Scaling limits of k-ary growing trees

Robin Stephenson

joint work with Bénédicte Haas

CEREMADE, Université Paris-Dauphine

The model

Construction algorithm.

Fix an integer $k \geqslant 2$. We define a sequence $(T_n(k), n \geqslant 0)$ of random k-ary trees by the following recursion:

Construction algorithm.

Fix an integer $k \geqslant 2$. We define a sequence $(T_n(k), n \geqslant 0)$ of random k-ary trees by the following recursion:

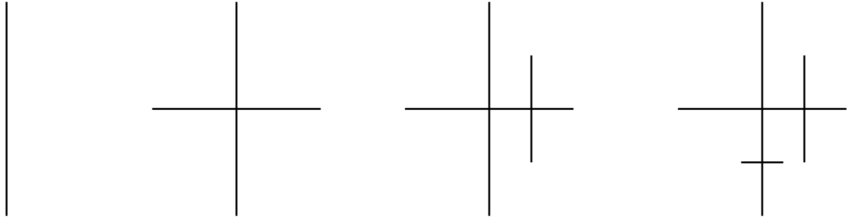
- $T_0(k)$ is the tree with a single edge and two vertices, a root and a leaf.

Construction algorithm.

Fix an integer $k \geqslant 2$. We define a sequence $(T_n(k), n \geqslant 0)$ of random k-ary trees by the following recursion:

- $T_0(k)$ is the tree with a single edge and two vertices, a root and a leaf.
- given $T_n(k)$, to make $T_{n+1}(k)$, choose uniformly at random one of its edges, add a new vertex in the middle, thus splitting this edge in two, and then add $k-1$ new edges starting from the new vertex.

$T_n(3)$ for $n = 0, 1, 2, 3.$



A few observations

- $T_n(k)$ has n internal nodes, $kn+1$ edges and $(k-1)n+1$ leaves

A few observations

- $T_n(k)$ has n internal nodes, $kn+1$ edges and $(k-1)n+1$ leaves
- When $k=2$, we recover a well-known algorithm of Rémy, used to generate uniform binary trees. It is then well-known that, when rescaled by $\sqrt{n}$, the tree $T_n(2)$ converges almost surely to a scalar multiple of Aldous' Brownian continuum random tree.

The main convergence theorem

Let $\mu_n(k)$ be the uniform measure on the set of leaves of $T_n(k)$.

Theorem

As n tends to infinity, we have

$$\left(\frac{T_n(k)}{n^{1/k}}, \mu_n(k)\right) \xrightarrow{\mathbb{P}} (\mathcal{T}_k, \mu_k),$$

where $(\mathcal{T}_k, \mu_k)$ is a random compact k-ary $\mathbb{R}$-tree with a measure μ_k on the set of its leaves, and the convergence is a convergence in probability for the Gromov-Hausdorff-Prokhorov metric.

Preliminaries

$\mathbb{R}$-trees

An $\mathbb{R}$-tree is a metric space $(\mathcal{T}, d)$ which satisfies the following two conditions:

- for all $x, y \in \mathcal{T}$, there exists a unique isometric map $\varphi_{x,y}\colon [0, d(x, y)] \to \mathcal{T}$ such that $\varphi_{x,y}(0) = x$ and $\varphi_{x,y}(d(x, y)) = y$.

$\mathbb{R}$-trees

An $\mathbb{R}$-tree is a metric space $(\mathcal{T}, d)$ which satisfies the following two conditions:

- for all $x, y \in \mathcal{T}$, there exists a unique isometric map $\varphi_{x,y}\colon [0, d(x, y)] \to \mathcal{T}$ such that $\varphi_{x,y}(0) = x$ and $\varphi_{x,y}(d(x, y)) = y$.
- for any continuous self-avoiding path $c\colon [0, 1] \to \mathcal{T}$, we have $c([0, 1]) = \varphi_{x,y}([0, d(x, y)])$, where $x = c(0)$ et $y = c(1)$.

Informally, there exists a unique continuous path between any two points, and this path is isometric to a line segment.

In practise, we will only want to look at rooted and measured trees: these are objects of the form $(\mathcal{T}, d, \rho, \mu)$ where ρ is a point on $\mathcal{T}$ called the root and μ is a Borel probability measure on $\mathcal{T}$. Since d and ρ will never be ambiguous, we shorten the notation to $(\mathcal{T}, \mu)$.

All our trees will also be compact.

