The Annals of Applied Probability
2014, Vol. 24, No. 6, 2297–2339
DOI: 10.1214/13-AAP978
© Institute of Mathematical Statistics, 2014

CUTTING DOWN TREES WITH A MARKOV CHAINSAW

BY LOUIGI ADDARIO-BERRY, NICOLAS BROUTIN AND CECILIA HOLMGREN

McGill University, Inria Paris-Rocquencourt and Stockholm University

We provide simplified proofs for the asymptotic distribution of the number of cuts required to cut down a Galton–Watson tree with critical, finite-variance offspring distribution, conditioned to have total progeny n. Our proof is based on a coupling which yields a precise, nonasymptotic distributional result for the case of uniformly random rooted labeled trees (or, equivalently, Poisson Galton–Watson trees conditioned on their size). Our approach also provides a new, random reversible transformation between Brownian excursion and Brownian bridge.

1. Introduction. The subject of cutting down trees was introduced by Meir and Moon [39, 40]. One is given a rooted tree T which is pruned by random removal of edges. At each step, only the portion containing the root is retained (we refer to the portions not containing the root as the *pruned* portions), and the process continues until eventually the root has been isolated. The main parameter of interest is the random number of cuts necessary to isolate the root. The dual problem of isolating a leaf or a node with a specific label has been considered by Kuba and Panholzer [32, 33].

The procedure has been studied on different deterministic and random trees. Essentially two kinds of random models have been considered for the tree: *recursive trees* with typical inter-node distances of order $\log n$ [22, 25, 26, 41] and trees arising from critical, finite-variance branching processes conditioned to have size n, with typical distances of order $\sqrt{n}$ [23, 27, 28, 43, 44]. In this paper, we are interested in the latter family, and will refer to such trees as *conditioned trees* for short.

For conditioned trees emerging from a progeny distribution with variance $\sigma^2 \in (0, \infty)$, once divided by $\sigma\sqrt{n}$, the number of cuts required to isolate the root of a conditioned tree of size n converges in distribution to a Rayleigh random variable with density $xe^{-x^2/2}$ on $[0, \infty)$. In this form, under only a second moment assumption, this was proved by Janson [28]; below we discuss earlier, partial results in this direction. The fact that the Rayleigh distribution appears here with a $\sqrt{n}$ scaling in a setting involving conditioned trees struck us as deserving of explanation. The Rayleigh distribution also arises as the limiting distribution of the

Received May 2013; revised October 2013.
MSC2010 subject classifications. Primary 60C05, 60F17, 05C05; secondary 11Y16.
Key words and phrases. Cutting down, Galton–Watson tree, real tree, continuum random tree, Gromov–Hausdorff convergence.

length of a path between two uniformly random nodes in a conditioned tree, after appropriate rescaling.

In this paper we show that the existence of a Rayleigh limit in both cases is not fortuitous. We will prove using a coupling method that the number of cuts and the distance between two random vertices are asymptotically equal in distribution (modulo a constant factor σ^2). This approach yields as a by-product very simple proofs of the results concerning the distribution of the number of cuts obtained in [23, 27, 28, 43]; this is explained in Section 6.

At the heart of our approach is a coupling which yields the *exact* distribution of the number of cuts for every fixed n, for the special case of uniform Cayley trees (uniformly random labeled rooted trees). Given a rooted tree t and a sequence $S = (v_1, \ldots, v_k)$ of not necessarily distinct nodes of t, consider an edge-removal procedure defined as follows. The *planting of t at S*, denoted $t\langle S\rangle$, is obtained from t by creating a new node w_i for each $1 \le i \le k$, whose only neighbor is v_i. (If the v_i's are not all distinct, then the procedure results in multiple new vertices being connected to the same original vertex; if $v_i = v_j$ for $i \ne j$, then $w_i \ne w_j$ are both connected to $v_i = v_j$.) Let $W = \{w_1, \ldots, w_k\}$ be the set of new vertices (it may be more natural to take W as a sequence, since S is a sequence, but taking W as a set turns out to be notationally more convenient later). For a subgraph t' of $t\langle S\rangle$ and a vertex v, we write $C(v, t')$ for the connected component of t' containing v; let also $C(V, t')$ be the (minimal) set of connected components containing all the vertices in a set V.

Let $F^{(0)} = t\langle S\rangle$, and for $j \ge 0$, let $F^{(j+1)}$ be obtained from $F^{(j)}$ by removing a uniformly random edge from among all edges of $C(W, F^{(j)})$, if there are any such edges. The procedure stops at the first time j at which $C(W, F^{(j)})$ simply consists of the set of new vertices $\{w_1, \ldots, w_k\}$. We call this procedure *planted cutting of S in t*. We remark that Janson [27] already introduced the planted cutting procedure in the case $k = 1$. Note that if t is a rooted tree with root r, then $t\langle\{r\}\rangle$ contains only one node which is not a node of t, and in this case the cutting procedure is almost identical to that described in the first paragraph of the Introduction; see, however, the remark just before Theorem 3.1. Write $M = M(t, S)$ for the (random) total number of edges removed in the above procedure. We remark that for each $0 \le i \le M$, $F^{(i)}$ has $i + 1$ connected components, each of which is a tree.

THEOREM 1.1. *Fix $n \ge 1$ and $k \ge 1$, let T_n be a uniform Cayley tree on nodes $[n] = \{1, \ldots, n\}$, let $V_1, \ldots, V_k$ be independent, uniformly random nodes of T_n and write $S_k = (V_1, \ldots, V_k)$. Then $M(T_n, S_k) - k$ is distributed as the number of edges spanned by the root plus k independent, uniformly random nodes in a uniform Cayley tree of size n.*

For $k \ge 1$, let χ_k be a chi random variable with $2k$ degrees of freedom; the distribution of χ_k is given by

$$\mathbf{P}(\chi_k \le x) = \int_0^x \frac{2^{1-k} s^{2k-1} e^{-s^2/2}}{(k-1)!}\, ds.$$

COROLLARY 1.2. *For any fixed k, as $n \to \infty$, $M(T_n, S_k)/\sqrt{n}$ converges to χ_k in distribution.*

The fact that, after rescaling, the number of edges spanned by the root and k random vertices in T_n converges to χ_k in distribution is well known; see, for example, Aldous [7], Lemma 21. In Appendix A we sketch one possible proof of Corollary 1.2 and briefly discuss stronger forms of convergence.

REMARKS.

⋆ In the special case $k = 1$, Theorem 1.1 states that the number of edges required to isolate the planted node in a planted uniform Cayley tree of size n is identical in distribution to the number of vertices on the path between two uniformly random nodes in a uniform Cayley tree of size n. For the case $k = 1$, Chassaing and Marchand [19] have also announced a simple bijective proof of this result, based on linear probing hashing.

⋆ After the current results were announced [3], and independently of our results, Bertoin [13] used powerful recent results of Haas and Miermont [24] to establish the distributional convergence in Corollary 1.2. Bertoin's results give a different explicit interpretation of the number of cuts as the asymptotic distance between two nodes. Bertoin and Miermont [14] also study the genealogy of the fragmentation resulting from the removal of edges in a random order.

⋆ The original analyses by Meir and Moon [39] include asymptotics for the mean and variance of the number of cuts. In recent years, the subject of distributional asymptotics has been revisited by several researchers. Panholzer [43] and Fill, Kapur and Panholzer [23] have studied the somewhat simpler case where, the laws of the trees (as n varies), satisfy a certain consistency relation. More precisely, if μ_n is the law of the n-vertex tree, the consistency condition requires that after one step of the cutting procedure, conditional on the size k of the pruned fragment, the pruned fragment and the remaining tree are independent, with respective laws μ_k and μ_{n-k}. The class of random trees which satisfy this property includes uniform Cayley trees. For this class, they obtained the limiting distribution of various functionals of the number of cuts using the method of moments, and gave an analytic treatment of the recursive equation describing the cutting procedure. Janson [27, 28] used a representation of the number of cuts in terms of generalized records in a labeled tree to extend some of these results to all the family trees of critical branching processes with offspring distribution having a finite variance. His method is also based on the calculation of moments.

In the case $k = 1$, our coupling approach also allows us to describe the joint distribution of the sequence of pruned trees. In this paper, a *forest* is a sequence of rooted labeled trees $\mathbf{f} = (t_1, \ldots, t_j)$ with pairwise disjoint sets of labels. In the notation of Theorem 1.1 and of the paragraph which precedes it, write $M = M(T_n, S_1)$

and write $(T^{(1)}, \dots, T^{(M)})$ for the connected components of $F^{(M)}$, listed in the order they are created during the edge-removal procedure on $T_n\langle S_1\rangle$. Note that the edge-removal procedure stops at the first time that w_1 is isolated, so necessarily $T^{(M)}$ consists simply of the single vertex w_1. For each $1 \le i \le M$, $T^{(i)}$ is a tree, which we view as rooted at whichever node of $T^{(i)}$ was closest to w_1 in $T_n\langle S_1\rangle$; in particular, necessarily $T^{(M-1)}$ is rooted at V_1.

THEOREM 1.3. *The forest $(T^{(1)}, \dots, T^{(M-1)})$ is distributed as a uniformly random forest on $[n]$.*

The analysis which leads to Theorem 1.3 will also yield as a by-product the following result.

THEOREM 1.4. *Let $F^n = (T_1, \dots, T_\kappa)$ be a uniformly random forest on $[n]$. For each $i \in [\kappa - 1]$, add an edge from the root of T_i to a uniformly random node from among all nodes in $T_{i+1}, \dots, T_\kappa$. Call the resulting tree T, and view T as rooted at the root of T_κ. Then T is distributed as a uniform Cayley tree on $[n]$.*

It turns out that our coupling approach allows us to prove results about a natural "continuum version" of the random cutting procedure which takes place on the Brownian continuum random tree (CRT). Our main result about randomly cutting the CRT is Theorem 5.1, below. Although we work principally in the language of $\mathbb{R}$-trees, Theorem 5.1 can be viewed as a new, invertible random transformation between Brownian excursion and reflecting Brownian bridge. Though the precise statement requires a fair amount of set-up, if this set-up is taken for granted the result can be easily described. (For the reader for whom the following three paragraphs are opaque, all the below terminology will be re-introduced and formally defined later in the paper.)

Let $(\mathcal{T}, d)$ be a CRT with root ρ and mass measure μ, write $\mathrm{skel}(\mathcal{T})$ for its skeleton, and let $\mathcal{P}$ be a homogeneous Poisson point process on $\mathrm{skel}(\mathcal{T}) \times [0, \infty)$ with intensity measure $\ell \otimes dt$, where ℓ is the length measure on the skeleton. We think of the second coordinate as a time parameter. View each point (p, τ) of $\mathcal{P}$ as a potential cut, but only make a cut at p if no previous cut has fallen on the path from the root ρ to p. At each time $0 \le t < \infty$, this yields a forest of countably many rooted $\mathbb{R}$-trees; we write $\mathcal{T}_t$ for the component of this forest containing ρ. Run to time *infinity*, this process again yields a countable collection of rooted $\mathbb{R}$-trees, later called $(f_i, i \in I_\infty)$. Furthermore, each element f_i of the collection comes equipped with a time index τ_i (the time at which it was cut).

For $0 \le t < \infty$, let $L(t) = \int_0^t \mu(\mathcal{T}_s)\, ds$, and let $L(\infty) = \lim_{t\to\infty} L(t)$. It turns out that $L(\infty)$ is almost surely finite. Next, create a single compact $\mathbb{R}$-tree $(\mathcal{T}', d')$ from the collection $(f_i, i \in I_\infty)$ and the closed interval $[0, L(\infty)]$ by identifying the root of f_i with the point $L(\tau_i) \in [0, L(\infty)]$, for each $i \in I_\infty$, then taking the completion of the resulting object. Let μ' be the push-forward of μ under the transformation described above.

THEOREM 1.5. *The triples* $(\mathcal{T}', d', \mu')$ *and* $(\mathcal{T}, d, \mu)$ *have the same distribution. Furthermore,* $0 \in \mathcal{T}'$ *and* $L(\infty) \in \mathcal{T}'$ *are independent and both have law* μ'.

Using the standard encoding of the CRT by a Brownian excursion, we may take the triple $(\mathcal{T}, d, \mu)$, together with the point ρ, to be encoded by a Brownian excursion. Similarly, it is possible to view the triple $(\mathcal{T}', d', \mu')$, together with the points 0 and $L(\infty)$, as encoded by a reflecting Brownian bridge; see Section 10 of [11] (this is also closely related to the "forest floor" picture of [15]). From this perspective, the transformation from $(\mathcal{T}, \rho)$ to $(\mathcal{T}', 0, L_\infty)$ becomes a new, random transformation from Brownian excursion to reflecting Brownian bridge. When expressed in the language of Brownian excursions and bridges, this theorem and our "inverse transformation" result, Theorem 1.7, below, have intriguing similarities to results from Aldous and Pitman [11]; we briefly discuss this in Appendix B.

As an immediate consequence of the above development, we will obtain the following result. Let $\nu(t)$ be the mass of the tagged fragment in the Aldous–Pitman [11] fragmentation at time t. Then, $(\nu(t), t \geq 0)$ is distributed as $(\mu(\mathcal{T}_t), t \geq 0)$ and we have the following.

COROLLARY 1.6. *The random variable* $\int_0^\infty \nu(t)\,dt$ *has the standard Rayleigh distribution.*

A different proof of this fact appears in a recent preprint by Abraham and Delmas [2]. We also note that the identity in Theorem 1.5 has been generalized to the case of Lévy trees in [1].

We are also able to explicitly describe the inverse of the transformation of Theorem 1.5, and we now do so. Let $(\mathcal{T}, d, \mu)$ be a measured CRT, and let ρ, ρ' be independent random points in $\mathcal{T}$ with law μ. Let B be the set of branch points of $\mathcal{T}$ on the path from ρ to ρ'. For each $b \in B$ let $\mathcal{T}_b$ be the set of points $x \in \mathcal{T}$ for which the path from x to ρ contains a point $b' \in B$ with $d(\rho, b') > d(\rho, b)$. In words, $\mathcal{T}_b$ is the set of points in subtrees that "branch off the path from ρ to ρ' after b." Then, independently for each point $b \in B$, let y_b be a random element of $\mathcal{T}_b$, with law $\mu/\mu(\mathcal{T}_b)$. Delete all nonbranch points on the path between ρ and ρ'; then, for each $b \in B$, identify the points b and y_b. Write $(\mathcal{T}', d')$ for the resulting tree, and μ' for the push-forward of μ to $\mathcal{T}'$.

THEOREM 1.7. *The triples* $(\mathcal{T}, d, \mu)$ *and* $(\mathcal{T}', d', \mu')$ *have the same distribution. Furthermore, the point* $\rho' \in \mathcal{T}'$ *has law* μ'.

We remark that it is not *a priori* obvious the inverse transformation should a.s. yield a connected metric space, let alone what the distribution of the resulting space should be. Theorems 1.5 and 1.7 together appear as Theorem 5.1, below.

PLAN OF THE PAPER. In Section 2 we gather definitions and state our notational conventions. In Section 3 we prove all finite distributional identities related to the case $k = 1$, in particular proving Theorems 1.3 and 1.4, and in Section 4 we

prove Theorem 1.1. Our results on cutting the CRT, notably Theorem 5.1, appear in Section 5; finally, in Section 6 we explain how our results straightforwardly imply the distributional convergence results obtained in [27, 28, 43].

2. Notation and definitions. We note that the terminology introduced in Sections 2.2 and 2.3 is not used until Section 5, and the reader may wish to correspondingly postpone their reading of these sections.

2.1. *Finite trees and graphs.* Given any finite graph G, we write $v(G)$ for the set of vertices (or *nodes*) of G and $e(G)$ for the set of edges of G, and write $|G|$ for the size (number of vertices) of G. If we say that G *is a graph on* S, we mean that $v(G) = S$. Given a graph G and $w \in v(G)$, we write $C(w, G)$ for the connected component of G containing w. Given a graph G and $S' \subset e(G)$, we sometimes write $G \setminus S'$ for the graph $(v(G), e(G) \setminus S')$.

Practically all graphs in this paper will be rooted trees and be denoted t or T. When we write "tree" we mean a rooted tree unless we explicitly say otherwise.

Given a rooted labeled tree t, we write $r(t)$ for the root of t. For a vertex u of t write $t(u)$ for the subtree of t rooted at u, write $h_t(u)$ for the number of edges on the path from $r(t)$ to u, and write $a(u) = a(u, t)$ for the parent of u in t, with the convention that $a(r(t)) = r(t)$. At times we view the edges of t as oriented toward $r(t)$. In other words, if we state that (u, v) is an oriented edge of t, or write $(u, v) \in e(t)$, we mean that $\{u, v\} \in e(t)$ and $v = a(u)$. In this case we call u the *tail* of $\{u, v\}$ and v the *head* of $\{u, v\}$. It is also sometimes useful to view $r(t)$ as both the head and tail of a directed loop $(r(t), r(t))$; we will mention this again when it arises.

Given a set $S = \{v_1, \dots, v_k\}$ of nodes of t, we write $t[\![S]\!]$ or $t[\![v_1, \dots, v_k]\!]$ for the subtree of t obtained by taking the union of all shortest paths between elements of S, and call $t[\![S]\!]$ *the subtree of* t *spanned by* S; if $r(t) \in S$ then we consider $t[\![S]\!]$ as rooted at $r(t)$. Given a single node $v \in t$, we write $t^{r \leftrightarrow v}$ to denote the tree obtained from t by rerooting at v. As mentioned in the Introduction, in this paper an *ordered forest* is a sequence of rooted labeled trees $\mathbf{f} = (t_1, \dots, t_k)$ with pairwise disjoint sets of labels. If we write $\mathbf{f} = (t_1, \dots, t_k)$ *is an ordered forest on* S we mean that $v(t_1) \cup \cdots \cup v(t_k) = S$.

Given a finite set S, by a *uniform Cayley tree on* S we mean a rooted tree chosen uniformly at random from among all rooted trees t on S; there are $|S|^{|S|-1}$ such trees. Given a rooted or unrooted tree t, and an ordered sequence $S = (v_1, \dots, v_k)$ of elements of $v(t)$, we recall the definition of $t\langle S\rangle$ (the *planting of* t *at* S) from the Introduction: for each $1 \le i \le k$, create a new node w_i and add a single edge between w_i and v_i. Given a set $U \subset v(t\langle S\rangle)$, we write $|U|$ for the number of nodes of $U \setminus \{w_1, \dots, w_k\}$. In other words, the nodes $w_1, \dots, w_k$ *are not included* when performing node counts in $t\langle S\rangle$.

2.2. *Metric spaces and real trees.* In this paper all metric spaces are assumed to be separable. Given a metric space $\mathrm{X} = (X, d)$, and a real number $c > 0$, we

write $c\mathrm{X}$ for the metric space obtained by scaling all distances by c. In other words, if $x, y \in X$, then the distance between x and y in $c\mathrm{X}$ is $cd(x, y)$. We also write $\mathrm{diam}(\mathrm{X}) = \sup\{d(x, y) : x, y \in X\} \in [0, \infty]$.

