Depth of high-degree vertices in Random recursive trees

Laura Eslava joint work with Louigi Addario-Berry

McGill University

AMS Fall Eastern Sectional Meeting 2016

Underlying idea





Random recursive trees/ Kingman's coalescent (Union-Find tree) **duality**.

 T_n is a rooted labelled tree on $V(T_n) = \{0, \ldots, n-1\}.$

- \triangleright Add a vertex labelled *n*,
- $\triangleright \text{ Select } j \in V(T_n) \\ \text{ uniformly at random,} \\$
- ▷ Add the edge $n \rightarrow j$.

 T_n is a rooted labelled tree on $V(T_n) = \{0, \ldots, n-1\}.$

- \triangleright Add a vertex labelled *n*,
- $\triangleright \text{ Select } j \in V(T_n) \\ \text{ uniformly at random,} \\$
- ▷ Add the edge $n \rightarrow j$.



 T_n is a rooted labelled tree on $V(T_n) = \{0, \ldots, n-1\}.$

- \triangleright Add a vertex labelled *n*,
- $\triangleright \text{ Select } j \in V(T_n) \\ \text{ uniformly at random,} \\$
- ▷ Add the edge $n \rightarrow j$.



 T_n is a rooted labelled tree on $V(T_n) = \{0, \ldots, n-1\}.$

- \triangleright Add a vertex labelled *n*,
- $\triangleright \text{ Select } j \in V(T_n) \\ \text{ uniformly at random,} \\$
- ▷ Add the edge $n \rightarrow j$.



 T_n is a rooted labelled tree on $V(T_n) = \{0, \ldots, n-1\}.$

- \triangleright Add a vertex labelled *n*,
- ▷ Select $j \in V(T_n)$ uniformly at random,
- ▷ Add the edge $n \rightarrow j$.



 T_n is a rooted labelled tree on $V(T_n) = \{0, \ldots, n-1\}.$

Given T_n , construct T_{n+1} :

- \triangleright Add a vertex labelled *n*,
- $\triangleright \text{ Select } j \in V(T_n) \\ \text{ uniformly at random,} \\$
- ▷ Add the edge $n \rightarrow j$.



Facts. - Vertex labels are increasing along root-to-leaf paths. - There are (n - 1)! possible outcomes for T_n .

Degree and depth

1 / / -

For
$$i \in V(T_n)$$
, let

$$deg_n(i) = \#\{j > i : j \to i \text{ in } T_n\},$$

$$ht_n(i) = dist(0, i)$$



Depth of final vertex

Theorem. (Devroye, 1988, Mahmoud 1991) As $n \to \infty$,

$$\frac{\operatorname{ht}_n(n-1)-\ln n}{\sqrt{\ln n}} \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}(0,1).$$

Proof's idea: Theory of records

$$v_0 = n - 1, v_{i+1} = \lfloor v_i U_i \rfloor,$$

$$ht_n(n-1) = \min\{i : v_i = 0\}$$
$$\stackrel{\ell}{=} \sum_{k=1}^{n-1} Ber(1/k)$$

The Bernoulli variables are independent.



Depth of final vertex

Theorem. (Devroye, 1988, Mahmoud 1991) As $n \to \infty$,

$$\frac{\operatorname{ht}_n(n-1)-\ln n}{\sqrt{\ln n}} \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}(0,1).$$

Proof's idea: Theory of records

$$v_0 = n - 1, v_{i+1} = \lfloor v_i U_i \rfloor,$$

$$ht_n(n-1) = \min\{i : v_i = 0\}$$
$$\stackrel{\mathcal{L}}{=} \sum_{k=1}^{n-1} Ber(1/k)$$

The Bernoulli variables are independent.



Depth of a uniformly chosen vertex

Theorem. (Devroye, 1988, Mahmoud 1991) As $n \to \infty$,

$$\frac{\operatorname{ht}_n(n-1)-\ln n}{\sqrt{\ln n}} \stackrel{\mathcal{L}}{\longrightarrow} N(0,1).$$

Corollary. Choose **uniformly** a vertex $\mathbf{u} \in T_n$. As $n \to \infty$,

$$\frac{\operatorname{ht}_n(\mathbf{u}) - \ln n}{\sqrt{\ln n}} \xrightarrow{\mathcal{L}} N(0,1)$$

What can we say about high-degree vertices?

What can we say about high-degree vertices?

