THE BROWNIAN CONTINUUM RANDOM TREE AS THE UNIQUE SOLUTION TO A FIXED POINT EQUATION

MARIE ALBENQUE AND CHRISTINA GOLDSCHMIDT

Abstract. In this note, we provide a new characterization of Aldous’ Brownian continuum random tree as the unique fixed point of a certain natural operation on continuum trees (which gives rise to a recursive distributional equation). We also show that this fixed point is attractive.

INTRODUCTION

The Brownian continuum random tree (BCRT), which was introduced and first studied by Aldous [3–5], is the prototypical example of a random continuum random tree. Its importance derives from the fact that it is the scaling limit of a large class of discrete trees including: all critical Galton–Watson trees with finite offspring variance [3, 5], unordered binary trees [17], uniform unordered trees [14], uniform unlabelled unrooted trees [23], critical multitype Galton–Watson trees [18], random trees with a prescribed degree sequence satisfying certain conditions [8] and random dissections [10].

Many of these convergence results are proved using some sort of functional coding. However, particularly in the case of unordered trees, a natural functional coding whose distributional properties are easily understood is not always available. In such settings, an alternative approach is desirable.

By a recursive distributional equation for a random variable $X$ taking values in some Polish space $S$, we mean an equation of the form

$$X \overset{d}{=} f((\xi_i, X_i), i \geq 1),$$

where $X_1, X_2, \ldots$ are i.i.d. copies of $X$, independent of the family $(\xi_i)_{i \geq 1}$, and $f$ is a suitable $S$-valued function. We can, of course, think of this equation in terms of probability distributions: if $\mu$ is the distribution of $X$ and $F(\mu)$ is the distribution of the right-hand side then $\mu$ is a fixed point of the operator $F$.

For families of random variables which satisfy a natural recursive distributional equation, the so-called contraction method has been demonstrated to be a powerful tool for proving convergence results. Suppose that $(M_n)$ is a sequence of distributions for which we wish to prove that there exists a limit $M$. The basic idea is as
follows. Suppose that $M_n$ can be described recursively in terms of $M_m$ for $m < n$. This equation often allows one to guess a limiting version, in which $M$ is described in terms of itself. In other words, $M$ should be the fixed point of some operator $\mathfrak{F}$. Suppose that, in addition, $\mathfrak{F}$ is a contraction in a suitable metric on the space of probability measures. Then Banach’s fixed point theorem tells us that there exists a fixed point and that $M_n \to M$ as $n \to \infty$ in the sense of that metric. 

This is straightforward in principle, but usually the recursive equation for $M_n$ does not have precisely the same form as the limiting operator. Moreover, finding a metric in which $\mathfrak{F}$ is a contraction (but which also yields weak convergence) is often highly non-trivial. In practice, this method has been applied very successfully for sequences of random variables (see, for example, [21, 22, 19]), but so far there is only one result for the more complicated setting of convergence of stochastic processes [20].

It is often the case that families of discrete trees have a recursive definition or description. Aldous [6] proved that the BCRT is a fixed point for a natural operation on continuum trees. With these two facts in mind, it is natural to ask if a contraction method can be established for random trees. This seems an ambitious aim, and there are several technical issues to be overcome (not least the choice of metric). But our original motivation stems from the fact that, if such a principle were to be established, then the characterization of possible limits should be the first step. In this article, we prove that the BCRT is the unique fixed point of an appropriate operator, and that this fixed point is attractive for a certain natural class of measures on continuum trees.

The rest of this note is organised as follows. In Section 1, we provide an overview of the various definitions of the BCRT which already exist in the literature. This also enables us to introduce various concepts we will need in the sequel. We then set up our fixed point equation. In Section 2, we prove that it has a unique solution. In Section 3, we show that repeatedly applying the fixed point operator to any suitable law on continuum trees gives convergence to the law of the BCRT in the sense of the Gromov–Prokhorov topology. Section 4 contains some concluding remarks.

1. Overview of definitions of the BCRT

We begin by introducing the notion of an $\mathbb{R}$-tree.

**Definition 1.1.** A compact metric space $(T, d)$ is a $\mathbb{R}$-tree if for all $x, y \in T$

- there exists a unique geodesic from $x$ to $y$ i.e. there exists a unique isometry $f_{x,y} : [0, d(x, y)] \to T$ such that $f_{x,y}(0) = x$ and $f_{x,y}(d(x, y)) = y$. The image of $f_{x,y}$ is called $[[x, y]]$;
- the only non-self-intersecting path from $x$ to $y$ is $[[x, y]]$ i.e. if $q : [0, 1] \to T$ is continuous and injective and such that $q(0) = x$ and $q(1) = y$ then $q([0, 1]) = [[x, y]]$.

An element $x \in T$ is called a vertex. A rooted $\mathbb{R}$-tree is an $\mathbb{R}$-tree $(T, d)$ with a distinguished vertex $\rho$ called the root. The height of a vertex $x$ is $d(\rho, x)$. The
degree \( \deg(x) \) of a vertex \( x \) is the number of connected components of \( T \setminus \{ x \} \). By a leaf, we mean a vertex of degree 1; write \( \mathcal{L}(T) \) for the set of leaves of \( T \). The tree \( T \) is leaf-dense if \( T \) is the closure of \( \mathcal{L}(T) \). We will often want to endow an \( \mathbb{R} \)-tree with a probability measure \((\mu, \text{say})\), which allows us to pick random points in the tree.

A measured metric space \((X, d, \mu)\) is a complete metric space \((X, d)\) equipped with a Borel probability measure \(\mu\) on \((X, d)\). Define a first equivalence relation by declaring two such spaces \((X, d, \mu)\) and \((X', d', \mu')\) to be GHP-equivalent if there exists an isometry \( f : X \to X' \) such that the image of \( \mu \) under \( f \) is \( \mu' \). Let \( S \) denote the space of GHP-equivalence classes of compact measured metric spaces.

