcal{C}_c(T,T')$ is the set of all *correspondences*, compact subsets $\mathcal{R} \subset T \times T'$ with coordinate projections T and T' and *distortion* $\mathrm{dis}(\mathcal{R}) = \sup\{|d(x,y) - d'(x',y')|\colon (x,x'),(y,y') \in \mathcal{R}\}$. Also

$$d_{\mathrm{GHP}}((T,d,\rho,\mu),(T',d',\rho',\mu')) = \frac{1}{2} \inf_{\mathcal{R}\in\mathcal{C}_c(T,T'),\nu\in\mathcal{M}_p(\mu,\mu')\colon (\rho,\rho')\in\mathcal{R}} \mathrm{dis}(\mathcal{R}) \vee 2\nu(\mathcal{R}^c),$$

where $\mathcal{M}_p(\mu,\mu')$ is the set of all *partial couplings* ν, finite measures on $(T \times T') \cup T \cup T'$ with marginals μ and μ' in the sense that $\nu(\,\cdot \times T') + \nu(\,\cdot \cap T) = \mu$ and $\nu(T \times \cdot\,) + \nu(\,\cdot \cap T') = \mu'$. We interpret the component on $T \times T'$ as the coupled part and the components on T and T' as uncoupled. This extends Miermont's expression based on full couplings of probability measures.

Lemma A.1. *The association $(\varepsilon, T) \mapsto (T, \ell_T|_{R_\varepsilon(T)})$ induces a continuous map $(0,\infty)\times\mathbb{T} \to \mathbb{T}_{\mathrm{w}}$.*

Proof. Fix a rooted compact real tree (T, d, ρ) and $\varepsilon \in (0,\infty)$. Denote by n the number of leaves of $R_{\varepsilon/2}(T)$. Let $\delta > 0$ be so small that $\delta < \varepsilon/16$ and that the shortest branch (between adjacent branch points or between a branch point and a leaf) of $R_\varepsilon(T)$ has length greater than 50δ.

Let (T', d', ρ') be such that $d_{\mathrm{GH}}(T,T') < \delta$ and ε' so that $|\varepsilon - \varepsilon'| < \delta$. Then

$$d_{\mathrm{GH}}(R_\varepsilon(T), R_{\varepsilon'+4\delta}(T')) \leq d_{\mathrm{GH}}(R_\varepsilon(T), R_\varepsilon(T')) + d_{\mathrm{GH}}(R_\varepsilon(T'), R_{\varepsilon'+4\delta}(T')) < 3\delta + |\varepsilon'+4\delta-\varepsilon| < 8\delta.$$

Hence, there is a correspondence $\mathcal{R}$ between $R_\varepsilon(T)$ and $R_{\varepsilon'+4\delta}(T')$ of distortion less than 16δ. In particular, the midpoint of each branch of $R_\varepsilon(T)$ is in correspondence with points on a *corresponding branch* of $R_{\varepsilon'+4\delta}(T')$, and indeed all branches of $R_{\varepsilon'+4\delta}(T')$ of length greater than 48δ are attained by this *branch correspondence* and are arranged in the same shape as in $R_\varepsilon(T)$. Note that $R_{\varepsilon'+4\delta}(T')$ may have further internal branches of shorter length, but only at most $k-2$

branches for every branch point of $R_\varepsilon(T)$ of degree $k \geq 3$. While $R_{\varepsilon'}(T')$ may also have further external subtrees, these are all of height less than $3\delta + \varepsilon - \varepsilon'$, hence not part of $R_{\varepsilon'+4\delta}(T')$.

We now define a new correspondence $\mathcal{R}'$ between $R_\varepsilon(T)$ and $R_{\varepsilon'+4\delta}(T')$, in which we include (x, y') whenever there is $(y, y') \in \mathcal{R}$ with $d(x, y) \leq 24\delta$ or $(x, x') \in \mathcal{R}$ with $d'(x', y') \leq 24\delta$. Then it is not hard to see that for each pair of corresponding branches, $\mathcal{R}'$ includes the following *special pairs*: the endpoints closer to the respective root and the points at distance s from this endpoint for each s up to and including the length of the shorter of the two branches. Furthermore, the special pairs couple most of the length. The uncoupled length is bounded by $(2n-1)48\delta$ for each of $R_\varepsilon(T)$ and $R_{\varepsilon'+4\delta}(T')$. This specifies a partial coupling ν' with $\nu'((\mathcal{R}')^c) < (2n-1)96\delta$.

