Random Trees and Their Limits

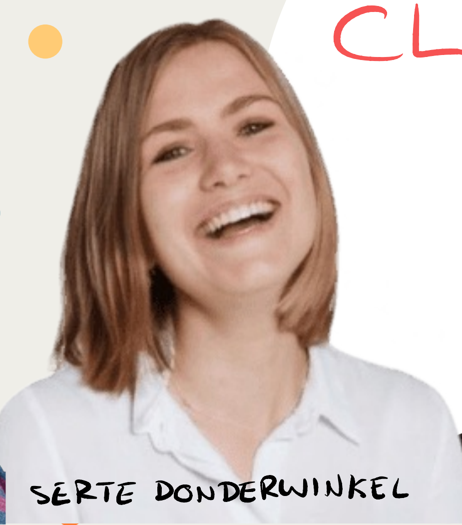
July 10
2023

XVI
CLAPEM

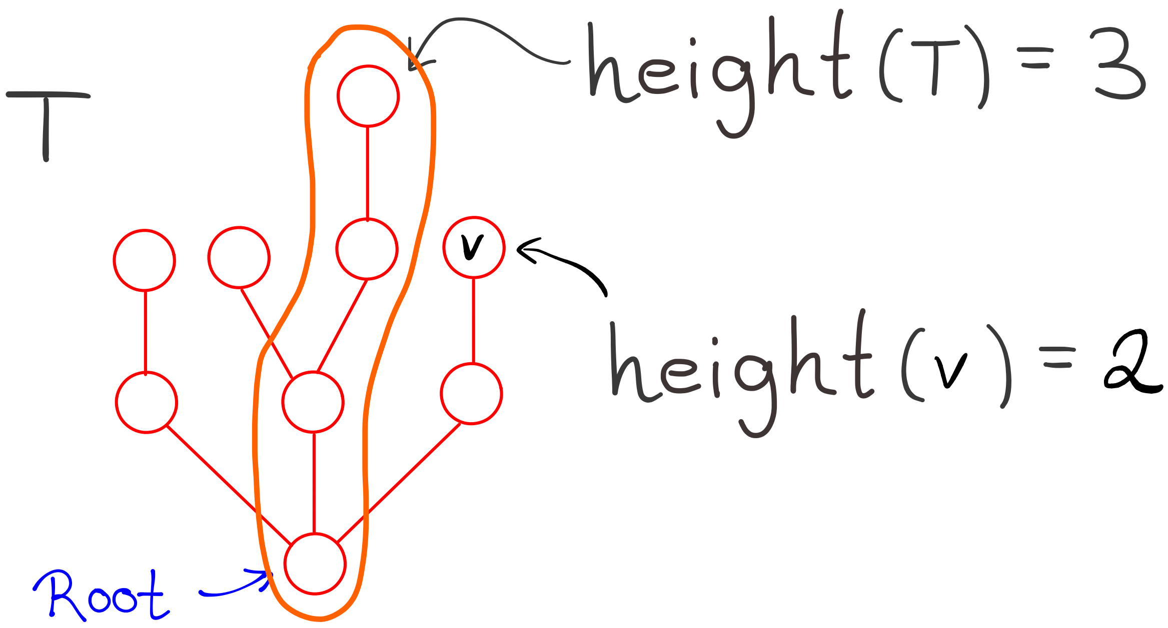

JAMES MARTIN


ARTHUR BLANC-RENAUDIE


IGOR KORTCHEMSKI


SERTE DONDERWINKEL

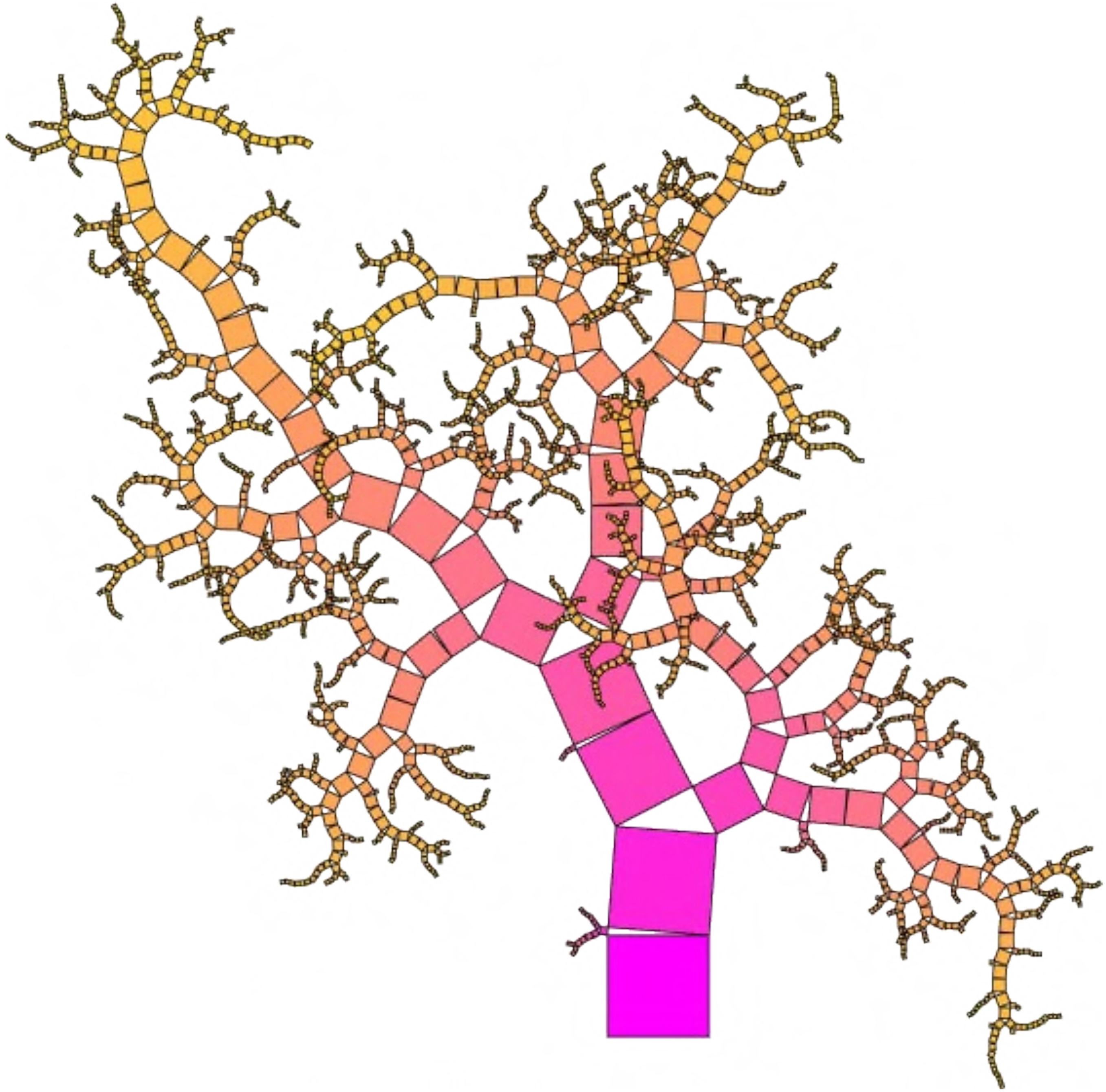

MIKAËL MAAZOUN



Height: Greatest distance from any node to the root

Q: How tall are random trees?