Then \( S \) is Polish when endowed with the Gromov–Hausdorff–Prokhorov topology \([1]\). Define a second equivalence relation by declaring \((X, d, \mu)\) and \((X', d', \mu')\) to be GP-equivalent if there exists an isometry \( g : \text{supp}(\mu) \to X' \) such that the image of \( \mu \) under \( f \) is \( \mu' \), where \( \text{supp}(\mu) \) denotes the topological support of \( \mu \). Let \( S' \) denote the space of GP-equivalence classes of compact measured metric spaces. Then \( S' \) is Polish when endowed with the Gromov–Prokhorov topology \([12]\).

1.1. The BCRT as an \( \mathbb{R} \)-tree encoded by a Brownian excursion. A standard way to generate \( \mathbb{R} \)-trees is via functional encoding. Suppose that \( h : [0, \infty) \to [0, \infty) \) is a continuous function of compact support such that \( h(0) = 0 \). Use it to define a pseudo-metric \( \tilde{d} \) by
\[
\tilde{d}(x, y) = h(x) + h(y) - 2 \inf_{x \land y \leq t \leq x \lor y} h(t), \quad x, y \geq 0.
\]
Define an equivalence relation \( \sim \) by letting \( x \sim y \) if \( \tilde{d}(x, y) = 0 \). Let \( T = [0, \infty)/\sim \), denote by \( \tau : [0, \infty) \to T \) the canonical projection and let \( d \) be the metric induced on \( T \) by \( \tilde{d} \). If \( \sigma \) is the supremum of the support of \( h \) then note that \( \tau(s) = 0 \) for all \( s \geq \sigma \). This entails that \( T = \tau([0, \sigma]) \) is compact. The metric space \((T, d)\) can then be shown to be an \( \mathbb{R} \)-tree (see Le Gall \([15]\)). The tree \( T \) can be naturally rooted at \( \rho = \tau(0) \), the equivalence class of 0, and we will sometimes think of it as a rooted object and sometimes not. There is a natural measure \( \mu \) on \( T \) given by the push-forward of the uniform distribution on \([0, \sigma]\) under the projection \( \tau \).

We define the BCRT \((T, \_\_\_\_)\) to be the \( \mathbb{R} \)-tree encoded by
\[
h(t) = \begin{cases} 
2e(t) & 0 \leq t \leq 1, \\
0 & t > 1,
\end{cases}
\]
where \((e(t), 0 \leq t \leq 1)\) is a standard Brownian excursion. We usually endow \((T, \_\_\_\_)\) with the probability measure \(\mathfrak{m}\) which is the push-forward of the Lebesgue measure on \([0, 1]\).

1.2. The BCRT as a limit of discrete trees. Let \( T_n \) be the ordered rooted tree representing the genealogy of a Galton–Watson branching process with offspring distribution having mean 1 and variance \(\sigma^2 \in (0, \infty)\). Think of \( T_n \) as a metric space by endowing it with the graph distance \(d_{\text{gr}}\) (which puts neighbouring vertices
at distance 1). Let $\mu_n$ be the uniform measure on the vertices of $T_n$. Then

$$(T_n, n^{-1/2}d_{gr}, \mu_n) \overset{d}{\to} (T, \sigma^{-1}m)$$

as $n \to \infty$, in the Gromov–Hausdorff–Prokhorov sense. (This convergence in distribution is originally due to Aldous [3], although this formulation is closer to that of Le Gall [16].)

### 1.3. The BCRT via random finite-dimensional distributions.

We may also characterize the BCRT as the unique continuum random tree having certain distributional properties. We must first introduce properly what we mean by a continuum tree.

**Definition 1.2.** A continuum tree is a triple $(T,d,\mu)$ where $(T,d)$ is an (unrooted) $\mathbb{R}$-tree and $\mu$ is a probability measure on $T$ which is non-atomic and satisfies

- $\mu(L(T)) = 1$;
- for every $v \in T$ of degree $k = \deg(v) \geq 2$, let $T_1, \ldots, T_k$ be the connected components of $T \setminus \{v\}$; then $\mu(T_i) > 0$ for all $1 \leq i \leq k$.

**Definition 1.3.** A continuum random tree (CRT) is a random variable taking values in the set of continuum trees.

(In [5], Aldous makes slightly different definitions of these quantities which, in particular, use rooted trees.) It will be important in the sequel to observe that, if we consider the BCRT to be rooted at the equivalence class of 0 in the Brownian excursion construction, then the root has the same distribution as a uniform pick from $m$ on $T$.

Give a CRT $(T,d,\mu)$, let $V_1, V_2, \ldots$ be i.i.d. samples from the measure $\mu$. For $m \geq 2$, define the reduced tree $R(m)$ to be the subtree of $T$ spanned by $V_1, V_2, \ldots, V_m$. For every $m \geq 2$, $R(m)$ is a discrete tree with edge-lengths and labelled leaves, and so its distribution is specified by its tree-shape $t$, an unrooted tree with $m$ labelled leaves, and its edge-lengths $x_1, x_2, \ldots, x_{2m-3}$. Note that, given the leaf labels, the edges of $t$ can be uniquely identified (and so can be also given labels). The reduced trees are clearly consistent, in that $R(m)$ is a subtree of $R(m+1)$.

**Theorem 1.4** (Aldous [5]). The distribution of a CRT $(T,d,\mu)$ is specified entirely by its random finite dimensional distributions, that is, the distribution of $R(m)$ for all $m \geq 2$.

