Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2015, Vol. 51, No. 4, 1342–1368
DOI: 10.1214/14-AIHP628

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Convergence of bi-measure $\mathbb{R}$-trees and the pruning process

Wolfgang Löhr[a,1], Guillaume Voisin[a,b] and Anita Winter[a]

[a] *Fakultät für Mathematik, Universität Duisburg-Essen, Campus Essen, Universitätsstrasse 2, Essen 45141, Germany.*
E-mail: wolfgang.loehr@uni-due.de; anita.winter@uni-due.de
[b] *Laboratoire de Mathématiques, Université Paris Sud, 91405, Orsay, France. E-mail: guillaume.voisin1@u-psud.fr*

Received 19 April 2013; revised 21 January 2014; accepted 3 June 2014

Abstract. In (*Ann. Inst. Henri Poincaré Probab. Stat.* **34** (1998) 637–686) a tree-valued Markov chain is derived by pruning off more and more subtrees along the edges of a Galton–Watson tree. More recently, in (*Ann. Probab.* **40** (2012) 1167–1211), a continuous analogue of the tree-valued pruning dynamics is constructed along Lévy trees. In the present paper, we provide a new topology which allows to link the discrete and the continuous dynamics by considering them as instances of the same strong Markov process with different initial conditions. We construct this pruning process on the space of so-called bi-measure trees, which are metric measure spaces with an additional pruning measure. The pruning measure is assumed to be finite on finite trees, but not necessarily locally finite. We also characterize the pruning process analytically via its Markovian generator and show that it is continuous in the initial bi-measure tree. A series of examples is given, which include the finite variance offspring case where the pruning measure is the length measure on the underlying tree.

Résumé. Dans (*Ann. Inst. Henri Poincaré Probab. Stat.* **34** (1998) 637–686), les auteurs obtiennent une chaîne de Markov à valeurs arbres en élaguant de plus en plus de sous-arbres le long des nœuds d'un arbre de Galton–Watson. Plus récemment dans (*Ann. Probab.* **40** (2012) 1167–1211), un analogue continu de la dynamique d'élagage à valeurs arbres est construit sur des arbres de Lévy. Dans cet article, nous présentons une nouvelle topologie qui permet de relier les dynamiques discrètes et continues en les considérant comme des exemples du même processus de Markov fort avec des conditions initiales différentes. Nous construisons ce processus d'élagage sur l'espace des arbres appelés bi-mesurés, qui sont des espaces métriques mesurés avec une mesure d'élagage additionnelle. La mesure d'élagage est supposée finie sur les arbres finis, mais pas nécessairement localement finie. De plus, nous caractérisons analytiquement le processus d'élagage par son générateur infinitésimal et montrons qu'il est continu en son arbre bi-mesuré initial. Plusieurs exemples sont donnés, notamment le cas d'une loi de reproduction à variance finie où la mesure d'élagage est la mesure des longueurs sur l'arbre sous-jacent.

MSC: 60F05; 60B12; 60J25; 05C05; 05C10; 60G55

Keywords: Tree-valued Markov process; CRT; Real trees; Pruning procedure; Pointed Gromov-weak topology; Prohorov metric; Non-locally finite measures

1. Introduction and motivation

Let $\mathcal{G}_1$ be a rooted Galton–Watson tree with an offspring generating function g. For $0 \le u \le 1$, let $\mathcal{G}_u$ be the subtree of $\mathcal{G}_1$ obtained by retaining each edge with probability u. Lyons [40] showed that $\mathcal{G}_u$ is again a Galton–Watson tree which corresponds to the offspring generating function $g_u = g(1 - u + u\cdot)$. As one can couple the pruning procedures for several $u \in [0, 1]$ in such a way that $\mathcal{G}_{u'}$ is a rooted subtree of $\mathcal{G}_u$ whenever $u' \le u$, they give rise to a non-decreasing tree-valued Markov process $(\mathcal{G}_u)_{u\in[0,1]}$ which was further studied in Aldous and Pitman [12]. Recently,

[1] Supported in part by the European Science Foundation under the RGLIS short visit grant 5429.

Abraham, Delmas and He consider in [4] another pruning procedure on Galton Watson trees where cut points fall on the branch points to the effect that the subtree above is pruned. Here each node of the initial Galton–Watson tree is cut independently with probability $1 - u^{n-1}$ where n is the number of children of the node.

In the same spirit some authors consider continuum tree analogues of pruning dynamics. Compare, for example, [8,11] for a pruning proportional to the length on the skeleton of a Brownian CRT, [43] for a pruning on the infinite branch points of a stable Lévy tree, [1] for a pruning on the infinite branch points of a Lévy tree without Brownian part, [2,7] for a combined pruning proportional to the length and on the infinite branch points of a general Lévy tree.

In [2] it is conjectured that the pruning procedure presented in the same paper is the continuous analogue of a mixture of the pruning procedures suggested in [12] and [2], that is of pruning procedures on Galton–Watson trees where cut points fall on edges as well as on nodes. However, no precise link between the discrete and the continuum tree-valued dynamics has been given so far. The main goal of the present paper is to present ONE Markov process, which in the following is referred to as THE *pruning process*. We shall give an analytic characterization via a Markovian generator and provide with the so-called *leaf-sampling weak vague topology* a notion of convergence which allows to state convergence of the discrete tree-valued dynamics to the associated continuous tree-valued dynamics.

It had been a long tradition to encode trees via continuous excursions, and to use uniform topology as a notion of convergence. A more recent and conceptional approach is to think of trees as "tree-like" metric spaces, the so-called $\mathbb{R}$-trees, and to use the Gromov–Hausdorff topology as a notion of convergence (compare, for example, [22] for an introduction into $\mathbb{R}$-trees and [31,34] for details on the Gromov–Hausdorff distance). For a long time convergence of suitably rescaled Galton–Watson processes were established for very particular offspring distributions only. To be in a position to prove an invariance principle, Aldous developed in [9,10] a notion of convergence by encoding trees as closed subsets of l^1_+, the space of summable sequences of positive numbers which were additionally equipped with a sampling measure. Convergence was then proposed as the convergence of all subtrees spanned by finite samples from the tree. Once more, this very neat and powerful idea had been generalized to the more conceptional encoding of trees as metric probability measure spaces where the tree space was equipped with the so-called *Gromov-weak topology* (compare [33,34]). Further developments which combine the Gromov–Hausdorff and Gromov-weak topology and allow for sampling measures that are finite on bounded sets can be found, for example, in [5,30,44].

In the present paper, we provide a unified framework by regarding these pruning processes as the *same* Feller-continuous Markov process on a (non-locally compact) space of $\mathbb{R}$-trees with different initial conditions, and to establish convergence in Skorohod space whenever the initial distribution converges. For that purpose, we introduce *bi-measure* $\mathbb{R}$*-trees*, i.e., metric measure spaces (T, r, μ), which are additionally equipped with a so-called *pruning* measure, ν. Here, the so-called *sampling measure* μ is a finite measure (allowing for a varying total mass), while the pruning measure is only assumed to be finite on finite subtrees. As the pruning measure is already part of the state, we are in a position to construct one (universal) pruning process. This process is a pure jump process which, given a bi-measure $\mathbb{R}$-tree, lets rain down successively more and more cut points according to a Poisson process whose intensity measure is equal to the pruning measure. At each cut point, the subtree above is cut off and removed, and the sampling and pruning measures are simultaneously updated by simply restricting them to the remaining, pruned part of the tree.

A major difficulty is that important examples for the pruning measures, such as the length measure on the Brownian CRT, are not locally finite. Therefore, we introduce with the *leaf-sampling weak vague topology* a new topology on the spaces of bi-measure $\mathbb{R}$-trees. We give equivalent characterizations of convergence and provide convergence determining classes of functions.

Outline. The paper is organized as follows. In Section 2 we introduce the leaf-sampling weak-vague topology and give a characterization of convergence. In Section 3 we construct the pruning process, calculate its Markovian generator and verify that the law of the process on Skorohod space depends continuously on the initial condition. Finally, in Section 4 we apply our main result to obtain convergence of various pruning processes that appeared in the literature.

2. Bi-measure $\mathbb{R}$-trees and the LWV-topology

In this section we introduce the space of $\mathbb{R}$-trees equipped with a finite *sampling measure* and a *pruning measure* which is assumed to be finite on finite subtrees. Moreover, we define the *leaf-sampling weak vague topology* (LWV-topology) on this space of bi-measure $\mathbb{R}$-trees. The idea behind our topology is to first sample a finite number of points from the tree according to the sampling measure. These points span a finite subtree. In many relevant examples they are actually the leaves of this subtree. Then we equip this finite subtree with the restriction of the pruning measure and obtain a random metric measure tree. For convergence of bi-measure trees, we require that these random metric measure trees converge together with the sampled points as *n-pointed metric measure $\mathbb{R}$-trees* in the Gromov-weak topology.

We therefore recall in Section 2.1 the notion of Gromov-weak topology on metric measure spaces and extend it to the *n-pointed Gromov-weak topology*. In Section 2.2 we then define a stronger topology on n-pointed metric measure $\mathbb{R}$-trees, the subtree Gromov-weak topology. Finally, in Section 2.3 we define the LWV-convergence. It turns out that it can be characterized by both the pointed as well as the subtree Gromov-weak convergence of samples from the bi-measure $\mathbb{R}$-tree and defines a separable, metrizable topology. In Section 2.4, we introduce classes of test functions that induce the LWV-topology. One of them turns out to be convergence determining. Using these test functions, we derive several convergence results.

2.1. *The n-pointed Gromov-weak topology*

Greven, Pfaffelhuber and Winter [33] define the space of metric probability measure spaces equipped with the Gromov-weak topology. In this subsection, we define a slightly more general space using finite measures instead of probability measures and considering n-pointed metric measure spaces. We do not give proofs, because the extension is straightforward.

We start recalling basic notation. As usual, given a topological space X, we denote by $\mathcal{C}(X)$ ($\mathcal{C}_b(X)$) the space of (bounded) continuous, $\mathbb{R}$-valued functions on X, and by $\mathcal{M}_1(X)$ ($\mathcal{M}_f(X)$) the space of probability (finite) measures, defined on the Borel σ-algebra of X. For $x \in X$, $\delta_x \in \mathcal{M}_1(X)$ is the Dirac measure in the point x. "$\Rightarrow$" means weak convergence in $\mathcal{M}_1(X)$ or in $\mathcal{M}_f(X)$. Recall that the support of μ, $\mathrm{supp}(\mu)$, is the smallest closed set $X_0 \subseteq X$ such that $\mu(X_0) = \mu(X) =: \|\mu\|$. For $\mu \in \mathcal{M}_f(X)$, we denote the normalization by

$$\mu^\circ := \frac{\mu}{\|\mu\|} \in \mathcal{M}_1(X). \tag{2.1}$$

The *push forward* of μ under a measurable map ϕ from X into another topological space Z is the finite measure $\phi_*\mu \in \mathcal{M}_f(Z)$ defined by

$$\phi_*\mu(A) := \mu\big(\phi^{-1}(A)\big), \tag{2.2}$$

for all measurable subsets $A \subseteq Z$. For the integral of an integrable function φ with respect to μ, we sometimes use the notation

$$\langle \mu, \varphi \rangle := \int \varphi \, \mathrm{d}\mu. \tag{2.3}$$

A *metric measure space* is a triple (X, r, μ), where (X, r) is a metric space such that $(\mathrm{supp}(\mu), r)$ is complete and separable and $\mu \in \mathcal{M}_f(X)$ is a finite measure on $(X, \mathcal{B}(X))$. If $\mathrm{supp}(\mu)$ is separable but not complete, we simply identify it with its completion.

Branching trees such as Galton–Watson trees and the CRT are often rooted. We therefore define a *rooted metric measure space* (X, r, ρ, μ) as a metric measure space (X, r, μ) together with a distinguished point $\rho \in X$ which is referred to as the *root*. To avoid heavy notations, in the following we suppress the metric and the root, i.e. we abbreviate, for example,

$$X = (X, r, \rho), \quad (X, \mu) = (X, r, \rho, \mu). \tag{2.4}$$

The definition of metric measure spaces given in [33] can easily be extended to rooted metric measure spaces. In the context of metric spaces, rooted spaces are often referred to as *pointed spaces* (compare, for example, Section 8 in [19]).

We want to extend these rooted metric measure spaces (X, r, μ) by fixing n additional points $u_1, \dots, u_n \in X$, and call $(X, r, \rho, (u_1, \dots, u_n), \mu)$ a (rooted) *n-pointed metric measure space*. The support of an n-pointed metric measure space $(X, r, \rho, (u_1, \dots, u_n), \mu)$ is defined by

$$\mathrm{supp}\big((X, r, \rho, (u_1, \dots, u_n), \mu)\big) := \mathrm{supp}(\mu) \cup \{\rho, u_1, \dots, u_n\}. \tag{2.5}$$

In the following we identify two n-pointed metric measure spaces if there is a measure preserving isometry between their supports that also preserves the root and the fixed points.

Definition 2.1 (The space $\mathbb{M}_n$***).*** *Two n-pointed metric measure spaces $\mathcal{X} = (X, r, \rho, (u_1, \dots, u_n), \mu)$ and $\mathcal{X}' = (X', r', \rho', (u'_1, \dots, u'_n), \mu')$ are called* equivalent *if there exists an isometry ϕ between* $\mathrm{supp}(\mathcal{X})$ *and* $\mathrm{supp}(\mathcal{X}')$ *such that $\phi_*\mu = \mu'$, $\phi(\rho) = \rho'$ and $\phi(u_k) = u'_k$ for all $1 \le k \le n$. It is clear that this defines an equivalence relation.*

We denote by $\mathbb{M}_n$ the set of equivalence classes of n-pointed metric measure spaces.

Remark 2.2. *Notice that for a notion of equivalence of metric measure spaces (X, r, μ) and (X', r', μ') there are two canonical choices. Either we insist that the metric spaces (X, r) and (X', r') are isometric or we are satisfied with their supports to be isometric thereby neglecting sets of measure zero* (*compare, for example*, [46], *Section* 27). *Here we take the second approach which allows for a characterization of convergence in $\mathbb{M}_n$ through convergence determining classes of functions. The gap between such a notion of* (*weak*) *convergence and a stronger topology which also requires the convergence of supports of the measures is closed in* [13].

To simplify notations, we do not distinguish between an n-pointed metric measure space and its equivalence class. That is, we write

$$\mathcal{X} = \big(X, (x_1, \dots, x_n), \mu\big) \in \mathbb{M}_n. \tag{2.6}$$

Remark 2.3 (The space $\mathbb{M}_0$***).*** $\mathbb{M}_0$ *is the usual space of rooted metric measure spaces* (*with finite measures*).

For a rooted metric space X, we define a map R^X that associates to a sequence of points the matrix of their distances to the root and to each other, i.e.,

$$R^X: \begin{cases} X^{\mathbb{N}} \to \mathbb{R}_+^{\binom{\mathbb{N}_0}{2}}, \\ (x_i)_{i \ge 1} \mapsto (r(x_i, x_j))_{0 \le i < j} \quad \text{with } x_0 := \rho. \end{cases} \tag{2.7}$$

The *distance matrix distribution* of an n-pointed metric measure space $\mathcal{X} = (X, (u_1, \dots, u_n), \mu)$ is then given by

$$\boldsymbol{\upsilon}^{\mathcal{X}} := \|\mu\| \cdot \big(R^X\big)_* \left(\bigotimes_{k=1}^{n} \delta_{u_k} \otimes \big(\mu^{\circ}\big)^{\otimes \mathbb{N}} \right) \in \mathcal{M}_f\big(\mathbb{R}_+^{\binom{\mathbb{N}_0}{2}}\big), \tag{2.8}$$

which obviously depends only on the equivalence class. Vershik's proof of Gromov's reconstruction theorem (see [34], Subsection $3\frac{1}{2}$.7) directly carries over to n-pointed metric measure spaces. Therefore, $\mathcal{X} \in \mathbb{M}_n$ is uniquely determined by its distance matrix distribution $\boldsymbol{\upsilon}^{\mathcal{X}}$.

Definition 2.4 (pGw-topology). *A sequence of n-pointed metric measure spaces $\mathcal{X}_N \in \mathbb{M}_n$ converges n-pointed Gromov-weakly* (pGw) *to $\mathcal{X} \in \mathbb{M}_n$ if*

$$\boldsymbol{\upsilon}^{\mathcal{X}_N} \underset{N \to \infty}{\Longrightarrow} \boldsymbol{\upsilon}^{\mathcal{X}} \tag{2.9}$$

in the weak topology on $\mathcal{M}_f(\mathbb{R}_+^{\binom{\mathbb{N}_0}{2}})$.