Part One: Models



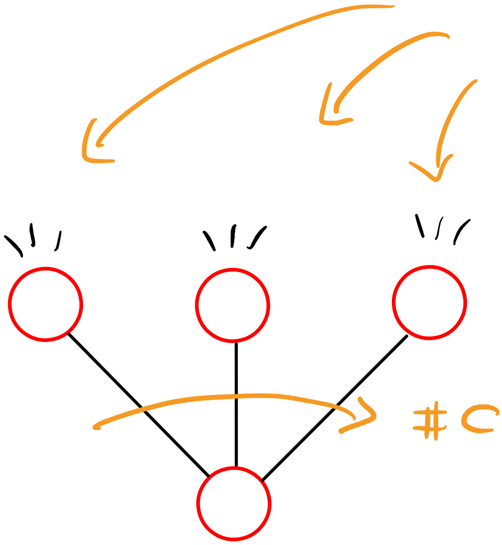
Models of Tree

① Bienaymé Trees*

* Family trees of branching processes

Offspring dist. μ

Let $X \sim \mu$, $T \sim \text{Bienaymé}(\mu)$



- $\mathbb{E}X > 1$

$$\mathbb{P}(ht(T) = \infty) > 0$$

$\mathbb{E}X = \lambda < 1$, $\mathbb{E}(X^2) < \infty$

$$\mathbb{P}(ht(T) \geq h) \sim c \cdot \lambda^h$$

(Kolmogorov, 1938)

$\mathbb{E}X = 1$, $\mathbb{E}(X^2) < \infty$

$$\mathbb{P}(ht(T) \geq h) \sim \frac{2}{Var(X)} \cdot \frac{1}{h}$$

(Kolmogorov, 1938; Kesten-Ney-Spitzer 1966)

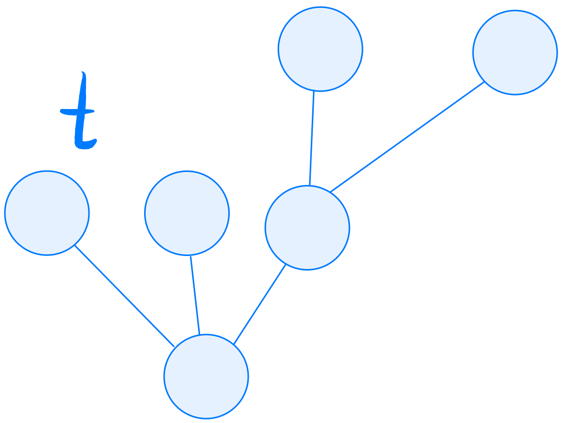
An aside about equiprobability

Type of tree t

$n_d(t) = \#\{v \in t : v \text{ has } d \text{ children}\}$

$\vec{n}(t) = (n_d(t), d \geq 0)$

$n_0(t) = \#$ leaves of t



$\vec{n}(t) = (4, 0, 1, 1, 0, 0, \ldots)$

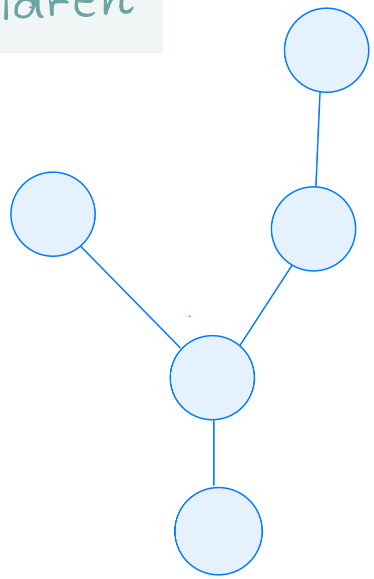
☆ "degree" := "# children"

If T is Bienaymé(μ)-distributed tree then for any fixed tree t,

$$\mathbb{P}(T = t) = \prod_{v \in t} \mu\{\deg_t(v)\} = \prod_{d \geq 0} \mu\{d\}^{n_d(t)}$$

so if $\vec{n}(t) = \vec{n}(t')$ then $\mathbb{P}(T = t) = \mathbb{P}(T = t')$.

e.g. $\mu = \frac{1}{3}\delta_0 + \frac{1}{2}\delta_1 + \frac{1}{6}\delta_2$



This tree has prob.

$\frac{1}{2} \cdot \frac{1}{6} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{216}$

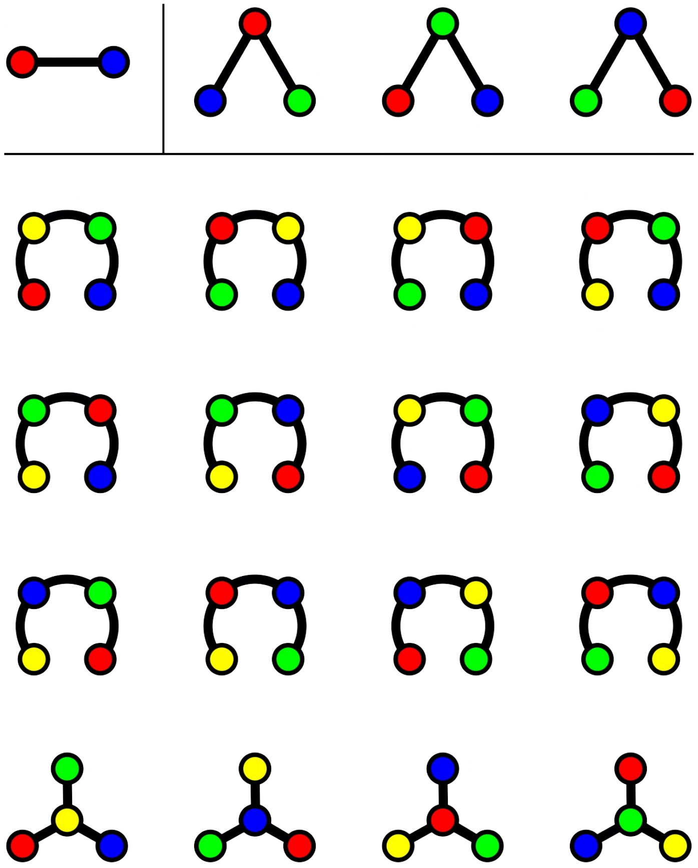
All trees of the same type are equiprob.