Finally considering the pair of $R_{\varepsilon/2}(T)$ and $R_{\varepsilon/2+4\delta}(T') \supset R_{\varepsilon'}(T')$, we argue as above to identify their tree shapes, so that $R_{\varepsilon'}(T') \setminus R_{\varepsilon'+4\delta}(T')$ consists of subtrees of height at most 4δ that together have fewer than $2n-1$ branches of length at most 4δ. Since $\mathrm{dis}(\mathcal{R}') < 40\delta$, we can extend $\mathcal{R}'$ to a correspondence $\mathcal{R}''$ between $R_\varepsilon(T)$ and $R_{\varepsilon'}(T')$ that has distortion $\mathrm{dis}(\mathcal{R}'') \leq 48\delta$ and extend the uncoupled part of ν' to get a partial coupling ν''. Then

$$d_{\mathrm{GH}}(T, T') < \delta,\ |\varepsilon - \varepsilon'| < \delta \quad \Rightarrow \quad d_{\mathrm{GHP}}((T, \ell_T|_{R_\varepsilon(T)}), (T', \ell_{T'}|_{R_{\varepsilon'}(T')})) < (2n-1)100\delta.$$

Since this bound does not depend on the isometry class of T, this establishes continuity of the induced map $(0, \infty) \times \mathbb{T} \to \mathbb{T}_{\mathrm{w}}$ at the pair ε in $(0, \infty)$ with the isometry class of T in $\mathbb{T}$. $\square$

B Three constant multiples of the Brownian CRT

Aldous [6, 7, 8] introduced the Brownian CRT $\mathcal{T}_{\mathrm{Ald}}$ via the line-breaking construction based on an inhomogeneous Poisson process of rate tdt. Since distances between consecutive points of the Poisson process are lengths in trees, intensity $ctdt$ yields $\mathcal{T}_{\mathrm{Ald}}/c$. Aldous's choice of intensity is such that the convergence of discrete uniform random trees with n vertices labelled $1, \ldots, n$ to $\mathcal{T}_{\mathrm{Ald}}$ is obtained when scaling edges by $\sqrt{n}$. Aldous shows in [8, Corollary 22] that $\mathcal{T}_{\mathrm{Ald}}$ has the same distribution as the tree $\mathcal{T}_{2B^{\mathrm{ex}}} = 2\mathcal{T}_{B^{\mathrm{ex}}}$, where $\mathcal{T}_{B^{\mathrm{ex}}}$ is the tree whose height function is the standard Brownian excursion B^{ex} of duration 1. He also shows that $\sigma\mathcal{G}_V^{(n)}/\sqrt{n} \to \mathcal{T}_{\mathrm{Ald}}$ in distribution as $n \to \infty$, when $\mathcal{G}_V^{(n)}$ is a Galton-Watson tree with finite variance non-arithmetic offspring distribution conditioned to have n vertices.

By Bertoin [11], the tree $\mathcal{T}_{B^{\mathrm{ex}}}$ in a Brownian excursion gives rise to a self-similar fragmentation at heights with binary dislocation measure $\nu_B(dx) = \sqrt{2/\pi}x^{-3/2}(1-x)^{-3/2}1_{[1/2,1)}(x)dx$. In the terminology of [27], this means that $\mathcal{T}_{B^{\mathrm{ex}}}$ is a self-similar CRT with dislocation measure ν_B. Haas and Miermont [28] reprove Aldous's Galton-Watson convergence result and refer to ν_B as the Brownian dislocation measure and to $\mathcal{T}_{B^{\mathrm{ex}}}$ as the Brownian continuum random tree, hence their choice is $\mathcal{T}_{\mathrm{HM}} := \mathcal{T}_{B^{\mathrm{ex}}} = \mathcal{T}_{\mathrm{Ald}}/2$.

