# Spaces of algebraic measure trees and triangulations of the circle

**Anita Winter**, University of Duisburg-Essen

joint work with **Wolfgang Löhr**, Bulletin de la SMF (2021)

**Workshop "Geometric and Topological Properties of Random Algebraic Varieties"**

Cologne, October $4^{\text{th}} - 6^{\text{th}}$, 2023

# From graph-theoretic trees to $\mathbb{R}$-trees

- Graph-theoretic trees have many applications from computer science to theoretical biology.

  $\rightsquigarrow$ How to define limit objects as the size of trees tend to infinity?

- Local structure approaches yield countably infinite graphs.

  $\rightsquigarrow$ Benjamini-Schramm approach: [Benjamini & Schramm (2001)], [Aldous & Lyons (2007)]

- We here seek a *global structure approach*: metric spaces with rescaled distances lead to $\mathbb{R}$-trees: [Tits (1977)]
  - study of isometry groups of hyperbolic spaces [Morgan & Shalen (1982)],
  - study of the fundamental groups of one-dimensional spaces [Mayer, Nikiel & Oversteegen (1992)]
  - characterization of the topological structure of one-dimensional spaces [Fisher & Zastrov (2013)], [Fabel (2015)]

# Our motivation in $\mathbb{R}$-trees

$\rightsquigarrow$ suitable state space for tree-valued stochastic processes that are continuum analogues of ...

- the *Aldous-Broder algorithm* for sampling a uniform spanning tree
  - of the complete graph ([Evans, Pitman & W. (2006)] or
  - the high-dimensional torus [Angtuncio-Hernandez, Berzunza-Ojeda & W. (in progress)])
- *subtree prune and regraft* algorithm used in reconstructing phylogenetic trees [Evans & W. (2006)]
- *evolving genealogies* in population genetics [(Greven, Pfaffelhuber & W. (2013)), [Greven, Sun & W. (2016)]]
- *pruning* procedures [Abraham, Delmas & Voisin (2010)], [Löhr, Voisin & W. (2013)], [He & Winkel (2019)], [Berzunza & W. (2023)]

$\rightsquigarrow$ all these encode trees as $\mathbb{R}$-trees and equip the tree space with Gromov-Hausdorff or Gromov-weak topology

# We want to shift focus from metric to tree structure

$\rightsquigarrow$ useful for several reasons

1. *The state space* of metric (measure) trees *is not compact*.
   - Tightness for random metric (measure spaces) are often not easy to check, or
   - some natural sequences of trees do not even converge as rescaled metric measure spaces, e.g. binary search tree
2. *The metric structure is often less canonical than the tree structure*, e.g. in situations where edge lengths are not of the same order.
3. Certain *functionals* of the tree structure might *not* be *continuous*.

   $\rightsquigarrow$ For example, the degree of a vertex.

# From topological structure to tree structure

$\rightsquigarrow$ ignore metric and focus on "tree-structure"

- Unlike for metric spaces, for general topological spaces no useful notion of convergence is known.
- For *tree spaces*, we can rely on the notion of a *branch point map* that sends any three points to a branch point.

  $\rightsquigarrow$ algebraic trees

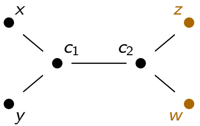
- We can also use the branch point map to define convergence of algebraic (measure) trees.

**Definition ([Tits (1977)], [Dress (1994)], [Chiswell (2001)])**

A metric space $(T, d)$ is an **ℝ-tree** iff

- $(T, d)$ is *connected*.
- $(T, d)$ is *0-hyperbolic*, i.e., $\forall\, x_1, x_2, x_3, x_4 \in T$,

$$d(x_1, x_2) + d(x_3, x_4) \leq \max\left\{d(x_1, x_3) + d(x_2, x_4), d(x_1, x_4) + d(x_2, x_3)\right\}.$$



**Definition ([Tits (1977)], [Dress (1994)], [Chiswell (2001)])**

A metric space $(T, d)$ is an **$\mathbb{R}$-tree** iff

- $(T, d)$ is *connected*.
- $(T, d)$ is *0-hyperbolic*, i.e., $\forall\, x_1, x_2, x_3, x_4 \in T$,

$$d(x_1, x_2)+d(x_3, x_4) \leq \max\big\{d(x_1, x_3)+d(x_2, x_4), d(x_1, x_4)+d(x_2, x_3)\big\}.$$

$\rightsquigarrow$ Intrinsic property

For $x_1, x_2, x_3 \in T$ there exists a *unique branch point $c(x_1, x_2, x_3) \in T$* with

$$\big[x_1, x_2\big] \cap \big[x_2, x_3\big] \cap \big[x_2, x_3\big] = \big\{c(x_1, x_2, x_3)\big\}.$$

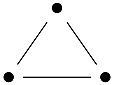
Note that

$$[x, y] := \big\{x' \in T :\ d(x, y) = d(x, x') + d(x', y)\big\}$$

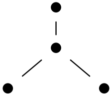
does not require the notion of *connectedness*.

# What are limit trees? metric trees

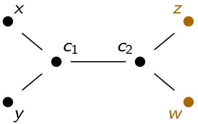
$\rightsquigarrow$ *drop connectedness*



I am not a tree :-(



I am a tree :-)




**Definition ([Athreya, Löhr, W. (2016)])**

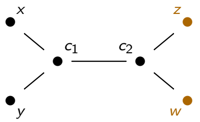
A metric space $(T, d)$ is a *metric tree* if it is (isometric to) a subset of an $\mathbb{R}$-tree with $c(x, y, z) \in T$ for all $x, y, z \in T$.