② Cayley Trees : Rooted labeled trees

$$\mathbb{T}_n = \{\text{rooted trees } t \text{ with } v(t) = \underbrace{\{1, 2, \dots, n\}}_{[n]}\}$$

Cayley's Formula:

$$|\mathbb{T}_n| = n^{n-1}$$



🔴 = 1

🔵 = 2

🟢 = 3

🟡 = 4

$H \overset{d}{=}$ supremum of standard Brownian excursion.

For $T_n \in_u \mathbb{T}_n$, $\mathrm{ht}(T_n)/n^{1/2} \xrightarrow{\text{dist}} 2H$ (Rényi + Szekeres, 1967)

$\exists\, c, C > 0$: $\mathbb{P}(\mathrm{ht}(T_n) \geq x n^{1/2}) \leq C \exp(-c x^2)$ (Łuczak 1995)

From Cayley to Bienaymé

For $T_n \in_u \mathbb{T}_n$, $\mathrm{ht}(T_n)/n^{1/2} \xrightarrow{\text{dist}} 2H$ (Renyi + Szekeres, 1967)

$\exists c, C > 0: \mathbb{P}(\mathrm{ht}(T_n) \geq x n^{1/2}) \leq C\exp(-cx^2)$ (Łuczak 1995)

These results also hold for (some) Bienaymé trees

Fix μ, $X \sim \mu$ with $\mathbb{E}X = 1$, $\mathrm{Var}(X) = \sigma^2 \in (0, \infty)$

Let $T_{n,\mu}$ be Bienaymé(μ) conditioned to have n vertices.

Then

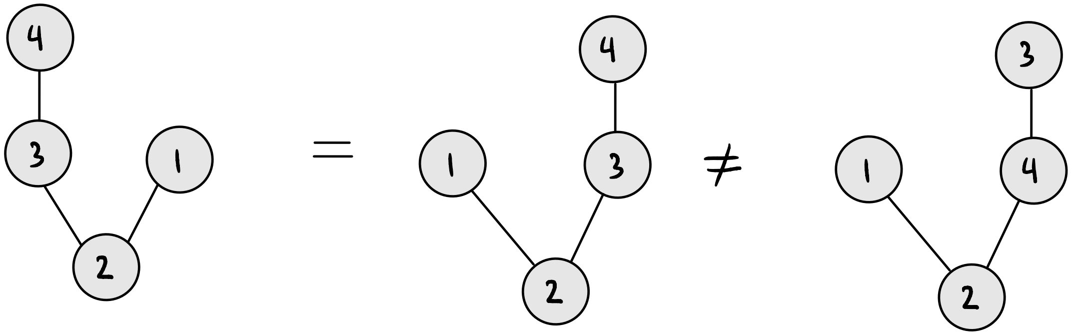
$\mathrm{ht}(T_{n,\mu})/n^{1/2} \xrightarrow{\text{dist}} \frac{2}{\sigma}H$ as $n \to \infty$ (Kolchin 1978)

$\exists c, C > 0: \mathbb{P}(\mathrm{ht}(T_{n,\mu}) \geq x n^{1/2}) \leq C\exp(-cx^2)$ (A-B, Devroye, Janson 2013)

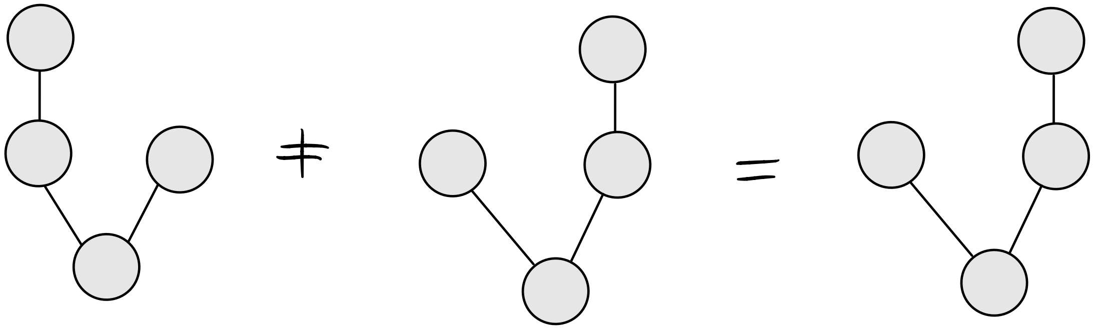
Conjecture: If $\sigma = \infty$ then $\mathrm{ht}(T_{n,\mu})/n^{1/2} \xrightarrow{\text{dist}} 0$ (Janson 2012)

The connection between Bienaymé and Cayley

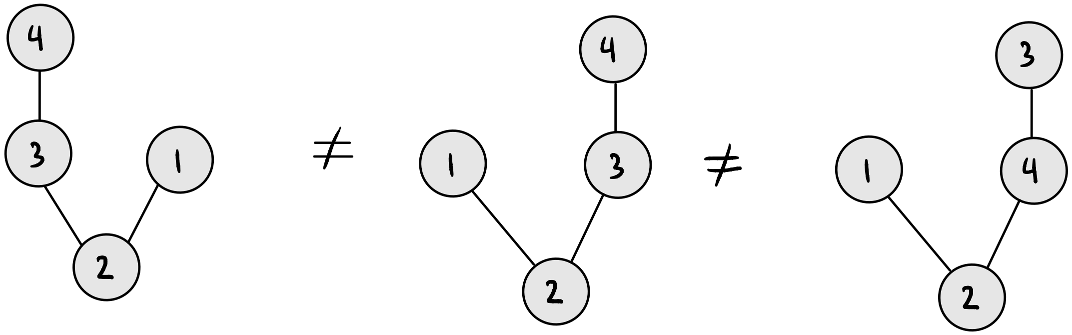
Cayley Trees: Rooted, labeled, unordered



Bienaymé Trees: Rooted, unlabeled, ordered



The connection: Rooted, labeled, ordered trees (with given degrees)



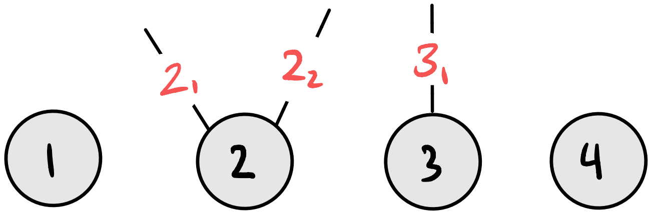
✯ "degree" := "# children"

