Tree heights and a new proof of Cayley's formula

Joint work with Serte Donderwinkel, Mikaël Maazoun, James Martin

Combinatorial tree

- Rooted labeled tree $t = (v(t), e(t))$

  $v(t) = [n] = \{1, 2, \ldots, n\}$; write $n = |t|$.

- For $v \in v(t)$ write $c(v) = c_t(v) = \#$ children of v in t,

  let $c_t = (c_t(v), v \in v(t)) = (c_t(1), \ldots, c_t(n))$.

For $c = (c(1), \ldots, c(n))$ non-negative integers, write

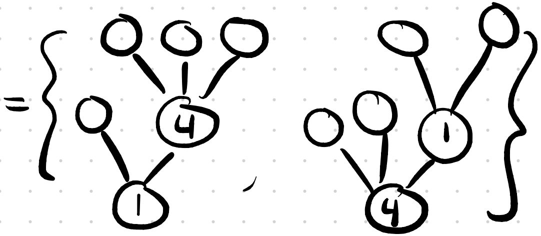
$$T_c = \{\text{combinatorial trees } t : |t| = n,\ c_t = c\}.$$

($|T_c| = 0$ unless $\sum_{u \in [n]} c(u) = n-1$.

Call such a seq. a child sequence.

e.g. $c = (c(1), c(2), c(3), c(4), c(5), c(6))$
$= (2, 0, 0, 3, 0, 0)$

$\mathcal{T}_c =$



4 labelings 6 labelings

Random combinatorial tree: A random tree T uniformly distributed over $\mathcal{T}_c$ for some c.

<u>Theorem</u> For all $\varepsilon > 0$ there are consts $a, A > 0$ s.t. the following holds.
Let T be a random combinatorial tree with $|T| = n \in \mathbb{N}$ and with at least εn leaves. Then for all $x > 1$,

$$\mathbb{P}(|T| > x\sqrt{n}) \le A \exp(-a x^2).$$

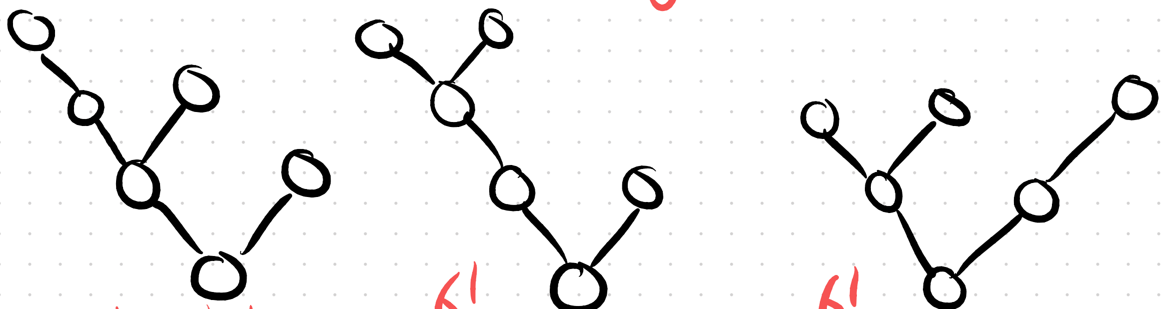
(In fact $a = \Theta(\varepsilon^2)$)

Path vertex: A vertex with exactly one child

Def For $\ell \geq 1$, $p \geq 0$, write $\mathcal{B}_{\ell,p}$ for the set of sub-binary combinatorial trees with ℓ leaves and p path vertices

(so $\ell - 1$ vertices with exactly 2 children)

$\mathcal{B}_{3,1}$



$6!$ labelings $\quad \frac{6!}{2}$ labelings $\quad \frac{6!}{2}$ labelings

Theorem: Let T be a random combinatorial tree with ℓ leaves, p path vertices. Let $B \in_u \mathcal{B}_{\ell,p}$. Then $\text{height}(T) \preccurlyeq_{st} \text{height}(B)$.