In both cases we can conclude

- Symmetry of $c(x, y, z)$ in $x, y, z$.
- (2-point condition) $c(x, y, y) = y$, and
- (3-point condition) $c(x, y, c(x, y, z)) = c(x, y, z)$
- (4-point condition) $c(x, y, z) \in \{c(x, y, w), c(x, z, w), c(y, z, w)\}$.

# Algebraic measure trees

characterize trees by map that sends three vertices to their branch point




**Definition (algebraic trees; [Löhr, W.])**

$(T, c)$ is an **algebraic tree** with **branch point map** $c = c_T : T^3 \to T$ if

- $c$ is *symmetric*: $c(x, y, z) = c(y, x, z) = c(x, z, y)$, for all $x, y, z \in T$.
- *2-point condition*: $c(x, y, y) = y$, $x, y \in T$.
- *3-point condition*: $c\big(x, y, c(x, y, z)\big) = c(x, y, z)$, for all $x, y, z \in T$.
- *4-point condition*: For all $x, y, z, w \in T$.

$$c(x, y, z) \in \big\{c(x, y, w), c(x, z, w), c(y, z, w)\big\}.$$

$\rightsquigarrow$ Finite algebraic trees correspond to finite graph-theoretic trees.

# Axioms are sufficient to capture the tree-structure

Let $(T, c)$ be an algebraic tree.

- The *path* between $x_1, x_2 \in T$ is given by

$$[x_1, x_2] := \{z \in T : z = c(x_1, x_2, z)\}.$$

  Then for all $x, y, z \in T$,

$$[x, y] \cap [x, z] \cap [y, z] = \{c(x, y, z)\}.$$

- $\{x_1, x_2\}$ is an *edge* if $[x_1, x_2] := \{x_1, x_2\}$.
- $A \subseteq T$ is a *subtree* if $c(A^3) \subseteq A$.
- Similarly, we can define *branch points*, *leaves*, ...

# Morphisms of algebraic trees

$\rightsquigarrow$ Like any decent algebraic structure, algebraic trees come with the notion of a structure-preserving morphism.

Let $(T, c)$ and $(T', c')$ be algebraic trees.

1. A map $f : T \to T'$ is called a *tree homomorphism*, if for all $x, y, z \in T$, $f\big(c(x, y, z)\big) = c\big(f(x), f(y), f(z)\big)$.
2. Equivalently, for all $x, y \in T$, $f\big([x, y]\big) \subseteq \big[f(x), f(y)\big]$.

# Morphisms of algebraic trees

$\rightsquigarrow$ Like any decent algebraic structure, algebraic trees come with the notion of a structure-preserving morphism.

Let $(T, c)$ and $(T', c')$ be algebraic trees.

1. A map $f : T \to T'$ is called a *tree homomorphism*, if for all $x, y, z \in T$, $f\big(c(x, y, z)\big) = c\big(f(x), f(y), f(z)\big)$.
2. Equivalently, for all $x, y \in T$, $f\big([x, y]\big) \subseteq \big[f(x), f(y)\big]$.

If a homomorphism $f : T \to T'$ is bijective, we refer to it as *isomorphism*.

# Algebraic trees as topological spaces

Let $(T, c)$ be an algebraic tree.

$\rightsquigarrow$ What is a good *notion of open sets*?

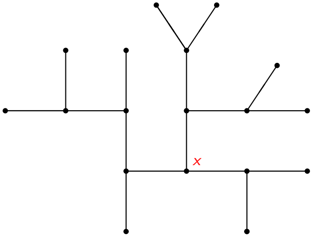
- Taking away $x \in T$, $T$ decomposes into *subtree components*.

# Algebraic trees as topological spaces

Let $(T, c)$ be an algebraic tree.

$\rightsquigarrow$ What is a good *notion of open sets*?

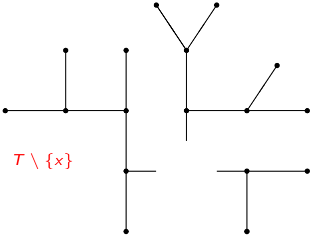
- Taking away $x \in T$, $T$ decomposes into *subtree components*.



Let $(T, c)$ be an algebraic tree.

$\rightsquigarrow$ What is a good *notion of open sets*?

- Taking away $x \in T$, $T$ decomposes into *subtree components*.

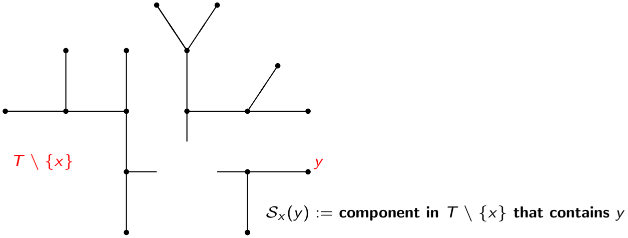


# Algebraic trees as topological spaces

Let $(T, c)$ be an algebraic tree.

$\rightsquigarrow$ What is a good *notion of open sets*?

- Taking away $x \in T$, $T$ decomposes into *subtree components*.

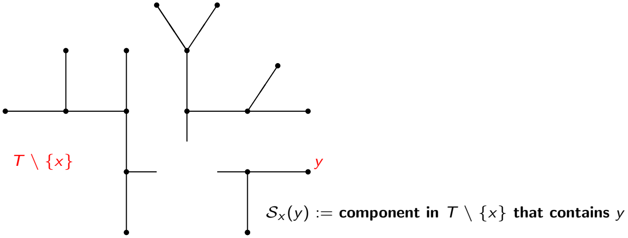


# Algebraic trees as topological spaces

Let $(T, c)$ be an algebraic tree.