Rooted, labeled, ordered trees (with given degrees)

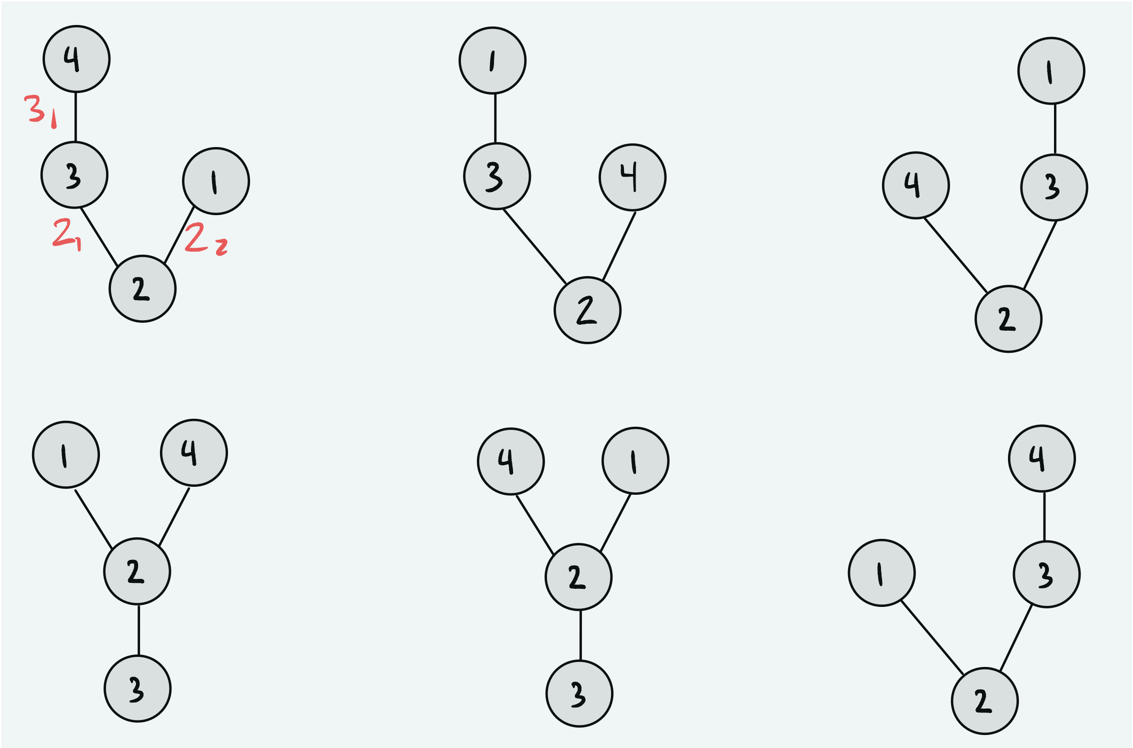
In a rooted, labeled, ordered tree with degrees $(d_1,\dots,d_n)$:

- Vertices labeled by $[n]$
- Edges labeled by $\bigcup_{i=1}^{n} \{ i_1, \dots, i_{d_i} \}$

Example:



$(d_1, d_2, d_3, d_4) = (0, 2, 1, 0)$



NB For $d = (d_1, \dots, d_n)$ to be the deg. seq. of a tree, necessary + sufficient that $d_1, \dots, d_n \in \mathbb{N}$, $\sum_{i=1}^{n} d_i = n-1$

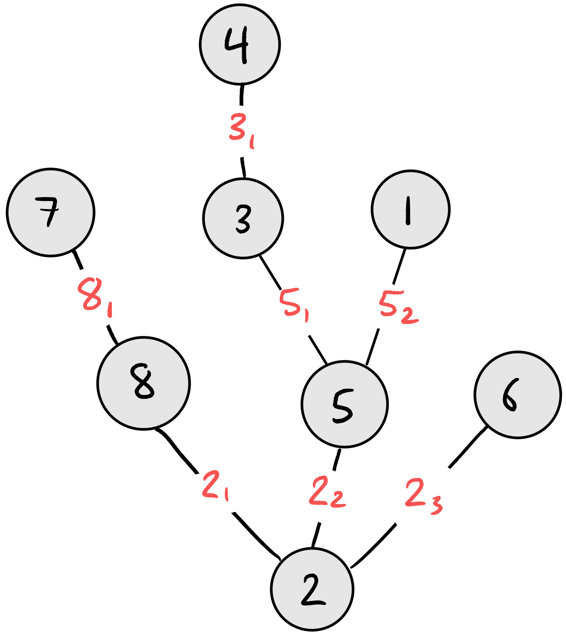
Prop: Fix $d=(d_1,\dots,d_n)$ s.t. $d_1,\dots,d_n \in \mathbb{N}$, $\sum_{i=1}^{n} d_i = n-1$, let

$T_n(d) = \{$rooted, labeled, ordered trees with degree sequence $d\}$.

Then $|T_n(d)| = (n-1)!$

Proof: Line-breaking construction

In example, $d=(d_1,\dots,d_8)=(0,3,1,0,2,0,0,1)$



Starting tree: Just the root

Rule: head from current tree to smallest-labeled leaf & record edges observed along branch.

First branch: $2_2\ 5_2$

Second branch: $5_1\ 3_1$

Third branch: 2_3

Fourth branch: $2_1\ 8_1$

Output: Permutation of tree edges

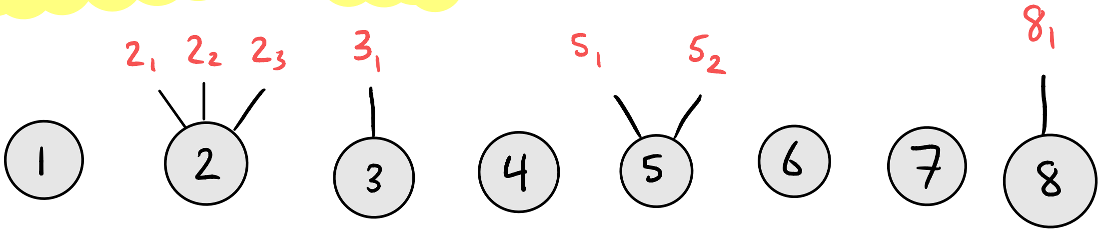
$\bigcup_{i=1}^{n} \{i_1,\dots,i_{d_i}\}$ $\quad 2_2\ 5_2\ 5_1\ 3_1\ 2_3\ 2_1\ 8_1$

Prop: Fix $d=(d_1,\dots,d_n)$ s.t. $d_1,\dots,d_n \in \mathbb{N}$, $\sum_{i=1}^{n} d_i = n-1$, let

$T_n(d) = \{$rooted, labeled, ordered trees with degree sequence $d\}$.