The reduced trees of the BCRT are binary almost surely. This entails that $R(m)$ has $2m - 2$ vertices and $2m - 3$ edges. Let its tree-shape be $t$ and its edge-lengths be $x_1, x_2, \ldots, x_{2m-3}$. Then $R(m)$ has density

$$f(t; x_1, x_2, \ldots, x_{2m-3}) = \left(\sum_{i=1}^{2m-3} x_i\right) \exp \left( -\frac{1}{2} \left(\sum_{i=1}^{2m-3} x_i\right)^2 \right).$$
Note that this implies that the tree-shape is, in fact, uniform on the set of binary tree-shapes with \( m \) leaves, and that the edge-lengths have an exchangeable distribution. We observe, for future reference, that the distance between two uniformly-chosen points of the BCRT has the Rayleigh distribution, with density \( xe^{-x^2/2} \) and expectation \( \sqrt{\pi/2} \).

(Note that in [5], Aldous restricts his discussion to binary trees, but the theory is easily extended; see Haas and Miermont [13].)

1.4. The BCRT as a fixed point. The principal contribution of this paper is a characterization of the BCRT as the unique fixed point of a certain operation on CRT’s. We need a couple of notational ingredients. We first recall the definition of the Dirichlet distribution.

**Definition 1.5.** A random variable taking values in the space \( \{ s = (s_1, s_2, \ldots, s_n) : s_i \geq 0, 1 \leq i \leq n, \sum_{i=1}^{n} s_i = 1 \} \) has the Dirichlet distribution with parameters \( (\alpha_1, \alpha_2, \ldots, \alpha_n) \) (written \( \text{Dir}(\alpha_1, \alpha_2, \ldots, \alpha_n) \)) if it has density

\[
\frac{\Gamma(\sum_{i=1}^{n} \alpha_i)}{\prod_{i=1}^{n} \Gamma(\alpha_i)} x_1^{\alpha_1-1} x_2^{\alpha_2-1} \cdots x_{n-1}^{\alpha_{n-1}-1} (1 - x_1 - x_2 - \cdots - x_{n-1})^{\alpha_n-1}
\]

with respect to \( (n-1) \)-dimensional Lebesgue measure.

Let \( \mathcal{M} \) be the set of probability distributions on (GHP-equivalence classes of) measured \( \mathbb{R} \)-trees. We define \( \mathfrak{F} : \mathcal{M} \rightarrow \mathcal{M} \) as follows: for \( M \in \mathcal{M} \),

- Sample independent trees \( (T_1, d_1, \mu_1) \), \( (T_2, d_2, \mu_2) \), \( (T_3, d_3, \mu_3) \) having distribution \( M \);
- For \( 1 \leq i \leq 3 \), pick a vertex \( X_i \in T_i \) according to the measure \( \mu_i \);
- Sample \( \Delta = (\Delta_1, \Delta_2, \Delta_3) \sim \text{Dir}(1/2, 1/2, 1/2) \) independently;
- Rescale the trees to obtain \( (T_1, \Delta_1^{1/2}d_1, \Delta_1\mu_1) \), \( (T_2, \Delta_2^{1/2}d_2, \Delta_2\mu_2) \), \( (T_3, \Delta_3^{1/2}d_3, \Delta_3\mu_3) \);
- Identify the vertices \( X_1, X_2 \) and \( X_3 \) in the rescaled trees to obtain a single larger tree \( (T^\circ, d) \) with a marked branch-point; the three measures \( \Delta_1\mu_1, \Delta_2\mu_2 \) and \( \Delta_3\mu_3 \) naturally give rise to a (probability) measure \( \mu \) on \( T^\circ \);
- Forget the marked branch-point in order to obtain \( (T, d, \mu) \); \( \mathfrak{F}(M) \) is the distribution of \( (T, d, \mu) \).
The operation on trees given by the function $\mathfrak{F}$ was first described by Aldous [6]. Let $M$ be the law of the BCRT. Theorem 2 of [6] implies, when rephrased in our terms, that $M$ is a solution of $M = \mathfrak{F}(M)$. Actually, what is shown in [6] is the following statement of the reverse of this construction: take a BCRT $(\mathcal{T}, \mathcal{m})$ and pick three points independently according to $\mathcal{m}$; the paths between pairs of these points intersect in a unique branch-point. Splitting at this branch-point then gives three BCRT’s, which have been randomly rescaled by $(\Delta_1, \Delta_2, \Delta_3)$ and depend on one another only through this rescaling. Moreover the former branch-point yields a point chosen independently from the mass measure of each of the three subtrees. (An expanded proof of Aldous’ Theorem 2 may be found in [2].) We will comment on this reversed perspective at the end of the paper.

Let $M$ be a solution to the fixed point equation. Write $(\Omega, \mathcal{F}, \mathbb{P})$ for the probability space on which all the forthcoming random objects are defined. In particular, under $\mathbb{P}$, let $(\mathcal{T}, d, \mu)$ be a continuum random tree sampled from the distribution $M$. Let $\alpha > 0$ be such that $\alpha\sqrt{\pi/2}$ is the expected distance between two points sampled independently from $\mathcal{T}$ according to $\mu$.

The first main result of this article is the following theorem, which is proved in the next section.

**Theorem 1.6.** Suppose that $M$ is a law on continuum trees which is a fixed point of $\mathfrak{F}$. Then $M = M$, the law of the BCRT.