$\rightsquigarrow$ What is a good *notion of open sets*?

- Taking away $x \in T$, $T$ decomposes into *subtree components*.



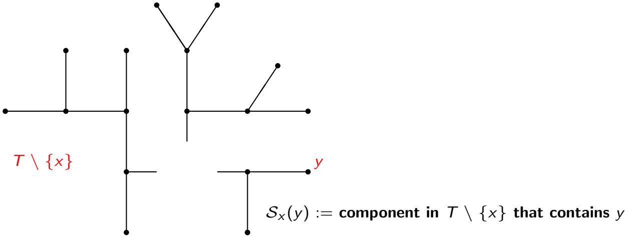

- $\tau := \tau(\{\mathcal{S}_x(y);\, x \neq y \in T\})$.

# Algebraic trees as topological spaces

Let $(T, c)$ be an algebraic tree.

$\rightsquigarrow$ What is a good *notion of open sets*?

- Taking away $x \in T$, $T$ decomposes into *subtree components*.



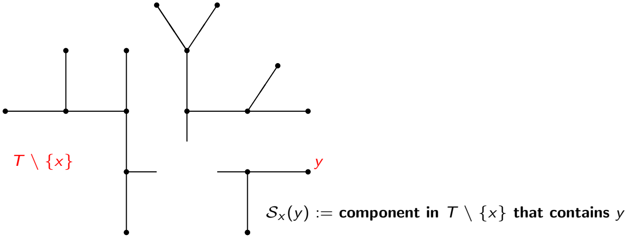

- $\tau := \tau(\{\mathcal{S}_x(y);\, x \neq y \in T\})$.
- This *component topology* is a Hausdorff topology.

# Algebraic trees as topological spaces

Let $(T, c)$ be an algebraic tree.

$\rightsquigarrow$ What is a good *notion of open sets*?

- Taking away $x \in T$, $T$ decomposes into *subtree components*.




- $\tau := \tau(\{\mathcal{S}_x(y);\, x \neq y \in T\})$.
- This *component topology* is a Hausdorff topology.
- The branch point map is continuous with respect to the component topology.

# On separability

$\rightsquigarrow$ *Intervals are closed sets.*

Indeed, with $\mathcal{S}_u(v)$ subtree component of $T \setminus \{u\}$ that contains $v$

$$\complement[x,y] = \bigcup\{\mathcal{S}_u(v) : \; u \in [x,y], v \in T, \mathcal{S}_u(v) \cap [x,y] = \emptyset\} \in \tau(C).$$

**Proposition (Löhr & W. (2021))**

*Let $(T,c)$ be an algebraic tree. Then the following are equivalent:*

1. *The topological space $(T,\tau)$ is separable, and has only countably many edges.*
2. *There exists a countable set $D$ such that for all $x,y \in T$ with $x \neq y$,*

$$D \cap [x,y) \neq \emptyset.$$

In what follows, we refer to $(T,c)$ as (order) separable iff the topological space $(T,c)$ is separable and $(T,c)$ has at most countably many edges.

# Metric representation of order separable algebraic trees

- Any separable metric tree $(T, r)$ defines an order separable algebraic tree $(T, c_{(T,r)})$ by letting $c_{(T,r)}$ send any triple $x, y, z \in T$ to its branch point in $(T, r)$.

- On the contrary, if $(T, c)$ is an order separable algebraic tree and $\nu$ is a measure on $\mathcal{B}(T, c)$, then

$$r_\nu(x, y) := \nu\big([x, y]\big) - \tfrac{1}{2}\nu\big(\{x\}\big) - \tfrac{1}{2}\nu\big(\{y\}\big), \quad x, y \in T$$

  defines a *pseudo-metric* with $c = c_{(T,r_\nu)}$.

- $r_\nu$ is a metric, if $\nu\big([x, y]\big) > 0$ for all $x, y \in T$.

  Such a measure always exists if $(T, c)$ is order separable. Take e.g. a probability measure supported on a countable set $D$ such that for all $x, y \in T$,

$$D \cap [x, y) \neq \emptyset.$$

# Algebraic measure trees

**Definition**

An **algebraic measure tree** $(T, c, \mu)$ consists of an *(order) separable* algebraic tree $(T, c)$ and a *probability measure* $\mu$ on $(T, \mathcal{B}(T, c))$.

Call two algebraic measure trees $(T, c, \mu)$ and $(T', c', \mu')$ **equivalent** iff $\mathrm{supp}(\mu)$ and $\mathrm{supp}(\mu')$ are subtrees, and there is an *isomorphism* $\varphi : \mathrm{supp}(\mu) \to \mathrm{supp}(\mu')$ such that

- for all $x, y, z \in T$, $\varphi\big(c(x, y, z)\big) = c\big(\varphi(x), \varphi(y), \varphi(z)\big)$, and
- $\mu \circ \varphi^{-1} = \mu'$.

$\mathbb{T} :=$ space of all equivalence classes of algebraic measure trees.

# Algebraic measure trees

**Definition**

An **algebraic measure tree** $(T, c, \mu)$ consists of an *(order) separable* algebraic tree $(T, c)$ and a *probability measure* $\mu$ on $(T, \mathcal{B}(T, c))$.

Call two algebraic measure trees $(T, c, \mu)$ and $(T', c', \mu')$ **equivalent** iff $\mathrm{supp}(\mu)$ and $\mathrm{supp}(\mu')$ are subtrees, and there is an *isomorphism* $\varphi : \mathrm{supp}(\mu) \to \mathrm{supp}(\mu')$ such that

- for all $x, y, z \in T$, $\varphi\big(c(x, y, z)\big) = c\big(\varphi(x), \varphi(y), \varphi(z)\big)$, and
- $\mu \circ \varphi^{-1} = \mu'$.