Gromov-Hausdorff-Prokhorov topology

Gromov-Hausdorff-Prokhorov topology

Let $(\mathcal{T}, d, \rho, \mu)$ et $(\mathcal{T}', d', \rho', \mu')$ be two rooted measured and compact $\mathbb{R}$-trees. Let

$$d_{GHP}\Big((\mathcal{T}, \mu), (\mathcal{T}', \mu')\Big) =$$

$$\inf \left[\max \Big(d_{\mathcal{Z},H}(\varphi(\mathcal{T}), \varphi'(\mathcal{T}')), d_{\mathcal{Z}}(\varphi(\rho), \varphi'(\rho')), d_{\mathcal{Z},P}(\varphi_*\mu, \varphi'_*\mu') \Big) \right],$$

where

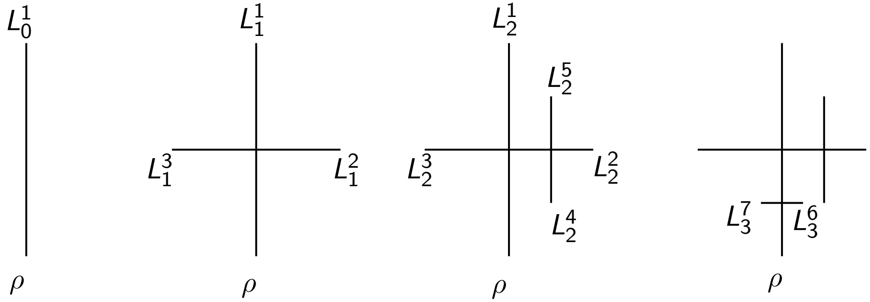
- the infimum is taken on all possible isometric embeddings φ and φ' of $(\mathcal{T}, d)$ and $(\mathcal{T}', d')$ in a common metric space $(\mathcal{Z}, d_{\mathcal{Z}})$
- $d_{\mathcal{Z},H}$ is the Hausdorff metric between nonempty closed subsets of $\mathcal{Z}$
- $d_{\mathcal{Z},P}$ being the Prokhorov metric between Borel probability measures on $\mathcal{Z}$.

This defines a well-behaved metric on the set of compact rooted measured trees.

Convergence in probability

Labelling the leaves of $T_n(k)$

Labelling the leaves of $T_n(k)$



Convergence of the distance between two leaves

Lemma

Let $i, j \in \mathbb{N}$. Then both

$$\frac{d(\rho, L_n^i)}{n^{1/k}}$$

and

$$\frac{d(L_n^i, L_n^j)}{n^{1/k}}$$

converge almost surely as n tends to infinity.

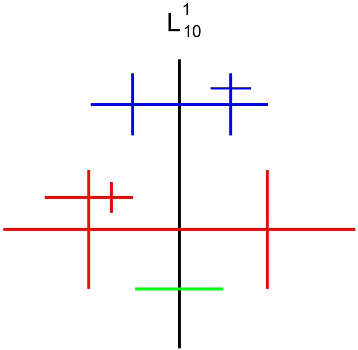
Proof of the lemma

We concentrate on the case of $d(\rho, L_n^1)$.

We use the theory of *Chinese restaurant processes*: imagine that $T_n(k)$ is a restaurant, and that its internal nodes are its clients, which enter successively. Say that two clients u and v are on the same table if the paths $[[\rho, u]]$ and $[[\rho, v]]$ branch out of $[[\rho, L_n^1]]$ at the same point.

The distance $d(\rho, L_n^1)$ is then equal to the number of tables, plus one.

The restaurant associated to $T_{10}(3)$



Proof of the lemma, continued

When we go from $T_n(k)$ to $T_{n+1}(k)$, we add a new client to the restaurant, and this client either sits at an existing table or we add a new table. The probabilities can be computed easily. In particular, assuming that there are $l \in \mathbb{N}$ tables at time n,

$$\mathbb{P}(\text{the } n+1\text{-th client sits at a new table}) = \frac{l+1}{kn+1}$$

Models of restaurant processes which satisfy this have been studied by Pitman, and it is well known that

$$\frac{\text{number of tables}}{n^{1/k}}$$

converges as n tends to infinity to a generalized Mittag-Leffler random variable.

The limiting tree $\mathcal{T}_k$

The limiting tree $\mathcal{T}_k$

Let $d_{i,j} = \lim_{n\to\infty} n^{-1/k} d(L_n^i, L_n^j)$ and $h_i = \lim_{n\to\infty} n^{-1/k} d(\rho, L_n^i)$.

The limiting tree $\mathcal{T}_k$

Let $d_{i,j} = \lim_{n \to \infty} n^{-1/k} d(L_n^i, L_n^j)$ and $h_i = \lim_{n \to \infty} n^{-1/k} d(\rho, L_n^i)$.

There exists a unique tree $\mathcal{T}_k$ equipped with a dense subset of leaves $\{L^i, i \in \mathbb{N}\}$ such that $d(L^i, L^j) = d_{i,j}$ and $d(\rho, L^i) = d_i$.