We see directly from the definition that functions of the form $\Phi: \mathbb{M}_n \to \mathbb{R}, \mathcal{X} \mapsto \langle \boldsymbol{\upsilon}^{\mathcal{X}}, f \rangle$ with $f \in \mathcal{C}_b(\mathbb{R}_+^{\binom{\mathbb{N}_0}{2}})$ are continuous. If f depends only on finitely many coordinates, Φ is called a *polynomial*, and there exists $m \in \mathbb{N}$, $\varphi \in \mathcal{C}_b(\mathbb{R}_+^{\binom{n+m+1}{2}})$ such that for $\mathcal{X} = (X, \underline{u}, \mu) \in \mathbb{M}_n$,

$$\Phi(\mathcal{X}) = \Phi^{m,\varphi}(\mathcal{X}) := \int_{X^m} \mu^{\otimes m}(\mathrm{d}\underline{v}) \varphi\big(R^X(\underline{u}, \underline{v})\big), \tag{2.10}$$

where $\underline{u} = (u_1, \dots, u_n)$, $\underline{v} = (v_1, \dots, v_m)$ and $(\underline{u}, \underline{v}) := (u_1, \dots, u_n, v_1, \dots, v_m)$. Note that

$$\Phi^{m,1}(\mathcal{X}) = \|\mu\|^m, \tag{2.11}$$

and, in particular, $\Phi^{0,1} \equiv 1$. Moreover, as polynomials are not bounded (compare (2.11)), we define a class $\Pi_n \subseteq \mathcal{C}_b(\mathbb{M}_n)$ of bounded test functions by

$$\begin{aligned}\Pi_n := \Big\{ &\Phi^{\gamma,m,\varphi}(\mathcal{X}) := \gamma\big(\|\mu\|\big) \cdot \Phi^{m,\varphi}(\mathcal{X}): \\ &\Phi^{m,\varphi} \text{ is a polynomial, } \gamma \in \mathcal{C}_b(\mathbb{R}_+), \lim_{x\to\infty} x^k \gamma(x) = 0, \forall k \in \mathbb{N} \Big\}.\end{aligned} \tag{2.12}$$

Recall the *Prohorov distance* d_{Pr} between two finite measures μ, ν on a metric space (Z, d),

$$\begin{aligned} &d_{\mathrm{Pr}}^{(Z,d)}(\mu, \nu) \\ &\quad := \inf\big\{\varepsilon > 0\colon \mu\big(A^\varepsilon\big) + \varepsilon \geq \nu(A), \nu\big(A^\varepsilon\big) + \varepsilon \geq \mu(A)\ \forall A \text{ closed}\big\},\end{aligned} \tag{2.13}$$

where $A^\varepsilon := \{x \in Z \mid d(x, A) < \varepsilon\}$.

***Definition 2.5** (**n-pointed Gromov–Prohorov distance**). We define the n-pointed Gromov–Prohorov distance between $\mathcal{X} = (X, \underline{u}, \mu)$ and $\mathcal{Y} = (Y, \underline{v}, \nu)$ in $\mathbb{M}_n$ by*

$$d_{\mathrm{pGP}}(\mathcal{X}, \mathcal{Y}) := \inf_d \left\{ d_{\mathrm{Pr}}^{(X \uplus Y, d)}(\mu, \nu) + d(\rho_X, \rho_Y) + \sum_{k=1}^n d(u_k, v_k) \right\}, \tag{2.14}$$

where the infimum is taken over all metrics d on the disjoint union $X \uplus Y$ that extends r_X and r_Y. If there is no confusion, we simply write d_{Pr} for $d_{\mathrm{Pr}}^{(X \uplus Y, d)}$.

Recall that a set $\mathcal{F} \subseteq \mathcal{C}_b(X)$ is *convergence determining* (on the topological space X) if, for probability measures μ_N, μ on X, the weak convergence $\mu_N \underset{N\to\infty}{\Longrightarrow} \mu$ is equivalent to $\int f \,\mathrm{d}\mu_N \underset{N\to\infty}{\longrightarrow} \int f \,\mathrm{d}\mu$ for all $f \in \mathcal{F}$.

Proposition 2.6 (**Π_n is convergence determining**). *Let $\mathcal{X}, \mathcal{X}_1, \mathcal{X}_2, \ldots \in \mathbb{M}_n$. The following conditions are equivalent*:

(i) $\mathcal{X}_N \xrightarrow{\mathrm{pGw}} \mathcal{X}$, *as* $N \to \infty$,
(ii) $\Phi(\mathcal{X}_N) \underset{N\to\infty}{\longrightarrow} \Phi(\mathcal{X})$, *for all polynomials* Φ,
(iii) $d_{\mathrm{pGP}}(\mathcal{X}_N, \mathcal{X}) \underset{N\to\infty}{\longrightarrow} 0$.

Furthermore, $\mathbb{M}_n$ is separable, d_{pGP} is a complete metric on $\mathbb{M}_n$, and the class $\Pi_n \subseteq \mathcal{C}_b(\mathbb{M}_n)$ is convergence determining on $\mathbb{M}_n$.

Proof. The proof of the equivalences is an obvious modification of that of Theorem 5 in [33]. Notice that μ° in the definition of the pGw-topology can be replaced by μ in the definition of polynomials because $\|\mu\| = \Phi^{1,1}(\mathcal{X})$. Separability and completeness follow in the same way as Proposition 5.6 in [33].

To see that Π_n induces the pGw-topology, note that $\Phi^{\gamma,0,1} \in \Pi_n$, and convergence of $\Phi^{\gamma,0,1}(\mathcal{X}_N) = \gamma(\|\mu_N\|)$ with $\gamma(x) = \mathrm{e}^{-x}$ implies the convergence of $\|\mu_N\|$. Hence, the topology induced by Π_n coincides with the topology induced by the polynomials. Using the fact that Π_n is multiplicatively closed, we see that it is convergence determining with the same proof as for the set of polynomials on the space of metric probability measure spaces (see [21,39]), or directly from Le Cam's theorem (see [38], [35], Lem. 4.1). □

2.2. *Measure $\mathbb{R}$-trees and subtree Gromov-weak topology*

In this subsection we define the *subtree Gromov-weak topology*. As "tree-like" metric spaces are 0-hyperbolic, throughout the paper we work with the subspaces

$$\mathbb{H}_n := \{\mathcal{X} \in \mathbb{M}_n \colon \mathcal{X} \text{ is 0-hyperbolic}\} \subseteq \mathbb{M}_n, \tag{2.15}$$

and

$$\mathbb{H} := \mathbb{H}_0 \subseteq \mathbb{M}_0, \tag{2.16}$$

where a metric measure space $\mathcal{X} \in \mathbb{M}_n$ is called 0-*hyperbolic* iff

$$r(x_1, x_2) + r(x_3, x_4) \le \max\bigl\{r(x_1, x_3) + r(x_2, x_4), r(x_1, x_4) + r(x_2, x_3)\bigr\}, \tag{2.17}$$

for all $x_1, x_2, x_3, x_4 \in \operatorname{supp}(\mathcal{X})$. It follows immediately from Theorem 2.5 in [31] that for each $n \in \mathbb{N}$, $(\mathbb{H}_n, d_{\mathrm{pGP}})$ is complete.

Recall that a 0-hyperbolic space is called an $\mathbb{R}$-*tree* if it is connected (see [23] for equivalent definitions and background on $\mathbb{R}$-trees). Given a (rooted) $\mathbb{R}$-tree (T, r, ρ), we denote the unique path between two points $x, y \in T$ by $[x, y]$, and $[x, y[:= [x, y] \setminus \{y\}$. The set of *leaves* of the tree is

$$\operatorname{Lf}(T) := T \Big\backslash \bigcup_{x \in T} [\rho, x[. \tag{2.18}$$

We also use the notation $[\![\underline{v}]\!]$ for the tree spanned by the root ρ and the vector $\underline{v} \in T^n$, i.e.,

$$[\![\underline{v}]\!] := \bigcup_{i=1}^{n} [\rho, v_i]. \tag{2.19}$$

Here and in the following we refer to any $\mathbb{R}$-tree of the form (2.19) as a *finite tree*.

Remark 2.7 **(0*-hyperbolic spaces are equivalent to* $\mathbb{R}$*-trees*).** *According to Theorem* 3.38 *of* [29], *every* 0*-hyperbolic space can be isometrically embedded into an* $\mathbb{R}$*-tree. Since our notion of equivalence of two* n*-pointed metric measure spaces* $\mathcal{X}$ *and* $\mathcal{X}'$ *requires only a* (*measure and point preserving*) *isometry between* $\operatorname{supp}(\mathcal{X})$ *and* $\operatorname{supp}(\mathcal{X}')$, *this means that every* n*-pointed* 0*-hyperbolic metric measure space is equivalent to an* n*-pointed*, *measured* $\mathbb{R}$*-tree. In the following we assume*, *without loss of generality*, *that* $\mathcal{X} \in \mathbb{H}_n$ *is an* $\mathbb{R}$*-tree*, *by choosing a connected representative of the equivalence class*.

Also note that, *given two* $\mathbb{R}$*-trees* (T, r), (T', r') *with subsets* $A \subseteq T$, $A' \subseteq T'$, *and an isometry* $\phi : A \to A'$, *there is a unique extension of* ϕ *to an isometry between the generated* $\mathbb{R}$*-trees*, $\overline{\phi} : [\![A]\!] \to [\![A']\!]$. *Indeed*, *for* $v \in [\![A]\!]$ *there exist* (*non-unique*) $x, y \in A$ *with* $v \in [x, y]$, *and a unique* $w_v \in [\phi(x), \phi(y)]$ *with* $r(x, v) = r'(\phi(x), w_v)$. *It is straightforward to check that* w_v *does not depend on the choice of* x, y *and* $\overline{\phi}(v) := w_v$ *is an isometry. In particular, for* $\mathcal{X} \in \mathbb{H}_n$, *the* $\mathbb{R}$*-tree* $[\![\operatorname{supp}(\mathcal{X})]\!]$ *is unique up to isometry.*

We now define a topology on $\mathbb{H}_n$ which requires that every subtree generated by a subset of the n distinguished points converges. For that purpose, we define a projection map which sends a list $\underline{u}$ to the sublist indexed by $I = \{i_1, \dots, i_k\}$ for given $1 \le i_1 < \cdots < i_k \le n$. That is,

$$\pi_I^n \colon \begin{cases} T^n \to T^k, \\ \underline{u} \mapsto (u_{i_1}, \dots, u_{i_k}). \end{cases} \tag{2.20}$$

The sublist $(u_1, \dots, u_k)$ of $\underline{u} \in T^n$ is simply denoted $\underline{u}^k$. With a slight abuse of notation, we also write

$$\pi_I^n(T, \underline{u}, \mu) := \bigl(\bigl[\!\bigl[\pi_I^n(\underline{u})\bigr]\!\bigr], \pi_I^n(\underline{u}), \mu\bigr) := \bigl(\bigl[\!\bigl[\pi_I^n(\underline{u})\bigr]\!\bigr], \pi_I^n(\underline{u}), \mu\restriction_{[\![\pi_I^n(\underline{u})]\!]}\bigr), \tag{2.21}$$

where the measure μ in the middle expression is tacitly understood to be restricted to the appropriate space, $[\![\pi_I^n(\underline{u})]\!]$.

Definition 2.8 (sGw-topology). *Consider n-pointed measure $\mathbb{R}$-trees $\mathcal{X}, \mathcal{X}_1, \mathcal{X}_2, \ldots \in \mathbb{H}_n$. We say that $(\mathcal{X}_N)_{N\in\mathbb{N}}$ converges subtree Gromov-weakly* (sGw) *to $\mathcal{X}$ iff $\mathcal{X}_N \xrightarrow[N\to\infty]{\text{pGw}} \mathcal{X}$ and*

$$\pi_I^n(\mathcal{X}_N) \xrightarrow[N\to\infty]{\text{pGw}} \pi_I^n(\mathcal{X}), \quad \forall I \subseteq \{1, \ldots, n\}. \tag{2.22}$$

Put

$$\widetilde{\mathbb{H}}_n := \{(T, \underline{u}, \mu) \in \mathbb{H}_n\colon \operatorname{supp}(\mu) \subseteq [\![\underline{u}]\!]\} \subseteq \mathbb{H}_n, \tag{2.23}$$

and note that $\widetilde{\mathbb{H}}_n$ consists only of finite trees with at most n leaves.

Remark 2.9 (Related topologies). *The sGw-topology is strictly stronger than the pGw-topology. On $\widetilde{\mathbb{H}}_n$, sGw-convergence implies measured Gromov–Hausdorff convergence* [32], *also known as weighted Gromov–Hausdorff convergence* [30,44].

Lemma 2.10 (Sufficient condition for sGw-convergence). *Consider random n-pointed measure $\mathbb{R}$-trees $\mathcal{X} = (T, \underline{u}, \mu), \mathcal{X}_N = (T_N, \underline{u}_N, \mu_N) \in \widetilde{\mathbb{H}}_n$, $N \in \mathbb{N}$ (in particular $T_N = [\![\underline{u}_N]\!]$). Assume that $(\mathcal{X}_N)_{N\in\mathbb{N}}$ converges almost surely* (a.s.) *to $\mathcal{X}$ in the n-pointed Gromov-weak topology, as $N \to \infty$. Furthermore, assume that there is a strictly increasing function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\psi(\|\mu\|)$ is integrable and*

$$\mathbb{E}\bigl[\psi\bigl(\mu_N\bigl([\![\pi_I^n(\underline{u}_N)]\!]\bigr)\bigr)\bigr] \underset{N\to\infty}{\longrightarrow} \mathbb{E}\bigl[\psi\bigl(\mu\bigl([\![\pi_I^n(\underline{u})]\!]\bigr)\bigr)\bigr], \quad \forall I \subseteq \{1, \ldots, n\}. \tag{2.24}$$

Then $(\mathcal{X}_N)_{N\in\mathbb{N}}$ converges also subtree Gromov-weakly to $\mathcal{X}$, a.s., as $N \to \infty$.

To prepare the proof, we state the following:

Remark 2.11 (pGw-convergence yields a tree homomorphism). *Consider a sequence of n-pointed measure $\mathbb{R}$-trees $\mathcal{X} = (T, \underline{u}, \mu)$, $\mathcal{X}_1 = (T_1, \underline{u}_1, \mu_1)$, $\mathcal{X}_2 = (T_2, \underline{u}_2, \mu_2), \ldots \in \widetilde{\mathbb{H}}_n$. Assume furthermore that $(\mathcal{X}_N)_{N\in\mathbb{N}}$ converges n-pointed Gromov-weakly to $\mathcal{X}$, a.s., as $N \to \infty$.*

For sufficiently large $N \in \mathbb{N}$, we can define a function $f_N : T_N \to T$ by sending the root to the root, letting $f_N(u_{N,k}) = u_k$ and $f_N(u_{N,k} \wedge u_{N,l}) = u_k \wedge u_l$, $k, l = 1, \ldots, n$, and then stretching linearly. Here, as usual, $u \wedge v$ denotes the unique branch point such that $[\rho, u \wedge v] = [\rho, u] \cap [\rho, v]$.

By construction, $\operatorname{dis}(f_N) \underset{N\to\infty}{\longrightarrow} 0$ *where*

$$\operatorname{dis}(f) := \sup_{x,y\in T} \bigl|r(x, y) - r'\bigl(f(x), f(y)\bigr)\bigr| \tag{2.25}$$

denotes the distortion *of a map $f : (T, r) \to (T', r')$.*

Proof of Lemma 2.10. Assume that N is large enough, such that the function $f_N : T_N \to T$ from Remark 2.11 is a tree homomorphism with $f_N(u_{N,k}) = u_k$, $k \in \{1, 2, \ldots, n\}$ and such that $\operatorname{dis}(f_N) \underset{N\to\infty}{\longrightarrow} 0$. We can therefore choose a metric d on $T_N \uplus T$ extending r_N and r such that $d(x, f_N(x)) \to 0$ for all $x \in T_N$ (compare, for example, [19], Corollary 7.3.28).

Thus $d_{\text{Pr}}((f_N)_*\mu_N, \mu_N) \leq \sup_x d(x, f_N(x)) \to 0$, as $N \to \infty$, and we obtain that

$$d_{\text{Pr}}\bigl((f_N)_*\mu_N, \mu\bigr) \leq d_{\text{Pr}}\bigl((f_N)_*\mu_N, \mu_N\bigr) + d_{\text{Pr}}(\mu_N, \mu) \underset{N\to\infty}{\longrightarrow} 0. \tag{2.26}$$

Fix now $I \subseteq \{1, \ldots, n\}$ and define the subtree

$$S := \bigl[\!\bigl[\pi_I^n(\underline{u})\bigr]\!\bigr] \subseteq T. \tag{2.27}$$

Because S is closed in T, we have $\limsup_{N\to\infty}(f_N)_*\mu_N(S)\le\mu(S)$ by the Portmanteau theorem (see Theorem 2.1 in [16]). Because ψ is increasing, this implies

$$\limsup_{N\to\infty}\psi\big((f_N)_*\mu_N(S)\big)\le\psi\big(\mu(S)\big). \tag{2.28}$$

By assumption (2.24),

$$\mathbb{E}\big[\psi\big((f_N)_*\mu_N(S)\big)\big]=\mathbb{E}\big[\psi\big(\mu_N\big(\big[\!\big[\pi_I^n(\underline{u}_N)\big]\!\big]\big)\big)\big]\underset{N\to\infty}{\longrightarrow}\mathbb{E}\big[\psi\big(\mu(S)\big)\big]. \tag{2.29}$$

(2.29) and (2.28) together yield $\psi((f_N)_*\mu_N(S))\underset{N\to\infty}{\longrightarrow}\psi(\mu(S))$, almost surely. Because ψ is strictly increasing, also $(f_N)_*\mu_N(S)\to\mu(S)$. Using once more the Portmanteau theorem and closedness of S, we obtain that

$$(f_N)_*\mu_N\restriction_S\underset{N\to\infty}{\Longrightarrow}\mu\restriction_S. \tag{2.30}$$

The inequality

$$\begin{aligned}&d_{\mathrm{pGw}}\big(\pi_I^n\big([\![\underline{u}_N]\!],\underline{u}_N,\mu_N\big),\pi_I^n\big([\![\underline{u}]\!],\underline{u},\mu\big)\big)\\&\quad\le d_{\mathrm{Pr}}\big(\mu_N\restriction_{[\![\pi_I^n(\underline{u}_N)]\!]},(f_N)_*\mu_N\restriction_S\big)+d_{\mathrm{Pr}}\big((f_N)_*\mu_N\restriction_S,\mu\restriction_S\big),\end{aligned} \tag{2.31}$$

then gives the sGw-convergence. □

As for the pGw-topology, we define an associated set of test functions $\tilde{\Phi}:\mathbb{H}_n\to\mathbb{R}$ by

$$\tilde{\Phi}(T,\underline{u},\mu):=\prod_{I\subseteq\{1,\dots,n\}}\Phi_I\big(\pi_I^n(T,\underline{u},\mu)\big), \tag{2.32}$$

where the Φ_I are polynomials on $\mathbb{H}_{\#I}$. Obviously, this class of test functions induces the sGw-topology on $\widetilde{\mathbb{H}}_n$, and together with the polynomials on $\mathbb{H}_n$, the sGw-topology on $\mathbb{H}_n$. We also define

$$\widetilde{\Pi}_n:=\Big\{\prod_{I\subseteq\{1,\dots,n\}}\Phi_I^{\gamma,m,\varphi}\circ\pi_I^n:\ \Phi_I^{\gamma,m,\varphi}\in\Pi_{\#I}\Big\}. \tag{2.33}$$

2.3. *The LWV-topology*

In this subsection we give the definition of *bi-measure* $\mathbb{R}$*-trees* and equip the space of equivalence classes of bi-measure $\mathbb{R}$-trees with the *leaf-sampling weak vague topology*, in the following referred to as the *LWV-topology*.