Before going further, we will briefly discuss the requirement that $M$ be a measure on continuum trees. Recall that the BCRT $(\mathcal{T}, \mathcal{m})$ is encoded by $(2e(t), 0 \leq t \leq 1)$, where $e$ is a standard Brownian excursion. Consider an independent Poisson point process (PPP) on $[0, 1] \times [0, \infty)$ with intensity $ds \otimes x^{-3}dx$. For each point $(s, x)$ of the PPP graft a massless branch of length $x$ to the point of $\mathcal{T}$ corresponding to $s$ under the canonical projection $\tau$ (note that $\tau(s)$ is almost surely a leaf). As there are almost surely only finitely many of these branches having length longer than any $\epsilon > 0$, this construction yields a compact metric space and, therefore, induces a probability distribution on the set of measured $\mathbb{R}$-trees. A simple computation shows that this distribution is a solution of the fixed point equation which is clearly not isometric to the BCRT. However, it seems reasonable to want to exclude such non-continuum tree-valued solutions.

Our second main result is as follows.

**Theorem 1.7.** Suppose that $M$ is a law on continuum trees such that if $(\mathcal{T}, d, \mu) \sim M$ and, given $(\mathcal{T}, d, \mu)$, $V_1, V_2$ are sampled independently from $\mu$, then $\mathbb{E}[d(V_1, V_2)]$ exists and is equal to $\sqrt{\pi/2}$. Let $M_n = F^n M$. Then

$$M_n \rightarrow M$$

as $n \rightarrow \infty$, in the sense of weak convergence of measures with the Gromov–Prohorov topology.

Note that if $\mathbb{E}[d(V_1, V_2)] = \alpha\sqrt{\pi/2}$ for some $\alpha \neq 1$ then the same result holds on rescaling the metric $d$ by $\alpha^{-1}$. Theorem 1.7 is proved in Section 3.
2. Uniqueness of the fixed point: proof of Theorem 1.6

We will prove Theorem 1.6 via random finite-dimensional distributions and Theorem 1.4. We start by thinking about the distance between two uniformly-chosen points. Throughout this section, we suppose that $M$ is a measure on continuum trees which is a fixed point of $\mathcal{S}$, and that $(T, d, \mu)$ is sampled from $M$. We write $\mathcal{S}(m), m \geq 2$ for the reduced trees of $(T, \alpha^{-1}d, \mu)$.

2.1. Two-point distances.

**Proposition 2.1.** The distance between two points of $(T, \alpha^{-1}d)$ sampled independently according to $\mu$ has the Rayleigh distribution.

**Proof.** Suppose that $(T_1, d_1, \mu_1), (T_2, d_2, \mu_2)$ and $(T_3, d_3, \mu_3)$ are sampled independently from $M$. Apply $\mathcal{S}$ with $\Delta = (\Delta_1, \Delta_2, \Delta_3)$ to obtain a new tree $(T, d, \mu) \sim M$. Suppose now that we sample two points independently according to $\mu$. Let $P_1, P_2$ and $P_3$ be the number of points falling in the subtrees of $T$ corresponding to $T_1, T_2$ and $T_3$ respectively. Then, conditional on $\Delta$, we have $(P_1, P_2, P_3) \sim \text{Multinomial}(2;\Delta_1, \Delta_2, \Delta_3)$. Let $I_k = \mathbb{1}_{\{P_k > 0\}}, k = 1, 2, 3$ and let $D$ be the distance between the two points. Then

$$D = \sqrt{\Delta_1 I_1} D_1 + \sqrt{\Delta_2 I_2} D_2 + \sqrt{\Delta_3 I_3} D_3,$$

where $D_1, D_2$ and $D_3$ are three independent copies of $D$, independent of everything else on the right-hand side, corresponding to the distances between two uniformly-chosen points in each of the three subtrees. Let $W_k = \sqrt{\Delta_k I_k}, k = 1, 2, 3$. Then this is precisely the setting of the smoothing transform studied by Durrett and Liggett [11]. In that paper, it is shown that the nature of the family of solutions to such distributional fixed point equations depends on the analytic properties of a certain function depending on the moments of $W_1, W_2, W_3$ for $s \geq 0$, let

$$\nu(s) = \log \left( \sum_{k=1}^{3} E \left[ W_k^s \mathbb{1}_{\{W_k > 0\}} \right] \right).$$

By symmetry, $\nu(s) = \log \left( 3E \left[ W_1^s \mathbb{1}_{\{W_1 > 0\}} \right] \right)$. Now, $\mathbb{P}(I_1 = 1|\Delta_1) = \Delta_1^2 + 2\Delta_1(1 - \Delta_1) = 2\Delta_1 - \Delta_1^2$. Since $\Delta_1 \sim \text{Beta}(1/2, 1)$, we obtain

$$E \left[ W_1^s \mathbb{1}_{\{W_1 > 0\}} \right] = 2E \left[ \Delta_1^{s+1} \right] - E \left[ \Delta_1^{s+2} \right] = \frac{2}{s+3} - \frac{1}{s+5}.$$

Hence,

$$\nu(s) = \log \left( \frac{3(s+7)}{(s+3)(s+5)} \right),$$

which is finite for all $s \geq 0$ and has its unique zero in $s \geq 0$ at $s = 1$. Moreover, $\nu'(1) = -7/24 < 0$. Theorems 1 and 2 of [11] then entail that the equation (3) has a unique fixed point, up to a constant scaling factor. Finally, the distance between two uniformly chosen points in a BCRT has the Rayleigh distribution and that must be a solution to (3). Since $E[D] = \alpha \sqrt{\pi/2}$ and the Rayleigh distribution has mean $\sqrt{\pi/2}$, the scaling factor must be $\alpha$. \qed
For future reference, we write \( \mathcal{F}_{3n} \) for the operator which takes the law of a non-negative real-valued random variable \( D' \) and associates to it the law of
\[
\sqrt{X_1} I_1 D'_1 + \sqrt{X_2} I_2 D'_2 + \sqrt{X_3} I_3 D'_3,
\]
where \( D'_1, D'_2 \) and \( D'_3 \) are three independent copies of \( D' \), independent of everything else on the right-hand side, and where \( I_1, I_2, I_3 \) are as above.