The limiting tree $\mathcal{T}_k$

Let $d_{i,j} = \lim_{n\to\infty} n^{-1/k} d(L_n^i, L_n^j)$ and $h_i = \lim_{n\to\infty} n^{-1/k} d(\rho, L_n^i)$.

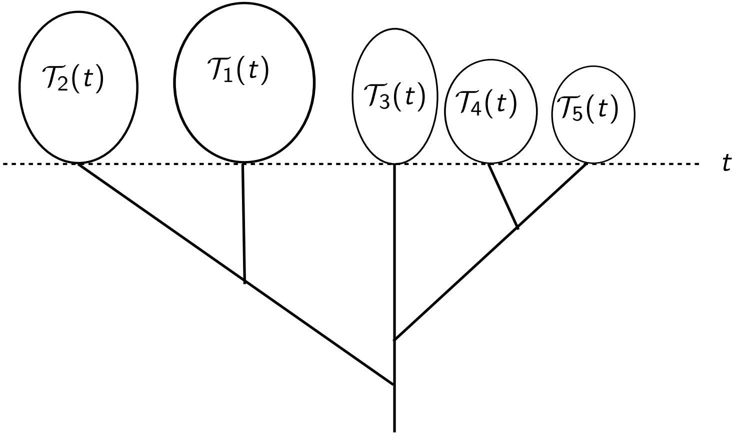
There exists a unique tree $\mathcal{T}_k$ equipped with a dense subset of leaves $\{L^i, i \in \mathbb{N}\}$ such that $d(L^i, L^j) = d_{i,j}$ and $d(\rho, L^i) = d_i$.

Convergence of $n^{-1/k} T_n(k)$ to $\mathcal{T}_k$ is then proven using the lemma and a tightness property.

The distribution of $\mathcal{T}_k$

Self-similar fragmentation trees

Let $\alpha < 0$, and $(\mathcal{T}, d, \rho, \mu)$ be a compact random tree.
For $t \geqslant 0$, we let $\mathcal{T}_1(t), \mathcal{T}_2(t), \ldots$ be the connected components of $\{x \in \mathcal{T}, d(\rho, x) > t\}$.



Self-similar fragmentation trees

We say that $(\mathcal{T}, \mu)$ is a self-similar fragmentation tree with index α if, for all $t \geqslant 0$, conditionally on $(\mu(\mathcal{T}_i(s);\ i \in \mathbb{N}, s \leqslant t))$:

- (Branching property) The subtrees $(\mathcal{T}_i(t), \mu_{\mathcal{T}_i(t)})$ are mutually independent.
- (Self-similarity) For any i, the tree $(\mathcal{T}_i(t), \mu_{\mathcal{T}_i(t)})$ has the same distribution as the original tree $(\mathcal{T}, \mu)$, multiplied by $\mu(\mathcal{T}_i(t))^{-\alpha}$.

The notation $\mu_{\mathcal{T}_i(t)}$ means the measure μ conditioned to the subset $T_i(t)$, which is a probability distribution.

Self-similar fragmentation trees

Linking these trees to the self-similar fragmentation processes of Bertoin shows that their distribution is characterized by three parameters:

- The index of self-similarity α.
- An *erosion coefficent* $c \geqslant 0$ which determines how μ is spread out on line segments.
- A *dislocation measure* ν, which is a σ-finite measure on the set

$$\mathcal{S}^{\downarrow} = \{\mathbf{s} = (s_i)_{i\in\mathbb{N}} : s_1 \geqslant s_2 \geqslant \ldots \geqslant 0, \sum s_i \leqslant 1\}.$$

  This measure determines how we allocate the mass when there is a branching point.

$\mathcal{T}_k$ is a self-similar fragmentation tree

Theorem

The tree $\mathcal{T}_k$ has the law of a self-similar fragmentation tree with:

- $\alpha = -\frac{1}{k}$.
- $c = 0$.
- *The measure ν_k is k-ary and conservative: it is supported on sequences such that $s_i = 0$ for $i \geqslant k+1$ and $\sum_{i=1}^k s_i = 1$, and we have*

$$\nu(\mathrm{d}\mathbf{s}) = \frac{(k-1)!}{k(\Gamma(\frac{1}{k}))^{(k-1)}} \prod_{i=1}^{k} s_i^{-(1-1/k)} \left(\sum_{i=1}^{k} \frac{1}{1-s_i} \right) \mathbf{1}_{\{s_1 \geqslant s_2 \geqslant \ldots \geqslant s_k\}} \mathrm{d}\mathbf{s}$$