Given a rooted measure $\mathbb{R}$-tree (T,μ), denote by

$$\mathrm{Sk}_\mu(T):=\bigcup_{v\in\mathrm{supp}(\mu)}[\rho,v[\,\cup\big\{v\in T:\ \mu\big(\{v\}\big)>0\big\} \tag{2.34}$$

the μ*-skeleton* of (T,μ), and by

$$\mathrm{Lf}_\mu(T):=\big[\!\big[\mathrm{supp}(\mu)\big]\!\big]\setminus\mathrm{Sk}_\mu(T) \tag{2.35}$$

the set of μ*-leaves* of (T,μ).

We call (T,μ,ν) a (rooted) *bi-measure* $\mathbb{R}$*-tree* if (T,μ) is a (rooted) measure $\mathbb{R}$-tree and ν is a (σ-finite) measure on T which satisfies the following two conditions:

(i) $\nu([\rho,u])$ is μ-a.s. finite for $u\in T$,
(ii) ν vanishes on the set of μ-leaves, i.e., $\nu(\mathrm{Lf}_\mu(T))=0$.

Note that (i) implies that ν is finite on subtrees of T with a finite number of leaves sampled with μ, a.s., and that $\nu\restriction_{\mathrm{Sk}_\mu(T)}$ is σ-finite (because our definition of measure $\mathbb{R}$-trees includes separability of $\mathrm{supp}(\mu)$). In many interesting cases, however, ν is not locally finite.

Definition 2.12 (The spaces $\mathbb{H}^{f,\sigma}$ and $\mathbb{H}^{K,\sigma}$). *Two bi-measure $\mathbb{R}$-trees (T,μ,ν) and (T',μ',ν') are called* equivalent *if there exists an isometry $\phi:[\![\mathrm{supp}(\mu)]\!]\to T'$ preserving the root and μ and preserving ν on the μ-skeleton, i.e., $\phi_*(\mu)=\mu'$ and $\phi_*(\nu\restriction_{\mathrm{Sk}_\mu(T)})=\nu'\restriction_{\mathrm{Sk}_{\mu'}(T')}$. In particular, (T,μ,ν) is equivalent to $(T,\mu,\nu\restriction_{\mathrm{Sk}_\mu(T)})$.*

We denote by $\mathbb{H}^{f,\sigma}$ the space of equivalence classes of (rooted) bi-measure $\mathbb{R}$-trees, and by $\mathbb{H}^{K,\sigma}:=\{(T,\mu,\nu)\in\mathbb{H}^{f,\sigma}\mid\|\mu\|\le K\}$, $K>0$, the subspace where the total mass of the sampling measure is bounded by K.

Similar to the distance matrix distribution $\boldsymbol{\nu}^{(T,\mu)}$ introduced in (2.8), which characterizes n-pointed measure $\mathbb{R}$-trees and is used to define the pGw-topology, we want to characterize bi-measure $\mathbb{R}$-trees by the so-called *subtree-vector-distribution*. To introduce this, consider for a given bi-measure $\mathbb{R}$-tree (T,μ,ν) the function

$$\tau_{(T,\mu,\nu)}:\begin{cases}\bigcup_{n\in\mathbb{N}}T^n\to\bigcup_{n\in\mathbb{N}}\mathbb{H}_n,\\ (u_1,u_2,\dots,u_n)\mapsto([\![u_1,\dots,u_n]\!],(u_1,\dots,u_n),\nu),\end{cases}\tag{2.36}$$

which sends a vector of n points in T to the n-pointed $\mathbb{R}$-tree spanned by these points and equipped with ν, which we tacitly understand to be restricted to the appropriate space, i.e. $[\![u_1,\dots,u_n]\!]$. We also define the function

$$\varsigma_{(T,\mu,\nu)}:\begin{cases}T^{\mathbb{N}}\to\prod_{n\in\mathbb{N}}\mathbb{H}_n,\\ \underline{u}\mapsto(\tau_{(T,\mu,\nu)}(u_1),\tau_{(T,\mu,\nu)}(u_1,u_2),\dots),\end{cases}\tag{2.37}$$

which sends a sequence of points to the sequence of pointed measure $\mathbb{R}$-trees spanned and pointed by the first 1, 2, etc. points and each of these is equipped with the appropriate restriction of ν. Note that $\tau_{(T,\mu,\nu)}$ does not depend on the measure μ and is in general not continuous.

Lemma 2.13 (Measurability). *Equip $\mathbb{H}_n$ with the n-pointed Gromov-weak topology, and $\prod_{n\in\mathbb{N}}\mathbb{H}_n$ with the product topology. Then the function $\varsigma_{\mathcal{X}}$ is measurable for all $\mathcal{X}\in\mathbb{H}^{f,\sigma}$.*

Proof. It is enough to show that $\tau_{\mathcal{X}}$ is measurable on T^n for each $n\in\mathbb{N}$. Fix therefore $n\in\mathbb{N}$.

Since $\mathbb{H}_n$ is separable (Proposition 2.6), and the space of all polynomials induces the n-pointed Gromov-weak topology on $\mathbb{H}_n$, it is enough to show that $\Phi\circ\tau_{\mathcal{X}}$ is measurable for every polynomial Φ (compare (2.10)). As for each $m\in\mathbb{N}$, $\varphi\in\mathcal{C}_b(\mathbb{R}_+^{\binom{m+n+1}{2}})$ and $\mathcal{X}=(T,\mu,\nu)$,

$$\Phi^{m,\varphi}\circ\tau_{\mathcal{X}}(\underline{u})=\int\nu^{\otimes m}(\mathrm{d}\underline{v})\mathbb{1}_{\{v_1,\dots,v_m\in[\![\underline{u}]\!]\}}\varphi\big(R^T(\underline{u},\underline{v})\big),\tag{2.38}$$

this follows from joint measurability of $(\underline{u},\underline{v})\mapsto\mathbb{1}_{\{v_1,\dots,v_m\in[\![\underline{u}]\!]\}}(\varphi\circ R^T)(\underline{u},\underline{v})$. □

We are now in a position to define the *subtree vector distribution*, $\varpi^{\mathcal{X}}$, of a bi-measure $\mathbb{R}$-tree $\mathcal{X}=(T,\mu,\nu)$ as

$$\varpi^{\mathcal{X}}:=\|\mu\|\cdot(\varsigma_{\mathcal{X}})_*\big((\mu^\circ)^{\otimes\mathbb{N}}\big)\in\mathcal{M}_f\bigg(\prod_{n\in\mathbb{N}}\mathbb{H}_n\bigg).\tag{2.39}$$

Definition 2.14 (LWV-topology). *We say that a sequence $(\mathcal{X}_N)_{N\in\mathbb{N}}$ converges to $\mathcal{X}$ in $\mathbb{H}^{f,\sigma}$ in the leaf-sampling weak vague topology* (LWV*-topology*) *if the corresponding subtree vector distributions converge, i.e.,*

$$\varpi^{\mathcal{X}_N}\underset{N\to\infty}{\Longrightarrow}\varpi^{\mathcal{X}},\tag{2.40}$$

where convergence is weak convergence of finite measures on $\prod_n(\mathbb{H}_n,\mathrm{pGw})$.

Remark 2.15. *Obviously, $\mathbb{H}^{K,\sigma}$ is closed in $\mathbb{H}^{f,\sigma}$ with LWV-topology, $\mathbb{H}^{f,\sigma}=\bigcup_{K\in\mathbb{N}}\mathbb{H}^{K,\sigma}$, and for every compact set $\mathbb{K}\subseteq\mathbb{H}^{f,\sigma}$ there exists $K\in\mathbb{N}$ with $\mathbb{K}\subseteq\mathbb{H}^{K,\sigma}$.*

Remark 2.16 (Relation with Gromov-weak topology).

(i) *LWV-convergence of* $(T_N, \mu_N, \nu_N)_{N\in\mathbb{N}}$ *implies Gromov-weak convergence of* $(T_N, \mu_N)_{N\in\mathbb{N}}$.
(ii) *Gromov-weak convergence of* $(T_N, \mu_N)_{N\in\mathbb{N}}$ *does not imply LWV-convergence of* $(T_N, \mu_N, \mu_N)_{N\in\mathbb{N}}$ (*compare Example* 2.21).

Recall from Definition 2.4 and Definition 2.8 the n-pointed Gromov-weak topology (pGw) and the subtree Gromov-weak topology (sGw), respectively. Let $(U_{N,k})_{k\in\mathbb{N}}$ be an i.i.d. sequence of μ_N°-distributed random variables, and $\underline{U}_N^n := (U_{N,1}, \dots, U_{N,n})$. The definition of LWV-convergence requires, in addition to convergence of $\|\mu_N\|$, the *joint* convergence in law with respect to the pGw-topology of $\tau_{\mathcal{X}_N}(\underline{U}_N^n)$, $n \in \mathbb{N}$. The next proposition shows that we can, on one hand, weaken this requirement to individual convergence of all $\tau_{\mathcal{X}_N}(\underline{U}_N^n)$, and, on the other hand, strengthen it to require convergence in law with respect to the sGw-topology.

Proposition 2.17 (Characterization of LWV-convergence). *Consider a sequence of bi-measure* $\mathbb{R}$*-trees* $\mathcal{X}_N = (T_N, \mu_N, \nu_N) \in \mathbb{H}^{f,\sigma}$ *and another bi-measure* $\mathbb{R}$*-tree* $\mathcal{X} \in \mathbb{H}^{f,\sigma}$ *such that* $\|\mu_N\| \to \|\mu\|$, *as* $N \to \infty$. *The three following statements are equivalent*:

(i) $\mathcal{X}_N \xrightarrow{\text{LWV}} \mathcal{X}$, *as* $N \to \infty$,
(ii) *for all* $n \in \mathbb{N}$,

$$(\tau_{\mathcal{X}_N})_*\big(\mu_N^\circ\big)^{\otimes n} \underset{N\to\infty}{\overset{\text{pGw}}{\Longrightarrow}} (\tau_{\mathcal{X}})_*\big(\mu^\circ\big)^{\otimes n}, \tag{2.41}$$

(iii) *equipping* $\prod_{n\in\mathbb{N}} \mathbb{H}_n$ *with the product topology* $\prod(\text{sGw})$,

$$(\varsigma_{\mathcal{X}_N})_*\big(\mu_N^\circ\big)^{\otimes\mathbb{N}} \underset{N\to\infty}{\overset{\prod(\text{sGw})}{\Longrightarrow}} (\varsigma_{\mathcal{X}})_*\big(\mu^\circ\big)^{\otimes\mathbb{N}}. \tag{2.42}$$

Proof. First remark that (iii) $\Rightarrow$ (i) $\Rightarrow$ (ii) is straightforward.

We prove that (ii) implies (iii). Fix therefore $n \in \mathbb{N}$. By Skorohod's representation theorem (Theorem 6.7 in [16]), there exists a list $\underline{U}^n = (U_1, \dots, U_n)$ of n i.i.d. random variables with common distribution μ° and $\underline{U}_N^n = (U_{N,1}, \dots, U_{N,n})$ i.i.d. random variables with distribution μ_N° such that

$$\tau_{\mathcal{X}_N}\big(\underline{U}_N^n\big) \underset{N\to\infty}{\overset{\text{pGw}}{\longrightarrow}} \tau_{\mathcal{X}}\big(\underline{U}^n\big), \quad \text{almost surely.} \tag{2.43}$$

In order to obtain sGw-convergence, by Lemma 2.10, it is sufficient to prove for all $I \subseteq \{1, \dots, n\}$ that $\nu_N([\![\pi_I^n(\underline{U}_N^n)]\!])$ converges weakly (as $\mathbb{R}_+$-valued random variable) to $\nu([\![\pi_I^n(\underline{U}^n)]\!])$. Because $\pi_I^n(\underline{U}_N^n)$ has the same distribution as $\underline{U}_N^{\#I}$, and similarly for $\underline{U}^n$ instead of $\underline{U}_N^n$, this follows from (2.41) for $n = \#I$, where we use that the total mass of an n-pointed measure $\mathbb{R}$-tree is continuous in the pGw-topology. Finally, we conclude from Lemma 2.10 that

$$\tau_{\mathcal{X}_N}\big(\underline{U}_N^n\big) \overset{\text{sGw}}{\longrightarrow} \tau_{\mathcal{X}}\big(\underline{U}^n\big), \quad \text{as } N \to \infty, \text{ almost surely.} \tag{2.44}$$

In particular, the one-dimensional marginals of $(\varsigma_{\mathcal{X}_N})_*(\mu_N^\circ)^{\otimes\mathbb{N}_0}$ converge as measures on $(\mathbb{H}_n, \text{sGw})$. In order to obtain convergence of laws on the product space, we have to show convergence of finite-dimensional marginals. This comes directly from the definition of sGw-convergence. □

We are now in a position to show that the subtree vector distribution characterizes bi-measure $\mathbb{R}$-trees uniquely.

Proposition 2.18 (Reconstruction theorem for $\mathbb{H}^{f,\sigma}$**).** *If* $\mathcal{X}, \mathcal{X}' \in \mathbb{H}^{f,\sigma}$ *are such that* $\varpi^{\mathcal{X}} = \varpi^{\mathcal{X}'}$, *then* $\mathcal{X} = \mathcal{X}'$.

Proof. Let $\mathcal{X} = (T, \mu, \nu)$, $\mathcal{X}' = (T', \mu', \nu') \in \mathbb{H}^{f,\sigma}$ with $\varpi^{\mathcal{X}} = \varpi^{\mathcal{X}'}$. It follows immediately that $\|\mu\| = \|\mu'\|$. Assume w.l.o.g. that $\|\mu\| = \|\mu'\| = 1$.

We will first adapt Vershik's proof of Gromov's reconstruction theorem for metric measure spaces to show that $(T,\mu)=(T',\mu')$ (compare [34], Subsection $3\frac{1}{2}.7$). Recall that a sequence $\underline{u}=(u_n)_{n\in\mathbb{N}}$ in T is called *μ-uniformly distributed* if

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{u_i}\underset{n\to\infty}{\Longrightarrow}\mu, \tag{2.45}$$

and note that, due to separability of T, $\mu^{\otimes\mathbb{N}}$-almost every sequence is μ-uniformly distributed (see, for example, [25], Theorem 11.4.1).

Of course, the corresponding statement is also true for μ' instead of μ, and as $\varpi^{\mathcal{X}}=\varpi^{\mathcal{X}'}$, we can find a μ-uniformly distributed sequence $\underline{u}=(u_n)_{n\in\mathbb{N}}$ in T, and a μ'-uniformly distributed sequence $\underline{u}'=(u'_n)_{n\in\mathbb{N}}$ in T' with $\varsigma_{\mathcal{X}}(\underline{u})=\varsigma_{\mathcal{X}'}(\underline{u}')$.

Put $u_0:=\rho$, $u'_0:=\rho'$. Then $f(u_k):=u'_k$, for all $k\in\mathbb{N}_0$, defines a root-preserving isometry from $\{u_0,u_1,\dots\}$ onto $\{u'_0,u'_1,\dots\}$, which can be extended to an isometry (still denoted by f) from $[\![\mathrm{supp}(\mu)]\!]$ onto $[\![\mathrm{supp}(\mu')]\!]$ (see Remark 2.7). Because the sequences are uniformly distributed and f_* is continuous,

$$\begin{aligned}f_*(\mu)&=f_*\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\delta_{u_i}\right)=\lim_{n\to\infty}f_*\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{u_i}\right)\\&=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\delta_{u'_i}=\mu'.\end{aligned} \tag{2.46}$$

We still need to show that $f_*(\nu')=\nu$ (on the μ-skeleton), or equivalently, $f_*(\nu')(S)=\nu(S)$ for all finite trees $S\subseteq\mathrm{Sk}_\mu(T)$. By definition of $\mathrm{Sk}_\mu(T)$ and the fact that $(u_n)_{n\in\mathbb{N}}$ is uniformly distributed, we have $S\subseteq[\![\underline{u}^n]\!]$ for sufficiently large n. Because $([\![\underline{u}^n]\!],\underline{u}^n,\nu)$ and $([\![\underline{u}'^n]\!],\underline{u}'^n,\nu')$ are equivalent as n-pointed metric measure spaces, $f_*(\nu')\restriction_{[\![\underline{u}^n]\!]}=\nu\restriction_{[\![\underline{u}^n]\!]}$. □

We can now immediately conclude that $\mathbb{H}^{f,\sigma}$ is separable and metrizable. We are not able to come up, however, with a complete metric. "Polishness" of the state space will not be used throughout the paper.