More strongly: take a child sequence

$$c = (c(1), c(2), \ldots, c(n))$$

Fix entries i, j with $c(i) > c(j) \geq 0$

and let

$$c = (c(1), \ldots, c(i), \ldots, c(j), \ldots, c(n))$$

$$\downarrow -1 \qquad \downarrow +1$$

$$c' = (c(1), \ldots, c(i)-1, \ldots, c(j)+1, \ldots, c(n))$$

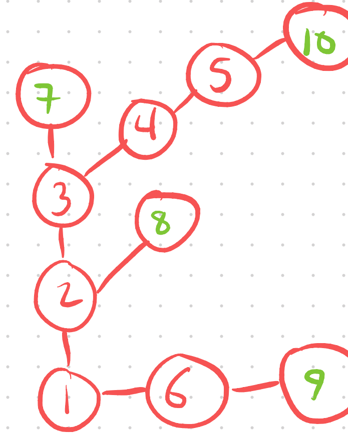
Theorem If $T \in_u \mathcal{T}_c$ and $T' \in_u \mathcal{T}_{c'}$

then $\text{height}(T) \preccurlyeq_{st} \text{height}(T')$

"Large degrees make short trees"

Proof technique: coding trees by sequences
(A-B, Donderwinkel, Maazoun, Martin)

Given tree t with vertices $v(t) = [n]$, root ρ, list leaves in ↑ order as $\ell_1, \dots, \ell_{r+1}$.



Let t_0 = one-vertex tree ρ.

For $1 \le i \le r+1$ let $p_i = (p_i(1), \dots, p_i(h_i))$

$p_i(1) \in t_{i-1}$, $p_i(h_i) = \ell_i$

be the branch from t_{i-1} to ℓ_i, and let

$$t_i = t_{i-1} + p_i$$

$p_1 = 1, 2, 3, 7$

$p_2 = 2, 8$

$p_3 = 3, 4, 5, 10$

$p_4 = 1, 6, 9$

Then t is coded by the sequence

$v(t) = 1, 2, 3, 2, 3, 4, 5, 1, 6$

$$v(t) = (p_1(1), \dots, p_1(h_1 - 1), p_2(1), \dots, p_2(h_2 - 1), \dots, p_{r+1}(1), \dots, p_{r+1}(h_{r+1} - 1))$$

Proof heuristic

Geometric decomposition of coding sequence

$v = (v_1 \, --- \, v_{n-1})$

Write $t(v;k)$ for tree built by first k elements of coding sequence

If $\text{height}(t(v)) \geq 2x\sqrt{n}$ then either $ht(t(v;x\sqrt{n}))$ is a path

or $ht(t(v;4^i x\sqrt{n})) - ht(t(v;4^{i-1}x\sqrt{n})) \geq \frac{x\sqrt{n}}{2^i}$

Birthday paradox if $v \in_u [n]^{n-1}$ then $\mathbb{P}(v_{k+1} \notin \{v_1, ---, v_k\}) = (1 - \frac{1}{n})^k = e^{-k/n}$

If $k = 4^i\sqrt{n}$ then $\mathbb{P}(v_{k+1}, --- v_{k+\ell} \text{ distinct}, \notin \{v_1, --- v_k\}) \leq (1 - \frac{\ell}{n})^k \approx e^{-k\ell/n}$

So expect a repetition when $\ell \approx \frac{\sqrt{n}}{4^i}$.

The method gives other results for conditioned Bienaymé trees (which are "averages" of random combinatorial trees):

Let B be an offspring dist, T_n be Bienaymé(B) conditioned to have n vertices.