### 2.2. A coupling

Having determined the distribution of \( S(2) \) (which, of course, has trivial tree-shape), we now want to determine the distribution of the reduced trees \( S(m), m \geq 3 \). In order to do so, we proceed by coupling a tree \( T \) distributed according to \( \mathcal{M} \) and a realisation \( \hat{T} \) of the BCRT, using the operator \( \mathcal{F} \). Before we can describe this coupling, we need to establish some notation. Let \( \Sigma = \cup_{i=0}^{\infty} \{1, 2, 3\}^i \) be the set of words on the alphabet \( \{1, 2, 3\} \), where, by convention, \( \{1, 2, 3\}^0 = \emptyset \) is the empty word. Let \( \Sigma_n = \cup_{i=0}^{n} \{1, 2, 3\}^i \) be the set of words with at most \( n \) letters. For \( i \in \Sigma \), write \( |i| \) for the length of the word \( i \). For \( 1 \leq m \leq |i| \), write \( i_m \) for the \( m \)th letter of \( i \) and \( i_{[m]} = i_1 \ldots i_m \) for the prefix consisting of the first \( m \) letters of \( i \).

Fix \( n \geq 0 \) and start from a family \( (T_i)_{i \in \{1, 2, 3\}^{n+1}} \) of \( 3^{n+1} \) independent copies of \( T \), and a family \( (\hat{T}_i)_{i \in \{1, 2, 3\}^{n+1}} \) of \( 3^{n+1} \) independent copies of the BCRT. We will refer to these as the input trees and will use them and successive applications of \( \mathcal{F} \) in order to build the trees \( T = T_\emptyset \) and \( \hat{T} = \hat{T}_\emptyset \). At each application of \( \mathcal{F} \), we will use the same scaling factors and glue together subtrees with the same labels. More precisely, let \( (\Delta^{(i)})_{i \in \Sigma_n} \) and \( (U^{(i)})_{i \in \Sigma_n} \) be independent families of independent random variables where, for each \( i \in \Sigma_n \), \( \Delta^{(i)} = (\Delta^{(i)}, \Delta^{(i)}_2, \Delta^{(i)}_3) \sim \text{Dir}(1/2, 1/2, 1/2) \) and \( U^{(i)} = (U^{(i)}_1, U^{(i)}_2, U^{(i)}_3) \), where \( U^{(i)}_1, U^{(i)}_2 \) and \( U^{(i)}_3 \) are independent uniform random variables on \([0, 1]\).

The families \( (T_i)_{i \in \Sigma_n} \) and \( (\hat{T}_i)_{i \in \Sigma_n} \) are constructed recursively as follows. The tree \( T_i \) (resp. \( \hat{T}_i \)) is constructed by applying \( \mathcal{F} \) to \( T_{i_1}, T_{i_2} \) and \( T_{i_3} \) (resp. \( \hat{T}_{i_1}, \hat{T}_{i_2} \) and \( \hat{T}_{i_3} \)) with scaling factors \( \Delta^{(i)}_1, \Delta^{(i)}_2 \) and \( \Delta^{(i)}_3 \), where we emphasize that the same scaling factors are used to construct both families. In each of the trees \( T_{i_1}, T_{i_2} \) and \( T_{i_3} \) (resp. \( \hat{T}_{i_1}, \hat{T}_{i_2} \) and \( \hat{T}_{i_3} \)), we need to pick a uniform point which tells us where to glue them together (once rescaled) to form \( T_i \) (resp. \( \hat{T}_i \)). But if \( |i| < n \), we will also want to keep track of where these uniform points sit in the trees at level \( n+1 \). We can split this problem into two parts: first finding the label of the subtree at level \( n+1 \) in which a particular uniform point lies, and then finding where precisely within that subtree it sits. We will use the random variables \( (U^{(i)})_{i \in \Sigma_n} \) to determine the label of the subtree, and the exact location of the point is then a uniform pick from that subtree. We will use the same labels in \( T \) and \( \hat{T} \) but independent picks from the respective subtrees chosen.

Let \( \Delta_k = \Delta^{(i)}_k \), for \( 1 \leq k \leq 3 \). For \( i \in \Sigma_n \setminus \{\emptyset\} \) and \( 1 \leq k \leq 3 \), recursively define \( \Delta_{ik} := \Delta_k \Delta^{(i)}_k \). In addition, for \( j \in \Sigma \), write \( \Delta^{(i)}_j := \Delta_{ij}/\Delta_i \). For \( i \in \Sigma_n \), let \( j \in \{1, 2, 3\}^{n-|i|+1} \). Then the probability that a uniform point in \( T_i \) belongs to the subtree \( T_{ij} \) is equal to \( \Delta^{(i)}_j \). For \( i \in \Sigma_n \) and \( 1 \leq k \leq 3 \), we define a word \( L^{(i,k)} \) of length \( n - |i| \) such that \( i k L^{(i,k)} \) represents the label of the input tree at level \( n+1 \).
in which the uniform point sampled in $T_{ik}$ sits. It is convenient to use a random recursive partition of the interval $[0, 1]$ to choose this point. The boundaries of this partition are defined by:

$$B_1^{(i,k)} = 0, \quad B_2^{(i,k)} = \Delta_1^{(ik)}, \quad B_3^{(i,k)} = \Delta_1^{(ik)} + \Delta_2^{(ik)}.$$  

Recursively, for $j \in \Sigma_{n-|i|} \setminus \{\emptyset\}$, let

$$B_{j1}^{(i,k)} = B_j^{(i,k)}, \quad B_{j2}^{(i,k)} = B_j^{(i,k)} + \Delta_{j1}^{(ik)}, \quad B_{j3}^{(i,k)} = B_j^{(i,k)} + \Delta_{j1}^{(ik)} + \Delta_{j2}^{(ik)}.$$
For \( 0 \leq \ell \leq n - |i| - 1 \), if \( L_{[\ell]}^{(i,k)} = j \) then let

\[
L_{[\ell+1]}^{(i,k)} = \begin{cases} 
1 & \text{if } B_{i}^{(1,k)} \leq U_{j}^{(i)} \leq B_{i}^{(2,k)} \\
2 & \text{if } B_{j}^{(2,k)} < U_{j}^{(i)} \leq B_{j}^{(3,k)} \\
3 & \text{if } B_{j}^{(3,k)} < U_{j}^{(i)}
\end{cases}
\]

Observe that the definition of \( L_{[\ell]}^{(i,k)} \) depends only on \((\Delta_{i})_{i\in\Sigma_n}\) and \((U_{j}^{(i)})_{i\in\Sigma_n}\).