Theorem (Devroye, Lu 1995) Let $\Delta_n = \max\{\deg_n(i) : i \in V(T_n)\}$. As $n \to \infty$, a.s.

$$\frac{\Delta_n}{\log_2 n} \to 1.$$

Asymptotic High-degree distribution

$$X_d^{(n)} = \#\{i \in [n] : \deg_n(i) = \lfloor \log_2 n \rfloor + d\}$$

Theorem (Addario-Berry,E. 2015⁺) Let $n = 2^k$. Jointly for all $d \in \mathbb{Z}$, as $k \to \infty$ $X_d^{(n)} \xrightarrow{\mathcal{L}} W_d$;

where the W_d are independent Poisson r.v.'s with mean $2^{-(d+1)}$.

$$\cdots \qquad \underbrace{\bullet}_{X_{-3}} \qquad \underbrace{\bullet}_{X_{-2}} \qquad \underbrace{\bullet}_{X_{-1}} \qquad \underbrace{\bullet}_{X_{0}} \qquad \underbrace{\bullet}_{X_{1}} \qquad \cdots \qquad$$

A marked point process on $\mathbb{Z}\times\mathbb{R}$

Let $2\alpha = \log_2 e$, then $2\alpha \ln n = \log_2 n$.

$$\bullet = \left(\deg_n(v) - \lfloor \log_2 n \rfloor, \frac{\operatorname{ht}_n(v) - (1 - \alpha) \ln n}{\sqrt{(1 - \alpha - \alpha^2/2) \ln n}} \right)$$
$$\mathcal{DH}^{(n)} : \bullet \bullet \underbrace{}_{-2} / \underbrace{\bullet \bullet}_{-1} / \underbrace{\bullet}_{0} / \underbrace{\bullet}_{1} / \underbrace{\bullet}_{2} / \bullet \bullet \bullet$$

A marked point process on $\mathbb{Z}\times\mathbb{R}$

Theorem (E. 2015⁺) Let $n = 2^k$. There is an explicit marked point process \mathcal{MP} , such that in the space of marked point processes on $\mathbb{Z} \times \mathbb{R}$, as $k \to \infty$

$$\mathcal{DH}^{(n)} \xrightarrow{\mathcal{L}} \mathcal{MP};$$

the marks are distributed as independent standard gaussian variables.

$$\bullet = \left(\deg_n(v) - \lfloor \log_2 n \rfloor, \frac{\operatorname{ht}_n(v) - (1 - \alpha) \ln n}{\sqrt{(1 - \alpha - \alpha^2/2) \ln n}} \right)$$
$$\mathcal{DH}^{(n)} : \cdots \underbrace{}_{-2} / \underbrace{}_{-1} / \underbrace{}_{0} / \underbrace{}_{1} / \underbrace{}_{2} / \cdots$$

Conditional depth of a high-degree vertex

Proposition (E. 2015⁺) Choose uniformly a vertex $\mathbf{u} \in V(T_n)$. Fix $m \in \mathbb{Z}$, conditional on deg_n $\mathbf{u} = \lfloor \log_2 n \rfloor + m$, as $n \to \infty$

$$\frac{\operatorname{ht}_{n}(\mathbf{u})-(1-\alpha)\ln n}{\sqrt{(1-\alpha+\alpha^{2}/2)\ln n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1);$$

where $2\alpha = \log_2 e$.

Fix $n \in \mathbb{N}$, for each $1 \le t \le n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

$$F_t = \{T_1^{(t)}, \dots, T_{n-t+1}^{(t)}\}$$

Given F_t , construct F_{t+1} :

- ▷ Uniformly choose two trees in F_t,
- Add an edge labelled t between the roots:

directed to either tree with equal probability.



Fix $n \in \mathbb{N}$, for each $1 \le t \le n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

$$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$$

Given F_t , construct F_{t+1} :

- ▷ Uniformly choose two trees in F_t,
- Add an edge labelled t between the roots:

directed to either tree with equal probability.



Fix $n \in \mathbb{N}$, for each $1 \le t \le n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

$$F_t = \{T_1^{(t)}, \dots, T_{n-t+1}^{(t)}\}$$

Given F_t , construct F_{t+1} :

- ▷ Uniformly choose two trees in F_t,
- Add an edge labelled t between the roots:

directed to either tree with equal probability.



Fix $n \in \mathbb{N}$, for each $1 \le t \le n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

$$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$$

Given F_t , construct F_{t+1} :

- ▷ Uniformly choose two trees in F_t,
- Add an edge labelled t between the roots:

directed to either tree with equal probability.



Fix $n \in \mathbb{N}$, for each $1 \le t \le n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

$$F_t = \{T_1^{(t)}, \dots, T_{n-t+1}^{(t)}\}$$

Given F_t , construct F_{t+1} :

- Uniformly choose two trees in F_t,
- Add an edge labelled t between the roots:

directed to either tree with equal probability.



Fix $n \in \mathbb{N}$, for each $1 \le t \le n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

$$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$$

Given F_t , construct F_{t+1} :

- ▷ Uniformly choose two trees in F_t,
- Add an edge labelled t between the roots:

directed to either tree with equal probability.