Given a metric space (X, d) and $x, y \in X$, a *geodesic between x and y* is an isometry $f : [0, d(x, y)] \to X$ such that $f(0) = x$ and $f(d(x, y)) = y$. In this case we call the image $\mathrm{Im}(f)$ a *shortest path between x and y*.

A metric space $\mathrm{T} = (T, d)$ is an $\mathbb{R}$-tree if for all $x, y \in T$ the following two properties hold:

(1) There exists a unique geodesic between x and y. In other words, there exists a unique isometry $f : [0, d(x, y)] \to T$ such that $f(0) = x$ and $f(d(x, y)) = y$.

(2) If $g : [0, d(x, y)] \to T$ is a continuous injective map with $g(0) = x$ and $g(d(x, y)) = y$, then $f([0, d(x, y)]) = g([0, d(x, y)])$.

Given an $\mathbb{R}$-tree (T, d) and $a, b \in T$, we write $[\![a, b]\!]$ for the image of the unique geodesic from a to b, and write $]\!]a, b[\![= [\![a, b]\!] \setminus \{a, b\}$. The *skeleton* $\mathrm{skel}(\mathrm{T})$ is defined as

$$\bigcup_{a,b \in T}]\!]a, b[\![.$$

(We could equivalently define $\mathrm{skel}(\mathrm{T})$ as the set of points whose removal disconnects the space.) Since (T, d) is separable by assumption, this may be re-written as a countable union, and so there is a unique σ-finite measure ℓ on T with $\ell(]a, b[) = d(a, b)$ for all $a, b \in T$ and such that $\ell(T \setminus \mathrm{skel}(T)) = 0$. We refer to ℓ as the *length measure* on T.

For a set $S \subset T$, write $T[\![S]\!]$ for the subspace of T spanned by $\bigcup_{x,y \in S}]\!]x, y[\![$ and d_S for its distance (the restriction of d to $T[\![S]\!]$), and note that $(T[\![S]\!], d_S)$ is again a real tree.

2.3. *Types of convergence*. Before proceeding to definitions, we remark that not all the terminology of this subsection is yet fully standardized. The Gromov–Hausdorff distance is by now well-established. The name "Gromov–Hausdorff–Prokhorov distance" seems to have first appeared in [48], Chapter 27, where it had a slightly different meaning. The probabilistic aspects of the Gromov–Hausdorff–Prokhorov distance were substantially developed in [24, 42]. In particular, it is shown in [42], Section 6.1, that the below definition of d_{GHP} is equivalent to a definition based on the more standard Prokhorov distance between measures.

Gromov–Hausdorff distance. Let $\mathrm{X} = (X, d_X)$ and $\mathrm{Y} = (Y, d_Y)$ be compact metric spaces. The *Gromov–Hausdorff distance* $d_{\mathrm{GH}}(\mathrm{X}, \mathrm{Y})$ between X and Y is defined as follows. Let $\mathcal{S}$ be the set of all pairs (ϕ, ψ), where $\phi : X \to Z$ and $\psi : Y \to Z$ are isometric embeddings into some common metric space (Z, d_Z). Then

$$d_{\mathrm{GH}}(\mathrm{X}, \mathrm{Y}) = \inf_{(\phi,\psi) \in \mathcal{S}} d_{\mathrm{H}}\big(\phi(X), \psi(Y)\big),$$

where d_{H} denotes Hausdorff distance in the target metric space. It can be verified that d_{GH} is indeed a distance and that, writing $\mathcal{M}$ for the set of isometry-equivalence classes of compact metric spaces, $(\mathcal{M}, d_{\mathrm{GH}})$ is a complete separable metric space. We say that a sequence $\mathrm{X}_n = (X_n, d_n)$ of compact metric spaces converges to a compact metric space $\mathrm{X} = (X, d)$ if $d_{\mathrm{GH}}(\mathrm{X}_n, \mathrm{X}) \to 0$ as $n \to \infty$. It is then obvious that X is uniquely determined up to isometry. There are two alternate descriptions of the Gromov–Hausdorff distance that will be useful and which we now describe.

Next, for compact metric spaces (X, d_X) and (Y, d_Y), and a subset C of $X \times Y$, the *distortion* $\mathrm{dis}(C)$ is defined by

$$\mathrm{dis}(C) = \sup\bigl\{\bigl|d_X(x, x') - d_Y(y, y')\bigr| : (x, y) \in C, (x', y') \in C\bigr\}.$$

A *correspondence* C between X and Y is a Borel subset of $X \times Y$ such that for every $x \in X$, there exists $y \in Y$ with $(x, y) \in C$ and vice versa. Write $\mathscr{C}(X, Y)$ for the set of correspondences between X and Y. We then have

$$d_{\mathrm{GH}}(\mathrm{X}, \mathrm{Y}) = \tfrac{1}{2}\inf\bigl\{r : \exists C \in \mathscr{C}(X, Y) \text{ such that } \mathrm{dis}(C) < r\bigr\}$$

and there is a correspondence which achieves this infimum.

Given a correspondence C between X and Y and $\varepsilon \geq 0$ write

$$C_\varepsilon = \bigl\{(x, y) \in X \times Y : \exists (x', y') \in C, d_X(x, x') \leq \varepsilon, d_Y(y, y') \leq \varepsilon\bigr\}$$

and note that C_ε is again a correspondence, with distortion at most $\mathrm{dis}(C) + 2\varepsilon$. We call C_ε the ε *blow-up of* C.

Let $\mathrm{X} = (X, d_X, (x_1, \ldots, x_k))$ and $\mathrm{Y} = (Y, d_Y, (y_1, \ldots, y_k))$ be metric spaces, each with an ordered set of k distinguished points (we call such spaces *k-pointed metric spaces*). When $k = 1$, we simply refer to pointed (rather than 1-pointed) metric spaces, and write (X, d_X, x) rather than $(X, d_X, (x))$. The *k-pointed Gromov–Hausdorff distance* is defined as

$$\begin{aligned} &d^k_{\mathrm{GH}}(\mathrm{X}, \mathrm{Y}) \\ &\quad = \tfrac{1}{2}\inf\bigl\{r : \exists C \in \mathscr{C}(X, Y) \text{ such that } (x_i, y_i) \in C, 1 \leq i \leq k \text{ and } \mathrm{dis}(C) < r\bigr\}. \end{aligned}$$

It is straightforward to verify that for each k, the space $(\mathcal{M}^k, d^k_{\mathrm{GH}})$ of marked isometry-equivalence classes of k-pointed compact metric spaces, endowed with the distance d^k_{GH}, forms a complete separable metric space.

Couplings and Gromov–Hausdorff–Prokhorov distance. Let (X, d, μ) and (X', d', μ') be two measured metric spaces, and let ν be a Borel measure on $X \times X'$. We say ν is a (*defective*) *coupling* between μ and μ' if $p_*\nu \leq \mu$ and $p'_*\nu \leq \mu'$, where $p : X \times X' \to X$ and $p' : X \times X' \to X'$ are the canonical projections. The *defect* of ν is defined as

$$D(\nu) = \max\bigl((\mu - p_*\nu)(X), \bigl(\mu' - p'_*\nu\bigr)\bigl(X'\bigr)\bigr).$$

We let $\mathcal{C}(\mu, \mu')$ be the set of couplings between μ and μ', and for $\varepsilon \geq 0$ we write $\mathcal{C}_\varepsilon(\mu, \mu') = \{\nu \in \mathcal{C}(\mu, \mu') : D(\nu) \leq \varepsilon\}$

The *Prokhorov distance* between two finite positive Borel measures μ, μ' on the same space (X, d) is

$$d_{\mathrm{P}}^{\circ}(\mu, \mu') = \inf\{\varepsilon > 0 : \mu(F) \leq \mu'(F^\varepsilon) + \varepsilon \text{ and } \mu'(F) \leq \mu(F^\varepsilon) + \varepsilon \text{ for every closed } F \subseteq X\},$$

where $F^\varepsilon = \{x \in X : \exists x' \in F, d(x, x') < \varepsilon\}$.

There is another distance which generates the same topology and lends itself more naturally to combination with the correspondences introduced above. We define

$$d_{\mathrm{P}}(\mu, \mu') = \inf\{\varepsilon > 0 : \exists \nu \in \mathcal{C}_\varepsilon(\mu, \mu'), \nu(\{(x, x') \in X \times X : d(x, x') \geq \varepsilon\}) < \varepsilon\}.$$

By analogy with the latter, the *Gromov–Hausdorff–Prokhorov* (GHP) distance between $\mathrm{X} = (X, d, \mu)$ and $\mathrm{X}' = (X', d', \mu')$ is defined as

$$d_{\mathrm{GHP}}(\mathrm{X}, \mathrm{X}') = \inf\left\{\varepsilon > 0 : \begin{array}{c} \exists \nu \in \mathcal{C}_\varepsilon(\mu, \mu') \text{ and } R \in \mathscr{C}(X, X') \text{ such that} \\ \nu(R^c) < \varepsilon, \operatorname{dis}(R) < 2\varepsilon \end{array}\right\}.$$

We always have $d_{\mathrm{GHP}}(\mathrm{X}, \mathrm{X}') \geq d_{\mathrm{GH}}(\mathrm{X}, \mathrm{X}')$. Similarly to before, the collection $\widehat{\mathcal{M}}$ of measured isometry-equivalence classes of compact metric spaces, endowed with the distance d_{GHP}, forms a complete separable metric space [42], Section 6.

Given $\mathrm{X} = (X, d_X, \mu, (x_1, \ldots, x_k))$ and $\mathrm{X}' = (X', d', \mu', (x'_1, \ldots, x'_k))$, two k-pointed measured metric spaces, we define the *k-pointed Gromov–Hausdorff–Prokhorov distance* as

$$d^k_{\mathrm{GHP}}(\mathrm{X}, \mathrm{X}') = \inf\left\{\varepsilon > 0 : \begin{array}{c} \exists \nu \in \mathcal{C}_\varepsilon(\mu, \mu') \text{ and } R \in \mathscr{C}(X, X') \text{ such that} \\ \nu(R^c) < \varepsilon, \operatorname{dis}(R) < 2\varepsilon \text{ and } (x_i, x'_i) \in R, 1 \leq i \leq k \end{array}\right\}.$$

Once again, we may define an associated complete separable metric space $(\widehat{\mathcal{M}}^k, d^k_{\mathrm{GHP}})$.

3. Cutting down uniform Cayley trees.

3.1. *The Aldous–Broder dynamics.* Given a simple random walk $\{X_n\}_{n \in \mathbb{N}}$ on a finite connected graph G, we may generate a spanning tree T of G by including all edges (X_k, X_{k+1}) with the property that $X_{k+1} \notin \{X_i\}_{0 \leq i \leq k}$. The resulting tree T is in fact almost surely a uniformly random spanning tree of G. (More generally, if G comes equipped with edge weights $\{w_e : e \in e(G)\}$, then the probability the simple random walk on the weighted graph G generates a specific spanning tree t is proportional to $\prod_{e \in e(t)} w_e$.) This fact was independently discovered by Broder [17] and Aldous [10], and the above procedure is commonly called the Aldous–Broder algorithm.

By reversibility, the tree T generated by the Aldous–Broder algorithm may instead be viewed as generated by a simple random walk $\{X_n\}_{n\leq 0}$ on G, started from stationarity at time $-\infty$; see [36], pages 127–128. If instead of stopping the walk at time zero we instead stop at time $i \geq 0$, then the walk $\{X_n\}_{n\leq i}$ gives another tree, say T_i. What we call the *Aldous–Broder dynamics* is the (deterministic) rule by which the sequence $\{T_i, i \geq 0\}$ is obtained from T_0 and from the sequence $\{X_n, n \geq 0\}$. In the current section, we explain these dynamics. In the next section, we introduce a modification of the Aldous–Broder dynamics, and use it to exhibit the key coupling alluded to in Section 1.

Recall that given a rooted tree t and $x \in v(t)$, $t(x)$ denotes the subtree of t rooted at x. Fix an integer $n \geq 1$ and a tree t on $[n]$, and let $\mathbf{x} = (x_i)_{i\in\mathbb{N}}$ be a sequence of elements of $[n] = \{1, 2, \ldots, n\}$.

We then form a sequence of trees $\{T^m(t, \mathbf{x}) : m \in \mathbb{N}\}$. First, $T^0 = t$. Then, for $m \geq 0$, we proceed as follows:

- if $x_{m+1} = r(T^m)$, then $T^{m+1} = T^m$;
- if $x_{m+1} \neq r(T^m)$, then form T^{m+1} by removing the unique edge of T^m with tail x_{m+1}, then adding the edge (x_m, x_{m+1}), and finally rerooting at x_{m+1}.

In all cases, $r(T^m) = x_m$ for all $m \geq 1$. We refer to this procedure as *the Aldous–Broder dynamics on t and* $\mathbf{x}$. One can equivalently think of the root vertex as being both the head and tail of a directed loop; then one always removes the unique edge with tail x_{m+1} in T^m and adds the directed edge (x_m, x_{m+1}). Taking this perspective, let $R_{m+1} = R_{m+1}(t, \mathbf{x})$ be the subtree of T^m rooted at x_{m+1}, so $R_{m+1} = T^m(x_{m+1})$. Let $K_{m+1} = K_{m+1}(t, \mathbf{x})$ be the other component created when removing the edge with tail x_{m+1}, which is empty if $x_{m+1} = x_m$ and otherwise contains x_m. In all cases T^{m+1} is obtained from R_{m+1} and K_{m+1} by adding an edge from x_m to x_{m+1}; see Figure 1.

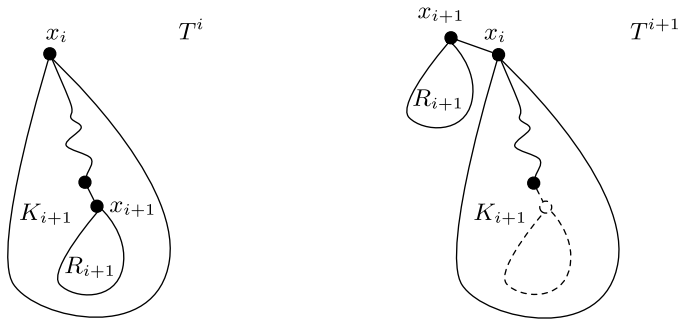

FIG. 1. *Two successive trees T^i and T^{i+1} built from the sequence construction: T^{i+1} is obtained from T^i by cutting above x_{i+1} and rearranging the parts in such a way that the subtree above the cut is appended as a child of the root x_{i+1} of the subtree R_{i+1} below the cut.*

3.2. *A modified Aldous–Broder dynamics.* Say that a sequence $\mathbf{x} \in [n]^{\mathbb{N}}$ is *good* if for each $k \in [n]$, $\sup\{i : x_i = k\} = \infty$. Fix a tree t on $[n]$ and a good sequence $\mathbf{x}$. We now describe a rule for removing a set of edges from t to obtain an ordered forest $\mathbf{F} = \mathbf{F}(t, \mathbf{x})$ on $[n]$. [Recall that an *ordered forest* is an ordered sequence $(t_1, \ldots, t_k)$ of rooted trees.]

In words, to build $\mathbf{F}(t, \mathbf{x})$ we start from the tree t and make the cuts that are dictated by the sequence $\mathbf{x}$, but *ignore* any such cuts that fall in a subtree we have already pruned at an earlier step. Since $\mathbf{x}$ is good, we will eventually prune the root $r(t)$ and so we will ignore all but finitely many of the cuts.

Formally, let $\sigma_0 = 0$ and, for $i \geq 1$, let

$$\sigma_i = \inf\left\{ m > \sigma_{i-1} : x_m \notin \bigcup_{j=1}^{i-1} t(x_{\sigma_j}) \right\}.$$

Then let $\kappa = \kappa(t, \mathbf{x}) = \inf\{i : x_{\sigma_i} = r(t)\}$. Note that we always have $\sigma_1 = 1$, that $\kappa < \infty$ since $\mathbf{x}$ is good, and that for all $j > \kappa$, $\sigma_j = \infty$. Recall that we write $t = (v(t), e(t))$, where $v(t)$ and $e(t)$ denote the vertex and edge set of t, respectively. After all the cuts in $\mathbf{x}$ have been made, we are left with a graph

$$f = \big(v(t), e(t) \setminus \{(x_{\sigma_i}, a(x_{\sigma_i})), 1 \leq i \leq \kappa\}\big).$$

For $1 \leq i \leq \kappa$, let $T_i = T_i(t, \mathbf{x}) = C(x_{\sigma_i}, f)$. Note that T_i is a tree, which we view as rooted at x_{σ_i}. We then take

$$\mathbf{F} = \mathbf{F}(t, \mathbf{x}) = (T_1, \ldots, T_\kappa).$$

Write $r_i = r_i(t, \mathbf{x})$ for the root of T_i and note that $r_\kappa = r(t)$. Finally, write $\widehat{T} = \widehat{T}(t, \mathbf{x})$ for the tree obtained from the forest $\mathbf{F}(t, \mathbf{x})$ by adding a directed edge from the root of T_{i+1} to the root of T_i, for each $i \in [\kappa - 1]$, and rooted at r_1 (as suggested by the orientation of the edges). These definitions are illustrated in Figure 2. We call this procedure the *modified Aldous–Broder dynamics on t and $\mathbf{x}$*.

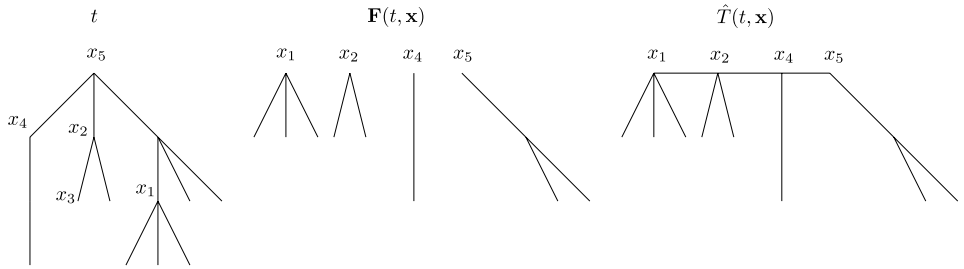

FIG. 2. *Left: a tree t, with node labels suppressed for readability; the first five nodes $x_1, \ldots, x_5$ of some good sequence are marked in the figure. Center: the forest $\mathbf{F}(t, \mathbf{x})$ built by applying the modified Aldous–Broder dynamics to t with any sequence $\mathbf{x}$ starting with $x_1, \ldots, x_5$. The trees are $T_1(t, \mathbf{x}), \ldots, T_4(t, \mathbf{x})$ are shown from left to right, and $r_1 = x_1, r_2 = x_2, r_3 = x_4, r_4 = x_5$. Right: the tree $\widehat{T}(t, \mathbf{x})$, which has root x_1.*

REMARK. The cutting procedure described above differs slightly from that used in much of the work on the subject. More precisely, it is more common to cut the tree by the removal of random edges rather than the selection of random vertices. However, there is a close correspondence between the vertex selection procedure and the edge selection procedure on a planted version of the same tree, which means results proved for one procedure have immediate analogues for the other. In particular, Janson ([27], Lemma 6.1) analyzed the difference between the two variants and showed that it is asymptotically negligible.