Kortchemski [34] does not use the term "Brownian CRT" except when referring to the work of Rizzolo [47] and then without identifying constants. [34, Remark 4.6] specifies the height function that encodes his standard limiting tree in the case $\alpha = 2$ as $H = \sqrt{2}B^{\mathrm{ex}}$. In particular, his limiting CRT is $\mathcal{T}_{\mathrm{Kor}} := \sqrt{2}\mathcal{T}_{B^{\mathrm{ex}}} = \sqrt{2}\mathcal{T}_{\mathrm{HM}} = \mathcal{T}_{\mathrm{Ald}}/\sqrt{2}$. Kortchemski's motivation is to align with other stable laws with Laplace exponent r^α and hence with the other stable trees of index $\alpha \in (1, 2)$. In this, he follows Duquesne and Le Gall [19, p.105] and Duquesne [18, p.1002], but they only make qualitative remarks and refer to "proportional" when comparing with Brownian excursions, as this is not important for their results.

Bertoin and Miermont [13] and Dieuleveut [17] use $\mathcal{T}_{\mathrm{Br}} := \mathcal{T}_{\mathrm{Ald}} = 2\mathcal{T}_{B^{\mathrm{ex}}} = 2\mathcal{T}_{\mathrm{HM}} = \sqrt{2}\mathcal{T}_{\mathrm{Kor}}$.

Acknowledgements

This work was started during a research visit of the second author to Beijing Normal University. We would like to thank Beijing Normal University and EPSRC (EP/K02979/1) for providing support for this research visit. H. He is supported by NSFC (No. 11671041, 11531001, 11371061).

We would like to thank the referee for carefully reading our paper, spotting several inaccuracies and for making valuable suggestions to improve the presentation.

References

[1] R. Abraham and J.-F. Delmas. Record process on the continuum random tree. *Ann. Probab.*, 40(3):1167–1211, 2012.

[2] R. Abraham and J.-F. Delmas. Local limits of conditioned Galton-Watson trees: the condensation case. *Electron. J. Probab.*, 19(56):1–29, 2014.

[3] R. Abraham, J.-F. Delmas, and H. He. Pruning Galton-Watson trees and tree-valued Markov processes. *Ann. Inst. Henri Poincaré Probab. Statist.*, 48(3):688–705, 2012.

[4] R. Abraham, J.-F. Delmas, and P. Hoscheit. A note on the Gromov-Hausdorff-Prokhorov distance between (locally) compact metric measure spaces. *Electron. J. Probab*, 18(14):1–21, 2013.

[5] L. Addario-Berry, N. Broutin, and C. Holmgren. Cutting down trees with a Markov chainsaw. *Ann. Appl. Probab.*, 24(6):2297–2339, 2014.

[6] D. Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991.

[7] D. Aldous. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pages 23–70. Cambridge Univ. Press, Cambridge, 1991.

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

[9] D. Aldous and J. Pitman. Tree-valued Markov chains derived from Galton-Watson processes. *Ann. Inst. Henri Poincaré Probab. Statist.*, 34(5):637–686, 1998.

[10] D. Aldous and J. Pitman. Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent. *Probab. Theory Related Fields*, 118(4):455–482, 2000.

[11] J. Bertoin. Self-similar fragmentations. *Ann. Inst. Henri Poincaré Probab. Statist.*, 38(3):319–340, 2002.

[12] J. Bertoin. Fires on trees. *Ann. Inst. Henri Poincaré Probab. Statist.*, 48(4):909–921, 2012.

[13] J. Bertoin and G. Miermont. The cut-tree of large Galton-Watson trees and the Brownian CRT. *Ann. Appl. Probab.*, 23(4):1469–1493, 2013.

[14] N. Broutin and M. Wang. Cutting down **p**-trees and inhomogeneous continuum random trees. *Bernoulli*, 23(4A):2380–2433, 2017.

[15] M. Camarri and J. Pitman. Limit distributions and random trees derived from the birthday problem with unequal probabilities. *Electron. J. Probab.*, 5(2):1–18, 2000.