Fix $n \in \mathbb{N}$, for each $1 \le t \le n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

$$F_t = \{T_1^{(t)}, \dots, T_{n-t+1}^{(t)}\}$$

Given F_t , construct F_{t+1} :

- ▷ Uniformly choose two trees in F_t,
- Add an edge labelled t between the roots:

directed to either tree with equal probability.



Fix $n \in \mathbb{N}$, for each $1 \le t \le n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

$$F_t = \{T_1^{(t)}, \dots, T_{n-t+1}^{(t)}\}$$

Given F_t , construct F_{t+1} :

- ▷ Uniformly choose two trees in F_t,
- Add an edge labelled t between the roots:

directed to either tree with equal probability.



Fix $n \in \mathbb{N}$, for each $1 \le t \le n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

$$F_t = \{T_1^{(t)}, \dots, T_{n-t+1}^{(t)}\}$$

Given F_t , construct F_{t+1} :

- ▷ Uniformly choose two trees in F_t,
- Add an edge labelled t between the roots:

directed to either tree with equal probability.



Fix $n \in \mathbb{N}$, for each $1 \le t \le n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

$$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$$

Given F_t , construct F_{t+1} :

- ▷ Uniformly choose two trees in F_t,
- Add an edge labelled t between the roots:

directed to either tree with equal probability.



Fix $n \in \mathbb{N}$, for each $1 \le t \le n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

$$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$$

Given F_t , construct F_{t+1} :

- Uniformly choose two trees in F_t,
- Add an edge labelled t between the roots:

directed to either tree with equal probability.

All choices are independent.



 F_n

Recursive trees: Via Kingman's Coalescent

Lemma. There is a mapping ϕ such that $\phi(F_n) \stackrel{\mathcal{L}}{=} T_n$; furthermore, ϕ preserves the shape of F_n .



Proof's idea.

- Vertex labels are exchangeable.
- Edge labels are decreasing along root-to-leaf paths.
- There are n!(n-1)! possible outcomes for F_n .

 $S = S^{(n)} = \{t \le n - 1 : \text{Tree containing } 1 \text{ merges at time } t\}$

(1)

- One tree's root increases its degree and
- vertices in the other tree increase their depth by 1.
- Vertex 1 starts as root.



 $S = S^{(n)} = \{t \le n - 1 : \text{Tree containing } 1 \text{ merges at time } t\}$

- One tree's root increases its degree and
- vertices in the other tree increase their depth by 1.
- Vertex 1 starts as root.



 $S = S^{(n)} = \{t \le n - 1 : \text{Tree containing } 1 \text{ merges at time } t\}$

- One tree's root increases its degree and
- vertices in the other tree increase their depth by 1.
- Vertex 1 starts as root.



 $S = S^{(n)} = \{t \le n - 1 : \text{ Tree containing 1 merges at time } t\}$ A favourable merge for 1 is when its tree's root increases its degree.

- One tree's root increases its degree and
- vertices in the other tree increase their depth by 1.
- Vertex 1 starts as root.



 $S = S^{(n)} = \{t \le n - 1 : \text{ Tree containing 1 merges at time } t\}$ A favourable merge for 1 is when its tree's root increases its degree.

- One tree's root increases its degree and
- vertices in the other tree increase their depth by 1.
- Vertex 1 starts as root.



 $S = S^{(n)} = \{t \le n - 1 : \text{ Tree containing 1 merges at time } t\}$ A favourable merge for 1 is when its tree's root increases its degree.

- One tree's root increases its degree and
- vertices in the other tree increase their depth by 1.
- Vertex 1 starts as root.



 $S = S^{(n)} = \{t \le n - 1 : \text{ Tree containing 1 merges at time } t\}$ A favourable merge for 1 is when its tree's root increases its degree.

- One tree's root increases its degree and
- vertices in the other tree increase their depth by 1.
- Vertex 1 starts as root.



 $S = S^{(n)} = \{t \le n - 1 : \text{ Tree containing 1 merges at time } t\}$ A favourable merge for 1 is when its tree's root increases its degree.

- One tree's root increases its degree and
- vertices in the other tree increase their depth by 1.
- Vertex 1 starts as root.



 $S = S^{(n)} = \{t \le n - 1 : \text{ Tree containing 1 merges at time } t\}$ A favourable merge for 1 is when its tree's root increases its degree.

- One tree's root increases its degree and
- vertices in the other tree increase their depth by 1.
- Vertex 1 starts as root.



 $S = S^{(n)} = \{t \le n - 1 : \text{ Tree containing 1 merges at time } t\}$ A favourable merge for 1 is when its tree's root increases its degree.

- One tree's root increases its degree and
- vertices in the other tree increase their depth by 1.
- Vertex 1 starts as root.