$\mathbb{T} :=$ space of all equivalence classes of algebraic measure trees.

$\rightsquigarrow$ Related approaches to continuum trees:

- didendritic systems; [Evans, Grübel & Wakolbinger (2017)]
- interval partition trees; exchangeable hierarchies [Forman (2020)]

# Algebraic measure trees

**Definition**

An **algebraic measure tree** $(T, c, \mu)$ consists of an *(order) separable* algebraic tree $(T, c)$ and a *probability measure* $\mu$ on $(T, \mathcal{B}(T, c))$.

Call two algebraic measure trees $(T, c, \mu)$ and $(T', c', \mu')$ **equivalent** iff $\mathrm{supp}(\mu)$ and $\mathrm{supp}(\mu')$ are subtrees, and there is an *isomorphism* $\varphi : \mathrm{supp}(\mu) \to \mathrm{supp}(\mu')$ such that

- for all $x, y, z \in T$, $\varphi\big(c(x, y, z)\big) = c\big(\varphi(x), \varphi(y), \varphi(z)\big)$, and
- $\mu \circ \varphi^{-1} = \mu'$.

$\mathbb{T} :=$ space of all equivalence classes of algebraic measure trees.

$\rightsquigarrow$ Three strategies for a notion of convergence in $\mathbb{T}$.

1. rely on Gromov-weak topology
2. use combinatorial notion of convergence
3. rely on convergence of subtree component masses

# Convergence of metric measure spaces

- *mm-space* $(X, d, \mu)$ = a Polish space $(X, d)$ + a probability measure $\mu$ on $\mathcal{B}(X)$

# Convergence of metric measure spaces

- *mm-space* $(X, d, \mu)$ = a Polish space $(X, d)$ + a probability measure $\mu$ on $\mathcal{B}(X)$
- Two metric measure spaces $(X, d, \mu)$ and $(X', d', \mu')$ *equivalent* if there exists an *isometry* $\psi : \mathrm{supp}(\mu) \to \mathrm{supp}(\mu')$ with $\mu' = \mu \circ \psi^{-1}$.

$\mathbb{H}$ := equivalence classes of mm-trees.

# Convergence of metric measure spaces

- *mm-space* $(X, d, \mu)$ = a Polish space $(X, d)$ + a probability measure $\mu$ on $\mathcal{B}(X)$
- Two metric measure spaces $(X, d, \mu)$ and $(X', d', \mu')$ *equivalent* if there exists an *isometry* $\psi : \mathrm{supp}(\mu) \to \mathrm{supp}(\mu')$ with $\mu' = \mu \circ \psi^{-1}$.

$\mathbb{H}$ := equivalence classes of mm-trees.

- We evaluate elements in $\mathbb{H}$ with *sample distance polynomials*, i.e.,

$$\Phi^{m,\phi}(T, d, \mu) := \int \mu^{\otimes m}(\mathrm{d}\underline{x})\, \phi\Big(\big(d(x_i, x_j)\big)_{1 \le i < j \le m}\Big)$$

where $m \in \mathbb{N}$, $\phi \in \mathcal{C}_b(\mathbb{R}_+^{\binom{m}{2}})$.

# Convergence of metric measure spaces

- *mm-space* $(X, d, \mu)$ = a Polish space $(X, d)$ + a probability measure $\mu$ on $\mathcal{B}(X)$
- Two metric measure spaces $(X, d, \mu)$ and $(X', d', \mu')$ *equivalent* if there exists an *isometry* $\psi : \mathrm{supp}(\mu) \to \mathrm{supp}(\mu')$ with $\mu' = \mu \circ \psi^{-1}$.

$\mathbb{H}$ := equivalence classes of mm-trees.

- We evaluate elements in $\mathbb{H}$ with *sample distance polynomials*, i.e.,

$$\Phi^{m,\phi}(T, d, \mu) := \int \mu^{\otimes m}(\mathrm{d}\underline{x})\, \phi\Big(\big(d(x_i, x_j)\big)_{1 \le i < j \le m}\Big)$$

where $m \in \mathbb{N}$, $\phi \in \mathcal{C}_b(\mathbb{R}_+^{\binom{m}{2}})$.

- (*Vershik's reconstruction theorem*) If $\mathcal{X}_1, \mathcal{X}_2 \in \mathbb{H}$ such that $\Phi(\mathcal{X}_1) = \Phi(\mathcal{X}_2)$ for all sample distance polynomial then $\mathcal{X}_1 = \mathcal{X}_2$.

[Greven, Pfaffelhuber & W. (2009)]

**Definition (Gromov-weak topology)**

A sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges *Gromov-weakly* to $\mathcal{X}$ in $\mathbb{H}$ iff

$$\Phi(\mathcal{X}_n) \longrightarrow \Phi(\mathcal{X}) \quad \text{for all sample distance polynomials } \Phi.$$

[Greven, Pfaffelhuber & W. (2009)]

### Definition (Gromov-weak topology)

A sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges *Gromov-weakly* to $\mathcal{X}$ in $\mathbb{H}$ iff

$$\Phi(\mathcal{X}_n) \longrightarrow \Phi(\mathcal{X}) \quad \text{for all sample distance polynomials } \Phi.$$

[Löhr (2013)]

### Corollary (Convergence determining)

*If $(\mathcal{X}_N)_{N\in\mathbb{N}}$ and $\mathcal{X}$ are $\mathbb{H}$-valued random variables with*

$$\mathbb{E}\big[\Phi(\mathcal{X}_N)\big] \underset{N\to\infty}{\longrightarrow} \mathbb{E}\big[\Phi(\mathcal{X})\big]$$

*for all sample distance polynomials. Then*

$$\mathcal{L}(\mathcal{X}_N) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}(\mathcal{X}).$$

# The branch point distribution distance

$\leadsto$ for an order separable tree $(T, c)$ there are many metric representations

*Question.* Is there an intrinsic choice?