Corollary 2.19 (Separability & metrizability). *The space $\mathbb{H}^{f,\sigma}$ equipped with the LWV-topology is separable and metrizable.*

Proof. As the map which sends a bi-measure $\mathbb{R}$-tree to its subtree vector distribution is injective, we can identify $\mathbb{H}^{f,\sigma}$ with a subspace of $\mathcal{M}_f(\prod_{n\in\mathbb{N}}\mathbb{H}_n)$. $\mathbb{H}_n$ is separable, metrizable according to Proposition 2.6, hence the same holds for the countable product and the space of finite measures on it (with weak topology). □

It is important to note that μ and ν play different rôles in the LWV-topology, even if ν happens to be finite and μ is supported on the skeleton. While the convergence is *weak* with respect to μ, it is *vague* with respect to ν in the sense that the total ν-mass is not preserved under convergence, but mass may get lost in the limit. We give two examples of this phenomenon.

Example 2.20. *Consider the (finite) $\mathbb{R}$-tree shown in Figure 1 and define the probability measures $\mu_N:=(1-\frac{1}{N})\delta_{y_1}+\frac{1}{N}\delta_{y_2}$. Then (T,μ_N) converges Gromov-weakly to $(\{\rho,y_1\},\delta_{y_1})$. We endow (T,μ_N) with a constant measure $\nu:=\delta_w$, then (T,μ_N,δ_w) converges in the LWV-topology to $(\{\rho,y_1\},\delta_{y_1},0)$.*

Example 2.21 (Figure 2). *We define a sequence of $\mathbb{R}$-trees*

$$T_N:=\{\rho,z,y,x_1,x_2,\dots,x_N\}$$

shown in Figure 2 where $r_N(\rho,z)=r_N(z,y)=1$ and $r(z,x_i)=\frac{1}{N}$, for all $i=1,\dots,N$. We define a probability measure μ_N on the leaves of T_N by $\mu_N=\lambda\delta_y+(1-\lambda)\sum_i\frac{1}{N}\delta_{x_i}$, then (T_N,μ_N) converges Gromov-weakly to

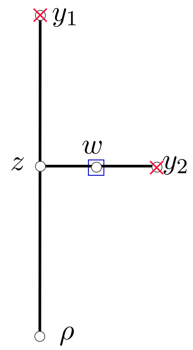

Fig. 1. A constant mass of the pruning measure at w gets lost in the limit if the mass of the sampling measure at y_2 tends to zero. The crosses $\times$ are μ-masses and the squares $\square$ are ν-masses.

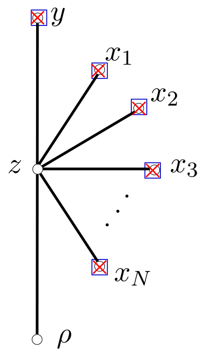

Fig. 2. If N tends to ∞ and the distances from x_i to z tend to zero, the masses of the pruning measure at x_i get lost in the limit, while the masses of the sampling measure at x_i are preserved.

$(\{\rho, z, y\}, \lambda\delta_y + (1-\lambda)\delta_z)$. *If we endow this measure* $\mathbb{R}$*-tree with the measure* $\nu_N = \mu_N$, *then* (T_N, μ_N, ν_N) *converges in the LWV-topology to* $(\{\rho, z, y\}, \lambda\delta_y + (1-\lambda)\delta_z, \lambda\delta_y)$.

2.4. *Convergence determining classes for the LWV-topology*

In this subsection, we introduce important classes of test functions and use them to obtain several convergence results. Namely, we consider functions $\Psi = \Psi^{\gamma,n,\Phi} : \mathbb{H}^{f,\sigma} \to \mathbb{R}$ of the form

$$\Psi(\mathcal{X}) := \Psi^{\gamma,n,\Phi}(\mathcal{X}) := \gamma\big(\|\mu\|\big) \cdot \int_{T^n} \mu^{\otimes n}(\mathrm{d}\underline{u})\Phi\big(\tau_{\mathcal{X}}(\underline{u})\big), \tag{2.47}$$

where $\gamma \in \mathcal{C}_b(\mathbb{R}_+)$ and $\Phi \in \mathcal{C}_b(\mathbb{H}_n)$.

Recall Π_n and $\widetilde{\Pi}_n$ from (2.12) and (2.33). As we will see later, the following subspaces of test functions are helpful in characterizing LWV-convergence. Put

$$\mathcal{F} := \big\{\Psi^{1,n,\Phi} \mid \Phi \in \Pi_n\big\}, \tag{2.48}$$

and

$$\tilde{\mathcal{F}}^1 := \big\{\Psi^{1,n,\tilde{\Phi}} \mid \tilde{\Phi} \in \widetilde{\Pi}_n\big\}, \tag{2.49}$$

and

$$\tilde{\mathcal{F}} := \Big\{\Psi^{\gamma,n,\tilde{\Phi}} \mid \tilde{\Phi} \in \widetilde{\Pi}_n, \lim_{x\to\infty} x^k\gamma(x) = 0\ \forall k \in \mathbb{N}\Big\}. \tag{2.50}$$

Lemma 2.22 (LWV-convergence via test functions). *Both $\mathcal{F}$ and $\tilde{\mathcal{F}}$ induce the LWV-topology, i.e., for a sequence of bi-measure $\mathbb{R}$-trees $\mathcal{X}_N \in \mathbb{H}^{f,\sigma}$ and another bi-measure $\mathbb{R}$-tree $\mathcal{X} \in \mathbb{H}^{f,\sigma}$, the following statements are equivalent*:

(i) $\mathcal{X}_N \xrightarrow{\text{LWV}} \mathcal{X}$, *as* $N \to \infty$,
(ii) $\Psi(\mathcal{X}_N) \to \Psi(\mathcal{X})$, *as* $N \to \infty$, *for all* $\Psi \in \mathcal{F}$,
(iii) $\tilde{\Psi}(\mathcal{X}_N) \to \tilde{\Psi}(\mathcal{X})$, *as* $N \to \infty$, *for all* $\tilde{\Psi} \in \tilde{\mathcal{F}}$.

Proof. The equivalence of (i) and (ii) is clear, as by Proposition 2.17, LWV-convergence is equivalent to the convergence of $\|\mu_N\| \to \|\mu\|$ together with $\langle (\tau_{\mathcal{X}_N})_*(\mu_N^\circ)^{\otimes n}, f\rangle \to \langle (\tau_{\mathcal{X}})_*(\mu^\circ)^{\otimes n}, f\rangle$, as $N \to \infty$, for all $n \in \mathbb{N}$ and for a class of functions f which determine the n-pointed Gromov-weak convergence. Moreover, by Proposition 2.6, Π_n is such a convergence determining class. As $\Psi(\mathcal{X}_N) = \langle (\tau_{\mathcal{X}_N})_*(\mu_N)^{\otimes n}, \Phi\rangle$, the claim follows.

By Proposition 2.17, $\tilde{\mathcal{F}}$ contains only functions which are continuous with respect to the LWV-topology, and thus (i) clearly implies (iii). To see that (iii) implies (ii), note that for $\gamma(x) := \mathrm{e}^{-x}$, convergence of $\Psi^{\gamma,0,1}(\mathcal{X}_N) = \gamma(\|\mu_N\|)$ implies convergence of $\|\mu_N\|$. Hence convergence of $\tilde{\Psi}(\mathcal{X}_N)$, for all $\tilde{\Psi} \in \tilde{\mathcal{F}}$, implies convergence of $\Psi(\mathcal{X}_N)$, for all $\Psi \in \mathcal{F}$. □

Proposition 2.23 (Convergence determining classes). *The following hold*:

(i) *The class of test functions $\tilde{\mathcal{F}}$ is convergence determining on $\mathbb{H}^{f,\sigma}$.*
(ii) *The class of test functions $\tilde{\mathcal{F}}^1$ is convergence determining on $\mathbb{H}^{K,\sigma}$ for all $K > 0$.*

Proof. We apply Theorem 6 from [17], a slight extension of Le Cam's theorem (see [38]) in the separable, metrizable case: if a set of bounded real-valued functions is multiplicatively closed and induces a separable, metrizable topology, then it is a convergence determining class with respect to this topology. By Lemma 2.22, $\tilde{\mathcal{F}}$ induces the LWV-topology, which is separable, metrizable by Corollary 2.19. We therefore need to verify that if $\tilde{\Psi}_1, \tilde{\Psi}_2 \in \tilde{\mathcal{F}}$, then $\tilde{\Psi}_1 \cdot \tilde{\Psi}_2 \in \tilde{\mathcal{F}}$. Let $\tilde{\Psi}_i = \Psi^{\gamma_i, n_i, \tilde{\Phi}_i}$ for some $n_i \in \mathbb{N}_0$, $\gamma_i \in \mathcal{C}_b(\mathbb{R}_+)$ with $\lim_{x\to\infty} x^k \gamma_i(x) = 0$, for all $k \in \mathbb{N}$, and $\tilde{\Phi}_i \in \widetilde{\Pi}_n$, $i = 1, 2$. Then

$$\begin{aligned}&\Psi^{\gamma_1,n_1,\tilde{\Phi}_1} \cdot \Psi^{\gamma_2,n_2,\tilde{\Phi}_2}(\mathcal{X}) \\ &\quad = (\gamma_1\gamma_2)\big(\|\mu\|\big) \cdot \int_{T^{n_1+n_2}} \mu^{\otimes(n_1+n_2)}(\mathrm{d}\underline{u}_1, \mathrm{d}\underline{u}_2)\tilde{\Phi}_1\big([\![\underline{u}_1]\!], \underline{u}_1, \nu\big)\tilde{\Phi}_2\big([\![\underline{u}_2]\!], \underline{u}_2, \nu\big).\end{aligned} \tag{2.51}$$

For $\underline{u} = (\underline{u}_1, \underline{u}_2)$, let $\tilde{\Phi}([\![\underline{u}]\!], \underline{u}, \nu) := \tilde{\Phi}_1([\![\underline{u}_1]\!], \underline{u}_1, \nu) \cdot \tilde{\Phi}_2([\![\underline{u}_2]\!], \underline{u}_2, \nu)$. As $\underline{u}_1$ and $\underline{u}_2$ are sublists of $\underline{u}$, $\tilde{\Phi} \in \widetilde{\Pi}_n$ and therefore $\tilde{\Psi}_1 \cdot \tilde{\Psi}_2 \in \tilde{\mathcal{F}}$.

To get the second statement in the same way, note that functions $\Psi^{1,n,\tilde{\Phi}} \in \tilde{\mathcal{F}}^1$ are bounded on $\mathbb{H}^{K,\sigma}$. □

An important fact about the LWV-topology is that Gromov-weak convergence of measure $\mathbb{R}$-trees implies LWV-convergence if the trees are additionally equipped with their respective length measures (see Example 2.24 for a definition of length measure and Proposition 2.25 for the statement). We obtain the same also for a slightly more general class of measures. Given a family $(T_i, \mu_i)_{i\in I}$ of measure $\mathbb{R}$-trees, we say that a family $(\nu_i)_{i\in I}$ of measures on respective T_i *depends continuously on the distances* if, for all $n \in \mathbb{N}$, there exists a continuous mapping $F_n : \mathbb{R}^{\binom{n+1}{2}} \to \mathbb{H}_n$, where $\mathbb{H}_n$ is endowed with the pGw-topology, such that

$$\big([\![\underline{u}]\!], \underline{u}, \nu_i\big) = F_n\big(R^{T_i}(\underline{u})\big), \quad \forall \underline{u} \in T_i^n, \forall i \in I. \tag{2.52}$$

Example 2.24 (Length measure). *The length measure, λ_T, on a separable 0-hyperbolic and connected metric space T generalizes the Lebesgue measure on $\mathbb{R}$ in an obvious way* (*compare* [31]). *Recall the set of leaves of T from* (2.18). *The length measure can be defined by the following two requirements*:

$$\forall x, y \in T: \quad \lambda_T\big([x, y]\big) = r(x, y) \quad \textit{and} \quad \lambda_T\big(\mathrm{Lf}(T)\big) = 0. \tag{2.53}$$

Obviously, the family of length measures $(\lambda_T)_{T\in\{\mathbb{R}\text{-trees}\}}$ depends continuously on the distances. The same is true if we replace λ_T by $\nu_T = f_T \cdot \lambda_T$, where f_T is a density that depends only on the height, i.e., $f_T(v) := h(r(\rho, v))$ for a bounded measurable function h (*which does not depend on T*).

We can relax the continuity of the F_n, $n \in \mathbb{N}$, a little. Let $(T, \mu) \in \mathbb{H}$. We say that a family $(\nu_i)_{i\in I}$ as above *depends $\boldsymbol{v}^{(T,\mu)}$-almost continuously on the distances* if it satisfies (2.52) with functions F_n that are not necessarily continuous, but where the set of discontinuity points is a null set with respect to the distance matrix distribution induced by (T, μ), i.e. $(R^T)_*\mu^{\otimes n}(\mathrm{Discont}(F_n)) = 0$.

Proposition 2.25 (LWV-convergence from Gromov-weak convergence). *Consider a sequence $(\mathcal{X}_N)_{N\in\mathbb{N}} := (T_N, \mu_N, \nu_N)_{N\in\mathbb{N}}$ and $\mathcal{X}_\infty := (T_\infty, \mu_\infty, \nu_\infty)$ in $\mathbb{H}^{f,\sigma}$ such that the measures $\nu_\infty, \nu_1, \nu_2, \dots$ depend $\boldsymbol{v}^{(T_\infty,\mu_\infty)}$-almost continuously on the distances.*

If $(T_N, \mu_N) \xrightarrow[N\to\infty]{\mathrm{Gw}} (T_\infty, \mu_\infty)$, then

$$(T_N, \mu_N, \nu_N) \xrightarrow[N\to\infty]{\mathrm{LWV}} (T_\infty, \mu_\infty, \nu_\infty). \tag{2.54}$$

In particular, the embedding defined by

$$\begin{aligned} &\mathbb{H} \to \mathbb{H}^{f,\sigma}, \\ &(T, \mu) \mapsto (T, \mu, \lambda_T), \end{aligned} \tag{2.55}$$

where λ_T is the length measure, is a homeomorphism onto its image.

Proof. Given $n \in \mathbb{N}$, fix a function $F_n : \mathbb{R}_+^{\binom{n+1}{2}} \to \mathbb{H}_n$ as in (2.52), such that the set of discontinuity points of F_n is a zero set with respect to $(R^{T_\infty})_*(\mu_\infty^{\otimes n})$. For $N \in \mathbb{N} \cup \{\infty\}$, let $\underline{U}_N$ be a random vector in T_N^n with distribution $(\mu_N^\circ)^{\otimes n}$. Then the assumed Gromov-weak convergence means that $\|\mu_N\| \to \|\mu_\infty\|$ and

$$R^{T_N}(\underline{U}_N) \underset{N\to\infty}{\overset{\mathcal{L}}{\Longrightarrow}} R^{T_\infty}(\underline{U}_\infty), \tag{2.56}$$

where $\overset{\mathcal{L}}{\Rightarrow}$ denotes convergence in law. By the continuous mapping theorem (see Theorem 5.1 in [16]), we obtain

$$\big([\![\underline{U}_N]\!], \underline{U}_N, \nu_N\big) = F_n\big(R^{T_N}(\underline{U}_N)\big) \underset{N\to\infty}{\overset{\mathcal{L}}{\Longrightarrow}} F_n\big(R^{T_\infty}(\underline{U}_\infty)\big) = \big([\![\underline{U}_\infty]\!], \underline{U}_\infty, \nu_\infty\big). \tag{2.57}$$

Using that $([\![\underline{U}_N]\!], \underline{U}_N, \nu_N)$ has law $(\tau_{\mathcal{X}_N})_*((\mu_N^\circ)^{\otimes n})$ for $N \in \mathbb{N} \cup \{\infty\}$, the claimed LWV-convergence $\mathcal{X}_N \xrightarrow{\mathrm{LWV}} \mathcal{X}_\infty$ now follows from Proposition 2.17. That (2.55) defines a homeomorphism onto its image is now obvious, because the length measure depends continuously on the distances (see Example 2.24). □

Corollary 2.26 (Sampling measure perturbation). *Consider two sequences of bi-measure $\mathbb{R}$-trees $\mathcal{X}_N^i := (T_N, \mu_N^i, \nu_N)$, $i = 1, 2$ that differ by their sampling measures μ_N^1 and μ_N^2. Assume that $\mathcal{X}_N^1 \xrightarrow[N\to\infty]{\mathrm{LWV}} \mathcal{X}$, and that the pruning measures $(\nu_N)_{N\in\mathbb{N}}$ depend $\boldsymbol{v}^{(T,\mu)}$-almost continuously on the distances. If $d_{\mathrm{Pr}}(\mu_N^1, \mu_N^2) \xrightarrow[N\to\infty]{} 0$, then also $\mathcal{X}_N^2 \xrightarrow[N\to\infty]{\mathrm{LWV}} \mathcal{X}$.*

Proof. As $\mathcal{X}_N^1 \xrightarrow[N\to\infty]{\mathrm{LWV}} \mathcal{X}$, implies that $(T_N, \mu_N^1) \xrightarrow[N\to\infty]{} (T, \mu)$ in the Gw-topology, we get $d_{\mathrm{pGP}}((T_N, \mu_N^1), (T, \mu)) \xrightarrow[N\to\infty]{} 0$ by Proposition 2.6. Since μ_N^1 and μ_N^2 are defined on the same space T_N, the latter implies that also

$$\lim_{N\to\infty} d_{\mathrm{pGP}}\big((T_N, \mu_N^2), (T, \mu)\big) = 0 \tag{2.58}$$

(compare (2.14)). Proposition 2.25 allows us to endow these metric measure spaces with the associated measures ν_N and some ν_∞ on T, defined by (2.52). Because of uniqueness of LWV-limits, we have $(T, \mu, \nu_\infty) = (T, \mu, \nu)$. □

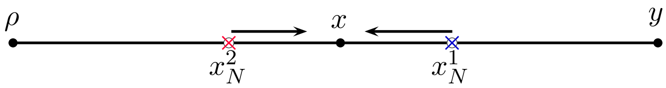

Fig. 3. The tree T_N with the two sequences x_N^1 and x_N^2 that converge to x.