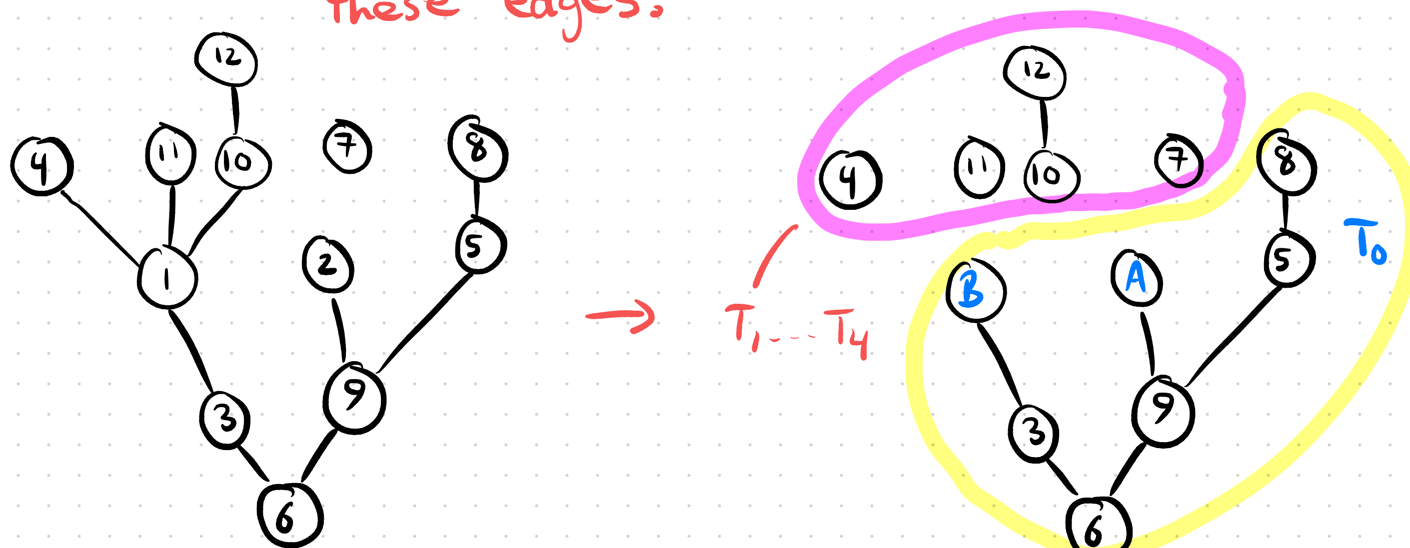
Theorem

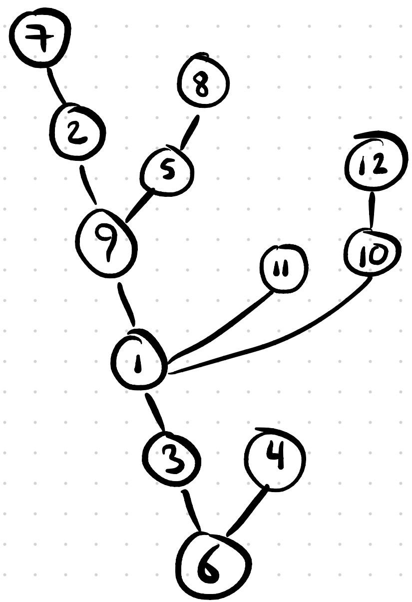
- If $\mathbb{E}B \leq 1$, $\mathbb{E}(B^2) = \infty$ then $\frac{ht(T_n)}{\sqrt{n}} \xrightarrow{P} 0$
- If $\mathbb{E}B < 1$, $\mathbb{E}[e^{tB}] = \infty$ $\forall t > 0$ then $\frac{ht(T_n)}{\sqrt{n}} \xrightarrow{P} 0$
- $\forall \varepsilon > 0$ $\exists a, A > 0$ s.t. if $\mathbb{P}(B \leq 1) \leq 1 - \varepsilon$ then for all $x > 1$, $\mathbb{P}\left(\frac{ht(T_n)}{\sqrt{n}} > x\right) \leq A e^{-ax^2}$