# The branch point distribution distance

$\rightsquigarrow$ for an order separable tree $(T, c)$ there are many metric representations

*Question.* Is there an intrinsic choice?

**Definition (Branch point distribution distance)**

Let $(T, c, \mu)$ be an algebraic measure tree. The probability measure

$$\nu_{(T,c,\mu)} := \mu^{\otimes 3} \circ c^{-1}$$

is called the *branch point distribution*. and the associated metric $r_\nu$ the *branch point distribution distance*.

We say that a sequence $(\mathcal{X}_N)_{\mathbb{N}}$ converges *globally* to in $\mathcal{X}$ in $\mathbb{T}$, if $(\mathcal{X}_N)_{\mathbb{N}}$ converges *Gromov-weakly* to $\mathcal{X}$ in $\mathbb{T}$ with respect to the branch point distribution distance.

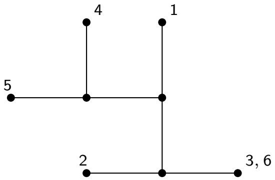
# A combinatorial notion of convergence

$\rightsquigarrow$ In many applications trees are binary.

$$\mathbb{T}_2 := \{\text{binary algebraic measure trees with } \textit{atoms only at leaves}\}$$

An *m-cladogram* is a combinatorial

- un-rooted, binary tree
- leaf-labeled, i.e., there is a surjective *labeling map* $\ell : \{1...m\}$ to the set of leaves.




Two $m$-cladograms $(C_1, c_1, \ell_1)$ and $(C_2, c_2, \ell_2)$ are **equivalent** if they are *label invariant isomorphic*.

# Sample shape

$\rightsquigarrow$ encode the shape of a sampled subtree as *cladogram*

**Definition (tree shape)**

Fix $(T, c_T) \in \mathbb{T}_2$, $m \in \mathbb{N}$, and non-branch points $u_1, ..., u_m \in T$. The **tree shape** $\mathfrak{s}_T(u_1, ..., u_m)$ is the unique (up to isomorphism) $m$-cladogram with

- *(sh1)* leaf set $\{u_1, ..., u_m\}$,
- *(sh2)* $u_i$ gets label $i$ for all $i = 1, ..., m$, and such that
- *(sh3)* the identity on the leave set extends to a tree homomorphism from $\mathfrak{s}_T(u_1, ..., u_m)$ onto $c_T\big(\{u_1, ..., u_m\}^3\big)$.

$\rightsquigarrow$ *(sh3)* tells us what to do when there are 3 points on a path.

# Sample shape distribution

Consider *sample shape polynomials*, i.e. functions $\Phi : \mathbb{T}_2 \to \mathbb{R}$ of the form

$$\Phi(x) = \int \mu^{\otimes m}(\mathrm{d}\underline{x})\phi\big(\mathfrak{s}_T(\underline{x})\big)$$

for some $m \in \mathbb{N}$ and $\phi : \mathcal{C}_m \to \mathbb{R}_+$.

# Sample shape distribution

Consider *sample shape polynomials*, i.e. functions $\Phi : \mathbb{T}_2 \to \mathbb{R}$ of the form

$$\Phi(x) = \int \mu^{\otimes m}(\mathrm{d}\underline{x})\phi\big(\mathfrak{s}_T(\underline{x})\big)$$

for some $m \in \mathbb{N}$ and $\phi : \mathcal{C}_m \to \mathbb{R}_+$.

**Proposition (Löhr & W. (2021))**

*For each $m \in \mathbb{N}$ and $\varphi \in \mathcal{C}(\mathbb{R}^{\binom{m}{2}})$, the sample distance polynomial*

$$\Phi(x) = \int \mu^{\otimes m}(\mathrm{d}\underline{x})\phi\big((r_\nu(x_i, x_j))_{1 \le i < j \le m}\big)$$

*can be approximated uniformly by sample shape polynomials.*

# Sample shape distribution

Main steps in proof:

1. Recall that $\nu = c_* \mu^{\otimes 3}$ and
$$r_\nu(x,y) = \nu([x,y]) - \tfrac{1}{2}\nu(\{x\}) - \tfrac{1}{2}\nu(\{y\}).$$
2. Approximate $\nu$ by *empirical branch point distributions*,
$$\nu_{n,\underline{u}} := \frac{1}{n}\sum_{i=0}^{n-1} \delta_{c(u_{3i+1},u_{3i+2},u_{3i+3})},$$
and note that $\Phi(\mathcal{X}) := \int \mu^{\otimes 3m}(\mathrm{d}\underline{u})\phi\big((r_{\nu_{n,\underline{u}}}(u_i,u_j))_{1\le i<j\le 3m}\big)$ is a sample shape polynomial because whether or not $u_i \in [u_k, u_l]$ depends on $\mathfrak{s}_T(\underline{u})$.
3. Use the uniform upper bound ([Vapnik & Chervonenkis (1971)])
$$\mathbb{E}\Big[\sup_{x,y\in T}\big|\nu([x,y]) - \nu_{n,u}([x,y])\big|\Big] \le 96\sqrt{\frac{\dim_{VC}(\{[x,y]; x,y\in T\})}{n}},$$
and note that $\dim_{VC}(\{[x,y]; x,y\in T\}) = 2$.