Example 2.27 (Counterexample). *We cannot extend the result of Corollary 2.26 to pruning measures which do not depend only on the distances.*

As illustrated in Figure 3, we consider a constant rooted metric space T and two fixed points $x, y \in T$ such that $x \in [\rho, y]$. We construct two sequences of points $(x_N^1)_{N\in\mathbb{N}}$ and $(x_N^2)_{N\in\mathbb{N}}$ that converge to x, the first from above, the second from below; i.e. $x_N^1 \in [x, y]$ and $x_N^2 \in [\rho, x]$ for all $N \in \mathbb{N}$, and $r(x_N^i, x) \underset{N\to\infty}{\longrightarrow} 0$ for $i = 1, 2$. We then define the two sequences of measures $\mu_N^i := \frac{1}{2}\delta_{x_N^i} + \frac{1}{2}\delta_y$ for $i = 1, 2$ and a constant measure $\nu_N = \nu = \delta_x$. Clearly, $\mathcal{X}_N^1 \overset{\text{LWV}}{\underset{N\to\infty}{\longrightarrow}} \mathcal{X}$ and $d_{\mathrm{Pr}}(\mu_N^1, \mu_N^2) \underset{N\to\infty}{\longrightarrow} 0$, but the sequence (T, μ_N^2, ν) does not converge, since the subtree $[\rho, x_N^2]$ never contains the point x, except at the limit. Thus $([\rho, x_N^2], \{x_N^2\}, \nu)$ does not converge pointed Gromov-weakly.

Lemma 2.28 (Sum of pruning measures). *Let $\mathcal{X}_N^i = (T_N, \mu_N, \nu_N^i) \in \mathbb{H}^{f,\sigma}$ with $(T_N, \mu_N, \nu_N^i) \overset{\text{LWV}}{\longrightarrow} \mathcal{X}^i = (T, \mu, \nu^i) \in \mathbb{H}^{f,\sigma}$, as $N \to \infty$, for $i = 1, 2$. If $(\nu_N^1)_{N\in\mathbb{N}}$ depends $\boldsymbol{\upsilon}^{(T,\mu)}$-almost continuously on the distances, we obtain*

$$\big(T_N, \mu_N, \nu_N^1 + \nu_N^2\big) \overset{\text{LWV}}{\underset{N\to\infty}{\longrightarrow}} \big(T, \mu, \nu^1 + \nu^2\big). \tag{2.59}$$

Proof. Fix $n \in \mathbb{N}$. Because $(\nu_N^1)_{N\in\mathbb{N}}$ depends $\boldsymbol{\upsilon}^{(T,\mu)}$-almost continuously on the distances, we can choose F_n as in (2.52). Let $\underline{U}_N$, $\underline{U}$ be random variables with distribution $\mu_N^{\otimes n}$, $\mu^{\otimes n}$, respectively. By the LWV-convergence and the Skorohod representation theorem, we can couple them such that $\tau_{\mathcal{X}_N^2}(\underline{U}_N) \overset{\text{pGw}}{\underset{N\to\infty}{\longrightarrow}} \tau_{\mathcal{X}^2}(\underline{U})$, a.s., which implies $R^{T_N}(\underline{U}_N) \underset{N\to\infty}{\longrightarrow} R^T(\underline{U})$. Because $R^T(\underline{U})$ is a.s. a continuity point of F_n, we also have

$$\tau_{\mathcal{X}_N^1}(\underline{U}_N) = F_n \circ R^{T_N}(\underline{U}_N) \overset{\text{pGw}}{\underset{N\to\infty}{\longrightarrow}} F_n \circ R^T(\underline{U}) = \tau_{\mathcal{X}^1}(\underline{U}), \quad \text{a.s.} \tag{2.60}$$

As explained in Remark 2.11, we can define functions $f_N : [\![\underline{U}_N]\!] \to [\![\underline{U}]\!]$ such that a.s. $f_N(\underline{U}_N) = \underline{U}$ for large enough N, $\mathrm{dis}(f_N) \underset{N\to\infty}{\longrightarrow} 0$, and $(f_N)_*(\nu_N^i) \underset{N\to\infty}{\Longrightarrow} \nu^i$. Then also $f_{N*}(\nu_N^1 + \nu_N^2) \underset{N\to\infty}{\Longrightarrow} \nu^1 + \nu^2$, which implies $([\![\underline{U}_N]\!], \underline{U}_N, \nu_N^1 + \nu_N^2) \overset{\text{pGw}}{\underset{N\to\infty}{\longrightarrow}} ([\![\underline{U}]\!], \underline{U}, \nu^1 + \nu^2)$, a.s. By Proposition 2.17, this implies the claimed LWV-convergence. □

Remark 2.29 (Assumption on $\boldsymbol{\upsilon}^{(T,\mu)}$-almost continuity is important). *In Lemma 2.28, we cannot drop the assumption that one of the measures depends $\boldsymbol{\upsilon}^{(T,\mu)}$-almost continuously on the distances, because then we cannot use the same coupling of $\underline{U}_N$ to get almost sure convergence of $\tau_{\mathcal{X}_N^i}(\underline{U}_N)$ for $i = 1$ and for $i = 2$.*

If we get LWV-convergence of a sequence of bi-measure $\mathbb{R}$-trees, the following lemma asserts that the limit is stable under a small perturbation of ν_N in a certain sense.

Lemma 2.30 (Pruning measure perturbation). *Consider two sequences of bi-measure $\mathbb{R}$-trees $\mathcal{X}_N^i := (T_N, \mu_N, \nu_N^i)$, $i = 1, 2$ that differ by their pruning measures ν_N^1 and ν_N^2. If the two pruning measures are Prohorov merging on subtrees sampled by $\mu_N^{\otimes n}$, i.e.,*

$$\lim_{N\to\infty} d_{\mathrm{Pr}}\big(\nu_N^1\restriction_{[\![U_N^n]\!]}, \nu_N^2\restriction_{[\![U_N^n]\!]}\big) = 0, \quad \mu_N^{\otimes n}\text{-a.s.}, \forall n \in \mathbb{N}, \tag{2.61}$$

then $\mathcal{X}_N^1 \overset{\text{LWV}}{\underset{N\to\infty}{\longrightarrow}} \mathcal{X}$, for some $\mathcal{X} = (T, \mu, \nu)$, implies $\mathcal{X}_N^2 \overset{\text{LWV}}{\underset{N\to\infty}{\longrightarrow}} \mathcal{X}$.

Proof. Let $\underline{U}_N$ and $\underline{U}$ be sequences of independent μ_N- and μ-distributed random variables in T_N and T, respectively. Because ν^1 and ν^2 are defined on the same measure $\mathbb{R}$-tree, the Prohorov distance in (2.61) is an upper bound for the pGP-distance, and we obtain

$$\begin{aligned} &d_{\mathrm{pGP}}\big(\tau_{\mathcal{X}_N^2}\big(\underline{U}_N^n\big), \tau_{\mathcal{X}}\big(\underline{U}^n\big)\big) \\ &\quad \le d_{\mathrm{Pr}}\big(\nu_N^1\restriction_{[\![\underline{U}_N^n]\!]}, \nu_N^2\restriction_{[\![\underline{U}_N^n]\!]}\big) + d_{\mathrm{pGP}}\big(\tau_{\mathcal{X}_N^1}\big(\underline{U}_N^n\big), \tau_{\mathcal{X}}\big(\underline{U}^n\big)\big) \underset{N\to\infty}{\longrightarrow} 0, \end{aligned} \tag{2.62}$$

almost surely, for all $n \in \mathbb{N}$. This implies $(\tau_{\mathcal{X}_N^2})_*(\mu_N^\circ)^{\otimes n} \overset{\mathrm{pGw}}{\Longrightarrow} (\tau_{\mathcal{X}})_*(\mu^\circ)^{\otimes n}$ for all $n \in \mathbb{N}$, and Proposition 2.17 gives the LWV-convergence. □

We conclude this section by giving a simple, sufficient (but far from necessary) condition for relative compactness of a set $\mathbb{K} \subseteq \mathbb{H}^{f,\sigma}$. Assume that for all $\mathcal{X}' = (T', \mu', \nu') \in \mathbb{K}$, there is an isometric embedding of T' into some common $\mathbb{R}$-tree T, and there are measures μ and ν on T dominating all the (push forwards of) μ' and ν', respectively. Further assume that $\mathcal{X} := (T, \mu, \nu) \in \mathbb{H}^{f,\sigma}$. In other words,

$$\mathbb{K} \subseteq \mathbb{S}_{\mathcal{X}} := \big\{\big(T, \mu', \nu'\big) \in \mathbb{H}^{f,\sigma} \mid \mu' \le \mu, \nu' \le \nu\big\}. \tag{2.63}$$

Then $\mathbb{K}$ is relatively compact, as the following lemma shows.

Lemma 2.31 (Compactness of $\mathbb{S}_{\mathcal{X}}$). *Let $\mathcal{X} = (T, \mu, \nu) \in \mathbb{H}^{f,\sigma}$. Then $\mathbb{S}_{\mathcal{X}}$, defined in (2.63), is compact in the LWV-topology.*

Proof. Consider measures $\mu_N \le \mu$, $\nu_N \le \nu$, $N \in \mathbb{N}$. We have to find a subsequence of $\mathcal{X}_N := (T, \mu_N, \nu_N)$ that converges in $\mathbb{S}_{\mathcal{X}}$. Fix finite subtrees $T_n \subseteq T$, $n \in \mathbb{N}$, with $T_n \subseteq T_{n+1}$ and $\bigcup_{n\in\mathbb{N}} T_n \supseteq \mathrm{Sk}_\mu(T)$.

Because the family $(\mu_N)_{N\in\mathbb{N}}$ is uniformly σ-additive and norm bounded, there exists a setwise convergent subsequence [18], Thm. 4.7.25. Assume w.l.o.g. that there is $\mu_\infty \in \mathcal{M}_f(T)$ with $\mu_N(A) \underset{N\to\infty}{\longrightarrow} \mu_\infty(A)$ for all measurable $A \subseteq T$. Similarly, using Cantor's diagonalization argument, we may assume that $\nu_N\restriction_{T_n}$ converges setwise to some $\hat{\nu}_n \in \mathcal{M}_f(T_n)$, for every $n \in \mathbb{N}$. Define

$$\nu_\infty(A) := \sup_{n\in\mathbb{N}} \hat{\nu}_n\big(T_n \cap A \cap \mathrm{Sk}_{\mu_\infty}(T)\big). \tag{2.64}$$

Because $\hat{\nu}_n\restriction_{T_{n-1}} = \hat{\nu}_{n-1}$, we can easily check that ν_∞ is a measure on T and $\mathcal{X}_\infty := (T, \mu_\infty, \nu_\infty) \in \mathbb{S}_{\mathcal{X}}$. Furthermore, for measurable $A \subseteq \mathrm{Sk}_{\mu_\infty}(T) \subseteq \mathrm{Sk}_\mu(T) \subseteq \bigcup_{n\in\mathbb{N}} T_n$, we obtain

$$\nu_\infty(A) = \sup_{n\in\mathbb{N}} \lim_{N\to\infty} \nu_N(A \cap T_n) \begin{cases} \le \liminf_{N\to\infty} \nu_N(A), \\ \ge \limsup_{N\to\infty} \nu_N(A) - \sup_{n\in\mathbb{N}} \nu(A \setminus T_n). \end{cases} \tag{2.65}$$

Using $A \subseteq \bigcup_{n\in\mathbb{N}} T_n$, this implies

$$\nu_\infty(A) = \lim_{N\to\infty} \nu_N(A). \tag{2.66}$$

We shall show that $\mathcal{X}_N \underset{N\to\infty}{\overset{\mathrm{LWV}}{\longrightarrow}} \mathcal{X}_\infty$. By Lemma 2.22, it is enough to show that $\Psi(\mathcal{X}_N) \to \Psi(\mathcal{X}_\infty)$ for all $\Psi \in \mathcal{F}$. Let

$$G := \big\{\underline{u} \in T^n \colon \nu\big([\![\underline{u}]\!] \setminus \mathrm{Sk}_{\mu_\infty}(T)\big) = 0\big\}. \tag{2.67}$$

Fix $\Psi = \Psi^{n,\Phi} \in \mathcal{F}$. Then (2.66) implies

$$\Phi \circ \tau_{\mathcal{X}_N}(\underline{u}) \underset{N\to\infty}{\longrightarrow} \Phi \circ \tau_{\mathcal{X}_\infty}(\underline{u}) \quad \forall \underline{u} \in G, \tag{2.68}$$

and with $B := T^n \setminus G$ we estimate

$$\begin{aligned}\left|\Psi(\mathcal{X}_N) - \Psi(\mathcal{X}_\infty)\right| &\le \mu_N^{\otimes n}(B) 2\|\Phi\|_\infty + \int_G |\Phi \circ \tau_{\mathcal{X}_N} - \Phi \circ \tau_{\mathcal{X}_\infty}| \, \mathrm{d}\mu_N^{\otimes n} \\ &\quad + \int |\Phi \circ \tau_{\mathcal{X}_\infty}| \, \mathrm{d}\big(\mu_N^{\otimes n} - \mu_\infty^{\otimes n}\big).\end{aligned} \tag{2.69}$$

The last term converges to zero because of the setwise convergence of μ_N to μ_∞, and the second term is bounded by $\int_G |\Phi \circ \tau_{\mathcal{X}_N} - \Phi \circ \tau_{\mathcal{X}_\infty}| \, \mathrm{d}\mu^{\otimes n}$, which converges to zero according to (2.68), using the dominated convergence theorem.

For every $(u_1, \ldots, u_n) \in \mathrm{supp}(\mu_\infty)^n \setminus G$, there is an index $k \in \{1, \ldots, n\}$ with $u_k \in \mathrm{At}(\nu) \setminus \mathrm{At}(\mu_\infty)$, where At denotes the set of atoms of a measure. Because $\mathrm{At}(\nu)$ is countable, this implies that B is a μ_∞-null set. Again using setwise convergence of μ_N, we obtain

$$\lim_{N\to\infty} \mu_N^{\otimes n}(B) = \mu_\infty^{\otimes n}(B) = 0. \qquad \square$$

3. The pruning process

In this section, we present the construction of the bi-measure valued pruning process, $(X_t)_{t\ge 0}$. In Section 3.1, we carry out an explicit construction given a realization of the Poisson point process which gives rise to a càdlàg path. We continue the construction in Section 3.2 by adding randomness and establishing that the stochastic process obtained this way has the strong Markov property. In Section 3.3, we establish the Feller property from which we can conclude that the law of the pruning process on Skorohod space is weakly continuous in the initial distribution on bi-measure $\mathbb{R}$-trees. Finally, in Section 3.4 we give an analytic characterization via the infinitesimal generator.

3.1. *Getting the construction started*: *Pruning moves*

It is convenient to introduce randomness later and work initially in a setting where the *cut times* and *cut points* are fixed. Given a bi-measure $\mathbb{R}$-tree, $(T, \mu, \nu) \in \mathbb{H}^{f,\sigma}$, consider a subset $\pi \subseteq \mathbb{R}_+ \times T$. Although π is associated with a particular class representative, it corresponds, of course, to a similar set for any representative of the same equivalence class by mapping across using the appropriate root invariant isometry. Then the set of cut points up to time t is the projection of $\pi \cap ([0, t] \times T)$ onto the tree, i.e.

$$\pi_t := \big\{v \in T \mid \exists s \le t : (s, v) \in \pi\big\}. \tag{3.1}$$

For every $v \in T$, the *tree pruned at* v is defined by

$$T^v := \big\{w \in T \mid v \notin [\rho, w]\big\}. \tag{3.2}$$

The pruned tree at the set $\pi_t \subseteq T$, T^{π_t}, is the intersection of the trees T^v pruned at $v \in \pi_t$, i.e.,

$$T^{\pi_t} := \bigcap_{v\in\pi_t} T^v. \tag{3.3}$$

We equip the pruned tree T^{π_t} with the restrictions of the measures μ and ν. As always, we write (T^{π_t}, μ, ν) instead of $(T^{\pi_t}, \mu\restriction_{T^{\pi_t}}, \nu\restriction_{T^{\pi_t}})$ and easily verify $(T^{\pi_t}, \mu, \nu) \in \mathbb{H}^{f,\sigma}$.

Lemma 3.1 (Càdlàg paths). *Fix $\mathcal{X} = (T, \mu, \nu) \in \mathbb{H}^{f,\sigma}$ and a set $\pi \subseteq \mathbb{R}_+ \times T$. The map $t \mapsto \mathcal{X}_t := (T^{\pi_t}, \mu, \nu)$ is càdlàg with respect to the LWV-topology.*

Proof. Let $0 < s < t$. As $T^{\pi_t} \subseteq T^{\pi_s}$, we obtain for all $\Psi = \Psi^{1,n,\Phi} \in \mathcal{F}$,

$$\begin{aligned}\big|\Psi(\mathcal{X}_s) - \Psi(\mathcal{X}_t)\big| &= \left|\int_{(T^{\pi_s})^n \setminus (T^{\pi_t})^n} \Phi \circ \tau_{\mathcal{X}} \, \mathrm{d}\mu^{\otimes n}\right| \le \|\Phi\|_\infty \cdot \mu^{\otimes n}\big(\big(T^{\pi_s}\big)^n \setminus \big(T^{\pi_t}\big)^n\big) \\ &\le \|\Phi\|_\infty \cdot n \cdot \|\mu\|^{n-1} \cdot \mu\big(T^{\pi_s} \setminus T^{\pi_t}\big). \end{aligned} \tag{3.4}$$

For fixed s, $\bigcap_{t>s} T^{\pi_s} \setminus T^{\pi_t} = \varnothing$, which implies that $\mu(T^{\pi_s} \setminus T^{\pi_t}) \to 0$, as $t \to s$ from the right. Because $\mathcal{F}$ induces the LWV-topology, this implies *right continuity*.