Now let $\mathcal{X} = (X_m)_{m\in\mathbb{N}}$ be a sequence of i.i.d. uniform $\{1,\dots,n\}$ random variables. It is easily seen that $\mathcal{X}$ is good with probability one. The following theorem is the key fact underlying almost all the results of the paper.

THEOREM 3.1. *Let T be a uniform Cayley tree on $[n]$. Then for any tree t on $[n]$ and any $w \in [n]$,*

$$\mathbf{P}\big(\widehat{T}(T,\mathcal{X}) = t \text{ and } r(T) = w\big) = n^{-n}.$$

Since there are n^{n-1} labeled rooted trees on $[n]$, there are n^n possible ways to choose a labeled rooted tree on $[n]$, plus an additional vertex of said tree. In other words, the theorem states that $\widehat{T}(T,\mathcal{X})$ is a uniform Cayley tree, and that $r(T)$ is uniform on $[n]$ and independent of $\widehat{T}(T,\mathcal{X})$ (the fact that $r(T)$ is uniform on $[n]$ is immediate from the fact that T is a uniform Cayley tree).

PROOF OF THEOREM 3.1. We proceed by induction on n, the case $n = 1$ being trivial. So we now suppose that $n > 1$. First, consider the case when $w = r(\widehat{T})$; we have $r(T) = r(\widehat{T})$ precisely if $X_1 = r(T)$ and in this case $\widehat{T} = T$. Thus, for any rooted tree t on $[n]$,

$$\mathbf{P}\big(\widehat{T} = t, r(T) = r(\widehat{T})\big) = \mathbf{P}\big(X_1 = r(T), T = t\big) = \frac{1}{n}\mathbf{P}(T = t) = \frac{1}{n^n},$$

since T is a uniform Cayley tree.

Next, fix a rooted tree t on $[n]$ and any $w \in [n]$, $w \neq r(t)$. Let $c = c(t,w)$ be the child of $r = r(t)$ for which the subtree of t rooted at c contains the node w. Let t_r and t_c be the subtrees containing r and c, respectively, when the edge (c,r) is removed from t. If we are to have $r(T) = w$ and $\widehat{T} = t$, then t_r must appear as a subtree of T, and we must additionally have $X_1 = r$. Since T is a uniform Cayley tree it follows that

$$\begin{aligned}
&\mathbf{P}\big(r(T) = w, \widehat{T} = t\big)\\
&\quad = \mathbf{P}\big(r(T) = w, \widehat{T} = t, t_r \text{ is a subtree of } T, X_1 = r\big) \qquad (1)\\
&\quad = \frac{(n - |t_r|)^{n-|t_r|}}{n^{n-1}} \cdot \frac{1}{n} \cdot \mathbf{P}\big(r(T) = w, \widehat{T} = t \,|\, t_r \text{ is a subtree of } T, X_1 = r\big).
\end{aligned}$$

Now let $\mathcal{X}' = (X'_i)_{i\in\mathbb{N}}$ be the subsequence of $\mathcal{X}$ consisting of the nodes of $K_1(T,\mathcal{X})$, the connected component of T containing the root after the edge above X_1 has been removed: for $i \in \mathbb{N}$, let

$$j_i = \min\bigl\{\ell : \bigl|\{X_1,\dots,X_\ell\} \cap v\bigl(K_1(T,\mathcal{X})\bigr)\bigr| = i\bigr\}$$

and set $X'_i = X_{j_i}$. Given that t_r is a subtree of T and $X_1 = r$, the entries of $\mathcal{X}'$ are independent, uniformly random elements of $v(t_c)$. Furthermore, under this conditioning we have that $\widehat{T}(T,\mathcal{X}) = t$ and $r(T) = w$ precisely if $\widehat{T}(K_1(T,\mathcal{X}),\mathcal{X}') = t_c$ and $r(K_1(T,\mathcal{X})) = w$. Since T is a uniform Cayley tree and $K_1(T,\mathcal{X})$ is obtained from T by removing the subtree rooted at X_1, it is immediate that conditional on its vertex set, $K_1(T,\mathcal{X})$ is again a uniform Cayley tree (and has less vertices than T). By induction, it follows that

$$\begin{aligned}
&\mathbf{P}\bigl(r(T) = w, \widehat{T} = t \mid t_r \text{ is a subtree of } T, X_1 = r\bigr)\\
&\quad = \mathbf{P}\bigl(\widehat{T}\bigl(K_1(T,\mathcal{X}),\mathcal{X}'\bigr) = t_c, r\bigl(K_1(T,\mathcal{X})\bigr) = w \mid t_r \text{ is a subtree of } T, X_1 = r\bigr)\\
&\quad = |t_c|^{-|t_c|}.
\end{aligned}$$

Since $|t_c| = n - |t_r|$, together with (1) this yields that $\mathbf{P}(\widehat{T}(T,\mathcal{X}) = t \text{ and } r(T) = w) = n^{-n}$, as required. □

We can transform the modified Aldous–Broder procedure for isolating the root into an edge-removal procedure, as follows. First, plant the tree to be cut at its root. Next, each time a node is selected for pruning, instead remove the parent edge incident to each selected vertex. The Aldous–Broder procedure then becomes the planted cutting procedure described in the Introduction, and $\kappa(T,\mathcal{X})$ is precisely the number of edges removed before the planted vertex is isolated. But $\kappa(T,\mathcal{X})$ is also the number of vertices on the path from $r(\widehat{T})$ to $r(T)$ in $\widehat{T}$. By Theorem 3.1, and from known results about the distance between the root and a uniformly random node in a uniform Cayley tree [4, 6, 7, 31, 41], the case $k = 1$ of Theorem 1.1 and of Corollary 1.2 follow immediately. By a well-known bijective correspondence between labeled rooted trees with a distinguished vertex and ordered labeled rooted forests (see, e.g., [11]), Theorem 1.3 also follows immediately (the forest consists of the sequence of trees obtained when removing the edges on the path between the root and the distinguished vertex).

REMARK. Aldous [5] studied the subtree rooted at a uniformly random node in a critical, finite variance Galton–Watson tree conditioned to have size n. In particular, he showed that such a subtree converges in distribution to an *unconditioned* critical Galton–Watson tree. It is then straightforward that, for fixed $k \geq 1$, the first k trees that are cut converge in distribution to a forest of k critical Galton–Watson trees. On the other hand, a critical Galton–Watson tree conditioned to be large converges locally (in the sense of local weak convergence of [9], i.e., inside balls

of arbitrary fixed radius k around the root) to an infinite path of nodes having a size-biased number of children (exactly one of which is again on the infinite path), where each nonpath node is the root of an unconditioned critical Galton–Watson tree. This is the incipient infinite cluster for critical, finite variance Galton–Watson trees [30]. Theorem 1.3 then appears as a strengthening of this picture, valid only for Poisson Galton–Watson trees, in which k is allowed to grow with n.

Recall that T is a uniform Cayley tree on $[n]$ and that $\mathcal{X} = (X_m)_{m\in\mathbb{N}}$ is a sequence of i.i.d. uniform elements of $[n]$. In the next proposition, which is essentially a time-reversed version of Theorem 3.1, we write $\mathbf{F}(T, \mathcal{X}) = \mathbf{F}$ for readability.

PROPOSITION 3.2. *For any forest* $\mathbf{f} = (t_1, \dots, t_k)$ *on* $[n]$, *given that* $\mathbf{F} = \mathbf{f}$, *independently for each* $i \in [k-1]$ *the parent* $a(r(t_i), T)$ *of* $r(t_i)$ *in* T *is a uniformly random element of* $\bigcup_{j=i+1}^{k} v(t_j)$.

PROOF. If $k = 1$, then there is nothing to prove. If $k > 1$, then fix any sequence $\mathbf{v} = (v_1, \dots, v_{k-1})$ with $v_i \in \bigcup_{j=i+1}^{k} v(t_j)$ for each $i \in [k-1]$. Write $t(\mathbf{f}, \mathbf{v})$ for the tree formed from $\mathbf{f}$ by adding an edge from $r(t_i)$ to v_i for each $i \in [k-1]$. In order that $\mathbf{F} = \mathbf{f}$ and that, for each $i \in [k-1]$, $a(r(t_i), T) = v_i$, it is necessary and sufficient that $T = t(\mathbf{f}, \mathbf{v})$ and that for each $i \in [k]$, $X_{\sigma_i} = r(t_i)$. The probability that $T = t(\mathbf{f}, \mathbf{v})$ is $n^{-(n-1)}$. Furthermore, since $(X_m)_{m\in\mathbb{N}}$ are i.i.d. elements of $[n]$,

$$\mathbf{P}\big(X_{\sigma_i} = r(t_i), 1 \le i \le k | T = t(\mathbf{f}, \mathbf{v})\big) = \prod_{i\in[k]} \frac{1}{|\bigcup_{j\ge i} v(t_j)|}.$$

It follows that

$$\mathbf{P}\big(\mathbf{F} = \mathbf{f} \text{ and } a\big(r(t_i), T\big) = v_i, 1 \le i < k\big) = \frac{1}{n^{n-1}} \cdot \prod_{i\in[k]} \frac{1}{|\bigcup_{j\ge i} v(t_j)|},$$

which proves the proposition since this expression does not depend on v_1, $\dots$, v_{k-1}. □

Theorem 1.4 is an immediate consequence of Proposition 3.2.

4. Isolating more than one vertex. In this section we describe how to generalize the arguments of Section 3.2 to obtain results on isolating sets of vertices of size greater than one. Recall that when performing the planted cutting of S in t, described in Section 1, we wrote $W = \{w_1, \dots, w_k\}$ for the set of new vertices, and wrote $M = M(t, S)$ for the (random) total number of edges removed. In order to study the random variable M, it turns out to be necessary to study a transformation of the planted cutting procedure. The modified procedure is defined via a canonical re-ordering of the sequence of removed edges. As such, it may be coupled with

the original procedure so that the final set of removed edges is the same in both. In particular, both procedures isolate the vertices of W, and the total number of cuts has the same distribution in both.

In the following, for an edge e and a connected component C, we write $e \in C$ to mean that *both* endpoints of e lie in C, or equivalently (since the connected components are trees) that the removal of e leaves C disconnected. Also, recall from Section 2 that given a set A of edges, we write $t \setminus A$ for the graph $(v(t), e(t) \setminus A)$.

Now fix a sequence $\mathbf{e} = (e_1, \ldots, e_m)$ of distinct edges of t. We say that $\mathbf{e}$ is a *possible cutting sequence* (for S in t) if:

- each edge $\{v_i, w_i\}$, $1 \le i \le k$ appears in $\mathbf{e}$ ($\mathbf{e}$ really isolates $w_1, \ldots, w_k$), and
- for each $1 \le j \le m$, one has $e_j \in C(W, t \setminus \{e_1, \ldots, e_{j-1}\})$, that is, each e_j indeed produces a cut.

We now describe a canonical re-ordering of $\mathbf{e}$, which we denote $\mathbf{e}^*$; this re-ordering operation gives rise to the modified cutting procedure. In $\mathbf{e}^*$, we first list all edges whose removal decreases the size of the component containing w_1 (in increasing order of arrival time). We then list all *remaining* edges whose removal decreases the size of the component containing w_2, again in increasing order of arrival time, and so on. (This is somewhat related to a size-biased reordering of an exchangeable random structure; see [45], Chapter 1. The next three paragraphs formalize this description.)

For $1 \le i \le k$, write

$$U_i = U_i(\mathbf{e}) = \big\{ j : e_j \in C\big(w_i, t \setminus \{e_1, \ldots, e_{j-1}\}\big) \big\}$$

and let $U_i^* = U_i \setminus (\bigcup_{j=1}^{i-1} U_j)$. In words, U_i^* is the set of times j at which the component containing w_i does not contain any of $w_1, \ldots, w_{i-1}$, and such that removing the current edge e_j decreases the size of this component.

Next, let $m(i) = m(i, t, \mathbf{e}) = |U_i|$, write $Z_i = Z_i(\mathbf{e}) = (z_{i,1}, \ldots, z_{i,m(i)})$ for the sequence obtained by listing the elements of U_i in increasing order, and define Z_i^* accordingly. Notice that once w_i is in a component distinct from $w_1, \ldots, w_{i-1}$, it can never rejoin such a component, and so writing $s(i) = s(i, t, \mathbf{e}) = \min\{\ell : z_{i,\ell} \in U_i^*\}$, we must have

$$Z_i^* = (z_{i,s(i)}, z_{i,s(i)+1}, \ldots, z_{i,m(i)}).$$

We then write

$$\begin{aligned}\mathbf{e}^* &= (e_{z_{1,s(1)}}, \ldots, e_{z_{1,m(1)}}, e_{z_{2,s(2)}}, \ldots, e_{z_{2,m(2)}}, \ldots, e_{z_{k,s(k)}}, \ldots, e_{z_{k,m(k)}}) \\ &= \big(e_1^*, \ldots, e_m^*\big),\end{aligned}$$

the latter equality constituting the definition of $e_1^*, \ldots, e_m^*$. For $1 \le i \le k$, let $a_i(t, \mathbf{e}^*) = 1 + \sum_{\ell=1}^{i-1}(m(\ell) - s(\ell) + 1)$ let $b_i(t, \mathbf{e}^*) = \sum_{\ell=1}^{i}(m(\ell) - s(\ell) + 1)$, and set

$$\mathbf{e}_i^* = \big(e_j^*, a_i \le j \le b_i\big) = \big(e_{z_{i,j}}, s(i) \le j \le m(i)\big).$$

We remark that necessarily $e_{z_{i,m(i)}} = \{w_i, v_i\}$, and so in particular the sequence $\mathbf{e}_i^*$ is nonempty for each $1 \le i \le k$.

Now write $\mathbf{E} = \mathbf{E}(t, S) = (E_1, \ldots, E_M)$ for the random sequence of removed edges (in the original planted cutting procedure), write $\mathbf{E}^* = \mathbf{E}^*(t, S) = (E_1^*, \ldots, E_M^*)$ for the rearrangement of $\mathbf{E}$ described above, and likewise define $\mathbf{E}_i^*$, for $1 \le i \le k$, as above.

It is easily seen that if $\mathbf{e}$ is not a possible cutting sequence, then $\mathbf{P}(\mathbf{E}(t, S) = \mathbf{e}) = 0$, and if $\mathbf{e}$ *is* a possible cutting sequence, then

$$\mathbf{P}\big(\mathbf{E}(t, S) = \mathbf{e}\big) = \prod_{j=1}^{m} \frac{1}{|e(C(W, t \setminus \{e_1, \ldots, e_{j-1}\}))|}. \tag{2}$$

For our purposes, it is in fact the expression for $\mathbf{P}(\mathbf{E}^*(t, S) = \mathbf{e}^*)$ given in the following lemma that will be more useful. Fix any sequence $\mathbf{f} = (f_1, \ldots, f_m)$ of edges of $t\langle S\rangle$. If there exists a possible cutting sequence $\mathbf{e} = (e_1, \ldots, e_m)$ for $S = (v_1, \ldots, v_k)$ in t such that $\mathbf{e}^* = \mathbf{f}$, then we say that $\mathbf{f}$ is *valid* (*for t and S*).

LEMMA 4.1. *Given any sequence $\mathbf{f} = (f_1, \ldots, f_m)$ that is valid for t and S, we have*

$$\mathbf{P}\big(\mathbf{E}^*(t, S) = \mathbf{f}\big) = \prod_{i=1}^{k} \prod_{j=a_i(t,\mathbf{f})}^{b_i(t,\mathbf{f})} \frac{1}{|e(C(w_i, t \setminus \{f_1, \ldots, f_{j-1}\}))|}.$$

PROOF. We prove the lemma by induction on $|e(t\langle S\rangle)|$. Fix $\mathbf{f}$ as in the statement of the lemma, write

$$\mathcal{E}(\mathbf{f}) = \mathcal{E}(\mathbf{f}, t, S) = \big\{\mathbf{e} : \mathbf{e} \text{ is a possible cutting sequence for } S \text{ in } t \text{ and } \mathbf{e}^* = \mathbf{f}\big\}$$

and note that $\mathbf{f} \in \mathcal{E}(\mathbf{f})$. For any $\mathbf{e} = (e_1, \ldots, e_m) \in \mathcal{E}(\mathbf{f})$ we necessarily have $e_1 = f_1$, and so

$$\mathbf{P}\big(\mathbf{E}_1^*(t, S) = f_1\big) = \mathbf{P}\big(\mathbf{E}_1(t, S) = f_1\big) = \frac{1}{|e(t\langle S\rangle)|}.$$

If $e_1 = \{v_1, w_1\}$, then writing $S' = (v_2, \ldots, v_k)$, we have

$$\begin{aligned}\mathbf{P}\big(\mathbf{E}^* = \mathbf{f} | \mathbf{E}_1^* = f_1\big) &= \mathbf{P}\big(\mathbf{E}^* = \mathbf{f} | \mathbf{E}_1 = f_1\big) \\ &= \mathbf{P}\big(\mathbf{E}^*\big(t, S'\big) = (f_2, \ldots, f_m)\big)\end{aligned}$$

and the result follows by induction since $t\langle S'\rangle$ has fewer edges than $t\langle S\rangle$.

If $e_1 \ne \{v_1, w_1\}$, then write $t_1 = C(w_1, t\langle S\rangle \setminus \{e_1\})$, and write t_2 for the other component of $t\langle S\rangle \setminus \{e_1\}$; each of these trees has fewer edges than $t\langle S\rangle$. Write $S_1 = (x_1, \ldots, x_{k_1})$ and $S_2 = (y_1, \ldots, y_{k_2})$ for the nodes of S within t_1 and t_2, respectively, listed in the same order as in S.