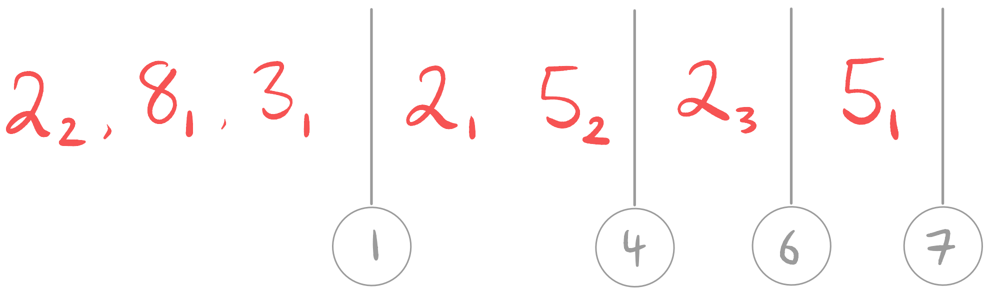
Then $|T_n(d)| = (n-1)!$

Inverse Process of line-breaking construction (diff. example)



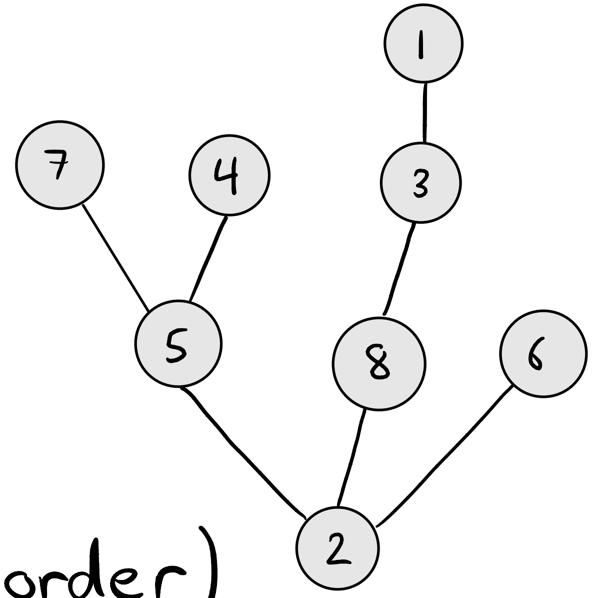
← $d = (0, 3, 1, 0, 2, 0, 0, 1)$

Fix a permutation of $\bigcup_{i=1}^{n} \{i_1, \dots, i_{d_i}\}$



This is a list of the edges of a tree

(rule: repetitions ⇝ edges to leaves, in ↑ order)



□

Corollary

\# rooted, labeled, <u>unordered</u> trees with degree sequence

$d = (d_1, \ldots, d_n)$ is $\binom{n-1}{d} := \dfrac{(n-1)!}{d_1! \cdots d_n!}$ ← Ignore ordering of children (subscripts)

Corollary

If $(n_i, i \geq 0)$ is s.t. $\sum_{i \geq 0} n_i = n$, $\sum_{i \geq 0} i n_i = n-1$, then

\# rooted, unlabeled, ordered trees with type $(n_i, i \geq 0)$ is

$$\frac{(n-1)!}{\prod_{i \geq 0} n_i!}$$

← Permuting names of vertices of same degree yields the same unlabeled tree

Corollary $|T_n| = n^{n-1}$

(Cayley's Formula)

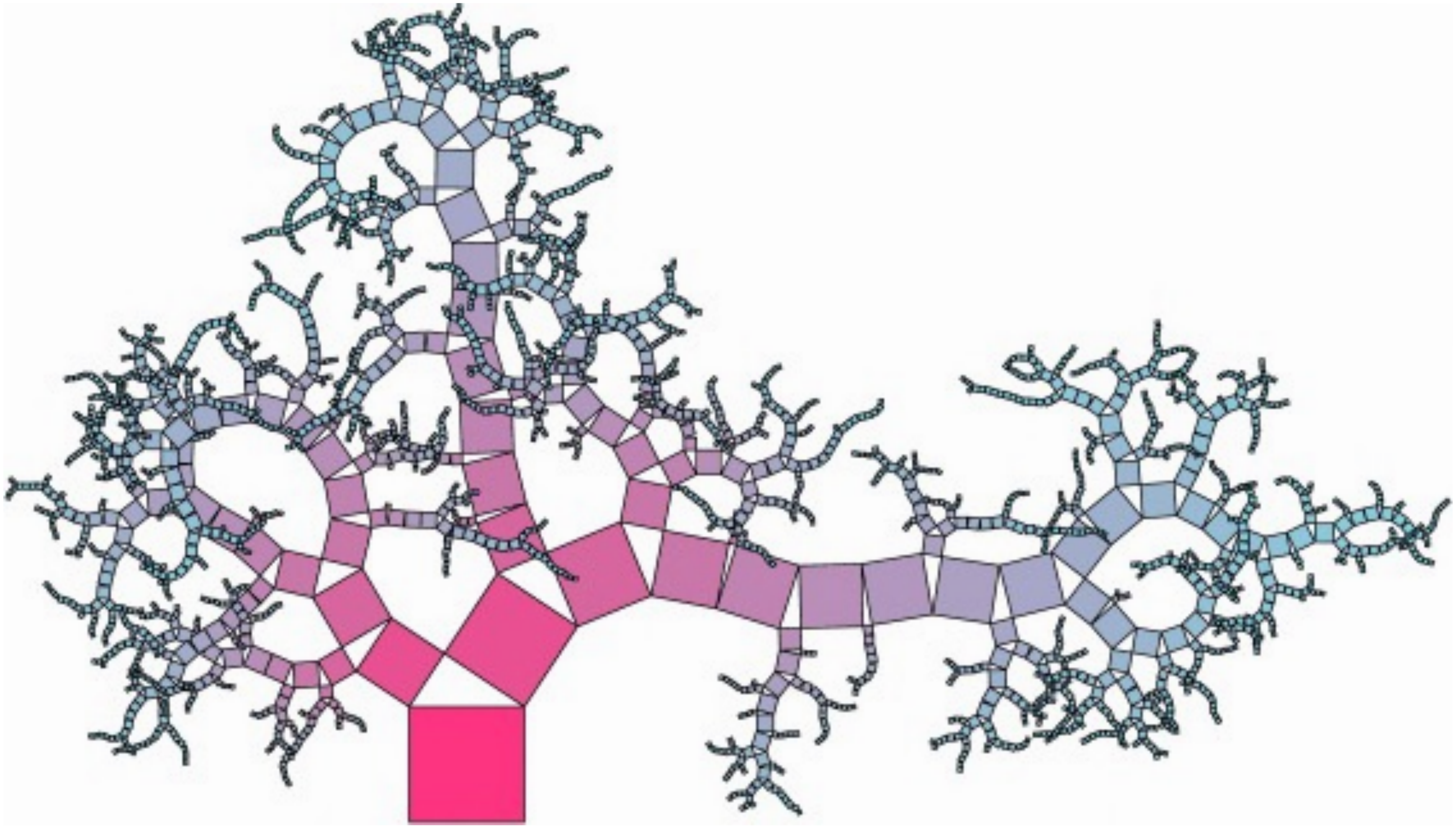
$T_n = \{$rooted trees t with $v(t) = [n]\}$

Proof: In the line-breaking construction, allow any integer from $[n]$ in each of the $n-1$ positions, and ignore subscripts. □

The combinatorics allow results to be easily transferred between models.