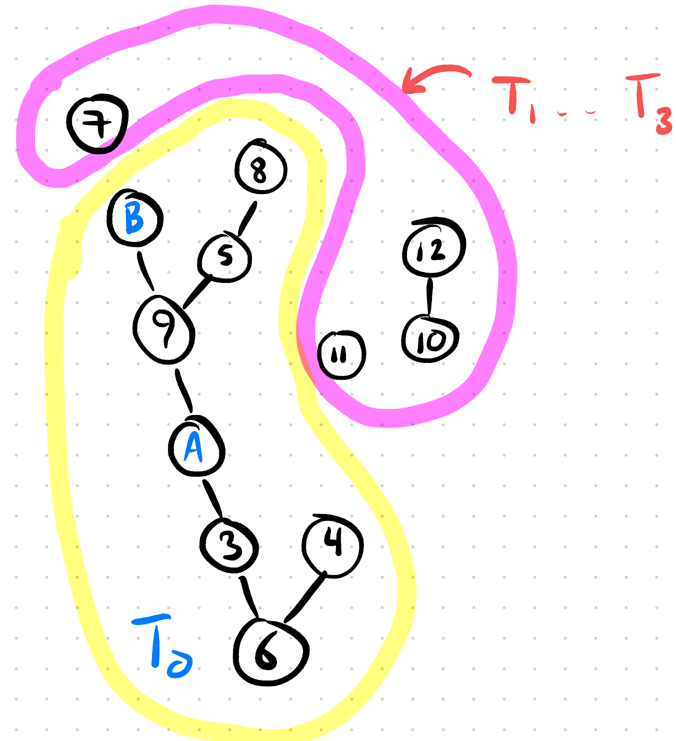
Proof technique (stochastic inequality)

Let T_0 be subtree of T obtained by removing edges from (i), (j) to children (excluding children which are ancestors of (i), (j)) with (i), (j) relabeled as (A), (B) unif. at random, containing the root.

Let $T_1, \ldots, T_m$ be the other subtrees created by removing these edges.







<u>NB</u> (I) If $A \not< B$ or $B \not< A$ then $m = c(i) + c(j) - 1$

In this case $\mathbb{P}(A \wedge B = i \mid \text{(I)}) = \frac{c(i)}{c(i)+c(j)}$

(II) Otherwise $m = c(i) + c(j)$. $\rightarrow$ In this case $\mathbb{P}(A = i \mid \text{(II)}) = \frac{1}{2}$.

From T_0 and $T_1, \ldots, T_m$,

To reconstruct T from $T_0, \ldots, T_m$ (in distribution)

- Randomly relabel A, B as ⓘ, ⓙ with the correct prob.
- If ⓘ ≼ ⓙ then attach a random size-$(c(i)-1)$ subset of $T_1, \ldots T_m$ to ⓘ, remaining trees to ⓙ
- If ⓙ ≼ ⓘ then attach a random size-$(c(j)-1)$ subset of $T_1, \ldots T_m$ to ⓘ, remaining trees to ⓘ
- Otherwise attach a random size-$c(i)$ subset of $T_1, \ldots T_m$ to ⓘ, remaining trees to ⓙ.

From T_0 and $T_1, \dots, T_m$,

To reconstruct T' from $T_0, \dots, T_m$ (in distribution)

- Randomly relabel A, B as ⓘ, ⓙ with the correct probab.
- If ⓘ ≼ ⓙ then attach a random size-$(c(i)-1)$ subset of $T_1, \dots T_m$ to ⓘ, remaining trees to ⓙ
- If ⓙ ≼ ⓘ then attach a random size-$(c(j)-1)$ subset of $T_1, \dots T_m$ to ⓘ, remaining trees to ⓘ
- Otherwise attach a random size-$c(i)$ subset of $T_1, \dots T_m$ to ⓘ, remaining trees to ⓙ.