Now fix any possible cutting sequence $\mathbf{e} = (e_1, \dots, e_m)$ with $e_1 = f_1$. Write $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$ for those edges in the sequence $(e_2, \dots, e_m)$ falling in t_1 and t_2, respectively, and listed in the same order as in $\mathbf{e}$. Then it is clear that, conditionally on $\mathbf{E}_1 = f_1$, the sequences $\mathbf{E}(t_1, S_1)$ and $\mathbf{E}(t_2, S_2)$ have the distribution of the planted cutting procedure on $t_1\langle S_1\rangle$ and $t_2\langle S_2\rangle$, respectively, and are independent. In other words,

$$\begin{aligned}&\mathbf{P}\big(\mathbf{E}(t_1, S_1) = \mathbf{e}^{(1)}, \mathbf{E}(t_2, S_2) = \mathbf{e}^{(2)} | \mathbf{E}_1 = f_1\big)\\ &\qquad = \mathbf{P}\big(\mathbf{E}(t_1, S_1) = \mathbf{e}^{(1)}\big) \cdot \mathbf{P}\big(\mathbf{E}(t_2, S_2) = \mathbf{e}^{(2)}\big).\end{aligned}$$

Furthermore, if $\mathbf{e} \in \mathcal{E}(\mathbf{f})$, then $e_1 = f_1$, and $\mathbf{e} \in \mathcal{E}(\mathbf{f})$ if and only if $\mathbf{e}^{(1)} \in \mathcal{E}(\mathbf{f}^{(1)}, t_1, S_1)$ and $\mathbf{e}^{(2)} \in \mathcal{E}(\mathbf{f}^{(2)}, t_2, S_2)$. [Note: this does not mean that the map from $\mathbf{e}$ to $(\mathbf{e}^{(1)}, \mathbf{e}^{(2)})$ is bijective! In fact, for a given pair $\mathbf{e}^{(1)} \in \mathcal{E}(\mathbf{f}^{(1)}, t_1, S_1)$ and $\mathbf{e}^{(2)} \in \mathcal{E}(\mathbf{f}^{(2)}, t_2, S_2)$, the number of pre-images in $\mathcal{E}(\mathbf{f})$ is precisely $\binom{m-1}{m_1}$, where m_1 is the length of $\mathbf{f}^{(1)}$.] Also, $\mathbf{f}^{(1)}$ (resp., $\mathbf{f}^{(2)}$) is valid for t_1 and S_1 (resp., for t_2 and S_2). It follows that

$$\begin{aligned}&\mathbf{P}\big(\mathbf{E}^* = \mathbf{f} | \mathbf{E}_1 = f_1\big)\\ &\quad = \sum_{\mathbf{e}\in\mathcal{E}} \mathbf{P}(\mathbf{E} = \mathbf{e} | \mathbf{E}_1 = f_1)\\ &\quad = \sum_{\mathbf{e}^{(1)}\in\mathcal{E}(\mathbf{f}^{(1)},t_1,S_1)} \sum_{\mathbf{e}^{(2)}\in\mathcal{E}(\mathbf{f}^{(2)},t_2,S_2)} \mathbf{P}\big(\mathbf{E}(t_1, S_1) = \mathbf{e}^{(1)}, \mathbf{E}(t_2, S_2) = \mathbf{e}^{(2)} | \mathbf{E}_1 = f_1\big)\\ &\quad = \sum_{\mathbf{e}^{(1)}\in\mathcal{E}(\mathbf{f}^{(1)},t_1,S_1)} \sum_{\mathbf{e}^{(2)}\in\mathcal{E}(\mathbf{f}^{(2)},t_2,S_2)} \mathbf{P}\big(\mathbf{E}(t_1, S_1) = \mathbf{e}^{(1)}\big) \cdot \mathbf{P}\big(\mathbf{E}(t_2, S_2) = \mathbf{e}^{(2)}\big)\\ &\quad = \mathbf{P}\big(\mathbf{E}^*(t_1, S_1) = \mathbf{f}^{(1)}\big) \cdot \mathbf{P}\big(\mathbf{E}^*(t_2, S_2) = \mathbf{f}^{(2)}\big)\end{aligned}$$

from which the result again follows by induction. □

The formula in the preceding lemma implies that removing edges in the order given by $\mathbf{E}^*$ corresponds to the following procedure. For each $1 \le i \le k$, in that order, remove edges of t uniformly at random from among those whose removal reduces the size of the component currently containing w_i, until w_i is isolated. We call this the *ordered cutting* of S in t.

For $1 \le i \le k$, write M_i for the random time at which w_i is isolated in the ordered cutting procedure

$$\begin{aligned}M_i = M_i(t, S) &= \max\big\{j : E_j^* \in C\big(w_i, t \setminus \{E_1^*, \dots, E_{j-1}^*\}\big)\big\}\\ &= \min\big\{j : \big|C\big(w_i, t \setminus \{E_1^*, \dots, E_j^*\}\big)\big| = 0\big\}\end{aligned}$$

(recall that the counting does not include planted vertices), and note that $M_1 < M_2 < \cdots < M_k \stackrel{d}{=} M$.

Now, let T be a uniform Cayley tree on $[n]$, let $V_1, \dots, V_k$ be independent, uniformly random elements of $[n]$, and let $S_k = (V_1, \dots, V_k)$. Then write $M_k = M(T, S_k)$ for the number of edges removed during the ordered cutting of S_k in t.

THEOREM 4.2. *$M_k - k$ is distributed as the number of edges spanned by the root plus k independent, uniformly random nodes in a uniform Cayley tree of size n.*

Theorem 1.1 follows immediately from Theorem 4.2 and the relationship between planted cutting and ordered cutting described above. To prove Theorem 4.2, we will exhibit a coupling which generalizes that of Section 3.2 and which we now explain. The coupling hinges upon the following, easy lemma, whose proof is omitted. Recall that if S is a set of nodes in a tree t, then $t[\![S]\!]$ is the subtree of S spanned by S.

LEMMA 4.3. *Fix $i \geq 1$. Let T be a uniform Cayley tree on $[n]$, let $V_1, \dots, V_{i+1}$ be independent, uniformly random elements of $[n]$, and let $S = \{r(T), V_1, \dots, V_i\}$. Let U be the most recent ancestor of V_{i+1} in T which is an element of $v(T[\![S]\!])$. Let R be the set of nodes whose path to V_{i+1} uses no edges of $T[\![S]\!]$ (such paths may pass through U). Let $T^+ = T[\![R]\!]$, let $T^- = T[\![([n] \setminus R) \cup \{U\}]\!]$ and root T^+ and T^- at U and at $r(T)$, respectively. Then conditionally on R, T^+ is a uniformly random labeled rooted tree on R, independent of T^- and of $V_1, \dots, V_i$, and V_{i+1} is a uniformly random element of R independent of T^+, T^- and $V_1, \dots, V_i$.*

The definitions in Lemma 4.3 are depicted in Figure 3.

PROOF OF THEOREM 4.2. We provide a coupling between the random sequence of edges $\mathbf{E}^*(T, (V_1, \dots, V_k))$ and a sequence $T_1, \dots, T_k$ of trees on $[n]$,

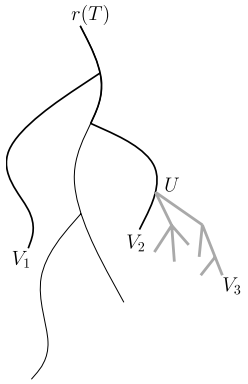

FIG. 3. *An example of the definitions of Lemma 4.3 in the case $i = 2$ [so $S = (r(T), V_1, V_2)$]. The subtree $T[\![S]\!]$ is in thicker black lines. The tree T^+ is in thick grey lines, and the tree T^- consists of all black lines (thick and thin).*

such that the following properties hold. First, for any rooted tree t on $[n]$, and any $v_1, \dots, v_i$ elements of $[n]$ (not necessarily distinct),

$$\mathbf{P}(T_i = t, V_1 = v_1, \dots, V_i = v_i) = n^{-(n-1+i)}. \tag{3}$$

Second, for each $1 \le i \le k$, the following holds:

($\star$) the forest obtained from $T\langle(V_1, \dots, V_i)\rangle$ by first removing all edges of $\{E_1^*, \dots, E_{M_i}^*\}$, then deleting $w_1, \dots, w_i$, is identical to the forest obtained from T_i by removing all edges of its subtree $T_i[\![r(T_i), V_1, \dots, V_i]\!]$.

Equation (3) says that T_i is a uniform Cayley tree and $V_1, \dots, V_i$ are independent of T_i, and ($\star$) then implies in particular (by considering only the case $i = k$) that $M_k - k$ is equal to the number of edges of $T_k[\![r(T_k), V_1, \dots, V_k]\!]$. This clearly implies the theorem, and so it remains to explain how we construct such a sequence.

Fix a sequence $\mathcal{X} = (X_i)_{i \ge 1}$ of i.i.d. uniform elements of $[n]$. Let T_1 be the tree built by running the modified Aldous–Broder dynamics on $T^{r \leftrightarrow V_1}$ (recall that this is the tree T, rerooted at node V_1) with the sequence $(X_i)_{i \ge 1}$. [In the notation of Section 3.2, $T_1 = \widehat{T}(T^{r \leftrightarrow V_1}, \mathcal{X})$.] By Theorem 3.1, for any tree t on $[n]$ and any $v \in [n]$, $\mathbf{P}(T_1 = t, V_1 = v) = n^{-n}$, so (3) holds in the case $i = 1$. Temporarily write $u_1, \dots, u_\ell$ for the nodes on the path in T_1 from $r(T_1)$ to V_1, in the same order they appear on that path. We must then have $u_\ell = V_1$, and $M_1 = \ell$. For $1 \le j \le \ell - 1$, let $E_j^* = \{u_j, a(u_j, T^{r \leftrightarrow V_1})\}$, and note that this is also an edge of T since T and $T^{r \leftrightarrow V_1}$ have the same edge set. Then let $E_{M_1}^* = \{u_\ell, w_1\} = \{V_1, w_1\}$. (An example of this construction is shown in Figure 4.) By construction, it is immediate that ($\star$) then holds in the case $i = 1$.

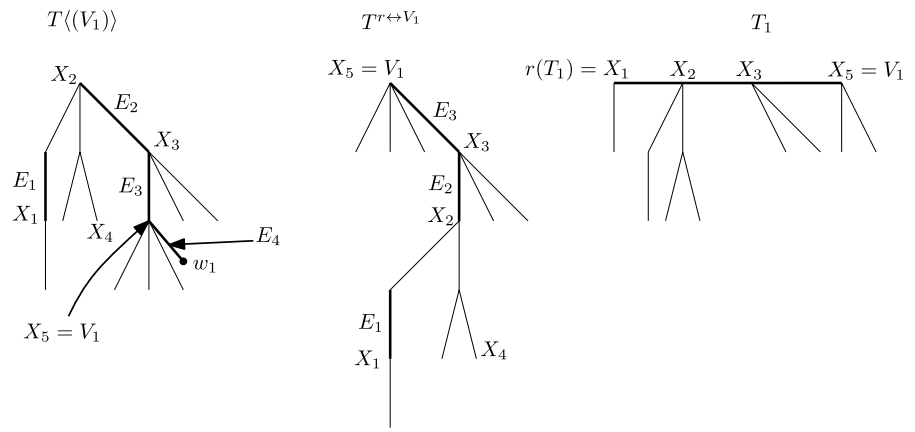

FIG. 4. *Left*: *the tree* $T\langle(V_1)\rangle$. *Center*: *the tree* $T^{r \leftrightarrow V_1}$, *planted at* V_1. *Right*: *the tree* T_1. *The vertex and edge labels provide an example of the construction in the proof of Theorem* 4.2, *in the case* $k = 1$. *For each of the three trees, the forest obtained by removing the bold edges* [*and, for* $T\langle(V_1)\rangle$, *then throwing away the vertex* w_1] *is identical.*

Now fix $1 \le j < k$, suppose that $T_1, \dots, T_j$ and $(E_1^*, \dots, E_{M_j}^*)$ are already defined and that (3) and ($\star$) both hold for each $1 \le i \le j$. As defined, V_{j+1} is independent of T_i and of $(E_1^*, \dots, E_{M_j}^*)$, and so for any tree t on $[n]$ and any sequence $u_1, \dots, u_{j+1}$ of elements of $[n]$, we have

$$\mathbf{P}(T_j = t, V_1 = u_1, \dots, V_{j+1} = u_{j+1}) = n^{-(n-1+i+1)}. \tag{4}$$

Let U be the most recent ancestor of V_{j+1} that lies in $T_j[\![r(T_i), V_1, \dots, V_j]\!]$, and define T^+ and T^- as in Lemma 4.3.

Now let $\mathcal{X}'$ be a random sequence such that conditionally on $v(T^+)$, the entries of $\mathcal{X}'$ are independent uniform elements of $v(T^+)$, independent of all preceding randomness. Then apply the modified Aldous–Broder dynamics to $T^{+,r\leftrightarrow V_{j+1}}$, and call the result T^*. By Theorem 3.1, given that $v(T^+) = S$, (T^+, V_{j+1}) and (T^*, V_{j+1}) are identically distributed. As above, let $u_1, \dots, u_\ell$ be the nodes on the path from $r(T^*)$ to V_{j+1}, and note that we must have $M_{j+1} = M_j + \ell$. For $1 \le i \le \ell - 1$ let $E_{M_i+j}^* = \{u_i, a(u_i, T^{+,r\leftrightarrow V_{j+1}})\}$, and let $E_{M_{j+1}}^* = \{V_{j+1}, w_{j+1}\}$. In words, we have applied exactly the same construction as in the case $i = 1$, but to the subtree T^+ of T (which contains V_{j+1}). Figures 3 and 4 may be useful as visual aids to these definitions.

Write P for the parent of U in T_j, and $C_1, \dots, C_\ell$ for the children of U in $T_j \setminus T^+$ (any such child is an ancestor of at least one of $V_1, \dots, V_j$). Now let T_{j+1} be the tree obtained from T_j by replacing T^+ by T^*. In other words, T_{j+1} is built from T_j by, first, removing all edges of T_j that are incident to nodes of T^+ and then, second, adding all edges of T^* as well as edges from the root of T^* to P and to each of $C_1, \dots, C_\ell$. With this construction, ($\star$) now holds for all $1 \le i \le j + 1$.

Finally, write $R = v(T^+)$. By Lemma 4.3 and by Theorem 3.1, (T^+, V_{j+1}) and (T^*, V_{j+1}) are identically distributed conditionally on their vertex sets, and both are independent of T^- and of $V_1, \dots, V_j$. It follows that (4) still holds with T_j replaced by T_{j+1}, and this verifies (3) and completes the proof by induction. □

5. A novel transformation of the Brownian CRT. In [28], Janson suggested that it should be possible to define a version of the cutting procedure directly on $\mathcal{T}$. In this section, we provide such a construction. This construction yields straightforward, "conceptual" proofs of some of the main results of [28], and also provides a novel, reversible transformation from $\mathcal{T}$ to another, doubly-rooted Brownian CRT. (We remark in passing that the results of this section can also be straightforwardly used to prove the first convergence result from Theorem 1.10 of [28].) Using the by now well-known coding of the Brownian CRT by a standard Brownian excursion, this transformation can be viewed as a new, invertible random transformation between Brownian excursion and Brownian Bridge.

We now describe the details of the construction, using the language of $\mathbb{R}$-trees. For the interested reader, we describe the corresponding transformation from Brownian excursion to reflecting Brownian bridge in Appendix B.

We begin with a quick, high-level description of the transformation. An initial compact real tree $\mathcal{T}$ distributed as the Brownian CRT will be cut by points falling on its skeleton. When a point arrives, the current tree is separated into two connected components; the one containing the root will suffer further cuts at later times, while the other one—the pruned tree—will no longer be cut. As in the discrete transformation of Section 3.2, the cut trees are rearranged by attaching their roots to a "backbone" so as to form a new real tree. We now describe the continuous transformation by first building the backbone that will eventually connect the roots of the pruned subtrees, and then specifying where these subtrees should be grafted along the backbone.

5.1. *The details of the transformation.* Let $\mathcal{P}$ be a Poisson process on $\mathrm{skel}(\mathcal{T}) \times [0, \infty)$ with intensity measure $\ell \otimes dt$, and for each $t \geq 0$, let

$$\mathcal{P}_t^\circ = \{x \in \mathcal{T} : \exists s, 0 \leq s \leq t, (x, s) \in \mathcal{P}\}.$$

In [8], Aldous and Pitman used the point process $\mathcal{P}$ to construct (what is now called) a self-similar fragmentation process on $\mathcal{T}$ [12]. For each $t \geq 0$, let $\mathcal{F}_t^\circ = \mathcal{T} \setminus \mathcal{P}_t^\circ$. In particular, two points $u, v \in \mathcal{T} \setminus \mathcal{P}_t^\circ$ are in the same component of $\mathcal{F}_t^\circ$ precisely if, in $\mathcal{T}$, the path $[\![u, v]\!]$ contains no element of $\mathcal{P}_t^\circ$. Aldous and Pitman [8] established many beautiful facts about how the collection of masses of the components of $\mathcal{F}_t^\circ$ evolve with t; one basic fact from [8] is that a.s., for each $t > 0$, $\mathcal{F}_t^\circ$ has only countably many components, and the total mass of all components of $\mathcal{F}_t^\circ$ is one. (This seems intuitively obvious, but note that it is a priori possible that for every $t > 0$, $\mathcal{F}_t^\circ$ contains uncountably many components, each of mass zero; consider $[0, 1] \setminus \mathbb{Q}$.)

DESCRIPTION OF THE BACKBONE. For $t \geq 0$, write $\widetilde{\mathcal{T}}_t$ for the component of $\mathcal{F}_t^\circ$ containing the root ρ at time t; then define a process $(L(t), t \geq 0)$ by setting

$$L(t) = \int_0^t \mu(\widetilde{\mathcal{T}}_s)\, ds. \tag{5}$$

The process $L(t)$ is the continuum analogue of the "number of cuts by time t"; the process $(L(t), t \geq 0)$ will code the distance along the backbone in the continuum transformation; see Theorem 5.5 and Corollary 5.6 below.

Theorem 6 of [8] states that if we define an increasing function $(X(t), t \geq 0)$ by

$$\big(\mu(\widetilde{\mathcal{T}}_t), t \geq 0\big) = \frac{1}{1 + X(t)}, \tag{6}$$

then $X(\cdot)$ is a stable subordinator of index $1/2$, or in other words, is distributed as the inverse local time process at zero of a standard reflecting Brownian motion. The function $X(\cdot)$ has almost sure quadratic growth, and it follows that $L(\infty) := \lim_{t \to \infty} L(t)$ is almost surely finite. [The proof of Corollary 5.6, below, contains a different proof that $L(\infty)$ is almost surely finite, using the principle of accompanying laws.]

THE PRUNED SUBTREES, AND THEIR GRAFTING ON THE BACKBONE. Since $\mathcal{P}$ is a countable set, we may enumerate its atoms as $((p_i, \tau_i), i \in \mathbb{N})$. For $t \geq 0$, let

$$I_t = \{i \in \mathbb{N} : 0 \leq \tau_i \leq t, \mu(\widetilde{\mathcal{T}}_{\tau_i}) < \mu(\widetilde{\mathcal{T}}_{\tau_i -})\}$$

and let

$$\mathcal{P}_t = \{p_i : i \in I_t\} \subseteq \mathcal{P}_t^\circ.$$

Let $P_\infty = \bigcup_{t \geq 0} \mathcal{P}_t$, and let $I_\infty = \bigcup_{t \geq 0} I_t$. Next, for $0 \leq t \leq \infty$, let $\widetilde{\mathcal{F}}_t = \mathcal{T} \setminus \mathcal{P}_t$, let $\tilde{d}_t$ be its intrinsic distance: for points x, y in the same component of $\widetilde{\mathcal{F}}_t$, we have $\tilde{d}_t(x, y) = d(x, y)$, while for x, y in distinct components of $\widetilde{\mathcal{F}}_t$, we have $\tilde{d}_t(x, y) = \infty$,[1] and let $\tilde{\mu}_t$ be the restriction of μ to $\widetilde{\mathcal{F}}_t$. Then let $(\mathcal{F}_t, d_t)$ be the metric space completion of $(\widetilde{\mathcal{F}}_t, \tilde{d}_t)$, and let μ_t be the extension of $\tilde{\mu}_t$ obtained by assigning measure zero to all points of $\mathcal{F}_t \setminus \widetilde{\mathcal{F}}_t$; note that there are only countably many such points.[2]

Next, write $\mathcal{T}_t$ for the component of $\mathcal{F}_t$ containing ρ. We then have that a.s. for all $t \geq 0$, $\widetilde{\mathcal{T}}_t$ is a connected component of $\widetilde{\mathcal{F}}_t$, and that a.s.

$$\big(\mu(\widetilde{\mathcal{T}}_t), t \geq 0\big) = \big(\mu_t(\mathcal{T}_t), t \geq 0\big). \tag{7}$$

By definition, a.s. for every $0 \leq s < t$, every component of $\widetilde{\mathcal{F}}_s$ not containing ρ is also a component of $\widetilde{\mathcal{F}}_t$. This naturally extends to the completions $\mathcal{F}_s$ and $\mathcal{F}_t$.