To construct the *left limit*, define $T_{t-} := \bigcap_{0 \le s < t} T^{\pi_s} \supseteq T^{\pi_t}$ for each $t > 0$, and define $\mathcal{Y}_t := (T_{t-}, \mu, \nu)$, which is obviously an element of $\mathbb{H}^{f,\sigma}$. Similarly as before, for all $0 < s < t$ and $\Psi \in \mathcal{F}$, there exists a constant $C = C^{\Psi}$ such that

$$\big|\Psi(\mathcal{X}_s) - \Psi(\mathcal{Y}_t)\big| \le C \cdot \mu\big(T^{\pi_s} \setminus T_{t-}\big). \tag{3.5}$$

As, for fixed t, $\bigcap_{s<t} T^{\pi_s} \setminus T_{t-} = \varnothing$, $\mathcal{Y}_t$ is indeed the left limit. □

3.2. *Continuing the construction*: *Adding randomness*

In this subsection we define, given a bi-measure $\mathbb{R}$-tree $\mathcal{X} = (T, \mu, \nu)$, the pruning process of $\mathcal{X}$, where π is now the (random) Poisson point measure with intensity $\lambda \otimes \nu$ on $\mathbb{R}_+ \times T$. Here, we identify an atomic measure $\mathfrak{m}$ on T with the set $\mathrm{At}(\mathfrak{m})$ of its atoms and define

$$T^{\mathfrak{m}} := T^{\mathrm{At}(\mathfrak{m})} = \bigcap_{v \in \mathrm{At}(\mathfrak{m})} T^{v}. \tag{3.6}$$

Definition 3.2 (The pruning process). *Fix a bi-measure $\mathbb{R}$-tree $\mathcal{X} := (T, \mu, \nu) \in \mathbb{H}^{f,\sigma}$. Let $\pi^{\mathcal{X}}$ be the Poisson point measure on $\mathbb{R}_+ \times T$ with intensity measure $\lambda \otimes \nu$, where λ is the Lebesgue measure on $\mathbb{R}_+$. We define* the pruning process, *$X := (X_t)_{t \ge 0}$, as the bi-measure $\mathbb{R}$-tree-valued process obtained by pruning $X_0 := \mathcal{X}$ at the points of the Poisson point process $\pi_t(\cdot) := \pi_t^{\mathcal{X}}(\cdot) := \pi^{\mathcal{X}}([0, t] \times \cdot)$, i.e.,*

$$X_t := \big(T^{\pi_t}, \mu, \nu\big) := \big(T^{\pi_t}, \mu\restriction_{T^{\pi_t}}, \nu\restriction_{T^{\pi_t}}\big). \tag{3.7}$$

$\mathbb{E}^{\mathcal{X}}$, or $\mathbb{E}$ if there is no confusion, denotes the distribution of the process X starting from $X_0 = \mathcal{X}$.

Lemma 3.3 (Strong Markov property). *The pruning process X is a strong Markov process.*

Proof. Denote by $(\mathcal{A}_t)_{t \ge 0}$ the filtration generated by the Poisson point process $(\pi_t)_{t \ge 0}$. Note that X is adapted to this filtration. Using the strong Markov property of the Poisson process, we get for every $t \ge 0$, stopping time σ, and $\underline{u} \in T^n$, $n \in \mathbb{N}$,

$$\mathbb{P}\big(\pi_{\sigma+t}\big([\![\underline{u}]\!]\big) = 0 \mid \mathcal{A}_\sigma\big) = \mathbb{1}_{\{\pi_\sigma([\![\underline{u}]\!]) = 0\}} \mathbb{P}\big(\pi_t\big([\![\underline{u}]\!]\big) = 0\big). \tag{3.8}$$

For every $\tilde{\Psi} = \Psi^{1,n,\tilde{\Phi}} \in \tilde{\mathcal{F}}^1$, this implies

$$\begin{aligned}\mathbb{E}\big[\tilde{\Psi}(X_{\sigma+t}) \mid \mathcal{A}_\sigma\big] &= \int_{T^n} \mu^{\otimes n}(\mathrm{d}\underline{u}) \mathbb{P}\big(\pi_{\sigma+t}\big([\![\underline{u}]\!]\big) = 0 \mid \mathcal{A}_\sigma\big) \cdot \tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big) \\ &= \int_{(T^{\pi_\sigma})^n} \mu^{\otimes n}(\mathrm{d}\underline{u}) \mathrm{e}^{-t\nu([\![\underline{u}]\!])} \cdot \tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big). \end{aligned} \tag{3.9}$$

On the other hand, we also have

$$\mathbb{E}^{X_\sigma}\big[\tilde{\Psi}(X_t)\big] = \int_{(T^{\pi_\sigma})^n} \mu^{\otimes n}(\mathrm{d}\underline{u}) \mathrm{e}^{-t\nu([\![\underline{u}]\!])} \cdot \tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big). \tag{3.10}$$

Because $X_t \in \mathbb{H}^{\|\mu\|,\sigma}$, for all $t \ge 0$, and $\tilde{\mathcal{F}}^1$ is a separating class on this space, we obtain the strong Markov property. □

3.3. *Continuity of the pruning process*

In this subsection we show that the law of X_t under $\mathbb{P}^{\mathcal{X}}$ is weakly continuous in the initial value $\mathcal{X}$ for each $t \geq 0$. This property is sometimes referred to as the *Feller property* of the corresponding semigroup $(S_t)_{t\geq 0}$, although this terminology is often restricted to the case of a locally compact state space and transition operators that map the space of continuous functions that vanish at infinity into itself. In the latter, more restrictive case, the Feller property implies that the law of the whole process (as random variable on Skorohod space) depends continuously on the initial value. If S_t maps only $\mathcal{C}_b$ into itself, this is no longer the case in general, and one needs an extra argument. The pruning process $(X_t)_{t\geq 0}$, however, does depend continuously on the initial condition (Theorem 3.6).

Let $(S_t)_{t\geq 0}$ be the semi-group associated to the pruning process $(X_t)_{t\geq 0}$, i.e. for $t \geq 0$ and a bounded measurable function $G : \mathbb{H}^{f,\sigma} \to \mathbb{R}$,

$$S_t G(\mathcal{X}) := \mathbb{E}^{\mathcal{X}}\big[G(X_t)\big]. \tag{3.11}$$

Proposition 3.4 (Feller continuity). *The process $X := (X_t)_{t\geq 0}$ is Feller continuous, i.e., $S_t(\mathcal{C}_b(\mathbb{H}^{f,\sigma})) \subseteq \mathcal{C}_b(\mathbb{H}^{f,\sigma})$.*

Proof. Consider the convergence of bi-measure $\mathbb{R}$-trees $\mathcal{X}_N \xrightarrow{\text{LWV}} \mathcal{X}$. Write $K := \sup\{\|\mu_N\|, N \in \mathbb{N}\}$, then the sequence converges in $\mathbb{H}^{K,\sigma}$. Because $\tilde{\mathcal{F}}^1$ is convergence determining on $\mathbb{H}^{K,\sigma}$ (see Proposition 2.23), it is enough to prove for all $\tilde{\Psi} \in \tilde{\mathcal{F}}^1$, $t > 0$ that

$$\mathbb{E}^{\mathcal{X}_N}\big[\tilde{\Psi}(X_t)\big] \underset{N\to\infty}{\longrightarrow} \mathbb{E}^{\mathcal{X}}\big[\tilde{\Psi}(X_t)\big]. \tag{3.12}$$

Fix therefore $\tilde{\Psi} = \Psi^{1,n,\tilde{\Phi}} \in \tilde{\mathcal{F}}^1$. Then

$$\begin{aligned}\mathbb{E}^{\mathcal{X}_N}\big[\tilde{\Psi}(X_t)\big] &= \mathbb{E}^{\mathcal{X}_N}\bigg[\int_{(T_N^{\pi_t})^n} \tilde{\Phi} \circ \tau_{\mathcal{X}_N}\, \mathrm{d}\mu_N^{\otimes n}\bigg] \\ &= \int_{T_N^n} \mu_N^{\otimes n}(\mathrm{d}\underline{u})\mathbb{P}\big(\pi_t^{\mathcal{X}_N}\big([\![\underline{u}]\!]\big) = 0\big) \cdot \tilde{\Phi}\big([\![\underline{u}]\!], \underline{u}, \nu_N\big).\end{aligned} \tag{3.13}$$

Using $\mathbb{P}(\pi_t^{\mathcal{X}_N}([\![\underline{u}]\!]) = 0) = \exp(-t\nu_N([\![\underline{u}]\!]))$, we see that $\mathbb{E}^{\mathcal{X}_N}[\tilde{\Psi}(X_t)] = \tilde{\Psi}'(\mathcal{X}_N)$ for some $\tilde{\Psi}' \in \tilde{\mathcal{F}}^1$. The convergence follows therefore from the LWV-convergence of $(\mathcal{X}_N)_{N\in\mathbb{N}}$. □

Consider a separable, metrizable space E and a contraction semigroup $S = (S_t)_{t\geq 0}$ on $\mathcal{C}_b(E)$. We define

$$\mathcal{D}(S) := \Big\{ f \in \mathcal{C}_b(E) \colon \lim_{t\to 0} \|S_t f - f\|_\infty = 0 \Big\}. \tag{3.14}$$

Note that $\mathcal{D}(S)$ is uniformly closed, S_t maps $\mathcal{D}(S)$ into itself, and the restriction of $(S_t)_{t\geq 0}$ to $\mathcal{D}(S)$ is a strongly continuous contraction semigroup. In particular, the restricted semigroup has a generator $\Omega_S : \mathcal{D}(\Omega_S) \to \mathcal{D}(S)$ with dense domain $\mathcal{D}(\Omega_S) \subseteq \mathcal{D}(S)$.

Lemma 3.5. *Let E be a separable, metrizable space, and $Y^x = (Y_t^x)_{t\geq 0}$, $x \in E$, an E-valued, Feller-continuous (time-homogeneous) Markov process with càdlàg paths and semigroup $S = (S_t)_{t\geq 0}$ on $\mathcal{C}_b(E)$. Assume that there is a set $\mathcal{G} \subseteq \mathcal{D}(S)$ that is multiplicatively closed and induces the topology of E. Then the map*

$$\begin{aligned}&\mathcal{M}_1(E) \to \mathcal{M}_1\big(D_E(\mathbb{R}_+)\big), \\ &\eta \mapsto \mathcal{L}\big(Y^\eta\big)\end{aligned} \tag{3.15}$$

is continuous, where $D_E(\mathbb{R}_+)$ is the space of càdlàg paths with Skorohod topology, $\mathcal{L}$ is the law of a process, and Y^η is the process with initial condition $\mathcal{L}(Y_0^\eta) = \eta$, i.e., $\mathcal{L}(Y^\eta) = \int \mathcal{L}(Y^x)\eta(\mathrm{d}x)$.

Proof. It is sufficient to prove that $\mathcal{L}(Y^{x_N}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}(Y^x)$ for every convergent sequence $x_N \underset{N\to\infty}{\longrightarrow} x$ in E. Because $\mathcal{G}$ induces the topology of E, it strongly separates points (see Lemma 1 in [17]). According to Theorem 10 of [17], it is therefore enough to prove that for all $f_1, \ldots, f_k \in \mathcal{G}$,

$$\big(f_1\big(Y_t^{x_N}\big), \ldots, f_k\big(Y_t^{x_N}\big)\big)_{t\geq 0} \overset{\mathcal{L}}{\underset{N\to\infty}{\Longrightarrow}} \big(f_1\big(Y_t^x\big), \ldots, f_k\big(Y_t^x\big)\big)_{t\geq 0} \tag{3.16}$$

in Skorohod space as $\mathbb{R}^k$-valued processes. The assumed Feller continuity implies f.d.d. convergence, hence it is enough to prove tightness.

To this end, we apply Theorem 3.9.4 of [28]. The linear span $C_a := \operatorname{span}(\mathcal{G})$ of $\mathcal{G}$ is an algebra contained in $\mathcal{D}(S)$, and the domain $\mathcal{D}(\Omega_S)$ of the generator Ω_S of S is dense in $\mathcal{D}(S)$. For every $f \in \mathcal{D}(\Omega_S)$, we define $Z_t^N := \Omega_S f(Y_t^{x_N})$. Then the following hold:

(i) the processes $(f(Y_t^{x_N}) - \int_0^t Z_s^N \,\mathrm{d}s)_{t\geq 0}$ are martingales,
(ii) for all $T \geq 0$, $\sup_{N\in\mathbb{N}} \mathbb{E}[\operatorname{ess\,sup}_{0\leq t\leq T} |Z_t^N|] \leq \|\Omega_S f\|_\infty < \infty$.

Now tightness of the processes $(f_1(Y_t^{x_N}), \ldots, f_k(Y_t^{x_N}))_{t\geq 0}$, $N \in \mathbb{N}$, for every fixed $f_1, \ldots, f_k \in C_a \supseteq \mathcal{G}$ follows from [28], Thm. 3.9.4. □

Theorem 3.6 (Continuity in the initial distribution). *The law of X on the Skorohod space depends continuously on the initial condition.*

Proof. It is sufficient to prove continuity for deterministic initial conditions. Every convergent sequence $\mathcal{X}_N \overset{\text{LWV}}{\underset{N\to\infty}{\longrightarrow}} \mathcal{X}$ in $\mathbb{H}^{f,\sigma}$ is contained in $\mathbb{H}^{K,\sigma}$ for some $K > 0$, and the pruning process stays a.s. in that subspace. We verify the conditions of Lemma 3.5 for the $\mathbb{H}^{K,\sigma}$-valued pruning process. It has càdlàg paths (Lemma 3.1), is Feller-continuous (Proposition 3.4), and $\tilde{\mathcal{F}}^1 \subseteq \mathcal{C}_b(\mathbb{H}^{K,\sigma})$ is multiplicatively closed and induces the LWV-topology. It remains to show that $\tilde{\mathcal{F}}^1 \subseteq \mathcal{D}(S)$, where S is the $\mathcal{C}_b(\mathbb{H}^{K,\sigma})$-semigroup.

For $\tilde{\Phi} \in \tilde{\Pi}_n$, $x \in \mathbb{R}_+$, we define

$$\gamma_{\tilde{\Phi}}(x) := \sup_{(T,\underline{u},\nu)\in\mathbb{H}_n, \|\nu\|=x} \big|\tilde{\Phi}(T, \underline{u}, \nu)\big| \tag{3.17}$$

and note that $\lim_{x\to\infty} \gamma_{\tilde{\Phi}}(x) = 0$. Using Fubini's theorem, we obtain for $\tilde{\Psi} = \Psi^{1,n,\tilde{\Phi}} \in \tilde{\mathcal{F}}^1$ and $\mathcal{X} = (T, \mu, \nu) \in \mathbb{H}^{K,\sigma}$

$$\begin{aligned} S_t\tilde{\Psi}(\mathcal{X}) &= \int \mathbb{P}^{\mathcal{X}}\big(\pi_t^{\mathcal{X}}\big([\![\underline{u}]\!]\big) = 0\big) \cdot \tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big)\mu^{\otimes n}(\mathrm{d}\underline{u}) \\ &= \int \mathrm{e}^{-t\nu([\![\underline{u}]\!])} \cdot \tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big)\mu^{\otimes n}(\mathrm{d}\underline{u}). \end{aligned} \tag{3.18}$$

Therefore,

$$\sup_{\mathcal{X}\in\mathbb{H}^{K,\sigma}} \big|S_t\tilde{\Psi}(\mathcal{X}) - \tilde{\Psi}(\mathcal{X})\big| \leq K^n \sup_{x\in\mathbb{R}_+} \gamma_{\tilde{\Phi}}(x)\big(1 - \mathrm{e}^{-tx}\big) \underset{t\to 0}{\longrightarrow} 0.$$

□

3.4. *The infinitesimal generator*

In this subsection we calculate the action of the generator on the test functions $\tilde{\Psi} \in \tilde{\mathcal{F}}^1$. For these functions to be bounded, we have to work on the space $\mathbb{H}^{K,\sigma}$. Note that $\mathbb{H}^{K,\sigma}$ is a good state space for the pruning process, as once started in $\mathbb{H}^{K,\sigma}$, it will never leave the space. In the following we write

$$\big(\Omega, \mathcal{D}(\Omega)\big) \quad \text{and} \quad \big(\Omega_K, \mathcal{D}(\Omega_K)\big) \tag{3.19}$$

for the infinitesimal generators of the pruning process with state spaces $\mathbb{H}^{f,\sigma}$ and $\mathbb{H}^{K,\sigma}$ respectively.