Main steps of the proof

A theorem of Haas and Miermont states that, to show that $\mathcal{T}_k$ is a fragmentation tree, we can show that $T_n(k)$ is a "discrete self similar tree", something called the *Markov branching property*:

Lemma

Let $T_n^1, \ldots, T_n^k$ be the k subtrees rooted at the first node of $T_n(k)$. We let $X_n^1, \ldots, X_n^k$ be their number of internal nodes, and we order these such that $X_n^1 \geqslant X_n^2 \geqslant \ldots \geqslant X_n^k$. Then, conditionally on $X_n^1, \ldots, X_n^k$,

- *$T_n^1, \ldots, T_n^k$ are independent.*
- *For all i, T_n^i has the same distribution as $T_{X_n^i}(k)$.*

This is proved by using an induction on n.

Main steps of the proof

Next, we have to prove that, properly renormalized, the "discrete dislocation measures" converge:

Lemma

Let $\bar{q}_n$ be the distribution of $\left(\frac{X_n^1}{n-1}, \ldots, \frac{X_n^k}{n-1}\right)$. We then have

$$n^{1/k}(1 - s_1)\bar{q}_n(\mathrm{d}\mathbf{s}) \underset{n\to\infty}{\longrightarrow} (1 - s_1)\nu_k(\mathrm{d}\mathbf{s})$$

in the sense of weak convergence of measures on $\mathcal{S}^{\downarrow}$.

This part can be done by explicitely computing $\bar{q}_n$.

Fractal dimension of $\mathcal{T}_k$

Corollary

The Hausdorff dimension of $\mathcal{T}_k$ is almost surely equal to k.

This is a consequence of well-known results on fragmentation trees.

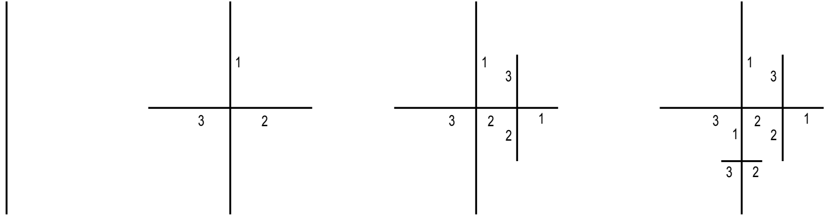
Stacking the trees

Labelling the edges

Labelling the edges

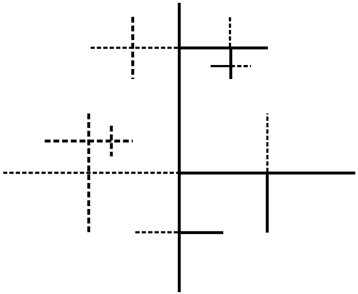
Each step of the algorithm creates k new edges. We give them labels 1 to k the following way:

- The upper half of the edge which was split in two is labeled 1
- The other new edges are labeled $2, \ldots, k$.



$T_n(k')$ inside $T_n(k)$

Consider an integer $k' < k$. Let $T_n(k, k')$ be the subset of $T_n(k)$ where we have erased all edges with labels $k'+1, k'+2, \ldots, k$ and all their descendants.



If we call I_n be the number of internal nodes which are in $T_n(k, k')$, then one can check that:

- Conditionally on I_n, $T_n(k, k')$ is distributed as $T_{I_n}(k')$.

$\mathcal{T}_{k'}$ inside $\mathcal{T}_k$

One can show that the sequence $\frac{I_n}{n^{k'/k}}$ converges a.s. to a random variable M. As a consequence we obtain:

Proposition

$$\frac{T_n(k,k')}{n^{1/k}} \xrightarrow{\mathbb{P}} M\,\mathcal{T}_{k,k'}$$

where $\mathcal{T}_{k,k'}$ is a version of $\mathcal{T}_{k'}$ hidden in $\mathcal{T}_k$, and is independent of M.

More on the stacking

- It is in fact possible to extract directly from $\mathcal{T}_k$ a subtree distributed as $\mathcal{T}_{k'}$, without going back to the finite case: at every branch point of $\mathcal{T}_k$, select only k' of the k branches at random with a well-chosen distribution.

More on the stacking

- It is in fact possible to extract directly from $\mathcal{T}_k$ a subtree distributed as $\mathcal{T}_{k'}$, without going back to the finite case: at every branch point of $\mathcal{T}_k$, select only k' of the k branches at random with a well-chosen distribution.
- Using the Kolmogorov extension theorem, one obtains the existence of a sequence $(J_k, k \geqslant 2)$ such that

$$\forall k' < k, J_{k'}\mathcal{T}_{k'} \subset J_k\mathcal{T}_k.$$

Thank you!