For $0 \leq t \leq \infty$, let $\tilde{\phi}_t$ be the identity map from $\widetilde{\mathcal{F}}_t$ to $\mathcal{T}$, and let ϕ_t be the unique extension of $\tilde{\phi}_t$ to $\mathcal{F}_t$ whose restriction to any component of $\mathcal{F}_t$ is a continuous function. With probability one, for each $i \in I_\infty$, p_i has degree two in $\mathcal{T}$ and also in $\mathcal{F}_{\tau_i -}$. It follows that almost surely, for each $i \in I_\infty$, $\mathcal{F}_{\tau_i} \setminus \mathcal{F}_{\tau_i -}$ contains precisely two points. Call these points x_i and y_i, labeled so that $x_i \notin \mathcal{T}_{\tau_i}$ and $y_i \in \mathcal{T}_{\tau_i}$. Write f_i for the component of $\mathcal{F}_{\tau_i}$ containing x_i. Necessarily, $x_i \in f_i \setminus \widetilde{\mathcal{F}}_t$ and $p_i = \phi_t(x_i)$ is the closest point of $\phi_t(f_i)$ to ρ; in other words, p_i is "the root of the subtree cut at time τ_i." Also, x_i and y_i are both leaves in $\mathcal{F}_{\tau_i}$. For distinct points $p_i, p_j \in I_t$ the trees f_i, f_j are disjoint, so in particular $x_i \neq x_j$.

The space $(\mathcal{F}_\infty, d_\infty, \mu_\infty)$ is the limiting analogue of the forest $\mathbf{F}$ from Section 3.2. We note that $(\mathcal{T}, \mu)$ can be recovered from $(\mathcal{F}_\infty, d_\infty, \mu_\infty)$ by identifying x_i and y_i for each $p_i \in I_\infty$, and taking as measure the corresponding push-forward of μ_∞.

[1] See [18], Sections 2.3 and 2.4, for the general definition of intrinsic distance for a subset of a metric space.

[2] The assiduous reader may ask: the forest $(\mathcal{F}_t, d_t, \mu_t)$ is meant to be a random element of what (Polish) space? One possible answer is to view this forest as given by some random function $e_t : [0, 1] \to [0, \infty)$ with $e_t(0) = e_t(1) = 0$, and with the "components" of the forest separated by the zeros of e_t; this perspective is elaborated in Appendix B. However, this forest itself is essentially introduced for expository purposes and plays no role in the sequel; as such, the details of how to formalize the definition of $(\mathcal{F}_t, d_t, \mu_t)$ are unimportant in the remainder of the paper.

For $0 \le t \le \infty$, let A_t be the real tree consisting of the line segment $[0, L(t)]$ with the standard distance. Then form a measured $\mathbb{R}$-tree $(\widehat{\mathcal{T}}_t, \hat{d}_t, \hat{\mu}_t)$ from A_t and $\mathcal{F}_t \setminus \mathcal{T}_t$, by identifying $x_i \in f_i$ and $L(\tau_i) \in [0, L(t)]$, for each $i \in I_t$, with measure $\hat{\mu}_t$ given by the push-forward of $\mu_t|_{\mathcal{F}_t \setminus \mathcal{T}_t}$. [We justify that $(\widehat{\mathcal{T}}_t, \hat{d}_t, \hat{\mu}_t)$ is indeed a well-defined random $\mathbb{R}$-tree, using a coding by excursions, in Appendix B.] We naturally view these spaces as increasing in t. Write $\widehat{\mathcal{T}} = \widehat{\mathcal{T}}_\infty$, $\hat{d} = \hat{d}_\infty$, $\hat{\mu} = \hat{\mu}_\infty$ and let $u = L(0)$ and $v = L(\infty)$. Almost surely both u and v are elements of $\widehat{\mathcal{T}}$.

The set of points of $[\![u, v]\!]$ of degree greater than two in $(\widehat{\mathcal{T}}, \hat{d})$ are precisely the images in $\widehat{\mathcal{T}}$ of the points $\{x_i, i \in I_\infty\}$ in $\mathcal{F}_\infty$, and if x is the image of such a point x_i, then $\hat{d}(u, x) = L(\tau_i)$. It follows that the set of times $\{\tau_i, i \in I_\infty\}$ is measurable with respect to $(\widehat{\mathcal{T}}, \hat{d}, \hat{\mu})$. Also, a.s. $\{y_i, i \in I_\infty\} \cap \{x_i, i \in I_\infty\} = \varnothing$, so none of the points $\{y_i, i \in I_\infty\}$ are identified with other points when forming $\widehat{\mathcal{T}}$. In other words, we may view the points $\{y_i, i \in I_\infty\}$ as points of $\widehat{\mathcal{T}}$ (rather than as *members* of equivalence classes of points).

Now recall the definition of $(\mathcal{T}, d, \mu, \rho)$ from the start of the section, and let ρ' be a point of $\mathcal{T}$ selected according to μ and independent of ρ.

THEOREM 5.1. *It holds that* $(\widehat{\mathcal{T}}, \hat{d}, \hat{\mu}, (u, v))$ *has the same distribution as* $(\mathcal{T}, d, \mu, (\rho, \rho'))$. *Furthermore, conditionally on* $(\widehat{\mathcal{T}}, \hat{d}, \hat{\mu}, (u, v))$, *the elements of* $\{y_i, i \in I_\infty\}$ *are mutually independent, and for all* $i \in I_\infty$, y_i *is distributed according to the probability measure* $\hat{\mu}|_{\widehat{\mathcal{T}} \setminus \widehat{\mathcal{T}}_{\tau_i}} / (1 - \hat{\mu}(\widehat{\mathcal{T}}_{\tau_i}))$.

We remark that Theorem 1.5 is an immediate consequence of the first assertion of the theorem. Likewise, Theorem 1.7 immediately follows from the definitions of $(\widehat{\mathcal{T}}, \hat{d}, \hat{\mu}, (u, v))$ and of the points $\{y_i, i \in I_\infty\}$ and from the second assertion of the theorem.

The remainder of Section 5 is devoted to the proof of Theorem 5.1. The proof of Theorem 5.1 relies on couplings with the construction for uniform Cayley trees, and we introduce these couplings in Section 5.2. In Section 5.3, we show that the process $(L(t), t \ge 0)$ is indeed the correct analogue of "number of cuts" in the discrete setting. Finally, we wrap up the proof of Theorem 5.1 in Section 5.4.

5.2. *Some couplings between discrete and continuous trees.* The couplings we introduce in this section are not specific to the case of uniform Cayley trees. This will be important in Section 6, when we extend our results to other finite-variance critical conditioned Galton–Watson trees.

Let $\xi = (\xi_i, i \ge 0)$ be a critical finite-variance offspring distribution, that is, a probability distribution on $\{0, 1, \ldots\}$ with

$$\sum_{i \ge 0} i \xi_i = 1 \quad \text{and} \quad \sigma^2 = \sum_{i \ge 0} i(i-1)\xi_i \in (0, \infty).$$

In the following, we consider only values of n such that a sum of n i.i.d. random variables with distribution ξ equals $n - 1$ with positive probability. For such $n \ge 1$, let T^n be a Galton–Watson tree with offspring distribution ξ, conditioned to have

n nodes. For $x, y \in T^n$ let $d^n(x, y)$ be $\sigma n^{-1/2}$ times the graph distance between x and y in T^n. Let ρ^n denote the root of T^n, let μ^n be the measure placing mass $1/n$ on each node of T^n and let ℓ^n be the measure placing mass $\sigma n^{-1/2}$ on each vertex of T^n (the "discrete, rescaled length measure"). Let next, $\mathcal{T}$ be the Brownian CRT with root ρ and distance metric d, let μ be its mass measure and let ℓ be the length measure on the skeleton of $\mathcal{T}$. We will use the following fundamental result heavily.

THEOREM 5.2 (Aldous [7], Le Gall [35]). *It holds that*

$$\big(T^n, d^n, \mu^n, \rho^n\big) \xrightarrow{d} (\mathcal{T}, d, \mu, \rho)$$

as $n \to \infty$, where convergence is in the 1-pointed Gromov–Hausdorff–Prokhorov sense.

Strictly speaking, neither of the above papers establishes Gromov–Hausdorff–Prokhorov convergence. However, deducing Theorem 5.2 from the earlier results is essentially immediate; we briefly sketch the line of the proof. First, by Proposition 10 of [42], to prove Theorem 5.2 it suffices to establish convergence of (T^n, d^n, μ^n) to $(\mathcal{T}, d, \mu)$ in the Gromov–Hausdorff–Prokhorov sense. Second, it is straightforward to verify that Gromov–Hausdorff–Prokhorov convergence is equivalent to Gromov–Hausdorff convergence plus convergence of all finite-dimensional marginals. The former convergence is established in [35], and the latter is established in [7]. (See also Theorem 8 of Haas and Miermont [24], who explicitly state Gromov–Hausdorff–Prokhorov convergence as an application of their results on Markov branching trees.)

First, by Skorohod's representation theorem (see, e.g., [16]), we may consider a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ in which we have the almost sure GHP convergence

$$\big(T^n, d^n, \mu^n, \rho^n\big) \to (\mathcal{T}, d, \mu, \rho).$$

In such a space, we may find a sequence of correspondences $(R_n, n \geq 1)$ between T^n and $\mathcal{T}$, such that $\operatorname{dis}(R_n) \to 0$ almost surely as $n \to \infty$. We may also find a sequence of couplings $(\nu_n, n \geq 1)$ between μ^n and μ such that the defect $D(\nu_n) \to 0$ almost surely as $n \to \infty$, and such that $\nu_n(R_n^c) \to 0$ almost surely as $n \to \infty$.

Next, let $(s_i, i \geq 1)$ be a random sequence of independent points of $\mathcal{T}$ distributed according to μ, and for each $n \in \mathbb{N}$ let $(s_i^n, i \geq 1)$ be a sequence of independent points of T^n distributed according to μ^n. Also, write $s_0 = \rho$ and $s_0^n = \rho^n$ for notational convenience, and for $k \geq 1$ write $S_k^n = \{s_0^n, \ldots, s_k^n\}$. The almost sure GHP convergence above implies [42], Proposition 10, that for each fixed $k \geq 1$,

$$\big(T^n, d^n, \mu^n, (s_0^n, \ldots, s_k^n)\big) \xrightarrow{d} \big(\mathcal{T}, d, \mu, (s_0, \ldots, s_k)\big),$$

in the sense of d_{GHP}^{k+1}, and Skorohod's theorem (applied once for each $k \geq 1$) then implies that we may work in a space in which almost surely, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \inf\big\{k : d_{\mathrm{GHP}}^{k+1}\big((T^n, \mu^n, (s_0^n, \ldots, s_k^n)), (\mathcal{T}, \mu, (s_0, \ldots, s_k))\big) \geq \varepsilon\big\} = \infty. \tag{8}$$

For each $n, k \geq 0$, recall that $T^n[\![S_k^n]\!]$ is the subtree of T^n spanned by S_k^n, and let ℓ_k^n be the restriction of ℓ^n to $T^n[\![S_k^n]\!]$. Also, let $\mathcal{T}[\![S_k]\!]$ be the subtree of $\mathcal{T}$ spanned by $S_k = \{s_0, \ldots, s_k\}$, and let ℓ_k be the length measure on $\mathcal{T}[\![S_k]\!]$. In the space in which (8) almost surely holds, we immediately have

$$\sup_{k\in\mathbb{N}} \lim_{n\to\infty} d_{\mathrm{GHP}}^{k+1}\big((T^n[\![S_k^n]\!], \ell_k^n, (s_0^n, \ldots, s_k^n)), (\mathcal{T}[\![S_k]\!], \ell_k, (s_0, \ldots, s_k))\big) = 0. \tag{9}$$

For each n let $\mathcal{P}^n$ be a Poisson process on $T^n \times [0, \infty)$ with intensity measure $\ell^n \otimes dt$. Then $\mathcal{P}^n$ converges in distribution to $\mathcal{P}$ in the sense of uniform convergence on sets of finite length measure [20], Chapter 11.

Recall that we have enumerated the atoms of $\mathcal{P}$ as $((p_i, \tau_i), i \in \mathbb{N})$; likewise, for each $n \in \mathbb{N}$ we list the atoms of $\mathcal{P}^n$ as $((p_i^n, \tau_i^n), i \in \mathbb{N})$. We noted above that a.s. for each $i \in \mathbb{N}$, p_i has degree two in $\mathcal{T}$ and in $\mathcal{F}_{\tau_i -}$. Since $\mathcal{T}$ is compact, yet another application of Skorohod's theorem then implies that we may find a space in which in addition to (8) and (9), almost surely for each $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \inf\{i : |\tau_i^n - \tau_i| > \varepsilon\} = \infty \tag{10}$$

for each $k \geq 0$ we have

$$\begin{aligned}\lim_{n\to\infty} \inf\{i \in \mathbb{N} : |\mathcal{T}[\![S_k]\!] \cap \{p_1, \ldots, p_i\}| \neq |T^n[\![S_k^n]\!] \cap \{p_1^n, \ldots, p_i^n\}|\} \\ = \infty\end{aligned} \tag{11}$$

and for any fixed $k \geq 0$, $i \geq 1$, writing

$$U_{k,i}^n = (s_0^n, \ldots, s_k^n, p_1^n, \ldots, p_i^n) \quad \text{and} \quad U_{k,i} = (s_0, \ldots, s_k, p_1, \ldots, p_i),$$

we a.s. have

$$d_{\mathrm{GHP}}^{k+1+i}\big((T^n, d^n, \mu^n, U_{k,i}^n), (\mathcal{T}, d, \mu, U_{k,i})\big) \to 0 \tag{12}$$

as $n \to \infty$.

To sum up: by a sequence of applications of Skorohod's theorem we have arrived at a space in which, after rescaling, the sequence T^n converge almost surely to a Brownian CRT $\mathcal{T}$. We have additionally coupled a sequence of random draws from the mass measure of $\mathcal{T}$ to its discrete counterpart, and a Poisson process on $\mathrm{skel}(\mathcal{T}) \times [0, \infty)$ to *its* discrete counterpart, in such a way that any finite collection of such points in the limiting space is arbitrarily closely approximated by a corresponding (in both the informal and the technical sense) collection of points in T^n, for n large enough. Furthermore, we have done so in such a manner that for any fixed $t > 0$ and $k \geq 1$, the operation of restricting the Poisson process to the set of points arriving before time t and falling within the subtree spanned by the first k random draws from the mass measure, commutes with taking the large-n limit.

5.3. *The convergence of the discrete backbones*. In this section we continue to assume that T^n is a conditioned Galton–Watson tree with critical, finite-variance offspring distribution ξ. Before proving Theorem 5.1, we also need to express the modified Aldous–Broder dynamics in the setting of conditioned Galton–Watson trees. The only minor issue which needs to be addressed is the fact that the modified Aldous–Broder dynamics should ignore points of $\mathcal{P}^n$ which fall in an already cut subtree.

First, we consider the planted tree $T^n\langle\rho^n\rangle$ and call the planted vertex w^n. We extend μ^n to $v(T^n\langle\rho^n\rangle)$ by setting $\mu^n(\{w^n\}) = 0$. Recall the notation $a(v)$ for the parent of vertex v. For each $0 \le t \le \infty$, let T_t^n be the component of

$$\big(v\big(T^n\langle\rho^n\rangle\big), e\big(T^n\langle\rho^n\rangle\big) \setminus \big\{\big(p_i^n, a\big(p_i^n\big)\big) : 0 \le \tau_i^n \le t\big\}\big)$$

containing w^n, and define T_{t-}^n accordingly. (The forest in the preceding equation is the finite-n analogue of $\mathcal{F}_t^\circ$, but will not be used in what follows.) Write

$$I_t^n = \big\{i \in \mathbb{N} : 0 \le \tau_i^n \le t, \mu^n\big(T_t^n\big) < \mu^n\big(T_{t-}^n\big)\big\}$$

for the indices corresponding to "effective" cuts up to time t, and let

$$\mathcal{P}_t^n = \big\{p_i^n : i \in I_t^n\big\}$$

be the set of locations of these cuts. For $i \in I_\infty^n$ let $x_i^n = p_i^n$ and let y_i^n be the parent of x_i^n in T^n (here we view ρ^n as its own parent). Then, for $0 \le t \le \infty$, let

$$\mathcal{F}_t^n = \big(v\big(T^n\big), e\big(T^n\big) \setminus \big\{(x_i, y_i) : 0 \le \tau_i \le t\big\}\big)$$

and for $i \in I_\infty^n$ write f_i^n for the component of $\mathcal{F}_\infty^n$ containing x_i^n. Note that f_i^n is in fact a component of $\mathcal{F}_t^n$ for all $\tau_i^n \le t \le \infty$.

Write $\kappa^n = |I_\infty^n|$, and write π^n for the permutation of I_∞^n that reorders the elements of I_∞^n in increasing order of the corresponding cut time, so that for $i, j \in I_\infty^n$, $\pi^n(i) < \pi^n(j)$ if and only if $\tau_i^n < \tau_j^n$. Also, write

$$u^n = x_{\pi^n(1)}^n \quad \text{and} \quad v^n = x_{\pi^n(\kappa^n)}^n.$$

Finally, let $\widehat{T}^n$ be the tree obtained from $\mathcal{F}_\infty^n$ by removing w^n, then adding the edges

$$\big(x_{\pi^n(i+1)}^n, x_{\pi^n(i)}^n\big) \qquad 1 \le i < \kappa^n.$$

We view $\widehat{T}^n$ as rooted at u^n.

REMARK. It is a standard fact that if ξ is a mean-one Poisson distribution (in fact, the mean does not matter), then T^n has the same distribution as the tree obtained from a uniform Cayley tree on $[n]$ by removing the vertex labels. In this case, Theorem 3.1 then implies that $\widehat{T}^n$ is distributed as a uniform Cayley tree with labels removed, and v^n is a uniformly random element of $v(\widehat{T}^n)$, independent of $\widehat{T}^n$. This fact will be used in the course of the proof of Theorem 5.1 in Section 5.4. However, it plays almost no role in the current section. In particular, all results presented in this section, with the exception of Corollary 5.6, are valid for general critical, finite-variance conditioned Galton–Watson trees.

LEMMA 5.3. *In the space where* (8)–(12) *hold, almost surely*

$$\big(\mu^n(T_t^n), t \geq 0\big) \to \big(\mu_t(\mathcal{T}_t), t \geq 0\big),$$

in the sense of uniform convergence on compacts for the Skorohod J_1 topology.