Proposition 3.7 (Infinitesimal generator). *For every $K > 0$, we have $\tilde{\mathcal{F}}^1 \subseteq \mathcal{D}(\Omega_K)$. Furthermore, for $\tilde{\Psi} = \Psi^{1,n,\tilde{\Phi}} \in \tilde{\mathcal{F}}^1$ and $\mathcal{X} = (T, \mu, \nu) \in \mathbb{H}^{K,\sigma}$,*

$$\Omega\tilde{\Psi}(\mathcal{X}) = \int \nu(\mathrm{d}v)\big[\tilde{\Psi}\big((T^v, \mu, \nu)\big) - \tilde{\Psi}(\mathcal{X})\big] \tag{3.20}$$

$$= -\int \mu^{\otimes n}(\mathrm{d}\underline{u})\nu\big([\![\underline{u}]\!]\big)\tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big). \tag{3.21}$$

Proof. Using Formula (3.18), we obtain for $\tilde{\Psi} = \Psi^{1,n,\tilde{\Phi}} \in \tilde{\mathcal{F}}^1$, $\mathcal{X} \in \mathbb{H}^{K,\sigma}$,

$$\frac{1}{t}\big(S_t\tilde{\Psi}(\mathcal{X}) - \tilde{\Psi}(\mathcal{X})\big) = -\frac{1}{t}\int \mu^{\otimes n}(\mathrm{d}\underline{u})\big(1 - \mathrm{e}^{-t\nu([\![\underline{u}]\!])}\big)\tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big). \tag{3.22}$$

Note that $|1 - \mathrm{e}^{-x} - x| \le x^2$, for all $x \ge 0$ and recall the definition of $\gamma_{\tilde{\Phi}}$ from (3.17). Comparing (3.22) to (3.21), we see that

$$\sup_{\mathcal{X}\in\mathbb{H}^{K,\sigma}}\left|\frac{1}{t}\big(S_t\tilde{\Psi}(\mathcal{X}) - \tilde{\Psi}(\mathcal{X})\big) + \int \mu^{\otimes n}(\mathrm{d}\underline{u})\nu\big([\![\underline{u}]\!]\big)\tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big)\right|$$
$$\le \sup_{\mathcal{X}\in\mathbb{H}^{K,\sigma}} t\cdot\int \mu^{\otimes n}(\mathrm{d}\underline{u})\nu\big([\![\underline{u}]\!]\big)^2\big|\tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big)\big| \le tK^n \sup_{x\in\mathbb{R}_+} x^2\gamma_{\tilde{\Phi}}(x). \tag{3.23}$$

Due to our assumptions on $\tilde{\Phi} \in \widetilde{\Pi}_n$, $x^2\gamma_{\tilde{\Phi}}(x)$ is bounded, and we obtain uniform convergence of $\frac{1}{t}(S_t\tilde{\Psi} - \tilde{\Psi})$ on $\mathbb{H}^{K,\sigma}$ for $t \to 0$. Hence $\tilde{\mathcal{F}}^1 \subseteq \mathcal{D}(\Omega_K)$ and Formula (3.21) are proven.

We next prove Formula (3.20). Notice that for all $\underline{u} \in T^n$,

$$\nu\big([\![\underline{u}]\!]\big) = \int_T \mathbb{1}_{\{v\in[\![\underline{u}]\!]\}}\nu(\mathrm{d}v) = \int_T 1 - \mathbb{1}_{\{\underline{u}\in(T^v)^n\}}\nu(\mathrm{d}v). \tag{3.24}$$

Inserting the latter into (3.21) and using Fubini's theorem yields

$$\Omega\tilde{\Psi}(\mathcal{X}) = \int_T \nu(\mathrm{d}v)\left(\int_{(T^v)^n} \mu^{\otimes n}(\mathrm{d}\underline{u})\tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big) - \int_{T^n} \mu^{\otimes n}(\mathrm{d}\underline{u})\tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big)\right), \tag{3.25}$$

which gives (3.20). □

4. Examples

In this section we want to apply Theorem 3.6 to obtain convergence of various pruning processes that appear in the literature. We first recall the excursion representation of a measure $\mathbb{R}$-tree. We denote by

$$\mathcal{E} := \big\{e : [0,1] \to \mathbb{R}_+ \mid e \text{ is l.s.c., } e(0) = e(1) = 0\big\} \tag{4.1}$$

the set of lower semi-continuous excursions on $[0, 1]$. From each excursion $e \in \mathcal{E}$, we can define a measure $\mathbb{R}$-tree in the following way:

- $r_e(x, y) := e(x) + e(y) - 2\inf_{[x,y]} e$ is a pseudo-distance on $[0, 1]$,
- $x, y \in [0, 1]$ are said to be equivalent, $x \sim_e y$, if $r_e(x, y) = 0$,
- the image of the projection $\pi_e : [0, 1] \to [0, 1]/\sim_e$ endowed with the push forward of r_e (again denoted r_e), i.e. $T_e := (T_e, r_e, \rho_e) := (\pi_e([0, 1]), r_e, \pi_e(0))$, is a 0-hyperbolic space (for example, [30], Lemma 3.1),
- we endow this space with the probability measure $\mu_e := \pi_{e*}\lambda_{[0,1]}$ which is the push forward of the Lebesgue measure on $[0, 1]$.

We denote by $g:\mathcal{E}\to\mathbb{H}_\rho$ the resulting "glue function",

$$g(e):=(T_e,\mu_e), \tag{4.2}$$

which sends an excursion to a rooted probability measure $\mathbb{R}$-tree. The map g is continuous if $\mathbb{H}_\rho$ is endowed with the Gromov-weak topology, and $\mathcal{E}$ with the uniform topology (see [6], Prop. 2.9 for the case of continuous excursions) or, more generally, with the weaker *excursion topology* introduced in [39] (see Theorem 4.8 there).

Example 4.1 (An approach via excursions). *Consider a sequence of random excursions $e_N=(e_N(s),s\in[0,1])\in\mathcal{E}$, $N\in\mathbb{N}$, that converges in distribution (with respect to the uniform, respectively the excursion topology) to $e\in\mathcal{E}$. For each $N\in\mathbb{N}$, we denote by $(X_t^N)_{t\geq 0}$ the pruning process started in the bi-measure tree $\mathcal{X}_{e_N}:=(T_{e_N},\mu_{e_N},\lambda_{T_{e_N}})\in\mathbb{H}^{f,\sigma}$, where $\lambda_{T_{e_N}}$ is the length measure on T_{e_N}, and similarly for $(X_t)_{t\geq 0}$ and $\mathcal{X}_e$.*

Due to continuity of g, we have that $g(e_N)$ converges Gromov-weakly in distribution to $g(e)$. By Proposition 2.25, we obtain the LWV-convergence in distribution of $\mathcal{X}_{e_N}$ to $\mathcal{X}_e$, and by Theorem 3.6, we get the Skorohod convergence

$$\left(X_t^N\right)_{t\geq 0}\overset{\mathrm{Sk}}{\Longrightarrow}(X_t)_{t\geq 0}$$

as $\mathbb{H}^{f,\sigma}$-valued processes with LWV-topology. Note that this, in particular, implies Skorohod convergence of the pruning processes $(T_{e_N}^{\pi_t},\mu_{e_N})_{t\geq 0}$ as measure $\mathbb{R}$-tree-valued processes in the usual Gromov-weak topology, where we do not keep track of the pruning measure.

We shall apply this example to Galton–Watson trees. Consider a critical or sub-critical Galton–Watson tree $\mathcal{G}$ with offspring distribution η on $\mathbb{N}_0$, i.e., every node in the discrete tree has a random number of children given independently by the distribution η, where $\mathbb{E}[\eta]\leq 1$. Encode $\mathcal{G}$ as a rooted $\mathbb{R}$-tree with unit length edges. For each $N\in\mathbb{N}$, let $\mathcal{G}_N$ be the tree $\mathcal{G}$ conditioned to have N nodes (in addition to the root). We consider two different sampling measures μ on $\mathcal{G}_N$: one is the *normalized length measure*

$$\mu_N^{\mathrm{ske}}:=\frac{1}{N}\lambda_{\mathcal{G}_N}, \tag{4.3}$$

and the second is the *uniform measure on the nodes*,

$$\mu_N^{\mathrm{nod}}:=\frac{1}{N}\sum_{i=1}^N\delta_{x_i}, \tag{4.4}$$

where $\{x_1,\dots,x_N\}$ are the nodes of $\mathcal{G}_N$. Notice that

$$\mu_N^{\mathrm{nod}}(A)=\sum_{x\in\mathrm{nod}(A)}\mu_N^{\mathrm{ske}}\big([x_-,x]\big)\leq\mu_N^{\mathrm{ske}}\big(\big\{v\in\mathcal{G}_N\mid r_{\mathcal{G}_N}(v,A)<1\big\}\big), \tag{4.5}$$

where $\mathrm{nod}(A)$ is the set of nodes in A and x_- is the parent of x.

In order to obtain convergence, we rescale the tree $\mathcal{G}_N$ to have edge lengths $a_N>0$, i.e., we leave the set unchanged and multiply the metric by a_N. We denote the rescaled tree by $a_N\mathcal{G}_N$. As

$$d_{\mathrm{Pr}}^{a_N\mathcal{G}_N}\big(\mu_N^{\mathrm{ske}},\mu_N^{\mathrm{nod}}\big)\leq a_N \tag{4.6}$$

on the rescaled tree by (4.5), μ_N^{nod} and μ_N^{ske} become arbitrary close whenever a_N converges to zero, as $N\to\infty$.

We also consider two different pruning measures ν: one is the *length measure on the rescaled tree*,

$$\nu_N^{\mathrm{ske}}:=\lambda_{a_N\mathcal{G}_N}=a_N\cdot N\cdot\mu_N^{\mathrm{ske}}, \tag{4.7}$$

and the second is a suitably rescaled *uniform measure on the nodes*,

$$\nu_N^{\mathrm{nod}}:=a_N\cdot N\cdot\mu_N^{\mathrm{nod}}. \tag{4.8}$$

In order to be in a position to apply Example 4.1, we associate the conditioned and rescaled bi-measure Galton–Watson tree with an excursion. That is, by the depth-first search algorithm we obtain a graph-theoretic path $\rho = y_0, y_1, \dots, y_{2N-1}, y_{2N} = \rho$ in the discrete tree, which traverses each edge exactly twice. The *contour process* $(C_N(t), 0 \le t \le 1)$ of $\mathcal{G}_N$ is the linear interpolation of $C_N(\frac{k}{2N}) := h(y_k) := r_{\mathcal{G}_N}(\rho, y_k)$, $k = 0, \dots, 2N$. Note that in our definition of C_N, the domain is normalized to [0, 1], and we obtain that

$$g(C_N) = \big(\mathcal{G}_N, \mu_N^{\mathrm{ske}}\big). \tag{4.9}$$

Example 4.2 (Brownian CRT). *Let the variance σ^2 of η be finite and choose*

$$a_N := \frac{\sigma}{\sqrt{N}}. \tag{4.10}$$

We know from Theorem 23 *in* [10] *that $(a_N C_N(t), 0 \le t \le 1)$ converges uniformly in distribution to $(2B(t), 0 \le t \le 1)$, where B is the standard Brownian excursion. We now apply Example* 4.1 *and get the LWV-convergence in distribution of the bi-measure $\mathbb{R}$-trees*

$$\left(\frac{\sigma}{\sqrt{N}}\mathcal{G}_N, \mu_N^{\mathrm{ske}}, \nu_N^{\mathrm{ske}}\right) \underset{N\to\infty}{\overset{\mathrm{LWV}}{\longrightarrow}} (\mathrm{CRT}, \mu, \lambda_{\mathrm{CRT}}), \tag{4.11}$$

where $(\mathrm{CRT}, \mu) = g(2B)$ is the $\mathbb{R}$-tree called Brownian continuum random tree, and λ_{CRT} is the length measure on the Brownian CRT.

By Corollary 2.26 *and Lemma* 2.30, *we also have the convergence*

$$\left(\frac{\sigma}{\sqrt{N}}\mathcal{G}_N, \mu_N, \nu_N\right) \underset{N\to\infty}{\overset{\mathrm{LWV}}{\longrightarrow}} (\mathrm{CRT}, \mu, \lambda_{\mathrm{CRT}}) \tag{4.12}$$

for all choices of $\mu_N \in \{\mu_N^{\mathrm{ske}}, \mu_N^{\mathrm{nod}}\}$ and $\nu_N \in \{\nu_N^{\mathrm{ske}}, \nu_N^{\mathrm{nod}}\}$. Finally we have the convergence of the pruning processes in Skorohod space:

$$\left(\frac{\sigma}{\sqrt{N}}\mathcal{G}_N^{\pi_t}, \mu_N, \nu_N\right)_{t\ge 0} \underset{\mathrm{LWV}}{\overset{\mathrm{Sk}}{\Longrightarrow}} \big(\mathrm{CRT}^{\pi_t}, \mu, \lambda_{\mathrm{CRT}}\big)_{t\ge 0}. \tag{4.13}$$

In particular,

$$\left(\frac{\sigma}{\sqrt{N}}\mathcal{G}_N^{\pi_t}, \mu_N^{\mathrm{nod}}\right)_{t\ge 0} \underset{\mathrm{Gw}}{\overset{\mathrm{Sk}}{\Longrightarrow}} \big(\mathrm{CRT}^{\pi_t}, \mu\big)_{t\ge 0}. \tag{4.14}$$

Notice that for $\nu_N = \nu_N^{\mathrm{ske}}$, the pruning process $(\mathcal{G}_N^{\pi_t})_{t\ge 0}$ is, up to the time transformation $u = \mathrm{e}^{-t/\sqrt{N}}$, the same as the pruning process $(\mathcal{G}_u^{\mathrm{AP}})_{u\in[0,1]}$ uniformly on the edges of Aldous and Pitman in [12]. *The process on the right hand side is the one considered by Aldous and Pitman* [11] *and by Abraham and Serlet* [8] *for example.*

Example 4.3 (α-stable Lévy tree). *We know from Theorem* 3.1 *of* [26] *that if η is in the domain of attraction of an α-stable distribution with $\alpha \in (1, 2]$, then there exists a sequence a_N such that $(a_N C_N(t), 0 \le t \le 1)$ converges uniformly in distribution to $(H(t), 0 \le t \le 1)$, where H is a continuous excursion that codes an α-stable Lévy tree, $(\mathrm{LT}_\alpha, \mu) := g(H)$. More precisely, for $\eta(k) \sim_{k\to\infty} Ck^{1-\alpha}$, we have $a_N = N^{-\bar\alpha}(\frac{\alpha(\alpha-1)}{C\Gamma(2-\alpha)})^{-1/\alpha}$ with $\bar\alpha = 1 - 1/\alpha$ (see Section* 1.2 *in* [20]). *As in Example* 4.2, *we obtain*

$$\big(a_N \mathcal{G}_N^{\pi_t}, \mu_N, \nu_N\big)_{t\ge 0} \underset{\mathrm{LWV}}{\overset{\mathrm{Sk}}{\Longrightarrow}} \big(\mathrm{LT}_\alpha^{\pi_t}, \mu, \lambda_{\mathrm{LT}_\alpha}\big)_{t\ge 0} \tag{4.15}$$

or more precisely

$$\left(\frac{1}{N^{\bar\alpha}}\mathcal{G}_N^{\pi_t}, \mu_N, \nu_N\right)_{t\ge 0} \underset{\mathrm{LWV}}{\overset{\mathrm{Sk}}{\Longrightarrow}} \left(\left(\frac{\alpha(\alpha-1)}{C\Gamma(2-\alpha)}\right)^{1/\alpha} \mathrm{LT}_\alpha^{\pi_t}, \mu, \lambda_{\mathrm{LT}_\alpha}\right)_{t\ge 0}, \tag{4.16}$$

where $\mu_N = \mu_N^{\mathrm{ske}}$ or μ_N^{nod} and $\nu_N = \nu_N^{\mathrm{ske}}$ or ν_N^{nod}.

Example 4.4 (Pruning at a height). *As before we consider the Gromov-weak convergence* $(a_N\mathcal{G}_N, \mu_N) \xrightarrow{\text{Gw}} (\mathrm{LT}_\alpha, \mu)$. *For* $a \geq 0$, *we define the pruning measure*

$$\nu_N^a := \sum_{x\in\mathcal{G}_N^a} \delta_x, \tag{4.17}$$

where $\mathcal{G}_N^a = \{x \in \mathcal{G}_N \mid r_N(\rho, x) = a\}$, *and the corresponding measure*

$$\nu_\infty^a := \sum_{x\in\mathrm{LT}_\alpha^a\cap\mathrm{Sk}_\mu(\mathrm{LT}_\alpha)} \delta_x \tag{4.18}$$

on LT_α. *Here, we restrict the pruning measure to the points of* LT_α *which are not leaves in order to ensure the condition* $\nu_\infty^a(\mathrm{Lf}_\mu(\mathrm{LT}_\alpha^a)) = 0$. *Because the probability that* $\mu(\mathrm{LT}_\alpha^a) \neq 0$ *is zero for fixed* a, *the sequence* $(\nu_N^a)_{N\in\mathbb{N}\cup\{\infty\}}$ *almost surely depends* $\boldsymbol{\upsilon}^{(\mathrm{LT}_\alpha,\mu)}$*-almost continuously on the distances, i.e.*

$$R^{\mathrm{LT}_\alpha}{}_*\mu^{\otimes n}\big(\mathrm{Discont}(F_n)\big) = 0 \quad a.s.,$$

for F_n *as in* (2.52). *We use Proposition* 2.25 *and the previous construction to get*

$$\big(a_N\mathcal{G}_N^{\pi_t}, \mu_N, \nu_N^a\big)_{t\geq 0} \stackrel{\text{Sk}}{\Longrightarrow} \big(\mathrm{LT}_\alpha^{\pi_t}, \mu, \nu_\infty^a\big)_{t\geq 0}. \tag{4.19}$$

It is easy to check that $(\mathrm{LT}_\alpha^{\pi_t}, \mu, \nu_\infty^a)$ *converges almost surely, as* $t \to \infty$, *in the LWV-topology to* $(\mathrm{LT}_\alpha^{\leq a}, \mu, 0)$ *where* $\mathrm{LT}_\alpha^{\leq a} = \{x \in \mathrm{LT}_\alpha \mid r(\rho, x) \leq a\}$. *This is the pruning construction at the height a of Miermont* [42].