So, finally, when we sample the uniform point in \( T_{i,k} \) (resp. \( \tilde{T}_{i,k} \)) needed to create \( T_{i} \) (resp. \( \tilde{T}_{i} \)), the value of \( L_{[\ell]}^{(i,k)} \) gives the index of the input tree in which the point sits. Then, conditionally on this choice, we sample the point uniformly from \( T_{i,k,L_{[\ell]}^{(i,k)}} \) (resp. \( \tilde{T}_{i,k,L_{[\ell]}^{(i,k)}} \)).

### 2.3. The reduced trees.

Now, consider \( S(3) \). Again, in this case, the tree-shape is deterministic. We will show that the lengths of the three branches can each be expressed as sums of rescaled distances between pairs of uniform points. We sample three new independent uniform points from our tree \( T \). If, when we decompose this tree into its three subtrees \( T_{1}, T_{2}, \text{ and } T_{3} \), the three points all happen to fall into different subtrees, then we have determined the branch-point between them (which it is natural to label by \( \emptyset \)). However, if at least two points fall into the same subtree, we must then further decompose that subtree in order to determine the location of the branch-point. Let \( N_{3} \geq 1 \) be the recursion depth we have to look at in order to determine the branch-point between our three points. More generally, let \( N_{m} \) be the depth to which we have to look in order to determine all of the branch-points between \( m \) uniform points.

**Proposition 2.2.** \( N_{m} < \infty \) almost surely for \( m \geq 3 \). In particular, \( N_{3} \sim \text{Geom}(2/35) \).

**Proof.** We proceed by induction on \( m \) and start with the case \( m = 3 \). There are three possibilities for the way in which the three points are distributed amongst the subtrees \( T_{1}, T_{2}, \text{ and } T_{3} \).

1. The three points fall in different subtrees.
2. All three points fall in the same subtree.
3. Two points fall in the same subtree and the remaining point falls in a different subtree.

In case (1), as observed above, the branch-point is necessarily \( \emptyset \). In case (2), we have a new independent copy of the original problem of finding the branch-point between three points chosen uniformly from a copy of \( T \). In case (3), the branch-point we seek is the same as that between \( \emptyset \) and the two uniform points which fell in the same subtree. But \( \emptyset \) is also a uniformly chosen point in that subtree. So again, it remains to find the branch-point between three points chosen uniformly from a copy of \( T \). Indeed, unless case (1) occurs, we recursively obtain a new (independent) copy of the original problem. (See Figure 3 for an illustration.) It follows that \( N_{3} \) has a geometric distribution. The probability that the three points
Figure 3. Finding the branch-point between three uniform points. One point falls in $T_2$ and two fall in $T_3$, so we must further decompose $T_3$. We have an independent copy of the original problem in $T_3$ (where one of the points considered is now $\emptyset$). One of the points falls in $T_{31}$ and the other two in $T_{31}$, so we must further decompose $T_{31}$. One of the points now falls in $T_{313}$ but the two others are in $T_{311}$, so we repeat in $T_{311}$. Finally, the three points fall in different subtrees of $T_{311}$ and so we obtain $N_3 = 4$.

fall in different subtrees at any step is given by

$$6E[\Delta_1 \Delta_2 \Delta_3] = \frac{2}{35}.$$  

For $m \geq 4$, there are again three possibilities for the distribution of the $m$ points amongst the subtrees $T_1$, $T_2$ and $T_3$:

1. At least two points fall in different subtrees from the rest.
2. All $m$ points fall in the same subtree.
3. $m - 1$ points fall in the same subtree and the remaining point falls in a different subtree.

In cases (2) and (3), we obtain again a new copy of the same problem. In case (1), we get two or three independent copies of a problem of strictly smaller size. Since case (1) occurs with positive probability, the result follows by (strong) induction. □

Proposition 2.3. $(S(m), m \geq 2)$ have the same joint distribution as the reduced trees $(R(m), m \geq 2)$ of the BCRT.

Proof. Fix $\epsilon > 0$ and $m \geq 2$. By Proposition 2.2, there exists $n$ sufficiently large that we have $P(N_m \leq n + 1) > 1 - \epsilon$. Consider the trees $T$ and $\tilde{T}$ constructed by the above coupling to recursion depth $n$, so that $T$ is distributed according to $M$ and $\tilde{T}$ according to the law of the BCRT. Consider $m$ points picked uniformly in $T$ and $T'$, where we couple the choice of these points in such a way that they fall
in subtrees with same label in $T$ and $T'$ (this is completely analogous to the way we couple the branch-points with the random variables $U_i$ in the previous section). On the event $\{N_m \leq n+1\}$, each branch-point of $S(m)$ corresponds to a point at which we have glued input trees together. In particular, the shapes of $S(m)$ and of the reduced tree $\hat{S}(m)$ in $\hat{T}$ are the same by construction. Moreover, the lengths of corresponding segments of $S(m)$ and $\hat{S}(m)$ are all made up of sums of scaled distances between pairs of uniform points in trees with the same labels at level $n+1$ and the same scaling factors. (Note that these scaling factors receive appropriate biases from the fact that uniform points have/have not fallen into the corresponding trees, but this affects only the scaling factors and not the underlying trees.) By Proposition 2.1 these have the same law in $T$ and $\hat{T}$. The result follows since $\epsilon$ was arbitrary and $\hat{S}(m) \overset{d}{=} \mathcal{R}(m)$. 