PROOF. Write ν_k for the uniform measure on points $s_0, \ldots, s_k$. In other words, given $\mathcal{T}$ and $s_0, \ldots, s_k$, ν_k assigns mass $1/(k+1)$ to each of the points $s_0, \ldots, s_k$. Similarly, write ν_k^n for the uniform measure on $s_0^n, \ldots, s_k^n$. By (12), for any fixed $i, k \geq 1$, almost surely

$$\lim_{n\to\infty} d_{\mathrm{GHP}}^{k+1+i}\big((T^n, d^n, \nu_k^n, U_{k,i}^n), (\mathcal{T}, d, \nu_k, U_{k,i})\big) = 0. \tag{13}$$

Also, by Theorem 8 of [8], for almost every realization of $\mathcal{T}$,

$$\lim_{k\to\infty} d_{\mathrm{P}}(\nu_k, \mu) = 0. \tag{14}$$

(In fact, in [8], only almost sure weak convergence is claimed, but the proof simply consists of an application of the Glivenko–Cantelli theorem and is easily seen to yield convergence with respect to d_{P}.) Since for all $t \geq 0$, $\mathcal{T}_t$ is a compact subspace of $\mathcal{T}$, and the $\mathcal{T}_t$ are decreasing in t, it follows that

$$\big(\nu_k(\mathcal{T}_t), t \geq 0\big) \to \big(\mu_t(\mathcal{T}_t), t \geq 0\big) \tag{15}$$

as $k \to \infty$. Combining (13) with (10) and (11), we obtain that for each $k \geq 0$, almost surely

$$\big(\nu_k^n(T_t^n), t \geq 0\big) \to \big(\nu_k(\mathcal{T}_t), t \geq 0\big) \tag{16}$$

as $n \to \infty$. Next, combining (14) with (12), we obtain that almost surely

$$\lim_{k\to\infty} \lim_{n\to\infty} d_{\mathrm{GHP}}^{k+1+i}\big((T^n, d^n, \mu^n, U_{k,i}^n), (\mathcal{T}, d, \nu_k, U_{k,i})\big) = 0,$$

which together with (13) implies that almost surely

$$\lim_{k\to\infty} \limsup_{n\to\infty} d_{\mathrm{P}}\big(\mu^n, \nu_k^n\big) = 0.$$

In view of (15) and (16), this proves the lemma. □

Next, for each $n \geq 1$, reorder the elements of $\mathcal{P}^n$ as $\{(p_{n,i}, t_{n,i}), i \geq 1\}$ so that $t_{n,i} < t_{n,i+1}$ for all $i \geq 1$. We emphasize that here we consider all atoms of $\mathcal{P}^n$, not only those that correspond to "effective cuts."

LEMMA 5.4. *In the space where* (8)–(12) *hold, a.s.*

$$\left(\frac{1}{\sigma\sqrt{n}} \sum_{\{j\,:\,t_{n,j}\leq t\}} \mu^n\big(T_{t_{n,j}}^n\big), t \geq 0\right) \to \big(L(t), t \geq 0\big),$$

in the sense of uniform convergence on compacts for the uniform distance, as $n \to \infty$.

PROOF. From Lemma 5.3 it is immediate that

$$\left(\int_0^t \mu^n(T_s^n)\,ds, t \geq 0\right) \to \left(\int_0^t \mu_s(\mathcal{T}_s)\,ds, t \geq 0\right)$$

as $n \to \infty$. Also, $\ell^n(T^n) = \sigma\sqrt{n}$, so the set $\{\tau_i^n, i \in \mathbb{N}\}$ forms a Poisson point process of intensity $\sigma\sqrt{n}$ on $[0, \infty)$, from which it follows straightforwardly that

$$\left(\frac{1}{\sigma\sqrt{n}} \sum_{\{j\,:\,t_{n,j} \leq t\}} \mu^n(T_{t_{n,j}}^n), t \geq 0\right) \to \left(\int_0^t \mu_s(\mathcal{T}_s)\,ds, t \geq 0\right)$$

and the result then follows from (7) and the definition of $L(t)$ in (5). □

Our next goal is to show that $(L(t), t \geq 0)$ is the limit of the discrete process which tracks the number of effective cuts up to time $t\sqrt{n}$. Write

$$L^n(t) = |\mathcal{P}_t^n| = \#\{s \leq t : \mu^n(T_s^n) < \mu^n(T_{s-}^n)\}$$

and note that, for every $n \geq 1$, $L^n(t)$ increases to $\kappa^n(T^n) = \#\{s > 0 : \mu^n(T_s^n) < \mu^n(T_{s-}^n)\}$, as $t \to \infty$.

THEOREM 5.5. *In the space in which* (8)–(12) *hold, a.s.*

$$\left(L^n(t)/(\sigma n^{1/2}), t \geq 0\right) \to \left(L(t), t \geq 0\right)$$

in the sense of uniform convergence on compacts for the uniform distance, as $n \to \infty$.

In proving Theorem 5.5 we will use a martingale inequality from [37], Theorem 3.15. Let $\{X_i\}_{i=0}^n$ be a bounded martingale with $X_0 = 0$, adapted to a filtration $\{\mathcal{G}_i\}_{i=0}^n$. Next let $V = \sum_{i=0}^{n-1} \mathbf{V}[X_{i+1}|\mathcal{G}_i]$, where

$$\mathbf{V}[X_{i+1}|\mathcal{G}_i] := \mathbf{E}\big[(X_{i+1} - X_i)^2|\mathcal{G}_i\big] = \mathbf{E}\big[X_{i+1}^2|\mathcal{G}_i\big] - X_i^2$$

is the predictable quadratic variation of X_{i+1}. Define

$$v = \operatorname{ess\,sup} V \quad \text{and} \quad b = \max_{0 \leq i \leq n-1} \operatorname{ess\,sup}(X_{i+1} - X_i|\mathcal{G}_i),$$

where for a random variable X, the essential supremum $\operatorname{ess\,sup} X$ is defined to equal $\sup\{x : \mathbf{P}(X \geq x) > 0\}$. Then we have the following bound [37]. For any $t \geq 0$,

$$\mathbf{P}\Big(\max_{0 \leq i \leq n} X_i \geq t\Big) \leq \exp\left(-\frac{t^2}{2v(1 + bt/(3v))}\right). \tag{17}$$

PROOF OF THEOREM 5.5. In a first part, we prove uniform convergence on compacts for which we do not need the trees T^n, $n \geq 1$, to be uniform Cayley trees. Fix $\delta > 0$ and $C > 0$. By Lemma 5.4, a.s.

$$\sup_{0 \leq t \leq C} \left| \frac{1}{\sigma\sqrt{n}} \sum_{\{j\,:\,t_{n,j} \leq t\}} \mu^n(T_{t_{n,j-1}}^n) - L(t) \right| \to 0$$

as $n \to \infty$. It follows that

$$\begin{aligned}(18)\qquad &\mathbf{P}\Big(\limsup_{n\to\infty} \sup_{0\le t\le C} \Big|\frac{1}{\sigma\sqrt{n}} L^n(t) - L(t)\Big| > \delta\Big) \\ &\quad \le \mathbf{P}\Big(\limsup_{n\to\infty} \sup_{0\le t\le C} \Big|L^n(t) - \sum_{\{j\,:\,t_{n,j}\le t\}} \mu^n\big(T^n_{t_{n,j-1}}\big)\Big| > \sigma\delta n^{1/2}\Big).\end{aligned}$$

Also, since $\mathcal{P}^n$ has intensity measure $\ell^n \otimes dt$ and $\ell^n(v(T^n)) = \sigma n^{1/2}$, we have that

$$\lim_{x\to\infty} \mathbf{P}\Big(\liminf_{n\to\infty} t_{n,\lfloor x\sqrt{n}\rfloor} > C\Big) = 1,$$

which implies that the probability in (18) is at most

$$(19)\qquad \lim_{x\to\infty} \mathbf{P}\Big(\limsup_{n\to\infty} \max_{i\le x\sqrt{n}} \Big|L^n(t_{n,i}) - \sum_{1\le j\le i} \mu^n\big(T^n_{t_{n,j-1}}\big)\Big| > \sigma\delta n^{1/2}\Big).$$

For $i \ge 1$, write

$$X_i = \mathbf{1}_{\{\mu^n(T^n_{t_{n,i}}) < \mu^n(T^n_{t_{n,i-1}})\}}.$$

Also, for each $i \ge 1$, let $\mathcal{P}^n_i = \{p_{n,1}, \dots, p_{n,i}\}$. Taking $\mathcal{G}_{n,i}$ to be the sigma field generated by T^n and $\mathcal{P}^n_i$, then $(X_i, i \ge 1)$ is adapted to $(\mathcal{G}_{n,i}, i \ge 1)$. Note that

$$\mathbf{E}[X_i|\mathcal{G}_{n,i-1}] = \mu^n\big(T^n_{t_{n,i-1}}\big) = \mathbf{E}\big[X_i^2|\mathcal{G}_{n,i-1}\big],$$

so in all cases $\operatorname{Var}[X_i|\mathcal{G}_{n,i-1}] \le 1/4$. Also, for all $i \ge 1$ we have $\sum_{j=1}^i X_j = L^n(t_{n,i})$. By (17), for any fixed $x > 0$ and $n \ge 1$, we thus have

$$\begin{aligned}(20)\qquad &\mathbf{P}\Big(\max_{i\le x\sqrt{n}} \Big|\sum_{j\le i} X_j - \sum_{j\le i} \mu^n\big(T^n_{t_{n,j-1}}\big)\Big| \ge y\Big) \\ &\quad \le 2\exp\Big(-\frac{2y^2}{x\sqrt{n}(1 + 4y/(3x\sqrt{n}))}\Big).\end{aligned}$$

Applying this bound with $y = \delta\sqrt{n}$ and summing over n, it follows by Borel–Cantelli that

$$\mathbf{P}\Big(\limsup_{n\to\infty} \max_{i\le x\sqrt{n}} \Big|L^n(t_{n,i}) - \sum_{1\le j\le i} \mu^n\big(T^n_{t_{n,j-1}}\big)\Big| > \delta n^{1/2}\Big) = 0,$$

which together with (19) shows that $(L^n(t)/(\sigma n^{1/2}), 0 \le t \le C) \to (L(t), 0 \le t \le C)$ almost surely for the uniform distance. □

COROLLARY 5.6. *If ξ is the* Poisson(1) *distribution then in the space in which* (8)–(12) *hold,*

$$\big(L^n(t)/n^{1/2}, t \ge 0\big) \to \big(L(t), t \ge 0\big)$$

in probability in the sense of uniform convergence on $[0, \infty)$.

PROOF. If ξ is the Poisson(1) distribution, then $\sigma = 1$. Uniform convergence on compacts follows from Theorem 5.5. Furthermore, as noted in the remark just before Lemma 5.3, in this case T^n is distributed as a uniform Cayley tree on $[n]$ with labels removed. Also, $\widehat{T}^n$ is again distributed as a uniform Cayley tree with labels removed, and $\kappa^n(T^n)$ is the distance between u^n and v^n in $\widehat{T}^n$, it follows from Theorem 3.1 that $\kappa^n(T^n)/n^{1/2}$ converges in distribution to a Rayleigh random variable.

For any $t, \delta > 0$, given that $\mu^n(T_t^n) \le \delta$, the difference $\kappa^n(T^n) - L^n(t)$ is dominated by the number of cuts required to isolate the root of a uniform Cayley tree on $\lfloor \delta n \rfloor$ vertices. It follows that for any $\varepsilon > 0$,

$$\lim_{t\to\infty} \limsup_{n\to\infty} \mathbf{P}\big(\kappa^n(T^n) - L^n(t) > \varepsilon n^{1/2}\big) = 0. \tag{21}$$

By the principle of accompanying laws (Theorem 9.1.13 of [47]), in the space in which (8)–(12) hold, we have

$$\frac{\kappa^n(T^n)}{n^{1/2}} \xrightarrow{p} L(\infty) = \lim_{t\to\infty} L(t),$$

which together with (21) implies uniform convergence on $[0, \infty)$. [This also yields a second proof that $L(\infty)$ is almost surely finite, as promised just after (5).] $\square$

Before proving Theorem 5.1 we note one consequence of Corollary 5.6, stated in the Introduction as Corollary 1.6. A different proof of this result can be found in Abraham and Delmas [2].

PROOF OF COROLLARY 1.6. In proving Corollary 5.6 we showed the existence of a space in which

$$L(\infty) \stackrel{p}{=} \lim_{n\to\infty} \frac{\kappa^n(T^n)}{n^{1/2}}$$

and the latter limit is Rayleigh distributed by Theorem 3.1 The lemma then follows from the definition of $L(t)$ in (5) and (7). $\square$

5.4. *The proof of Theorem* 5.1. In this section, in order to use the discrete results of Section 3, we assume that ξ is the Poisson(1) distribution, or equivalently (see the remark just before Lemma 5.3) that T^n is a uniform Cayley tree on $[n]$ with its labels removed. In particular, this implies that $\sigma = 1$.

Recall the definitions of the trees $\{f_i, i \in I_\infty\}$ and $\{f_i^n, i \in I_\infty^n\}$ from pages 2318 and 2322 (here we simply view each f_i as a subset of $\mathcal{T}$). Also, write $\hat{d}_n$ for $n^{-1/2}$ times the standard graph distance on $\widehat{T}^n$, and write $\hat{\mu}^n$ for the uniform probability measure on $v(\widehat{T}^n)$.

We work in a space where (8)–(12) all hold. For any $\varepsilon > 0$, let

$$J_\varepsilon = \big\{i \in I_\infty : \mu_\infty(f_i) > \varepsilon\big\}.$$

The set J_ε is necessarily finite (it has size at most ε^{-1}), so $K^{(\varepsilon)} := \sup\{i : i \in J_\varepsilon\}$ is a.s. finite. By (11), for all n sufficiently large we in particular have that $J_\varepsilon \subset I_\infty^n$, and we hereafter assume that inclusion indeed holds.

Let $S = \{u, v\} \cup \bigcup_{i\in J_\varepsilon} f_i$, and let $\widehat{T}_\varepsilon = \bigcup_{x,y\in S} [\![x, y]\!]$. In words, $\widehat{T}_\varepsilon$ is the minimal subtree of $\widehat{T}$ which contains each of the subtrees f_i, $i \in J_\varepsilon$ and also contains the distinguished nodes u and v. Likewise, let

$$\widehat{T}_\varepsilon^n = \widehat{T}^n \Big[\!\!\Big[\{u^n, v^n\} \cup \bigcup_{i\in J_\varepsilon} v(f_i^n) \Big]\!\!\Big].$$

We let $\hat{d}_\varepsilon = \hat{d}|_{\widehat{T}_\varepsilon}$, and define $\hat{\mu}_\varepsilon, \hat{d}_\varepsilon^n, \hat{\mu}_\varepsilon^n$ accordingly.

The set I_∞ is countable and $J_\varepsilon \uparrow I_\infty$ as $\varepsilon \downarrow 0$. Also, it follows from the result of Aldous and Pitman [8] mentioned earlier that $\sum_{i\in I_\infty} \mu_\infty(f_i) = 1$ a.s., and we thus a.s. have

$$\lim_{\varepsilon\downarrow 0} \sum_{i\notin J_\varepsilon} \mu_\infty(f_i) = 0.$$

Since $\mathcal{T}$ is compact and each f_i can be viewed as a subtree of $\mathcal{T}$, we must also a.s. have

$$\lim_{\varepsilon\downarrow 0} \sup_{i\notin J_\varepsilon} \operatorname{diam}(f_i) = 0.$$

(Otherwise, there would exist $\delta > 0$ and an infinite set $S \subset I_\infty$ such that for each $i \in S$, f_i has height greater than δ. For $i \in S$, letting q_i be any point in f_i whose distance to the root p_i of f_i is at least δ, the set $\{q_i, i \in S\}$ is infinite and its elements have pairwise distance at least δ, contradicting compactness.) By these facts and by (12), for any $\delta > 0$ there is $N = N(\varepsilon, \delta)$ which is almost surely finite, such that for all $n \geq N$ and $i \in J_\varepsilon$,

$$d_{\mathrm{GHP}}^{k+1+i}\big((T^n, d^n, \mu^n, U_{k,i}^n), (\mathcal{T}, d, \mu, U_{k,i})\big) < \delta \tag{22}$$

and additionally $\sum_{i\notin J_\varepsilon} \mu_\infty(f_i) < \delta$ and $\sup_{i\notin J_\varepsilon} \operatorname{diam}(f_i) < \delta$. We fix a correspondence $C \in \mathscr{C}((T^n, d^n, \mu^n, U_{k,i}^n), (\mathcal{T}, d, \mu, U_{k,i}))$ with $\operatorname{dis}(C) < 2\delta$ and containing the appropriate pairs of points from $U_{k,i}^n$ and $U_{k,i}$. It follows from the fact that $\sup_{i\notin J_\varepsilon} \operatorname{diam}(f_i) < \delta$ that

$$d_{\mathrm{GHP}}^2\big((\widehat{T}, \hat{d}, \hat{\mu}, (u, v)), (\widehat{T}_\varepsilon, \hat{d}_\varepsilon, \hat{\mu}_\varepsilon, (u, v))\big) < \delta \tag{23}$$

and that

$$\sup_{i\in I_\infty^n \setminus J_\varepsilon} n^{-1/2} \operatorname{diam}(f^n) < 3\delta. \tag{24}$$

Next, write $m_\delta = \sup_{x\in\mathcal{T}} \mu(B(x, \delta))$, where $B(x, \delta)$ is the ball of radius δ around x in $\mathcal{T}$. We have $m_\delta \downarrow 0$ a.s. as $\delta \to 0$. Choose $0 < \delta < \varepsilon^2$ small enough that

$m_{4\delta} < \varepsilon^2$. Then for $n \geq N(\varepsilon, \delta)$, and for all $i \in J_\varepsilon$, by considering the δ blow-up C_δ of the correspondence C, we see that

$$d^1_{\mathrm{GHP}}\big((f_i^n, d_\infty^n|_{f_i^n}, \mu_\infty^n|_{f_i^n}), (f_i, d_\infty|_{f_i}, \mu_\infty|_{f_i})\big) < 2\delta + m_{4\delta} < 2\varepsilon^2. \tag{25}$$

In particular, for each $i \in J_\varepsilon$, $|\mu_\infty^n(f_i^n) - \mu_\infty(f_i)| < 2\varepsilon^2$, so

$$\sum_{i \in J_\varepsilon} \big|\mu_\infty^n(f_i^n) - \mu_\infty(f_i)\big| < 2\varepsilon^2 |J_\varepsilon| < 2\varepsilon \tag{26}$$

and

$$\sum_{i \in I_\infty^n \setminus J_\varepsilon} \mu_\infty^n(f_i^n) \leq 2\varepsilon + \delta < 3\varepsilon. \tag{27}$$

By (24) and (27), it follows that for all n sufficiently large,

$$d^2_{\mathrm{GHP}}\big((\widehat{T}^n, \hat{d}^n, \hat{\mu}^n, (u^n, v^n)), (\widehat{T}_\varepsilon^n, \hat{d}_\varepsilon^n, \hat{\mu}_\varepsilon^n, (u^n, v^n))\big) < 3(\delta + \varepsilon) < 6\varepsilon.$$

For each $i \in I_\infty$, $L(\tau_i) = \hat{d}(u, x_i)$ and for each $i \in I_\infty^n$, $n^{-1/2} L^n(\tau_i^n) = \hat{d}(u^n, x_i^n)$. By Corollary 5.6, it follows that for all $i \in J_\varepsilon$, for all n sufficiently large, $|\hat{d}(u, x_i) - \hat{d}^n(u^n, x_i^n)| < \delta$. Together with (25) and (26), this implies that

$$d^2_{\mathrm{GHP}}\big((\widehat{T}_\varepsilon^n, \hat{d}_\varepsilon^n, \hat{\mu}_\varepsilon^n, (u^n, v^n)), (\widehat{T}_\varepsilon, \hat{d}_\varepsilon, \hat{\mu}_\varepsilon, (u, v))\big) < \max\big(\delta + 2\varepsilon^2, 2\varepsilon\big) < 3\varepsilon.$$

By the two preceding inequalities, (23) and the triangle inequality, we obtain that a.s. for all n sufficiently large,

$$d^2_{\mathrm{GHP}}\big((\widehat{T}^n, \hat{d}^n, \hat{\mu}^n, (u^n, v^n)), (\widehat{T}, \hat{d}, \hat{\mu}, (u, v))\big) < 9\varepsilon + \delta < 10\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the first assertion of the theorem then follows from Theorem 3.1.