Example: Facts about Bienaymé trees can be derived from facts about labeled, ordered trees by "averaging over the type," since trees of the same type are equiprobable under the Bienaymé dist.s

Part Two: Results



The height of random trees: universal bounds

Theorem (A-B, Donderwinkel 2023+)

Let T be a random tree with type $(n_i, i \geq 0)$, write $n = \sum_{i \geq 0} n_i < \infty$. Set $\delta = 1 - \frac{n_1}{n}$.

Then $\mathbb{P}(ht(T) \geq x n^{1/2}) \leq 5 \exp\left(-\delta \frac{x^2}{9000}\right)$

Theorem (A-B, Donderwinkel 2023+)

$\forall \varepsilon > 0$, $\exists\, c, C > 0$, $n_0 > 0$ s.t. the following holds for all $n \geq n_0$.

Fix μ an offspring distribution with $\mu_0 + \mu_1 \leq 1 - \varepsilon$, $\frac{\mu_0}{\mu_0 + \mu_1} > \varepsilon$.

Let $T_{n,\mu}$ be Bienaymé(μ) conditioned to have n vertices.

Then $\forall x > 0$,

$$\mathbb{P}(ht(T_{\mu,n}) \geq x n^{1/2}) \leq C \exp(-c x^2)$$

The height of random trees: variance-dependant bounds

Theorem (A-B, Donderwinkel 2023+)

Let T be a random tree with type $(n_i, i \geq 0)$, write $n = \sum_{i \geq 0} n_i < \infty$.

Set $\sigma^2 = \sum i(i-1)\frac{n_i}{n}$, $(\sigma')^2 = \frac{n}{n-n_1}\sigma^2$. Then $\forall x > 2^{14}$,

$$\mathbb{P}\left(ht(T) \geq x n^{1/2} \frac{\log(\sigma'+1)}{\sigma}\right) \leq 4 \exp\left(-x \log(\sigma'+1)/2^{14}\right)$$

Theorem (A-B, Donderwinkel 2023+)

Let μ be an offspring dist. with $\sum d\mu(d) \leq 1$, $\sum d^2\mu(d) = \infty$,

let $T_{n,\mu}$ be Bienaymé(μ) conditioned to have n vertices.

Then $\mathbb{E}(ht(T_{n,\mu})/\sqrt{n}) \longrightarrow 0$.

Theorem (A-B, Donderwinkel 2023+)

Let μ be an offspring dist. with $\sum d\mu(d) < 1$ s.t.

$\sum e^{td}\mu(d) = \infty$ for all $t > 0$. Then $\mathbb{E}(ht(T_{n,\mu})/\sqrt{n}) \longrightarrow 0$.

The height of random trees: stochastic comparison.

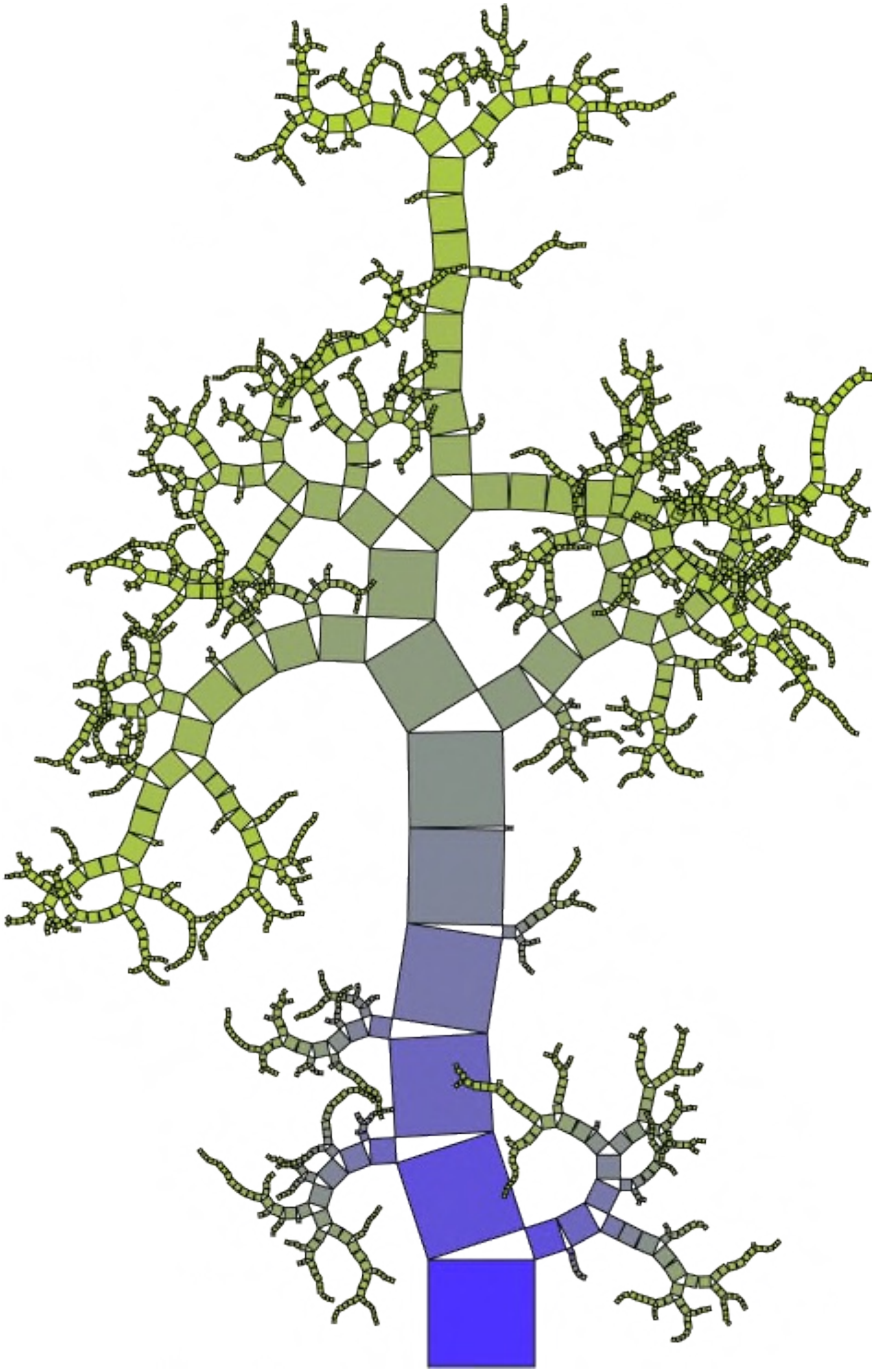
Theorem (A-B, Donderwinkel 2023+)

Let μ be any offspring dist. with $\mu(\{1\})=0$.