# Sample shape distribution

Consider *sample shape polynomials*, i.e. functions $\Phi : \mathbb{T}_2 \to \mathbb{R}$ of the form

$$\Phi(x) = \int \mu^{\otimes m}(\mathrm{d}\underline{x})\phi\big(\mathfrak{s}_T(\underline{x})\big)$$

for some $m \in \mathbb{N}$ and $\phi : \mathcal{C}_m \to \mathbb{R}_+$.

**Proposition (Löhr & W. (2021))**

*For each $m \in \mathbb{N}$ and $\varphi \in \mathcal{C}(\mathbb{R}^{\binom{m}{2}})$, the sample distance polynomial*

$$\Phi(x) = \int \mu^{\otimes m}(\mathrm{d}\underline{x})\phi\big((r_\nu(x_i, x_j))_{1\le i<j\le m}\big)$$

*can be approximated uniformly by sample shape polynomials.*

# Sample shape convergence

*Distance polynomials* can be approximated uniformly by *shape polynomials*.

---

Consequently, if $x, x' \in \mathbb{T}_2$ then

$$x = x' \text{ iff } \Phi^{m,\phi}(x) = \Phi^{m,\phi}(x'), \forall m \in \mathbb{N}, \phi : \mathfrak{C}_m \to \mathbb{R}.$$

# Sample shape convergence

*Distance polynomials* can be approximated uniformly by *shape polynomials*.

---

Consequently, if $x, x' \in \mathbb{T}_2$ then

$$x = x' \text{ iff } \Phi^{m,\phi}(x) = \Phi^{m,\phi}(x'), \forall m \in \mathbb{N}, \phi : \mathfrak{C}_m \to \mathbb{R}.$$

**Definition (Sample shape convergence)**

We say that a sequence $(x_n)_{n\in\mathbb{N}}$ converges in *sample shape* to $x$ in $\mathbb{T}_2$ if all sample shape polynomials converge.

# Sample shape convergence

*Distance polynomials* can be approximated uniformly by *shape polynomials*.

---

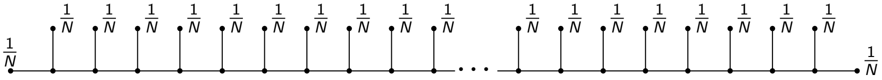
Consequently, if $x, x' \in \mathbb{T}_2$ then

$$x = x' \text{ iff } \Phi^{m,\phi}(x) = \Phi^{m,\phi}(x'), \forall m \in \mathbb{N}, \phi : \mathfrak{C}_m \to \mathbb{R}.$$

**Definition (Sample shape convergence)**

We say that a sequence $(x_n)_{n\in\mathbb{N}}$ converges in *sample shape* to $x$ in $\mathbb{T}_2$ if all sample shape polynomials converge.

$\rightsquigarrow$ For example, the *comb tree* with $N$ leaves and uniform distribution on the leaves *converges in sample shape to a line segment* with a continuous mass distribution.



# Sample subtree mass convergence

$\rightsquigarrow$ yet another notion of convergence -:)

# Sample subtree mass convergence

$\rightsquigarrow$ yet another notion of convergence -:)

given $(T, c, \mu) \in \mathbb{T}_2$, pick $U_1$, $U_2$, ... i.i.d. $\sim \mu$

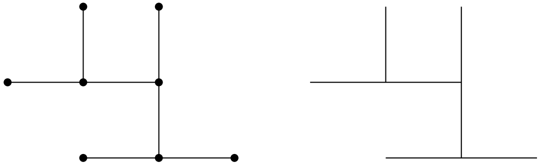
- Consider the subtree $S_{(T,c)}(U_1, ..., U_m)$, and label the generated branch points according to the order of their appearance.
- Evaluate the masses of the subtrees branching off.

# Sample subtree mass convergence

$\rightsquigarrow$ yet another notion of convergence -:)

given $(T, c, \mu) \in \mathbb{T}_2$, pick $U_1$, $U_2$, ... i.i.d. $\sim \mu$

- Consider the subtree $S_{(T,c)}(U_1, ..., U_m)$, and label the generated branch points according to the order of their appearance.
- Evaluate the masses of the subtrees branching off.

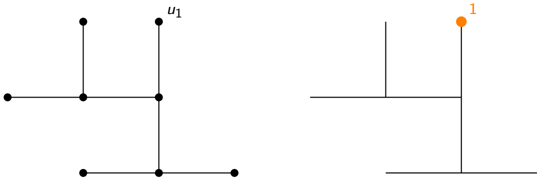


# Sample subtree mass convergence

$\rightsquigarrow$ yet another notion of convergence -:)

given $(T, c, \mu) \in \mathbb{T}_2$, pick $U_1$, $U_2$, ... i.i.d. $\sim \mu$

- Consider the subtree $S_{(T,c)}(U_1, ..., U_m)$, and label the generated branch points according to the order of their appearance.
- Evaluate the masses of the subtrees branching off.

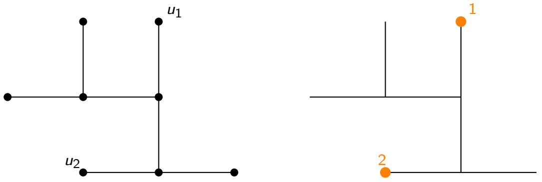


# Sample subtree mass convergence

$\rightsquigarrow$ yet another notion of convergence -:)

given $(T, c, \mu) \in \mathbb{T}_2$, pick $U_1$, $U_2$, ... i.i.d. $\sim \mu$

- Consider the subtree $S_{(T,c)}(U_1, ..., U_m)$, and label the generated branch points according to the order of their appearance.
- Evaluate the masses of the subtrees branching off.