Finally, since the distribution of the collection $\{y_i, i \in I_\infty\}$ is determined by its finite-dimensional distributions, the assertion in the statement of Theorem 5.1 about the collection $\{y_i, i \in I_\infty\}$ then follows from Lemma 5.7, below, whose straightforward proof is omitted.

LEMMA 5.7. *Fix $n \geq 1$, $k \geq 1$, let $K = \{i \in I_\infty^n : p_i^n \in T^n[\![S_k^n]\!]\}$ and let $j \in K$ be the element $i \in K$ which minimizes τ_i^n. Suppose that T^n is a uniform Cayley tree on $[n]$. Then for any $S \subset v(T^n)$, any tree t with $v(t) = S$, and any $y \in S$,*

$$\mathbf{P}\big(T^n_{\tau_j^n} = t \text{ and } y_j^n = y | v(T^n_{\tau_j^n}) = S\big) = |S|^{-|S|}.$$

6. Conditioned Galton–Watson trees with finite variance. We now want to prove that the picture that we have obtained for the process in the case of uniform Cayley trees is also valid when one considers conditioned Galton–Watson trees with critical, finite-variance offspring distribution. Fix an offspring distribution $\xi = (\xi_0, \xi_1, \ldots)$ with

$$\sum_{i\geq 0} i\xi_i = 1 \quad \text{and} \quad \sum_{i\geq 0} i(i-1)\xi_i = \sigma^2 \in (0,\infty).$$

THEOREM 6.1. *Let T^n be distributed as a Galton–Watson tree with offspring distribution ξ, conditioned to have n vertices. Then after rescaling, the number of cuts $\kappa(T^n)$ required to isolate the root of T^n is asymptotically Rayleigh distributed,*

$$\lim_{n\to\infty} \mathbf{P}\big(\kappa\big(T^n\big) \geq \sigma x\sqrt{n}\big) = e^{-x^2/2}.$$

Under a finite-variance assumption, Galton–Watson trees conditioned on their size have the same scaling limit as uniform Cayley trees, so when looking at a $(n, \sqrt{n})$ rescaling for time and space, the cutting process will essentially look the same. Completing the argument then boils down to showing that once the left-over tree has size $o(n)$ the number of cuts needed to completely destroy it is $o(\sqrt{n})$. The following lemma shows that this is indeed the case. (Although the factor $\varepsilon^{1/6}$ is certainly not best possible, it is sufficient for our needs.)

LEMMA 6.2. *Suppose that $\mathbf{E}\xi = 1$ and $\mathrm{Var}[\xi] = \sigma^2 \in (0,\infty)$. Let T^n be a Galton–Watson tree with progeny distribution ξ, conditioned on having size n. Let also $\tau^n(\varepsilon) = \inf\{t : \mu^n(T^n_t) < \varepsilon\}$. Then*

$$\limsup_{n\to\infty} \mathbf{P}\big(\kappa\big(T^n_{\tau^n(\varepsilon)}\big) \geq \varepsilon^{1/6}\sqrt{n}\big) \underset{\varepsilon\to 0}{\to} 0.$$

PROOF. Recall that for a rooted tree T and a node v of T, we write $h_T(v)$ for the *height* of v in T, which is the number of edges on the path from the root to v. We also write $h(T) = \max_{v\in v(T)} h_T(v)$, and call $h(T)$ the height of T. Finally, for $i \geq 0$ write $w_i(T) = \#\{v \in v(T) : h_T(v) = i\}$.

For any $x, y > 0$ we have

$$\begin{aligned}\mathbf{P}\big(\kappa\big(T^n_{\tau^n(\varepsilon)}\big) \geq y\sqrt{n}\big) &\leq \mathbf{P}\big(\kappa\big(T^n_{\tau^n(\varepsilon)}\big) \geq y\sqrt{n}, h\big(T^n_{\tau^n(\varepsilon)}\big) \leq x\sqrt{n}\big) \\ &\quad + \mathbf{P}\big(h\big(T^n_{\tau^n(\varepsilon)}\big) > x\sqrt{n}\big).\end{aligned} \tag{28}$$

The first term above is easily bounded using Markov's inequality. We use Janson's representation of the number of cuts as records in the tree [27, 28]. Given a tree t, rooted at r, one can assign extra labels to the vertices using a random permutation of $\{1, 2, \ldots, |t|\}$. This random permutation determines the order in which the vertices are considered for cutting. In this representation, a vertex u will actually produce a cut if and only if the path $[\![r, u]\!]$ between u and the root has not been

previously cut. This happens precisely if u has the minimum label of all vertices on $[\![r,u]\!]$. In particular, conditional on the height $h_t(u)$ of u in t, the probability that a vertex u produces a cut is $(h_t(u)+1)^{-1}$. It follows that

$$\begin{aligned}
&\mathbf{P}\big(\kappa\big(T^n_{\tau^n(\varepsilon)}\big)\geq y\sqrt{n}, h\big(T^n_{\tau^n(\varepsilon)}\big)\leq x\sqrt{n}\big)\\
&\quad\leq \frac{1}{y\sqrt{n}}\cdot\mathbf{E}\big[\kappa\big(T^n_{\tau^n(\varepsilon)}\big)\mathbf{1}_{\{h(T^n_{\tau^n(\varepsilon)})\leq x\sqrt{n}\}}\big]\\
&\quad\leq \frac{1}{y\sqrt{n}}\cdot\mathbf{E}\Big[\sum_{u\in T^n_{\tau^n(\varepsilon)}}\frac{1}{1+h_{T^n}(u)}\mathbf{1}_{\{h(T^n_{\tau^n(\varepsilon)})\leq x\sqrt{n}\}}\Big]\\
&\quad\leq \frac{1}{y\sqrt{n}}\cdot\mathbf{E}\Big[\sum_{0\leq i\leq x\sqrt{n}}\sum_{\{u:h_{T^n}(u)=i\}}\frac{1}{1+h_{T^n}(u)}\mathbf{1}_{\{h(T^n_{\tau^n(\varepsilon)})\leq x\sqrt{n}\}}\Big]\\
&\quad\leq \frac{1}{y\sqrt{n}}\cdot\sum_{0\leq i\leq x\sqrt{n}}\frac{\mathbf{E}[w_i(T^n)]}{1+i}\\
&\quad\leq \frac{1}{y\sqrt{n}}\cdot\sum_{0\leq i\leq x\sqrt{n}}\frac{Ci}{1+i}\leq\frac{Cx}{y},
\end{aligned}\tag{29}$$

we used the fact that $\mathbf{E}[w_k(T^n)]\leq Ck$ uniformly in $k\geq 0$ and $n\geq 0$ (see Devroye and Janson [21]) to obtain the second-to-last inequality.

To bound the second term, we relate the finite-n trees T^n to their limit $\mathcal{T}$. We work in a space in which (8)–(12) all hold, and recall from Section 5.2 the definitions of the collections of points $(s_i, i\geq 1)$ and $\{p^n_i : i\in\mathbb{N}\}$, and of their finite-$n$ counterparts $(s^n_i, i\geq 1)$ and $\{p^n_i : i\in\mathbb{N}\}$. In particular, recall the definitions of the sequences S_k, S^n_k, from page 2321.

We now use that for all $\delta>0$,

$$\lim_{k\to\infty}\mathbf{P}\big(d^1_{\mathrm{GH}}\big((\mathcal{T},d,\rho),(\mathcal{T}[\![S_k]\!],d|_{\mathcal{T}[\![S_k]\!]},\rho)>\delta\big)\big)=0.$$

By (8), we then also have that

$$\lim_{k\to\infty}\limsup_{n\to\infty}\mathbf{P}\big(d_{\mathrm{GH}}\big((T^n,d^n,\rho^n),\big(T^n[\![S^n_k]\!],d^n|_{T^n[\![S^n_k]\!]},\rho^n\big)>\delta\sqrt{n}\big)\big)=0.$$

Equations (10), (11) and (12) provide a coupling of the cuts falling on $T^n[\![S^n_k]\!]$ with those falling on $\mathcal{T}[\![S_k]\!]$ so that for any fixed $t>0$ and for all sufficiently large n, the cuts falling within $T^n[\![S^n_k]\!]$ and within $\mathcal{T}[\![S_k]\!]$ occur at essentially the same times and at essentially the same locations. [This is precisely formalized by (10), (11) and (12).] It then follows that in this space, for any $\varepsilon>0$ and $\delta>0$,

$$\limsup_{n\to\infty}\mathbf{P}\big(d^1_{\mathrm{GH}}\big((T^n_{\tau^n(\varepsilon)},\sigma n^{-1/2}d^n|_{T^n_{\tau^n(\varepsilon)}},\rho^n),(\mathcal{T}_{\tau(\varepsilon)},d|_{\mathcal{T}_{\tau(\varepsilon)}},\rho)\big)>\delta\big)=0.$$

Taking $\delta = x\sqrt{\varepsilon}$, from this we immediately obtain that

$$\begin{aligned}\limsup_{n\to\infty} \mathbf{P}\left(h\big(T^n_{\tau^n(\varepsilon)}\big) \geq \frac{x}{\sigma}\sqrt{\varepsilon n}\right) &\leq \mathbf{P}\big(h(\mathcal{T}_{\tau(\varepsilon)}) \geq x\sqrt{\varepsilon}\big) \\ &\leq e^{-\alpha x^2}\end{aligned} \tag{30}$$

for some constant $\alpha > 0$. The last inequality holds since: conditional on its mass, $\mathcal{T}_{\tau(\varepsilon)}$ is a Brownian CRT (see [8], equation (44)); we have $\mu(\mathcal{T}_{\tau(\varepsilon)}) \leq \varepsilon$; the height of a Brownian CRT is distributed as the supremum of a Brownian excursion; and the supremum of a Brownian excursion has Gaussian tails [29].

Then, choosing, for instance, $x = \varepsilon^{1/3}$ in (28) and $y = \varepsilon^{1/6}$ and using the bounds in (29) and (30) to bound (28) proves the result. □

Putting together Corollary 1.6 and the following lemma then yields Theorem 6.1.

LEMMA 6.3. *Let T^n be a Galton–Watson tree with offspring distribution ξ conditioned to have size n, and let $\mathcal{T}$ be a Brownian CRT. If $\mathbf{E}\xi = 1$ and $\mathrm{Var}(\xi) = \sigma^2 \in (0, \infty)$ then*

$$\frac{\kappa(T^n)}{\sigma\sqrt{n}} \underset{n\to\infty}{\overset{d}{\to}} \int_0^\infty \mu(\mathcal{T}_t)\,dt.$$

PROOF. Write T^n_t for the subtree containing the root at time t of the cutting process, and as in Section 5.3 write

$$L^n(t) = \#\big\{s \leq t : \mu^n\big(T^n_t\big) < \mu^n\big(T^n_{t-}\big)\big\}$$

for the number of cuts occurring before time t, Theorem 5.5 implies that for any fixed $t \in [0, \infty)$

$$\frac{L^n(t)}{\sigma\sqrt{n}} \xrightarrow{d} \int_0^t \mu(\mathcal{T}_t)\,dt \tag{31}$$

as $n \to \infty$.

Recall that $\tau^n(\varepsilon) = \inf\{t : \mu^n(T^n_t) < \varepsilon\}$. Since $\tau^n(\varepsilon) < \infty$ almost surely and additionally $\tau^n(\varepsilon) \to \tau(\varepsilon)$ in distribution jointly with the convergence in (31), we have

$$\frac{L^n(\tau^n(\varepsilon))}{\sigma\sqrt{n}} \xrightarrow{d} \int_0^{\tau(\varepsilon)} \mu(s)\,ds.$$

On the other hand,

$$\frac{\kappa(T^n) - L^n(\tau^n(\varepsilon))}{\sqrt{n}} \leq \frac{\kappa(T^n_{\tau^n(\varepsilon)})}{\sqrt{n}} \underset{\varepsilon\to 0}{\to} 0,$$

in probability, uniformly for all n sufficiently large, by Lemma 6.2. Since $\tau(\varepsilon) \to \infty$ almost surely as $\varepsilon \to 0$, it follows that

$$\frac{\kappa(T^n)}{\sigma\sqrt{n}} \xrightarrow{d} \int_0^\infty \mu(\mathcal{T}_t)\,dt$$

as $n \to \infty$, as claimed. □

APPENDIX A: COROLLARY 1.2: PROOF SKETCH AND DISCUSSION

Fix a rooted tree t with nodes $\{1, \dots, n\}$, and a sequence $(v_1, \dots, v_k) \in \{1, \dots, n\}^k$ of nodes of t. In t, view the children of a node as ordered so that node labels increase from left to right. Let t' be the subtree of t spanned by the root and $v_1, \dots, v_k$. Let the *reduced* tree t^* be obtained from t' by suppressing degree-two vertices (so in t^*, the parent of v_i corresponds to the most recent common ancestor of v_i and any of the v_j with $v_j \neq v_i$) and suppressing vertex labels (but keeping the plane tree structure). Since t^* has no nodes of degree 2, it has at most $2k-1$ edges, with equality precisely if it is binary and $v_1, \dots, v_k$ are distinct. Write e for the number of edges of t^*.

Given the tree t^*, one may recover t by listing an ordered rooted forest $f_1, \dots, f_m$, together with a weak composition $(c_1, \dots, c_e)$ of m into e parts. To do so, list the edges of t^* according to their order of first traversal by a contour exploration of t^*. Then glue the roots of $f_1, \dots, f_{c_1}$ along the first edge, $f_{c_1+1}, \dots, f_{c_1+c_2}$ along the second edge, and so on. A result of Riordan [46] states that the number of ordered rooted forests on vertices $\{1, \dots, n\}$ with m components is

$$B_{n,m} := m! \binom{n-1}{m-1} n^{n-m}.$$

It follows that the number of trees t with reduced tree t^* and such that t' has m vertices, is

$$A_k(t^*) B_{n,m} \binom{m+e-1}{e-1},$$

where $\binom{m+e-1}{e-1}$ is the number of weak compositions of m into e parts, and $A_k(t^*)$ is a combinatorial factor counting the possible locations of $v_1, \dots, v_k$ in t^*. More precisely, $A_k(t^*)$ is the number of multi-sets of vertices from t^* of size k (with multiplicity) containing all leaves of t^*. In particular, if t^* has k leaves then $A_k(t^*) = 1$. Since the total number of k-marked rooted trees on $[n]$ is n^{n+1-k}, and the number of binary plane trees with k leaves is given by the $(k-1)$'st Catalan number, straightforward approximations then prove Corollary 1.2.

It seems worthwhile to further observe that for any $p \in [1, \infty)$, the collection of laws of the random variables $((n^{-1/2} M(T_n, S_k))^p, n \geq 1)$ forms a uniformly

integrable family. To see this, using the notation of Theorem 1.1, let E_n be number of edges in the subtree of T_n spanned by its root $r = r(T_n)$ plus $V_1, \dots, V_k$. By Theorem 1.1,

$$M(T_n, S_k) \stackrel{d}{=} E_n + k.$$

Furthermore, writing d_n for graph distance in T_n, we have

$$E_n + k \le \sum_{i=1}^{k} d_n(r, V_i) + k.$$

Since $(d_n(r, V_i), 1 \le i \le k)$ are i.i.d. it follows by a union bound that for $x > 0$,

$$\mathbf{P}\big((n^{-1/2} M(T_n, S_k))^p \ge x\big) \le k\mathbf{P}\big(d_n(r, V_1) \ge n^{1/2} x^{1/p}/k - 1\big). \tag{32}$$

But the law of $d_n(r, V_1)$ is well-known (see [38] for an early derivation): we have, for $\ell \ge 1$,

$$\mathbf{P}\big(d_n(r, V_1) \ge \ell\big) = \prod_{j=1}^{\ell-1} \left(1 - \frac{j}{n}\right) \le \exp\left(-\frac{\ell(\ell-1)}{2n}\right). \tag{33}$$

Using (32) and (33), standard manipulations imply that for all $p \ge 1$,

$$\begin{aligned} &\lim_{K\to\infty} \sup_{n\ge 1} \mathbf{E}\big[(n^{-1/2} M(T_n, S_k))^p \mathbf{1}_{\{(n^{-1/2} M(T_n, S_k))^p \ge K\}}\big] \\ &\quad = \lim_{K\to\infty} \sup_{n\ge 1} \sum_{\ell \ge K} \mathbf{P}\big(n^{-1/2} M(T_n, S_k) \ge \ell^{1/p}\big) \\ &\quad = 0. \end{aligned}$$

This establishes the claimed uniform integrability.

Finally, note that convergence in distribution and the above uniform integrability imply that in any space in which (a sequence of random variables with the laws of) $(n^{-1/2} M(T_n, S_k), n \ge 1)$ converges *in probability* to χ_k (a chi random variable with $2k$ degrees of freedom), we additionally have convergence in L^p. This follows by standard arguments, for example, Theorem 13.7 of [49].

APPENDIX B: EXCURSIONS, BRIDGES, TREES AND FORESTS

In this section, we describe the transformations of Section 5 in the language of excursions. This perspective on the results serves two purposes. First, in the excursion framework, a similarity is immediately apparent, between the results of the current paper and results of Aldous and Pitman [11] on scaling limits of random mappings and on decompositions of reflecting Brownian bridge. Though there seems to be no direct link between the main results of the two papers, the idea that they may possess a common strengthening is intriguing. Second, as noted in

the body of Section 5, a careful reader may have had questions about the precision of the definitions of some of the random objects under consideration, and the excursion-theoretic description clarifies such matters.

Let $e = (e(t), 0 \le t \le 1)$ be a standard Brownian excursion, and write $\mathcal{T}_e$ for the $\mathbb{R}$-tree coded by e. (We recall that the points of $\mathcal{T}_e$ are equivalence classes $\{[x], 0 \le x \le 1\}$, where points $x, y \in [0,1]$ are equivalent if $e(x) = e(y) = \inf\{e(z) : x \le z \le y\}$, and refer the reader to [34] for more details of this standard construction.)

