Scaling limits of k-ary growing trees

Robin Stephenson

joint work with Bénédicte Haas

CEREMADE, Université Paris-Dauphine

Colloque JPS

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:

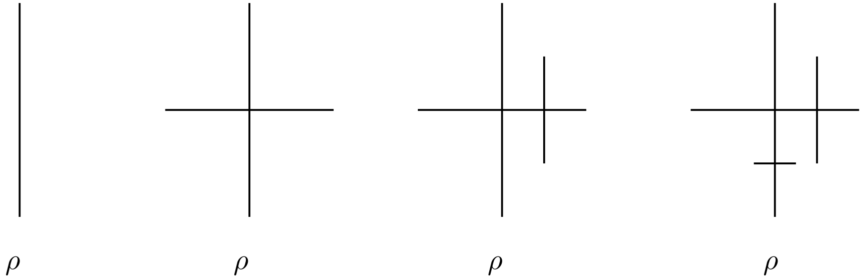
- $T_0(k)$ is the tree with a single edge and two vertices, a root and a leaf.

Construction algorithm.

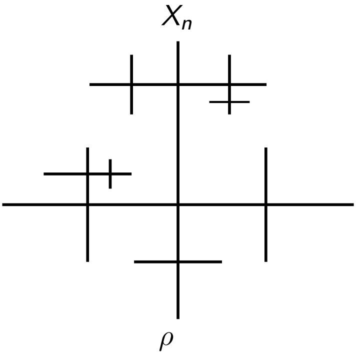
Fix an integer $k \geqslant 2$. We define a sequence $(T_n(k), n \geqslant 0)$ of random k-ary trees by the following recursion:

- $T_0(k)$ 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 $n = 0, 1, 2, 3$.



$n = 10$



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 the “size" of $T_n(2)$ is of order $\sqrt{n}$.

Size of $\mathcal{T}_n(k)$ and convergence

Distance between one point and the root

Let X_n be the point of $T_n(k)$ corresponding to the leaf of $T_0(k)$. Its distance to the root is of order $n^{1/k}$.

Lemma

$d(\rho, X_n)$ is of order $n^{1/k}$ as n goes to infinity, and in fact the quotient

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

converges a.s.

Distance between one point and the root

Let X_n be the point of $T_n(k)$ corresponding to the leaf of $T_0(k)$. Its distance to the root is of order $n^{1/k}$.

Lemma

$d(\rho, X_n)$ is of order $n^{1/k}$ as n goes to infinity, and in fact the quotient

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

converges a.s.

The proof of this lemma is simple. The process $(d(\rho, X_n), n \in \mathbb{Z}_+)$ is in fact a Markov chain: if we know that $d(\rho, X_n) = L$ then

$$d(\rho, X_{n+1}) = \begin{cases} L+1 \text{ with probability } \frac{L}{(k-1)n+1} \\ L \text{ with probability } 1 - \frac{L}{(k-1)n+1} \end{cases}$$

Martingale methods then easily give us the existence of the limit we want.

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

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

What kind of object is $\mathcal{T}_k$?

$\mathbb{R}$-trees

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

$\mathbb{R}$-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, an $\mathbb{R}$-tree is a connected union of line segments with no loops.

Graph-theoretical trees can be interpreted as $\mathbb{R}$-trees by considering each edge as a line segment and giving some length to each edge. In our case, the notation $\frac{T_n(k)}{n^{1/k}}$ means that we give to each edge of $T_n(k)$ a length equal to $n^{-1/k}$.

Graph-theoretical trees can be interpreted as $\mathbb{R}$-trees by considering each edge as a line segment and giving some length to each edge. In our case, the notation $\frac{T_n(k)}{n^{1/k}}$ means that we give to each edge of $T_n(k)$ a length equal to $n^{-1/k}$.

However, the structure of an $\mathbb{R}$-tree can be very complex, notably because the leaves or the branch points can form a dense subset.

Graph-theoretical trees can be interpreted as $\mathbb{R}$-trees by considering each edge as a line segment and giving some length to each edge. In our case, the notation $\frac{T_n(k)}{n^{1/k}}$ means that we give to each edge of $T_n(k)$ a length equal to $n^{-1/k}$.

However, the structure of an $\mathbb{R}$-tree can be very complex, notably because the leaves or the branch points can form a dense subset.

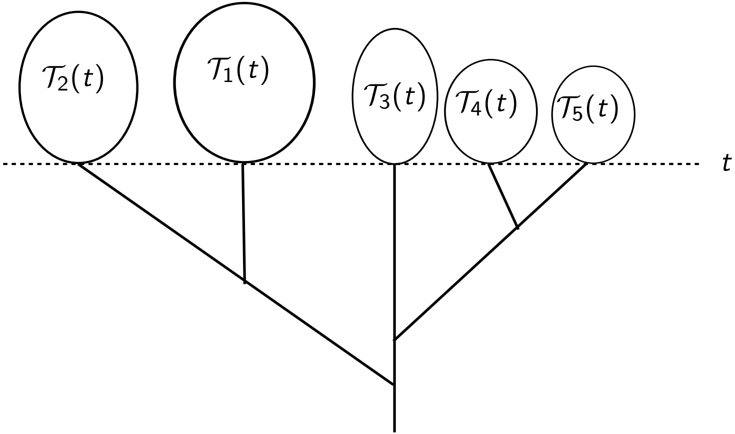
In practise, we will only want to look at rooted and measured trees: these are objects of the form $(\mathcal{T}, d, \rho, \mu)$ where ρ is a point on $\mathcal{T}$ called the root and μ is a Borel probability measure on $\mathcal{T}$. Since d and ρ will never be ambiguous, we shorten the notation to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

All our trees will also be compact.

Self-similar fragmentation trees

Self-similar fragmentation trees

Let $\alpha < 0$, and $(\mathcal{T}, d, \rho, \mu)$ be a compact random $\mathbb{R}$-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 $\left(\mu\Big(\mathcal{T}_i(s)\Big);\ i \in \mathbb{N}, s \leqslant t\right)$:

- (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 $(\mathcal{T}, \mu)$, rescaled by $\mu\Big(\mathcal{T}_i(t)\Big)^{-\alpha}$.

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*

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

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.

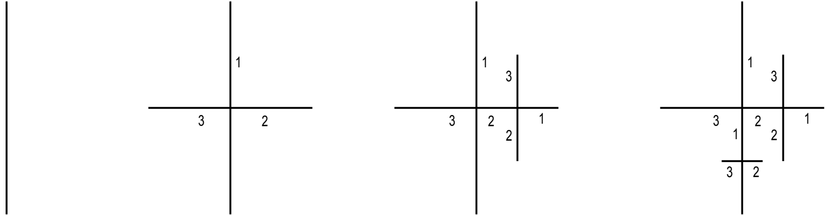
Stacking the trees

Labelling the edges

Labelling the edges

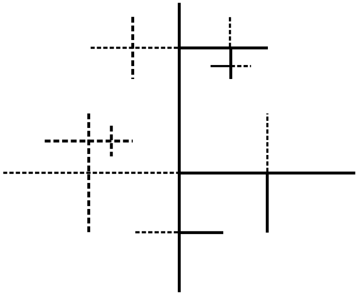
Each step of the algorithm creates k new edges. We give them labels 1 to k the following way:

- 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 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.

Thank you!