Remark 4.5 (Pruning based on other scaling results). *Some authors give other convergence of Galton–Watson trees to continuous trees. For example a sequence of Galton–Watson trees* $(\mathcal{G}_N)_{N\in\mathbb{N}}$ *conditioned to have maximum height at least* $\gamma_N T$ *converges to a general Lévy tree conditioned to have maximum height at least* T, *see Proposition* 2.5.2 *in* [27]. *Or a sequence of Galton–Watson trees that converges to a forest of Lévy trees, see Theorem* 2.4.1 *in* [27]. *In the first case, the previous results clearly applies. In the second case, in general we do not have an excursion with finite length anymore, i.e., the measure* μ^{ske} *might become infinite. However, if we restrict the domain of the contour processes to a finite interval, we can still apply the previous results.*

Example 4.6 (More general pruning). *A non-uniform pruning process on the branch points of a general Galton–Watson tree has been defined by Abraham, Delmas and He* [4]: *they cut a branch point* v *and its subtree above independently with probability* $1 - u^{c(v)-1}$, *where* $c(v)$ *is the number of children of* v. *This corresponds to taking the pruning measure* ν_N^{ADH} *on* $\mathcal{G}_N$ *that is supported on the branch points and satisfies*

$$\nu_N^{\text{ADH}}\big(\{v\}\big) := c(v) - 1. \tag{4.20}$$

A pruning process on the infinite branch points of a Lévy tree has been defined by Abraham and Delmas [2]: *they cut each infinite branch point and its subtree above independently with probability* $1 - \mathrm{e}^{-t\Delta_x}$ *where* Δ_x *is the weight of the node* x *that can be defined using the jumps of the Lévy process. This corresponds to taking a measure* ν^{AD} *on the infinite branch points of the Lévy tree.*

Because we know that a properly renormalized sequence of conditioned Galton–Watson trees converges to a Lévy tree, we conjecture that there exists a sequence b_N *such that*

$$\big(a_N\mathcal{G}_N, \mu_N^{\text{nod}}, \nu_N^{\text{ske}} + b_N\nu_N^{\text{ADH}}\big) \xrightarrow[N\to\infty]{\text{LWV}} \big(\mathrm{LT}, \mu^{\text{nod}}, \nu^{\text{ske}} + \nu^{\text{AD}}\big), \tag{4.21}$$

where LT *is a Lévy tree or at least an* α*-stable Lévy tree with* b_N *of the order* $N^{-1/\alpha}$ *up to a slowly varying function. The Poisson point process with intensity* $\nu^{\text{ske}} + \nu^{\text{AD}}$ *used in the pruning of the Lévy tree is the Poisson point process given in Section* 4.2 *of* [47].

Example 4.7 (Cutting down trees). *Random deconstruction of trees is an old topic which has recently gained a lot of attention (compare, [14,15,24,36,37,41,45]). The main result of [37] is the following. Given a finite-variance Galton–Watson tree conditioned to have N nodes, select an edge at random and delete the subtree above. Repeat the procedure until the root is isolated. Then the suitably rescaled number of cuts needed converges jointly with the rescaled tree to some random couple (Z_T, T). It is known that the limiting tree T is the Brownian CRT, while (unconditioned) Z_T is Rayleigh distributed. In a very recent paper, Abraham and Delmas [3] used a pruning with the length measure on the Brownian CRT (compare Example 4.2) and showed that given T, Z_T equals in distribution the averaged time it takes to separate a point from the root. The latter quantity was used in the proof given by Janson [37]. In this example, we show that whenever bi-measure $\mathbb{R}$-trees converge – provided some extra tightness conditions hold – Janson's quantities converge as well.*

Let $(\mathcal{G}_N, \mu_N, \nu_N)_{N\in\mathbb{N}\cup\{\infty\}}$ be a sequence of random bi-measure $\mathbb{R}$-trees such that

$$(\mathcal{G}_N, \mu_N, \nu_N) \underset{N\to\infty}{\overset{\text{LWV}}{\Longrightarrow}} (\mathcal{G}_\infty, \mu_\infty, \nu_\infty). \tag{4.22}$$

For each $N \in \mathbb{N}\cup\{\infty\}$, let the pruning process $(X_t^N)_{t\geq 0}$ start in $X_0^N = (\mathcal{G}_N, \mu_N, \nu_N)$. Denote by Θ_N the averaged time until a point gets separated from the root ρ_N, *where the average is taken with respect to the sampling measure μ_N. Given a realization $\mathcal{X} \in \mathbb{H}^{f,\sigma}$ of X_0^N, consider for each $u \in \operatorname{supp}(\mu_N)$ the (random) time $\mathcal{E}_{\mathcal{X}}^u$ until u gets separated from ρ_N, i.e., until a cut point falls on $[\rho_N, u[$. We abbreviate $\mathcal{E}_N^u := \mathcal{E}_{X_0^N}^u$ and obtain*

$$\Theta_N = \int_{\mathcal{G}_N} \mu_N(\mathrm{d}u)\mathcal{E}_N^u. \tag{4.23}$$

For all finite subsets $\{u_1, \ldots, u_n\} \subseteq \mathcal{G}_N$ and $t_1, \ldots, t_n \geq 0$, the distribution of $\mathcal{E}_N^{u_1}, \ldots, \mathcal{E}_N^{u_n}$ is given by

$$\mathbb{P}\big(\mathcal{E}_N^{u_1} \geq t_1, \ldots, \mathcal{E}_N^{u_n} \geq t_n \mid (\mathcal{G}_N, \mu_N, \nu_N)\big) = \prod_{l=1}^n \mathrm{e}^{-t_{p(l)}\cdot\nu_N(S_l\setminus S_{l+1})}, \tag{4.24}$$

where $p:\{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$ is any permutation such that $t_{p(1)} \leq \cdots \leq t_{p(n)}$, and $S_l := [\![u_{p(l)}, \ldots, u_{p(n)}]\!]$. Then for all $n \in \mathbb{N}$,

$$\begin{aligned}\mathbb{E}\big[\Theta_N^n\big] &= \mathbb{E}\left[\int_{\mathcal{G}_N^n} \mu_N^{\otimes n}\big(\mathrm{d}(u_1,\ldots,u_n)\big)\mathbb{E}\left[\prod_{l=1}^n \mathcal{E}_N^{u_l} \mid (\mathcal{G}_N, \mu_N, \nu_N)\right]\right] \\ &= n!\cdot\mathbb{E}\left[\int_{\mathcal{G}_N^n} \mu_N^{\otimes n}\big(\mathrm{d}(u_1,\ldots,u_n)\big)\prod_{j=1}^n \frac{1}{\nu_N([\![u_1,\ldots,u_j]\!])}\right],\end{aligned} \tag{4.25}$$

where the last equality is obtained by using (4.24) and easy computations with the formula $\mathbb{E}[\prod_{i=1}^n X_i] = \int_{\mathbb{R}_+^n} \mathbb{P}(X_i > t_i, \forall i)\,\mathrm{d}(t_1\ldots t_n)$.

Now assume the following:

(i) *for all $n \in \mathbb{N}$, $\varepsilon > 0$ there is an $M > 0$ such that*

$$\sup_{N\in\mathbb{N}} \mathbb{E}\left[\int_{\mathcal{G}_N^n} \mu_N^{\otimes n}\big(\mathrm{d}(u_1,\ldots,u_n)\big)\left(\prod_{j=1}^n \frac{1}{\nu_N([\![u_1,\ldots,u_j]\!])} - M\right)^+\right] \leq \varepsilon, \tag{4.26}$$

(ii) *there is only one probability measure $\mathbb{Q}$ on $\mathbb{R}_+$ with moments*

$$\int_{\mathbb{R}_+} \mathbb{Q}(\mathrm{d}\theta)\theta^n = n!\cdot\mathbb{E}\left[\int \mu_\infty^{\otimes n}\big(\mathrm{d}(u_1,\ldots,u_n)\big)\prod_{j=1}^n \frac{1}{\nu_\infty([\![u_1,\ldots,u_j]\!])}\right] \tag{4.27}$$

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

Note that these assumptions are in particular satisfied in the case of conditioned finite variance Galton–Watson trees converging to the Brownian CRT if ν_N is the length measure and μ_N the uniform distribution on the nodes (*see*, *e.g.*, [37], *proof of Lem*. 4.5, *Thm*. 1.9).

For each $n, M \in \mathbb{N}$, define $\gamma_M^n : \mathbb{H}_n \to \mathbb{R}_+$ by

$$\gamma_M^n\big(T, (u_1, \dots, u_n), \nu\big) := M \wedge \prod_{j=1}^n \nu\big([\![u_1, \dots, u_j]\!]\big)^{-1}. \tag{4.28}$$

Then $\gamma_M^n \in \mathcal{C}_b(\mathbb{H}_n)$ if $\mathbb{H}_n$ is equipped with the sGw-topology, and the LWV-convergence (4.22) *together with Proposition* 2.17 *implies that*

$$\mathbb{E}\bigg[\int \gamma_M^n \circ \tau_{(\mathcal{G}_N, \mu_N, \nu_N)}\, \mathrm{d}\mu_N^{\otimes n}\bigg] \underset{N\to\infty}{\longrightarrow} \mathbb{E}\bigg[\int \gamma_M^n \circ \tau_{(\mathcal{G}_\infty, \mu_\infty, \nu_\infty)}\, \mathrm{d}\mu_\infty^{\otimes n}\bigg]. \tag{4.29}$$

Thus, we also have $\mathbb{E}[\Theta_N^n] \underset{N\to\infty}{\longrightarrow} \mathbb{E}[\Theta_\infty^n]$ for each $n \in \mathbb{N}$, provided that (4.26) *holds. By assumption* (ii), *the moments of Θ_∞ determine its distribution uniquely, and therefore the method of moments yields*

$$\Theta_N \underset{N\to\infty}{\Longrightarrow} \Theta_\infty.$$

Acknowledgement

The authors would like to thank the anonymous referees for several helpful remarks and references.

References

[1] R. Abraham and J.-F. Delmas. Fragmentation associated with Lévy processes using snake. *Probab. Theory Related Fields* **141** (2008) 113–154. MR2372967

[2] R. Abraham and J.-F. Delmas. A continuum-tree-valued Markov process. *Ann. Probab.* **40** (3) (2012) 1167–1211. MR2962090

[3] R. Abraham and J.-F. Delmas. Record process on the continuum random tree. *ALEA Lat. Am. J. Probab. Math. Stat.* **10** (1) (2013) 225–251. MR3083925

[4] R. Abraham, J.-F. Delmas and H. He. Pruning Galton–Watson trees and tree-valued Markov process. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** (3) (2012). MR2976559

[5] R. Abraham, J.-F. Delmas and P. Hoscheit. A note on the Gromov–Hausdorff–Prokhorov distance between (locally) compact metric measure spaces. *Electron. J. Probab.* **18** (14) (2013) 1–21. MR3035742

[6] R. Abraham, J.-F. Delmas and P. Hoscheit. Exit times for an increasing Lévy tree-valued process. *Probab. Theory Related Fields* **159** (1–2) (2014) 357–403. MR3201925

[7] R. Abraham, J.-F. Delmas and G. Voisin. Pruning a Lévy continuum random tree. *Electron. J. Probab.* **15** (46) (2010) 1429–1473. MR2727317

[8] R. Abraham and L. Serlet. Poisson snake and fragmentation. *Electron. J. Probab.* **7** (2002) 1–15. MR1943890

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

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

[11] D. Aldous and J. Pitman. The standard additive coalescent. *Ann. Probab.* **26** (4) (1998) 1703–1726. MR1675063

[12] D. Aldous and J. Pitman. Tree-valued Markov chains derived from Galton–Watson processes. *Ann. Inst. Henri. Poincaré Probab. Stat.* **34** (5) (1998) 637–686. MR1641670

[13] S. Athreya, W. Löhr and A. Winter. The gap between Gromov-vague and Gromov–Hausdorff-vague topology. Preprint, 2014. Available at arXiv:1407.6309.

[14] J. Bertoin. Fires on trees. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** (4) (2012) 909–921. MR3052398

[15] J. Bertoin and G. Miermont. The cut-tree of large Galton–Watson trees and the Brownian CRT. *Ann. Appl. Probab.* **23** (4) (2013) 1469–1493. MR3098439

[16] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1999. MR1700749

[17] D. Blount and M. A. Kouritzin. On convergence determining and separating classes of functions. *Stochastic Process. Appl.* **120** (10) (2010) 1898–1907. MR2673979

[18] V. I. Bogachev. *Measure Theory, Vol. I*. Springer, Berlin, 2007. MR2267655

[19] D. Burago, Y. Burago and S. Ivanov. *A Course in Metric Geometry*. *Graduate Studies in Mathematics* **33**. Amer. Math. Soc., Providence, RI, 2001. MR1835418

[20] N. Curien and B. Haas. The stable trees are nested. *Probab. Theory Related Fields* **157** (2012) 847–883. MR3129805

[21] A. Depperschmidt, A. Greven and P. Pfaffelhuber. Marked metric measure spaces. *Electron. Commun. Probab.* **16** (2011) 174–188. MR2783338

[22] A. W. M. Dress and W. F. Terhalle. The real tree. *Adv. Math.* **120** (1996) 283–301. MR1397084

[23] A. W. M. Dress, V. Moulton and W. F. Terhalle. T-theory: An overview. *European J. Combin.* **17** (2–3) (1996). MR1379369

[24] M. Dromota, A. Iksanov, M. Mohle and U. Rosler. A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. *Random Structures Algorithms* **34** (2009) 319–336. MR2504401

[25] R. M. Dudley. *Real Analysis and Probability*. Cambridge Univ. Press, Cambridge, 2002. MR1932358

[26] T. Duquesne. A limit theorem for the contour process of conditioned Galton–Watson trees. *Ann. Probab.* **31** (2) (2003) 996–1027. MR1964956

[27] T. Duquesne and J.-F. Le Gall. Random trees, Lévy processes and spatial branching processes. *Astérisque* **281** (2002) vi+147. MR1954248

[28] S. N. Ethier and T. Kurtz. *Markov Processes. Characterization and Convergence*. Wiley, New York, 1986. MR0838085

[29] S. N. Evans. Probability and real trees. In *École d'Été de Probabilités de Saint Flour XXXV-2005*. *Lecture Notes in Mathematics* **1920**. Springer, Berlin. 2007. MR2351587

[30] S. N. Evans and A. Winter. Subtree prune and re-graft: A reversible real-tree valued Markov chain. *Ann. Probab.* **34** (3) (2006) 918–961. MR2243874

[31] S. N. Evans, J. Pitman and A. Winter. Rayleigh processes, real trees, and root growth with re-grafting. *Probab. Theory Related Fields* **134** (1) (2006) 81–126. MR2221786

[32] K. Fukaya. Collapsing of Riemannian manifolds and eigenvalues of Laplace operator. *Invent. Math.* **87** (3) (1987) 517–547. MR0874035

[33] A. Greven, P. Pfaffelhuber and A. Winter. Convergence in distribution of random metric measure spaces (The Λ-coalescent measure tree). *Probab. Theory Related Fields* **145** (1–2) (2009) 285–322. MR2520129

[34] M. Gromov. *Metric Structures for Riemannian and Non-Riemannian Spaces*. *Progress in Mathematics* **152**. Birkhäuser, Boston, MA, 1999. MR1699320

[35] J. Hoffmann-Jørgensen. Probability in Banach spaces. In *École d'Été de Probabilités de Saint Flour VI-1976*. *Lecture Notes in Mathematics* **598**. Springer, Berlin, 1977. MR0461610

[36] C. Holmgren. Random records and cuttings in binary search trees. *Combin. Probab. Comput.* **19** (2010) 391–424. MR2607374

[37] S. Janson. Random cuttings and records in deterministic and random trees. *Random Structures Algorithms* **29** (2006) 139–179. MR2245498

[38] L. Le Cam. Convergence in distribution of stochastic processes. *Univ. Calif. Publ. Statist.* **2** (1957) 207–236. MR0086117

[39] W. Löhr. Equivalence of Gromov–Prohorov- and Gromov's $\square_\lambda$-metric on the space of metric measure spaces. *Electron. Commun. Probab.* **18** (2013) no. 17, 10. MR3037215

[40] R. Lyons. Random walks, capacities, and percolation on trees. *Ann. Probab.* **20** (1992) 2043–2088. MR1188053

[41] A. Meir and J. W. Moon. Cutting down random trees. *J. Aust. Math. Soc.* **11** (1970). MR0284370

[42] G. Miermont. Self-similar fragmentations derived from the stable tree I: Splitting at heights. *Probab. Theory Related Fields* **127** (2003) 423–454. MR2018924

[43] G. Miermont. Self-similar fragmentations derived from the stable tree II: Splitting at nodes. *Probab. Theory Related Fields* **131** (2005) 341–375. MR2123249

[44] G. Miermont. Tessellations of random maps of arbitrary genus. *Ann. Sci. Ec. Norm. Super. (4)* **42** (2009) 725–781. MR2571957

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

[46] C. Villani. *Optimal Transport*. *Grundlehren der Mathematischen Wissenschaften* **338**. Springer, Berlin, 2009. MR2459454

[47] G. Voisin. Dislocation measure of the fragmentation of a general Lévy tree. *ESAIM Probab. Stat.* **15** (2011) 372–389. MR2870521