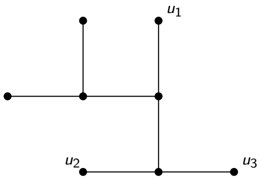


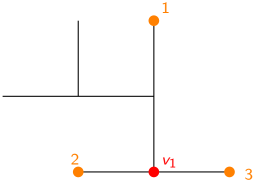
# Sample subtree mass convergence

$\rightsquigarrow$ yet another notion of convergence -:)

given $(T, c, \mu) \in \mathbb{T}_2$, pick $U_1$, $U_2$, ... i.i.d. $\sim \mu$

- Consider the subtree $S_{(T,c)}(U_1, ..., U_m)$, and label the generated branch points according to the order of their appearance.
- Evaluate the masses of the subtrees branching off.



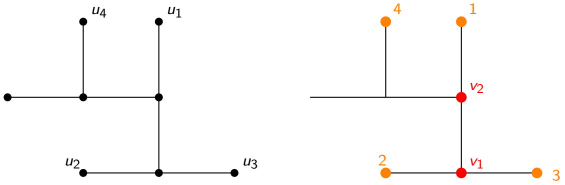



# Sample subtree mass convergence

$\rightsquigarrow$ yet another notion of convergence -:)

given $(T, c, \mu) \in \mathbb{T}_2$, pick $U_1$, $U_2$, ... i.i.d. $\sim \mu$

- Consider the subtree $S_{(T,c)}(U_1, ..., U_m)$, and label the generated branch points according to the order of their appearance.
- Evaluate the masses of the subtrees branching off.

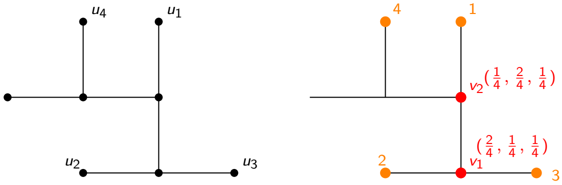


# Sample subtree mass convergence

$\rightsquigarrow$ yet another notion of convergence -:)

given $(T, c, \mu) \in \mathbb{T}_2$, pick $U_1, U_2, \ldots$ i.i.d. $\sim \mu$

- Consider the subtree $S_{(T,c)}(U_1, \ldots, U_m)$, and label the generated branch points according to the order of their appearance.
- Evaluate the masses of the subtrees branching off.

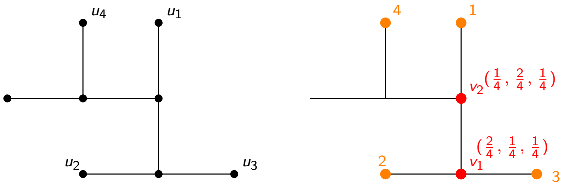


# Sample subtree mass convergence

$\rightsquigarrow$ yet another notion of convergence -:)

given $(T, c, \mu) \in \mathbb{T}_2$, pick $U_1, U_2, \ldots$ i.i.d. $\sim \mu$

- Consider the subtree $S_{(T,c)}(U_1, \ldots, U_m)$, and label the generated branch points according to the order of their appearance.
- Evaluate the masses of the subtrees branching off.




**Definition (Subtree mass convergence)**

$(T_n, c_n, \mu_n)$ converges to $(T, c, \mu)$ in the sense of *convergence of sample subtree masses* if for every $m \in \mathbb{N}$ the vectors of masses branching off the branch points in the subtree spanned by a sample of size $m$ converge.

# Main result

*Obvious inclusions*

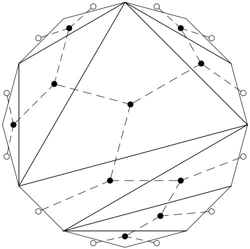
- Sample shape convergence and sample subtree masses convergence imply global convergence.
- Sample shape convergence implies sample subtree masses convergence.

**Theorem (Löhr & W. (2021))**

1. *All three notions of convergence are equivalent on* $\mathbb{T}_2$.
2. $\mathbb{T}_2$ *is compact.*

$\rightsquigarrow$ Proof uses correspondence to **(sub-)triangulations of the circle**.

[Aldous'94]



| | | |
|---|---|---|
| branch points | = | (open) triangles |
| edges connecting two vertices | = | separating side of the triangles |
| leaves | = | sides of the polygon |

$\rightsquigarrow$ Continuum limits (= triangulations of the circle) encode binary algebraic trees (without the measure); [Curien and Le Gall (2011)],[Curien, Haas, and Kortchemski (2015)],...

$\rightsquigarrow$ encode also the measure and unify discrete and continuum trees

**Definition ((sub-)triangulations; Aldous (1994), Löhr & W. (2021))**

A closed subset $C \neq \emptyset$ of the closed disc $\mathbb{D}$ is called **a sub-triangulation of the circle** if and only if the following two conditions hold:

**(T1)** $\mathrm{conv}(C) \setminus C$ is the disjoint union of open interiors of triangles.

**(T2)** $C$ is the union of non-crossing (non-intersecting except at endpoints), (possibly degenerated) closed straight line segments with endpoints in the circle $\mathbb{S} := \partial\mathbb{D}$.

# Sub-triangulations of the circle

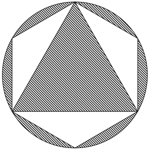
**Definition ((sub-)triangulations; Aldous (1994), Löhr & W. (2021))**

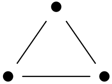
A closed subset $C \neq \emptyset$ of the closed disc $\mathbb{D}$ is called **a sub-triangulation of the circle** if and only if the following two conditions hold:

**(T1)** $\mathrm{conv}(C) \setminus C$ is the disjoint union of open interiors of triangles.

**(T2)** $C$ is the union of non-crossing (non-intersecting except at endpoints), (possibly degenerated) closed straight line segments with endpoints in the circle $\mathbb{S} := \partial\mathbb{D}$.