Let ν be the offspring dist. with $\nu(0)=\frac{1}{2}=\nu(2)$.

Then $\mathrm{ht}(T_{n,\mu}) \leqslant_{st} \mathrm{ht}(T_{n+1,\nu})$

Part Three: Proofs



Proof idea: Line-breaking

(*assume no deg 1 vertices for simplicity)

$(d_1,\dots,d_n) \rightsquigarrow$ deg. seq. with n_d entries $= d$.

$\tau = (\tau_1,\dots,\tau_{n-1})$

$\rightsquigarrow$ u. rand. perm. of $\bigcup_{i=1}^{n} \{i_1,\dots,i_{d_i}\}$

- Let $\ell(k) :=$ Label of τ_k

(if $\tau_k = i_j$ then $\ell(k) = i$)

- $L :=$ Length of first branch

$$= \min(t : \ell(t) \in \{\ell(1),\dots,\ell(t-1)\})$$

Theorem (A-B, Donderwinkel 2023+)

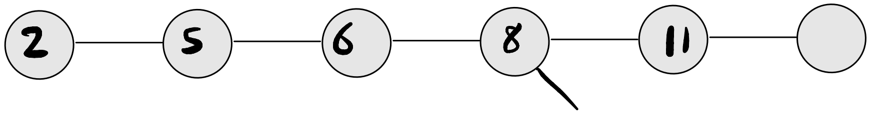
Let T be a random tree with type $(n_i, i \geq 0)$, write $n = \sum_{i \geq 0} n_i < \infty$. Set $\delta = 1 - \frac{n_1}{n}$.

Then $\mathbb{P}(ht(T) \geq x n^{1/2}) \leq 5 \exp(-\delta \frac{x^2}{9000})$

$2_1\ 5_3\ 6_1\ 8_2\ 11_5\ 8_1\ 4_4\ 12_1\ 3_2\ \dots$

$\downarrow$

2 5 6 8 11 8 4 12 3



$L = 6$

Prob. of ending a branch increases at least linearly in time :

$$\mathbb{P}(L=t \mid \ell(1),\dots,\ell(t-1))$$
$$= \frac{1}{n-t} \cdot \sum_{i\le t-1} (d_{\ell(i)}-1) \cdot \mathbb{1}_{\substack{\ell(1),\dots,\ell(t-1)\\ \text{distinct}}}$$

So, $\mathbb{P}(L=t \mid L>t-1)$

$$= \mathbb{P}\big(\ell(t)\in\{\ell(1),\dots,\ell(t-1)\} \mid \ell(1),\dots,\ell(t-1) \text{ distinct}\big)$$
$$\geqslant \frac{t-1}{n-t}$$

$$\Rightarrow \mathbb{P}(L>t) \leqslant \prod_{i=1}^{t-1}\left(1-\frac{i}{n-1-i}\right)$$
$$\leqslant \exp\left(-(t-1)^2/2(n-1)\right)$$

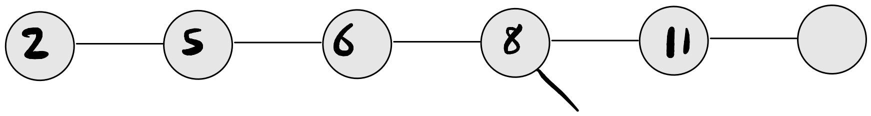
$$\mathbb{P}(L \geqslant x\sqrt{n-1}+1) \leqslant \exp\left(-\frac{x^2}{2}\right)$$

Sub-Gaussian tails for length of first branch.

$$2_1\ 5_3\ 6_1\ 8_2\ 11_5\ 8_1\ 4_4\ 12_1\ 3_2 \dots$$

$$\Big\downarrow$$

$$2\quad 5\quad 6\quad 8\quad 11\quad 8\quad 4\quad 12\quad 3$$



$L=6$

To extend from "first branch" to the height: chaining.

Branches attached between time $(2^i\sqrt{n}, 2^{i+1}\sqrt{n}]$ have typical length $\lesssim \sqrt{n}/2^i$.

We prove a maximal inequality to control lengths of all branches attached between time $(2^i\sqrt{n}, 2^{i+1}\sqrt{n}]$ simultaneously.

$$\rightsquigarrow \mathbb{P}\left(\text{height grows by} > x\sqrt{n}/2^{i/2} \text{ between time } (2^i\sqrt{n}, 2^{i+1}\sqrt{n}]\right) \leq \exp(-ix^2/2)$$

Then sum the bounds.

$\square$

Theorem (A-B, Donderwinkel 2023+)

Let T be a random tree with type $(n_i, i \geq 0)$, write $n = \sum_{i \geq 0} n_i < \infty$.

Set $\sigma^2 = n^{-1} \sum i(i-1) n_i$, $(\sigma')^2 = \frac{n}{n-n_1} \sigma^2$. Then $\forall x > 2^{14}$,

$$\mathbb{P}\left(ht(T) \geq x n^{1/2} \frac{\log(\sigma'+1)}{\sigma}\right) \leq 4 \exp\left(-x \log(\sigma'+1)/2^{14}\right)$$

Proof idea: line-breaking again. (Assume $n_1 = 0$ so $\sigma = \sigma'$)

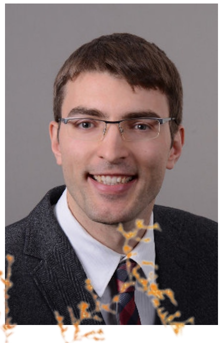
$$\mathbb{P}(L = t \mid \ell(1), \dots, \ell(t-1)) = \frac{1}{n-t} \cdot \sum_{i \leq t-1} \left(d_{\ell(i)} - 1\right) \cdot \mathbb{1}_{\substack{\ell(1), \dots, \ell(t-1) \\ \text{distinct}}}$$

- $\tau = (\tau_1, \dots, \tau_{n-1}) \rightsquigarrow$ u. rand. perm. of $\bigcup_{i=1}^{n} \{i_1, \dots, i_{d_i}\}$
- $\ell(k) :=$ Label of τ_k

★ $\sum_{i \leq t-1} \left(d_{\ell(i)} - 1\right)$ typically grows like $t\sigma$ so first branch length $L \asymp \sqrt{n}/\sigma$.