Theorem 1.6 follows.

3. Convergence to the fixed point: proof of Theorem 1.7

For $n \geq 0$, let $(T_n,d_n,\mu_n) \sim \mathfrak{S}^n M$ and, conditionally on $(T_n,d_n,\mu_n)$, let $V_1^n, V_2^n, \ldots$ be i.i.d. points of $T_n$ sampled according to $\mu_n$. Similarly, let $(T,\mathfrak{m}) \sim \overline{M}$ and, conditionally on $(T,\mathfrak{m})$, let $V_1, V_2, \ldots$ be i.i.d. points of $T$ sampled according to $\mathfrak{m}$. Write $\mu_n$ for the law of $d_n(V_1^n, V_2^n)$ and $\mathfrak{m}$ for the law of $(V_1, V_2)$ (which is, of course, Rayleigh). Let $S_n(m)$ be the reduced tree of $T_n$ spanned by $V_1^n, \ldots, V_m^n$ and $\mathcal{R}(m)$ be the reduced tree of $T$ spanned by $V_1, \ldots, V_m$. Convergence in the Gromov–Prokhorov distance is then equivalent to the convergence

$$S_n(m) \overset{d}{\to} \mathcal{R}(m) \quad as \quad n \to \infty$$

for each $m \geq 2$ (see Greven, Pfaffelhuber and Winter [12], or the introduction to Bertoin and Miermont [7]). We will again use the coupling of Subsection 2.2 to prove this. Indeed, for fixed $m$, we must look to recursion depth $N_m$ in order to separate our $m$ uniform points. Work on the event $\{N_m = k\}$, where $k < n$. Then in order to obtain coupled trees distributed as $\overline{M}$ and $M_n$ respectively, we need to “plug in” $3^k$ input trees, sampled according to $\overline{M}$ or $M_{n-k}$, respectively. In order to determine the distances in the reduced tree, we in fact only need to know about the distance between two uniform points in each of the input trees. In the input trees sampled according to $M_{n-k}$, this is $\hat{\mu}_{n-k}$ which is equal to the the $(n-k)$-fold iterate of the smoothing transform $\mathfrak{S}_{sm}$ applied to the law $\hat{\mu}_0$ of the distance between two uniformly sampled points of $T_n$. But, by assumption, $\int_0^\infty x \hat{\mu}_0(dx) = \sqrt{\pi}/2$ and so, for fixed $k$, $\hat{\mu}_{n-k} = \mathfrak{S}_{sm}^n \hat{\mu}_0 \to \mathfrak{m}$ as $n \to \infty$, by Theorem 2(b) of Durrett and Liggett [11] (that all that is required here is that the law $\hat{\mu}_0$ has the same mean as $\mathfrak{m}$ is a consequence of the facts already observed in the proof of Proposition 2.1 that the function $\nu$ has its unique zero in $s \geq 0$ at $s = 1$ and that $\nu'(1) < 0$). The edge-lengths in $S_n(m)$ can then be described as sums of randomly rescaled independent random variables sampled from $\hat{\mu}_{n-k}$. It is clear (since we use the same random scaling factors in order to construct both) that the edge-lengths of $S_n(m)$ converge in distribution to those of $\mathcal{R}(m)$ on the event
\{N_m = k\} for any fixed \(k \geq 1\). In order to conclude the proof of Theorem 1.7, we observe as before that \(P(N_m > K) \to 0\) as \(K \to \infty\), since \(N_m < \infty\) almost surely.

4. Concluding remarks

4.1. Related work. As mentioned in Subsection 1.4, Aldous [6] shows that, in a sense, we can “reverse” the operator \(\mathcal{F}\). Indeed, we can decompose a BCRT by picking three uniform points and splitting at the branch-point between them; we obtain three independent BCRT’s, Brownian-rescaled by \((\Delta_1, \Delta_2, \Delta_3) \sim \text{Dir}(1/2, 1/2, 1/2)\). Each of these subtrees is doubly marked, one mark being the original uniform point and the other being the former branch-point. Moreover, these two marks are independent uniform picks from the relevant subtree. This operation is used recursively by Croydon and Hambly [9] to prove that the BCRT is homeomorphic to a certain deterministic fractal with a random self-similar metric, along with the naturally-associated measure. In the course of their proof, they show (Lemma 10(d) of [9]) that all of the randomness in the BCRT is contained in an i.i.d. family of \text{Dir}(1/2, 1/2, 1/2) scaling factors \((\Delta_i, i \in \Sigma)\).

Although we have referred to this decomposition of the BCRT as the reverse of our operator \(\mathcal{F}\), there is, in fact, a rather subtle difference concerning marking and labelling. The forward version of Croydon and Hambly’s splitting operator acts on \textit{doubly uniformly marked trees} and can be paraphrased as follows: take three independent BCRT’s, \(T_1, T_2, T_3\), each with two independent uniform points, labelled 1 and 2. Rescale these trees according to the appropriate Dirichlet random vector and glue them together at the points labelled 1. Now relabel the point labelled 2 in \(T_1\) by 1, keep the point labelled 2 in \(T_2\) and forget the point labelled 2 in \(T_3\) as well as the branch-point just created. Then this is again a doubly uniformly marked BCRT. This seems to us a much less natural “forward” operation on continuum trees than the one pursued in this paper, but it has the advantage that the recursive decomposition obtained by going backwards does not have any of the labelling issues encountered in Section 2.2. Indeed, in this version there is no randomness in which subtree attaches to which other subtree.