$\rightsquigarrow$ Aldous' original definition missed **(T2)** which can be shown to be equivalent for branch points to exist.





Remember: I am not a tree

# Sub-triangulations of the circle

**Definition ((sub-)triangulations; Aldous (1994), Löhr & W. (2021))**

A closed subset $C \neq \emptyset$ of the closed disc $\mathbb{D}$ is called **a sub-triangulation of the circle** if and only if the following two conditions hold:

**(T1)** $\mathrm{conv}(C) \setminus C$ is the disjoint union of open interiors of triangles.

**(T2)** $C$ is the union of non-crossing (non-intersecting except at endpoints), (possibly degenerated) closed straight line segments with endpoints in the circle $\mathbb{S} := \partial\mathbb{D}$.

$\mathcal{T} :=$ set of all subtriangulations of the circle

$$\mathcal{T}^{\mathrm{cont}} := \{C \in \mathcal{T} : \mathbb{S} \subseteq C\}$$

# Sub-triangulations of the circle

**Definition ((sub-)triangulations; Aldous (1994), Löhr & W. (2021))**

A closed subset $C \neq \emptyset$ of the closed disc $\mathbb{D}$ is called **a sub-triangulation of the circle** if and only if the following two conditions hold:

**(T1)** $\mathrm{conv}(C) \setminus C$ is the disjoint union of open interiors of triangles.

**(T2)** $C$ is the union of non-crossing (non-intersecting except at endpoints), (possibly degenerated) closed straight line segments with endpoints in the circle $\mathbb{S} := \partial\mathbb{D}$.

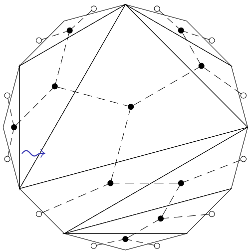
$$\mathcal{T} := \text{ set of all subtriangulations of the circle}$$

$$\mathcal{T}^{\mathrm{cont}} := \{C \in \mathcal{T} : \mathbb{S} \subseteq C\}$$

**Proposition (Löhr & W. (2021))**

*$\mathcal{T}$ and $\mathcal{T}^{\mathrm{cont}}$ are closed in the space of all closed subsets of $\mathbb{D}$ with respect to the Hausdorff metric and therefore compact.*

# From sub-triangulations to algebraic measure trees (2021)



| | | |
|---|---|---|
| branch points | = | (open) triangles |
| edges | = | sides of triangles |
| leaves | = | lie on circle |

Measure can be also encoded.

| | | |
|---|---|---|
| mass on leave | = | length on circle |

**Proposition (Löhr & W.)**

*There is a surjective map $\tau : \mathcal{T} \to \mathbb{T}_2$, which is continuous when $\mathbb{T}_2$ is equipped with the sample shape topology.*

$\rightsquigarrow$ From here one can conclude equivalence of the notions of convergence, compactness and metrizability of $\mathbb{T}_2$.

**Corollary**

*The sets of sample shape polynomials and of sample subtree mass polynomials are convergence determining for probability measures on $\mathbb{T}_2$.*

$\rightsquigarrow$ Stone-Weierstrass theorem; polynomials closed under multiplication

# Why is the result of interest?

1. Using global convergence allows to exploit well-known results about Gromov-weak convergence.
2. Showing convergence of graph theoretic tree-valued MCs as the number of vertices tends to infinity simplifies if we have *compact state space*:
   - [Löhr, Mytnik & W., *"Aldous chain on cladograms in the diffusion limit"*, (2020)]
   - [Nussbaumer & W., *"The algebraic $\alpha$-Ford tree under evolution"*, (arXiv2022)]
   - [Gambelin, *"The stable algebraic measure tree diffusion"*, part of his phD thesis]
3. The convergence of sample subtree mass tensor distributions allows to analyse the limit processes with stochastic analysis methods.

$\rightsquigarrow$ Approach is already generalized to two-level algebraic measure trees: [Nussbaumer, Tran Viet & W. (arXiv2022)]

# The end

Many thanks!

# References


- Dress; *"Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: a note on combinatorial properties of metric spaces"*, Adv. in Math. (1984)
- Greven, Pfaffelhuber and Winter; *Convergence of mm-spaces*, PTRF, (2009)
- Evans, Grübel & Wakolbinger; *"Doob-Martin boundary of Rémy tree growth chain"*, (2017)
- Fabel; *"A topological characterization of the underlying spaces of complete $\mathbb{R}$-trees"*, Michigan Math. J.(2015)
- Fisher & Zastrov; *"Combinatorial $\mathbb{R}$-trees as generalized Cayley graphs for fundamental groups of one-dimensional spaces"*, Geom. Dedicata, (2013)
- Forman; *"Exchangeable hierarchies and mass-structure of weighted real trees"*, EJP, (2020)
- Löhr; *Equivalence of Gromov-Prohorov- and Gromov's $\underline{\square}_\lambda$-metric on the space of metric measure spaces*, ECP, (2013)
- Löhr and Winter; *Algebraic measure trees and triangulations of the circle*, Bullein SFM (2021)
- Tits; *"A theorem of Lie-Kolchin for trees"*, (1977)
- Vapnik and Chervonenkis; *On the uniform convergence of relative frequencies of events to their probabilities*, Theor. Probability Appl. (1971)