★ $\log(\sigma'+1) = \log(\sigma+1)$ needed in chaining when σ large. □

Images by Benedikt Stufler



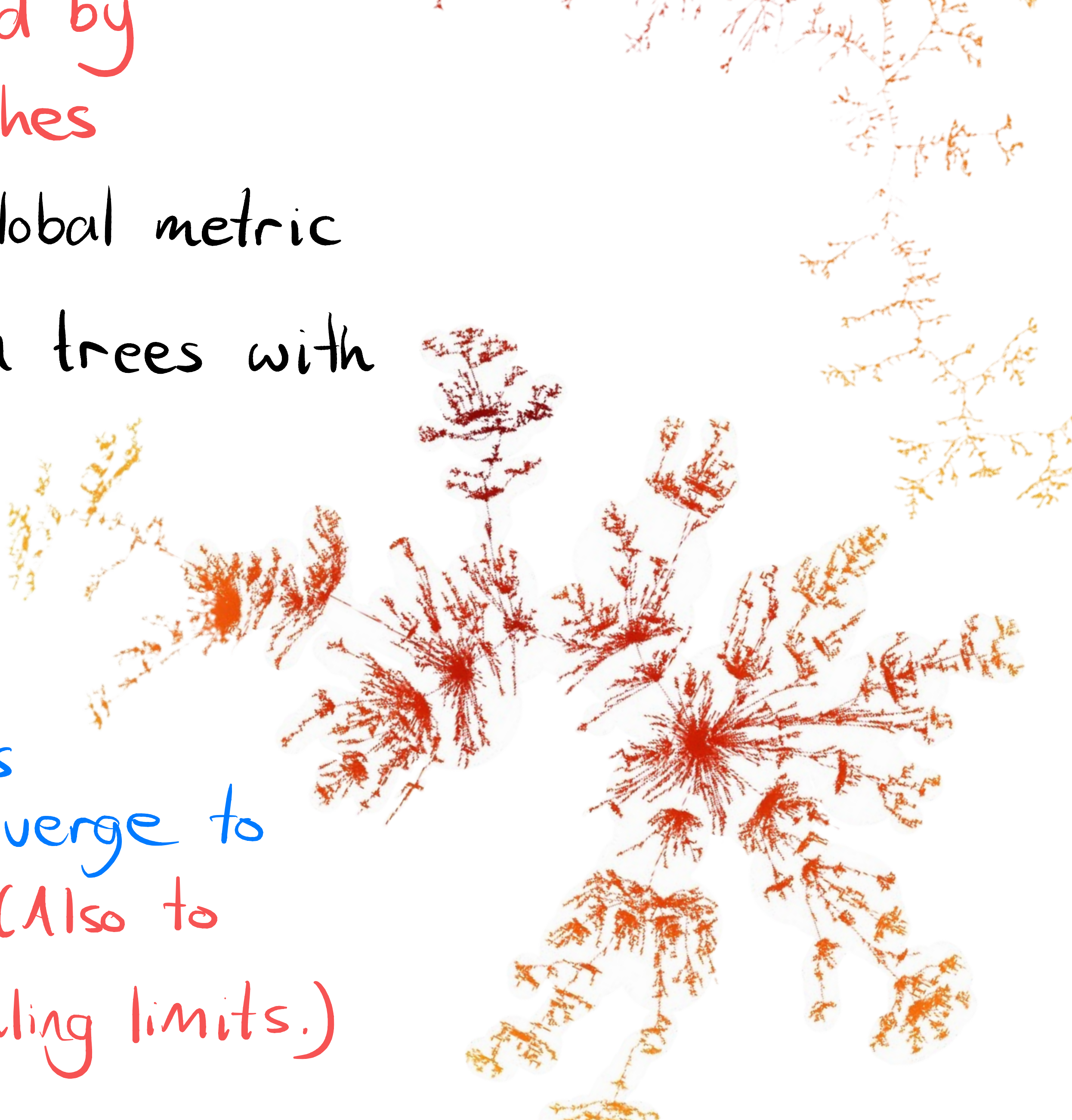
Aside:

Can also use line-breaking to study

$T(k) :=$ tree spanned by first k branches

Gives access to the global metric structure of random trees with given degrees.

Used by Arthur Blanc-Renaudie to prove that random trees with given degrees converge to Inhomogeneous CRTs. (Also to prove random graph scaling limits.)



Lower Bounds

Theorem (A-B, Donderwinkel, Kortchemski 2024+)
If μ is a critical offspring dist. then
$\mathrm{ht}(T_{\mu,n})/\log n \xrightarrow{\text{prob}} \infty$

Proof: In this case with high probability, in the line-breaking construction there will be a run of length $\gg \log n$ where only new nodes are seen
$\Rightarrow$ Branch of length $\gg \log n$ □

This result is optimal in that

a) $\exists$ subcritical μ s.t. $\mathbb{E}\,\mathrm{ht}(T_{\mu,n}) = O(\log n)$

b) If μ is supercritical then $\mathrm{ht}(T_{n,\mu}) = \Theta_{\mathbb{P}}(\sqrt{n})$

c) $\log n$ can not be replaced by $f(n)\log n$ with $f(n) \to \infty$. ☆

Open questions

- $\exists$? $a, A > 0$ s.t. the following holds.

  Fix critical branching dist. μ with variance $\sigma^2 \in (0,\infty)$

  Then $\forall x > 0$, $\forall n$, $\mathbb{P}(\mathrm{ht}(T_{n,\mu}) \geq x n^{1/2}) \leq A \exp(-a(\sigma x)^2)$.

  This bound would agree w/ the tail behaviour of $\frac{2}{\sigma} \cdot H$; recall that

  In our result have $\exp(-cx^2)$, $c = c(\mu_0, \mu_1)$.

$$\boxed{\mathrm{ht}(T_n)/n^{1/2} \xrightarrow{\text{dist}} \frac{2}{\sigma} H \text{ as } n \to \infty}$$

(Kolchin 1978)

- Fix $d = (d_1, \ldots, d_n)$ with $\sum d_i = 2m \geq 2(n-1)$, # neighbours

  let G_d be a u. random connected graph with degrees d.

  Then (?) $\mathbb{E}(\mathrm{diam}(G_d)) = O(n^{1/2})$

(Implicit constants should depend on # vertices of degree ≤ 2.)

References:

arXiv:2107.09726

arXiv:2201.11773

arXiv:2110.03378

THANKS!!