4.2. Convergence. The distributional convergence in Theorem 1.7 is in the sense of the Gromov–Prokhorov distance which, for example, does not distinguish between the BCRT and the BCRT decorated by the independent PPP discussed after Theorem 1.6. In particular, this convergence is equivalent to the convergence in distribution of the random finite dimensional distributions. It would be interesting to find conditions under which the convergence holds instead in the stronger Gromov–Hausdorff–Prokhorov sense; in particular, we would need a certain tightness condition to hold (see Corollary 19 of [5]).

Acknowledgments

The question answered in this paper was (to the best of our knowledge) first raised by Nicolas Curien at the YEP VII workshop at EURANDOM in March 2010. We
would like to thank Louigi Addario-Berry and Luc Devroye for inviting us to the Fifth Annual Workshop on Probabilistic Combinatorics and WVD at the Bellairs Institute of McGill University, Barbados, where we began thinking about it, and the Isaac Newton Institute in Cambridge for its invitation in March-April 2015 which enabled us to complete the paper. We would also like to thank David Croydon for detailed discussions relating to the paper [9] and Ralph Neininger for discussions about the contraction method. C.G.’s research was partially funded by EPSRC Postdoctoral Fellowship EP/D065755/1. M.A. acknowledges the support of the ERC under the agreement “ERC StG 208471 - ExploreMap” and the ANR under the agreement “ANR 12-JS02-001-01 - Cartaplus”.

References

1. R. Abraham, J.-F. Delmas, and P. Hoscheit, A note on the Gromov-Hausdorff-Prokhorov distance between (locally) compact metric measure spaces, Electron. J. Probab. 18 (2013), 1–21. MR 3035742
2. L. Addario-Berry, N. Broutin, and C. Goldschmidt, Critical random graphs: limiting constructions and distributional properties, Electron. J. Probab. 15 (2010), no. 25, 741–775. MR MR1085326 (91i:60024)
3. D. Aldous, The continuum random tree. I, Ann. Probab. 19 (1991), no. 1, 1–28. MR MR1085326 (91i:60024)
4. ______, The continuum random tree. II. An overview, Stochastic analysis (Durham, 1990), London Math. Soc. Lecture Note Ser., vol. 167, Cambridge University Press, Cambridge, 1991, pp. 23–70. MR MR1166406 (93f:60010)
5. ______, The continuum random tree. III, Ann. Probab. 21 (1993), no. 1, 248–289. MR MR1207226 (94c:60015)
6. ______, Recursive self-similarity for random trees, random triangulations and Brownian excursion, Ann. Probab. 22 (1994), no. 2, 527–545. MR MR1288122 (95i:60007)
7. J. Bertoin and G. Miermont, The cut-tree of large Galton-Watson trees and the Brownian CRT, Ann. Appl. Probab. 23 (2013), no. 4, 1469–1493. MR 3098439
8. N. Broutin and J.-F. Marckert, Asymptotics of trees with a prescribed degree sequence and applications, Random Structures Algorithms (to appear), arXiv:1110.5203 [math.PR].
9. D. Croydon and B. Hambly, Self-similarity and spectral asymptotics for the continuum random tree, Stochastic Process. Appl. 118 (2008), no. 5, 730–754.
10. N. Curien, B. Haas, and I. Kortchemski, The CRT is the scaling limit of random dissections, arXiv:1305.3534 [math.PR], 2013.
11. R. Durrett and T.M. Liggett, Fixed points of the smoothing transformation, Probab. Theory Related Fields 64 (1983), no. 3, 275–301.
12. A. Greven, P. Pfaffelhuber, and A. Winter, Convergence in distribution of random metric measure spaces (A-coalescent measure trees), Probab. Theory Related Fields 145 (2009), no. 1-2, 285–322. MR 2520129 (2011c:60008)
13. B. Haas and G. Miermont, The genealogy of self-similar fragmentations with negative index as a continuum random tree, Electron. J. Probab. 9 (2004), 57–97, paper no. 4.
14. ______, Scaling limits of Markov branching trees with applications to Galton-Watson and random unordered trees, Ann. Probab. 40 (2012), no. 6, 2589–2666.
15. J.-F. Le Gall, Random trees and applications, Probab. Surv. 2 (2005), 245–311. MR MR2203728 (2007h:60078)
16. ______, Random real trees, Ann. Fac. Sci. Toulouse Math. 15 (2006), no. 1, 35–62.
17. J.-F. Marckert and G. Miermont, The CRT is the scaling limit of unordered binary trees, Random Structures Algorithms 38 (2010), no. 4, 467–501.
18. G. Miermont, Invariance principles for spatial multitype Galton-Watson trees, Ann. Inst. Henri Poincaré Probab. Stat. 44 (2008), no. 6, 1128–1161. MR 2469338 (2010a:60292)
19. R. Neininger and L. Rüschendorf, A general limit theorem for recursive algorithms and combinatorial structures, Ann. Appl. Probab. 14 (2004), no. 1, 378–418.
20. R. Neininger and H. Sulzbach, On a functional contraction method, Ann. Probab. (to appear), arXiv:1202.1370 [math.PR].
21. U. Rösler, *A fixed point theorem for distributions*, Stochastic Process. Appl. 42 (1992), no. 2, 195–214.
22. U. Rösler and L. Rüschendorf, *The contraction method for recursive algorithms*, Algorithmica 29 (2001), no. 1, 3–33.
23. B. Stufler, *The continuum random tree is the scaling limit of unlabelled unrooted trees*, arXiv preprint [arXiv:1412.6333] (2014).