[16] L. c. de Raphélis. Scaling limit of multitype Galton-Watson trees with infinitely many types. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(1):200–225, 2017.

[17] D. Dieuleveut. The vertex-cut-tree of Galton-Watson trees converging to a stable tree. *Ann. Appl. Probab.*, 25(4):2215–2262, 2015.

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

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

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

[21] T. Duquesne and M. Winkel. Hereditary tree growth and Lévy forests. *arXiv:1211.2179*, 2012.

[22] T. Duquesne and M. Winkel. Hereditary tree growth and decompositions. *in preparation*, 2017.

[23] S. N. Evans, J. Pitman, and A. Winter. Rayleigh processes, real trees, and root growth with re-grafting. *Probab. Theory Related Fields*, 134(1):81–126, 2006.

[24] S. N. Evans and A. Winter. Subtree prune and regraft: a reversible real tree-valued Markov process. *Ann. Probab.*, 34(3):918–961, 2006.

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

[26] A. Greven, P. Pfaffelhuber, and A. Winter. Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Probab. Theory Related Fields*, 145(1-2):285–322, 2009.

[27] B. Haas and G. Miermont. The genealogy of self-similar fragmentations with negative index as a continuum random tree. *Electron. J. Probab.*, 9:57–97, 2004.

[28] B. Haas and G. Miermont. Scaling limits of Markov branching trees with applications to Galton-Watson and random unordered trees. *Ann. Probab.*, 40(6):2589–2666, 2012.

[29] H. He and M. Winkel. Invariance principles for pruning processes of Galton-Watson trees. *arXiv:1409.1014*, 2014.

[30] I. A. Ibragimov and Y. V. Linnik. *Independent and stationary sequences of random variables.* Wolters-Noordhoff Publishing, Groningen, 1971. With a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman.

[31] S. Janson. Random cutting and records in deterministic and random trees. *Random Structures Algorithms*, 29(2):139–179, 2006.

[32] O. Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[33] H. Kesten. Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. Henri Poincaré Probab. Statist.*, 22:425–487, 1987.

[34] I. Kortchemski. Invariance principles for Galton-Watson trees conditioned on the number of leaves. *Stochastic Process. Appl.*, 122(9):3126–3172, 2012.

[35] I. Kortchemski. Random stable laminations of the disk. *Ann. Probab.*, 42(2):725–759, 2014.

[36] A. Meir and J. W. Moon. Cutting down random trees. *J. Aust. Math. Soc.*, 11(03):313–324, 1970.

[37] G. Miermont. Invariance principles for spatial multitype Galton-Watson trees. *Ann. Inst. Henri Poincaré Probab. Statist.*, 44(6):1128–1161, 2008.

[38] G. Miermont. Tessellations of random maps of arbitrary genus. *Ann. Sci. Éc. Norm. Supér. (4)*, 42(5):725–781, 2009.

[39] J. Neveu. Erasing a branching tree. *Supplement to Adv. Appl. Prob.*, pages 101–108, 1986.

[40] J. Neveu and J. Pitman. Renewal property of the extrema and tree property of the excursion of a one-dimensional Brownian motion. In *Séminaire de Probabilités, XXIII*, volume 1372 of *Lecture Notes in Math.*, pages 239–247. Springer, Berlin, 1989.

[41] J. Neveu and J. W. Pitman. The branching process in a Brownian excursion. In *Séminaire de Probabilités, XXIII*, volume 1372 of *Lecture Notes in Math.*, pages 248–257. Springer, Berlin, 1989.

[42] A. Panholzer. Cutting down very simple trees. *Quaest. Math.*, 29(2):211–227, 2006.

[43] J. Pitman. Coalescent random forests. *J. Combin. Theory Ser. A*, 85(2):165–193, 1999.

[44] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002.

[45] J. Pitman and D. Rizzolo. Schröder's problems and scaling limits of random trees. *Trans. Amer. Math. Soc.*, 367(10):6943–6969, 2015.

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

[47] D. Rizzolo. Scaling limits of Markov branching trees and Galton-Watson trees conditioned on the number of vertices with out-degree in a given set. *Ann. Inst. Henri Poincaré Probab. Statist.*, 51(2):512–532, 2015.