Next, let $\mathcal{A}_e = \{(s,y) \in [0,1] \times \mathbb{R}^+ : 0 \le y \le e(s)\}$ be the set of points lying above the x-axis and below the graph of e. For each point (x,y) in A_e^o, the interior of A_e, let

$$\underline{s}(x,y) = \underline{s}(x,y,e) = \inf\{x' : x' \in (0,x), e(z) \ge y \ \forall z \in [x',x]\}$$

and let

$$\bar{s}(x,y) = \bar{s}(x,y,e) = \sup\{x' : x' \in (x,1), e(z) \ge y \ \forall z \in [x,x']\}.$$

In other words, the line segment $[\underline{s}(x,y), \bar{s}(x,y)] \times \{y\}$ is the maximal horizontal line segment through (x,y) contained in $\mathcal{A}_e$.

We wish to obtain an excursion-theoretic representation of the Poisson process on $\mathrm{skel}(\mathcal{T}_e) \times [0,\infty)$ with intensity measure $\ell \otimes \mathrm{Leb}_{[0,\infty)}$, where ℓ is the length measure on $\mathrm{skel}(\mathcal{T}_e)$ and $\mathrm{Leb}_{[0,\infty)}$ is Lebesgue measure on $[0,\infty)$. To do so, for $(x,y) \in \mathcal{A}_e^o$, we view the points of $[\underline{s}(x,y), \bar{s}(x,y)) \times \{y\}$ as representing the point $[\underline{s}(x,y)]$ of $\mathrm{skel}(\mathcal{T}_e)$. We then consider a process $\mathcal{P}_e^\circ$ which, conditional on e, is a Poisson process on $\mathcal{A}_e^o \times [0,\infty)$ with intensity measure at $((x,y),t)$ given by

$$\frac{d\mathrm{Leb}_{\mathcal{A}_e^o} \otimes d\mathrm{Leb}_{[0,\infty)}}{\bar{s}(x,y,e) - \underline{s}(x,y,e)}.$$

For $t \in [0,\infty)$, let

$$X_t = X_t(e, \mathcal{P}_e^\circ) = \{z \in [0,1] : \exists((x,y),s) \in \mathcal{P}_e^\circ, s \le t, z \in [\underline{s}(x,y), \bar{s}(x,y)]\}.$$

In words, the (equivalence classes of) points of X_t are the points of $\mathcal{T}_e$ lying in subtrees that have been cut by $\mathcal{P}_e^\circ$ by time t. We define X_{t-} accordingly, let $Y_t = [0,1] \setminus X_t$ and let $Y_{t-} = [0,1] \setminus X_{t-}$.

Next, for $0 \le t \le \infty$, let $m_t = \mathrm{Leb}_{[0,1]}(Y_t)$ be the Lebesgue measure of the points that are not yet cut at time t, and let $m_{t-} = \mathrm{Leb}_{[0,1]}(Y_{t-})$. Then let $\mathcal{P}_e = \{p = ((x,y),t) \in \mathcal{P}_e^\circ : m_t < m_{t-}\}$ for the set of points that reduce the measure of the "uncut subtree." We next explain how the points of $\mathcal{P}_e$ yield a family of transformations of the excursion e.

For $z \in \overline{Y_t}$, the closure of Y_t, let $v_t(z) = \mathrm{Leb}_{[0,1]}([0,z] \cap Y_t)$. The function $v_t : \overline{Y_t} \to [0, m_t]$ is nondecreasing. Furthermore, the results of [8] imply that $v_t(1) = m_t$ and that for $0 \le z < z' \le 1$ we have $v_t(z) = v_t(z')$ if and only if there exists $(x,y) \in \mathcal{A}_e^o$ such that $z = \underline{s}(x,y)$ and $z' = \bar{s}(x,y)$. In other words, $v_t(z) = v_t(z')$ precisely if $[z] = [z']$ is the root of a subtree that is cut before or at time t.

Let $e_t^0 : [1 - m_t, 1] \to [0, \infty)$ be given by setting $e_t^0(z) = e(v_t^{-1}(z - (1 - m_t)))$, where $v_t^{-1}(u) = \inf\{x : v_t(x) \geq u\}$ [we could in fact take $v_t^{-1}(u)$ to be any point in the pre-image of u under v_t; the comments of the preceding paragraph show that the value of $e(v_t^{-1}(u))$ does not depend on this choice]. Then Theorem 4 of [8], together with the comments in Section 3.5 of that paper, implies that conditional on m_t, if the function e_t^0 is translated to have domain $[0, m_t]$ then the result is distributed as a standard Brownian excursion of length m_t. We define m_{t-}, v_{t-} and the excursion e_{t-} similarly.

Next, for each point $p = ((x, y), t)$ of $\mathcal{P}_e$, we define a random function e^p with domain $[1 - m_{t-}, 1 - m_t]$ as follows. For $z \in [1 - m_{t-}, 1 - m_t]$, set

$$e^p(z) = e_{t-}\big(v_{t-}^{-1}\big(\underline{s}(x, y)\big) + z\big).$$

Notice that

$$(1 - m_t) - (1 - m_{t-}) = m_{t-} - m_t = v_{t-}\big(\bar{s}(x, y)\big) - v_{t-}\big(\underline{s}(x, y)\big).$$

Translated to have range $[0, v_{t-}(\bar{s}(x, y)) - v_{t-}(\underline{s}(x, y))]$, the excursion e^p then codes the tree cut by point p under the standard coding of trees by excursions.

Finally, for $t \in [0, \infty)$ let $e_t : [0, 1] \to [0, \infty)$ be the unique function such that $e_t|_{[1-m_t,1]} \equiv e_t^0$ and such that for each $p = (x, y, s) \in \mathcal{P}_e$ with $0 \leq s \leq t$,

$$e_t|_{[1-m_{s-},1-m_s]} \equiv e^p.$$

The function e_t is the "concatenation" of the functions

$$\big\{e^p, p = (x, y, s) \in \mathcal{P}_e : 0 \leq s \leq t\big\}$$

and of the function e_t^0. We define the function e_{t-} similarly. The function e_t is comprised of a countably infinite number of excursions away from zero; the trees coded by these excursions together comprise the $\mathbb{R}$-forest $(\mathcal{F}_t, d_t, \mu_t)$ of Section 5. A similar coding of a random continuum forest, by a reflecting Brownian bridge conditioned on its local time at zero, is described in [8], Section 3.5.

The random variables $(e_t, t \geq 0)$ are consistent in the sense that for any fixed $s \in [0, 1)$, there is an almost surely finite time t_0 such that for all $t' > t \geq t_0$, $e_{t'}|_{[0,s]} = e_t|_{[0,s]}$. It follows that the limit $e_\infty = \lim_{t\to\infty} e_t$ is almost surely well-defined.

In the current terminology, for $0 \leq t \leq \infty$, we have

$$L(t) = \int_0^t m_s \, ds.$$

We view $e_t|_{[0,1-m_t]} = e_\infty|_{[0,1-m_t]}$ as coding a random measured $\mathbb{R}$-tree with mass $1 - m_t$, as follows. Let $d_t^* : [0, 1 - m_t] \to [0, \infty)$ be given by setting

$$d_t^*(u, v) = e_t(v) + e_t(u) - 2 \inf_{u \leq s \leq v} e_t(s) + L(s_v) - L(s_u),$$

for all $0 \leq u \leq v \leq 1 - m_t$ such that there exist $s_u, s_v \in [0, t]$ for which $u \in [1 - m_{s_u-}, 1 - m_{s_u})$ and $v \in [1 - m_{s_v-}, 1 - m_{s_v})$.

Then the tree $(\widehat{\mathcal{T}}_t, \hat{d}_t, \hat{\mu}_t)$ of Section 5 may be defined as follows. Set $\widehat{\mathcal{T}}_t = \{[u], 0 \le u \le 1 - m_t\}$, where $[u]$ denotes the equivalence class of u : $[u] = \{0 \le v \le 1 - m_t : d_t^*(u, v) = 0\}$. Let $\hat{d}_t$ be the push-forward of d_t^* to $\widehat{\mathcal{T}}_t$, and let $\hat{\mu}_t$ be the push-forward of Lebesgue measure on $[0, 1 - m_t]$ to $\widehat{\mathcal{T}}_t$.

The content of the first assertion of Theorem 5.1 is that e_∞ is distributed as a reflecting Brownian bridge; we may see the equivalence between the first part of Theorem 5.1 and the latter statement as follows. First, a standard and trivial extension of Theorem 5.2, states that a uniformly random doubly-marked tree on $[n]$ converges to $(\mathcal{T}, d, \mu, (\rho, \rho'))$ with respect to d_{GHP}^2, where $(\mathcal{T}, d, \mu)$ is a Brownian CRT and ρ, ρ' are independent elements of $\mathcal{T}$, each with law μ. Next, recall the standard one-to-one map between doubly-marked trees on $[n]$ and ordered rooted forests on $[n]$ which "removes the edges on the path between the two marked vertices." Finally, results from [11]—in particular, the first two distributional convergence results in Theorem 8 of that paper, together with the remark in Section 10—imply that that the contour process of a uniformly random ordered rooted forest on $[n]$ converges after appropriate rescaling to a reflecting Brownian bridge. (We remark that a direct encoding of a doubly-rooted Brownian CRT by reflecting Brownian bridge, also mentioned in the Introduction, is given in [15]. The latter is closely related to, but distinct from, the encoding obtained by considering ordered rooted forests as above.)

Next, for each point $p = ((x, y), t) \in \mathcal{P}_e$, let

$$u_p = 1 - m_t + v_t\big(\underline{s}(x, y, e)\big) \in [1 - m_t, 1].$$

If we view e_t^0 as coding a tree, then the (equivalence class of the) point u_p is a leaf of this tree. Then let $y_p = y_p(e, \mathcal{P}_e)$ be the push-forward of u_p under the map that sends $e_t \to e_\infty$. In other words, let $p' = ((x', y'), t')$ be the a.s. unique point of $\mathcal{P}_e$ with $t' > t$, with $\underline{s}(x', y', e) < \underline{s}(x, y, e)$, with $\bar{s}(x', y', e) > \bar{s}(x, y, e)$, and minimizing t' subject to these constraints. Then we set

$$y_p(e, \mathcal{P}_e) = 1 - m_{t'-} + v_{t'-}\big(\underline{s}(x, y, e)\big) - v_{t'-}\big(\underline{s}(x', y', e)\big).$$

The second assertion of Theorem 5.1 is that conditional on e_∞, the law of $\{y_p, p \in \mathcal{P}_e\}$ is the same as that of the following family of random variables. Let $Z = \{z \in (0, 1) : e_\infty(z) = 0\}$. Then independently for each $z \in Z$ let Y_z be uniform on $[z, 1]$.

We remark that a related family of random variables plays a role in Theorem 8 of [11] (in particular in the third distributional convergence of that theorem). The latter theorem, which describes a distributional limit for uniformly random mappings of $[n]$, has several suggestive similarities to our main result. We do not see any direct relation between the distributional limits described in that paper and those established here. Establishing such a relation would certainly be of interest, and would likely yield insights in both the discrete and limiting settings.

REFERENCES

[1] ABRAHAM, R. and DELMAS, J.-F. (2013). The forest associated with the record process on a Lévy tree. *Stochastic Process. Appl.* **123** 3497–3517. MR3071387

[2] ABRAHAM, R. and DELMAS, J.-F. (2013). Record process on the continuum random tree. *ALEA Lat. Am. J. Probab. Math. Stat.* **10** 225–251. MR3083925

[3] ADDARIO-BERRY, L., BROUTIN, N. and HOLMGREN, C. (2010). Cutting down trees with a Markov chainsaw (with online slides). YEP VII Seminar (March 2010). http://www.eurandom.tue.nl/events/workshops/2010/YEPVII/YEVIIAbstracts.htm.

[4] ALDOUS, D. (1991). The continuum random tree. II. An overview. In *Stochastic* (*Durham*, 1990) (M. T. Barlow and N. H. Bingham, eds.) 23–70. Cambridge Univ. Press, Cambridge. MR1166406

[5] ALDOUS, D. (1991). Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.* **1** 228–266. MR1102319

[6] ALDOUS, D. (1991). The continuum random tree. I. *Ann. Probab.* **19** 1–28. MR1085326

[7] ALDOUS, D. (1993). The continuum random tree. III. *Ann. Probab.* **21** 248–289. MR1207226

[8] ALDOUS, D. and PITMAN, J. (1998). The standard additive coalescent. *Ann. Probab.* **26** 1703–1726. MR1675063

[9] ALDOUS, D. and STEELE, J. M. (2003). The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Discrete and Combinatorial Probability* (H. Kesten, ed.) 1–72. Springer, Berlin.

[10] ALDOUS, D. J. (1990). The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discrete Math.* **3** 450–465. MR1069105

[11] ALDOUS, D. J. and PITMAN, J. (1994). Brownian bridge asymptotics for random mappings. *Random Structures Algorithms* **5** 487–512. MR1293075

[12] BERTOIN, J. (2006). *Random Fragmentation and Coagulation Processes*. Cambridge Univ. Press, Cambridge. MR2253162

[13] BERTOIN, J. (2012). Fires on trees. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **48** 909–921. MR3052398

[14] BERTOIN, J. and MIERMONT, G. (2013). The cut-tree of large Galton–Watson trees and the Brownian CRT. *Ann. Appl. Probab.* **23** 1469–1493. MR3098439

[15] BERTOIN, J. and PITMAN, J. (1994). Path transformations connecting Brownian bridge, excursion and meander. *Bull. Sci. Math.* **118** 147–166. MR1268525

[16] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

[17] BRODER, A. (1989). Generating random spanning trees. In 30*th Annual Symposium on Foundations of Computer Science* 442–447. IEEE, New York.

[18] BURAGO, D., BURAGO, Y. and IVANOV, S. (2001). *A Course in Metric Geometry*. *Graduate Studies in Mathematics* **33**. Amer. Math. Soc., Providence, RI. MR1835418

[19] CHASSAING, P. and MARCHAND, R. (2009). Personal communication.

[20] DALEY, D. J. and VERE-JONES, D. (2007). *An Introduction to the Theory of Point Processes*. *Vol. II*: *General Theory and Structure*. Springer, New York.

[21] DEVROYE, L. and JANSON, S. (2011). Distances between pairs of vertices and vertical profile in conditioned Galton–Watson trees. *Random Structures Algorithms* **38** 381–395. MR2829308

[22] DRMOTA, M., IKSANOV, A., MOEHLE, M. and ROESLER, U. (2009). A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. *Random Structures Algorithms* **34** 319–336. MR2504401

[23] FILL, J. A., KAPUR, N. and PANHOLZER, A. (2006). Destruction of very simple trees. *Algorithmica* **46** 345–366. MR2291960

[24] HAAS, B. and MIERMONT, G. (2012). Scaling limits of Markov branching trees with applications to Galton–Watson and random unordered trees. *Ann. Probab.* **40** 2589–2666. MR3050512

[25] HOLMGREN, C. (2008). Random records and cuttings in split trees: Extended abstract. In *Fifth Colloquium on Mathematics and Computer Science*. 269–281. Assoc. Discrete Math. Theor. Comput. Sci., Nancy. MR2508793

[26] IKSANOV, A. and MÖHLE, M. (2007). A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Commun. Probab.* **12** 28–35. MR2407414

[27] JANSON, S. (2004). Random records and cuttings in complete binary trees. In *Mathematics and Computer Science. III*: *Algorithms*, *Trees*, *Combinatorics and Probability* (*Vienna*) (M. Drmota, P. Flajolet, D. Gardy and B. Gittenberger, eds.) 241–253. Birkhäuser, Basel. MR2090513

[28] JANSON, S. (2006). Random cutting and records in deterministic and random trees. *Random Structures Algorithms* **29** 139–179. MR2245498

[29] KENNEDY, D. P. (1976). The distribution of the maximum Brownian excursion. *J. Appl. Probab.* **13** 371–376. MR0402955

[30] KESTEN, H. (1986). Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. Henri Poincaré Probab. Stat.* **22** 425–487. MR0871905

[31] KOLCHIN, V. F. (1986). *Random Mappings*. Optimization Software Inc. Publications Division, New York. MR0865130

[32] KUBA, M. and PANHOLZER, A. (2008). Isolating a leaf in rooted trees via random cuttings. *Ann. Comb.* **12** 81–99. MR2401138

[33] KUBA, M. and PANHOLZER, A. (2008). Isolating nodes in recursive trees. *Aequationes Math.* **76** 258–280. MR2461893

[34] LE GALL, J.-F. (2005). Random trees and applications. *Probab. Surv.* **2** 245–311. MR2203728

[35] LE GALL, J.-F. (2006). Random real trees. *Ann. Fac. Sci. Toulouse Math.* (6) **15** 35–62. MR2225746

[36] LYONS, R. and PERES, Y. (2012). Probability on trees and networks. Unpublished manuscript.

[37] MCDIARMID, C. (1998). Concentration. In *Probabilistic Methods for Algorithmic Discrete Mathematics* (M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed, eds.) 195–248. Springer, Berlin. MR1678578

[38] MEIR, A. and MOON, J. W. (1970). The distance between points in random trees. *J. Combin. Theory* **8** 99–103. MR0263685

[39] MEIR, A. and MOON, J. W. (1970). Cutting down random trees. *J. Aust. Math. Soc.* **11** 313–324. MR0284370

[40] MEIR, A. and MOON, J. W. (1974). Cutting down recursive trees. *Math. Biosci.* **21** 173–181.

[41] MEIR, A. and MOON, J. W. (1978). On the altitude of nodes in random trees. *Canad. J. Math.* **30** 997–1015. MR0506256

[42] MIERMONT, G. (2009). Tessellations of random maps of arbitrary genus. *Ann. Sci. Éc. Norm. Supér.* (4) **42** 725–781. MR2571957

[43] PANHOLZER, A. (2003). Noncrossing trees revisited: Cutting down and spanning subtrees. In *Discrete Random Walks* (*Paris*, 2003) 265–276 (electronic). Assoc. Discrete Math. Theor. Comput. Sci., Nancy. MR2042393

[44] PANHOLZER, A. (2006). Cutting down very simple trees. *Quaest. Math.* **29** 211–227. MR2233368

[45] PITMAN, J. (2006). *Combinatorial Stochastic Processes*. *Lecture Notes in Math.* **1875**. Springer, Berlin. MR2245368

[46] RIORDAN, J. (1968). Forests of labeled trees. *J. Combin. Theory* **5** 90–103. MR0228386

[47] STROOCK, D. W. (2011). *Probability Theory. An Analytic View*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2760872

[48] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. Springer, Berlin. MR2459454
[49] WILLIAMS, D. (1991). *Probability with Martingales*. Cambridge Univ. Press, Cambridge. MR1155402

L. ADDARIO-BERRY
DEPARTMENT OF MATHEMATICS AND STATISTICS
MCGILL UNIVERSITY
805 SHERBROOKE STREET WEST
MONTREAL, CANADA, H3A 0B9
E-MAIL: louigi@gmail.com

N. BROUTIN
INRIA PARIS–ROCQUENCOURT
DOMAINE DE VOLUCEAU
78153 LE CHESNAY
FRANCE
E-MAIL: nicolas.broutin@inria.fr

C. HOLMGREN
DEPARTMENT OF MATHEMATICS
STOCKHOLM UNIVERSITY
114 18 STOCKHOLM
SWEDEN
E-MAIL: cecilia@math.su.se