 $S = S^{(n)} = \{t \le n - 1 : \text{ Tree containing 1 merges at time } t\}$

A favourable merge for 1 is when its tree's root increases its degree.

- One tree's root increases its degree and
- vertices in the other tree increase their depth by 1.
- Vertex 1 starts as root.



 $S = S^{(n)} = \{t \le n - 1 : \text{Tree containing } 1 \text{ merges at time } t\}$

A favourable merge for 1 is when its tree's root increases its degree.

- One tree's root increases its degree and
- vertices in the other tree increase their depth by 1.
- Vertex 1 starts as root.



 $S = S^{(n)} = \{t \le n - 1 : \text{Tree containing 1 merges at time } t\}$

A favourable merge for 1 is when its tree's root increases its degree.

- One tree's root increases its degree and
- vertices in the other tree increase their depth by 1.
- Vertex 1 starts as root.



Degree and depth of vertex 1 in F_n

 $S = S^{(n)} = \{t \le n-1 : \text{Tree containing } 1 \text{ merges at time } t\}$

A favourable merge for 1 is when its tree's root increases its degree.

Proposition.

Depth = Total # unfavourable merges.

 $\operatorname{ht}_{F_n}(1) \stackrel{\mathcal{L}}{=} \operatorname{Bin}(|\mathcal{S}|, 1/2).$

Degree = First streak favourable merges.

 $\deg_{F_n}(1) \stackrel{\mathcal{L}}{=} \min\{Geo(1/2), |\mathcal{S}|\}.$



Asymptotic normality of $|\mathcal{S}^{(n)}|$ and $\operatorname{ht}_{F_n}(1)$

Theorem.

$$\frac{|\mathcal{S}^{(n)}| - 2 \ln n}{\sqrt{2 \ln n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$
$$\frac{\operatorname{ht}_{F_n}(1) - \ln n}{\sqrt{\ln n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Proof. Lindeberg condititions for CLT are satisfied:

$$|\mathcal{S}^{(n)}| = \sum_{k=2}^n \operatorname{Ber}(2/k) \qquad \quad \operatorname{ht}_{F_n}(1) = \sum_{k=2}^n \operatorname{Ber}(1/k).$$

A heuristic: Normal approximation of $|\mathcal{S}|$ and a Binomial r.v.

 $\operatorname{ht}_{F_n}(1) \stackrel{\mathcal{L}}{=} \operatorname{Bin}(|\mathcal{S}|, 1/2)$

A heuristic: Normal approximation of $|\mathcal{S}|$ and a Binomial r.v.

$$egin{aligned} &\operatorname{ht}_{\mathcal{F}_n}(1) \stackrel{\mathcal{L}}{=} \operatorname{Bin}(|\mathcal{S}|, 1/2) \ &pprox rac{|\mathcal{S}|}{2} + rac{B_{|\mathcal{S}|}}{2} \sqrt{|\mathcal{S}|} \end{aligned}$$

$$B_m \xrightarrow{\mathcal{L}} N_1 \stackrel{\mathcal{L}}{=} N(0,1)$$

A heuristic: Normal approximation of $|\mathcal{S}|$ and a Binomial r.v.

$$\begin{aligned} \operatorname{ht}_{F_n}(1) \stackrel{\mathcal{L}}{=} \operatorname{Bin}(|\mathcal{S}|, 1/2) \\ &\approx \frac{|\mathcal{S}|}{2} + \frac{B_{|\mathcal{S}|}}{2}\sqrt{|\mathcal{S}|} \\ &\approx \frac{2\ln n + S_n\sqrt{2\ln n}}{2} + \frac{B_{2\ln n}}{2}\sqrt{2\ln n} \end{aligned} \qquad \begin{array}{l} B_m \stackrel{\mathcal{L}}{\longrightarrow} N_1 \stackrel{\mathcal{L}}{=} N(0, 1) \\ &S_n \stackrel{\mathcal{L}}{\longrightarrow} N_2 \stackrel{\mathcal{L}}{=} N(0, 1) \end{aligned}$$

A heuristic: Normal approximation of $|\mathcal{S}|$ and a Binomial r.v.

$$\begin{aligned} & \operatorname{ht}_{F_n}(1) \stackrel{\mathcal{L}}{=} \operatorname{Bin}(|\mathcal{S}|, 1/2) \\ & \approx \frac{|\mathcal{S}|}{2} + \frac{B_{|\mathcal{S}|}}{2} \sqrt{|\mathcal{S}|} \\ & \approx \frac{2 \ln n + S_n \sqrt{2 \ln n}}{2} + \frac{B_{2 \ln n}}{2} \sqrt{2 \ln n} \\ & = \ln n + \frac{S_n + B_{2 \ln n}}{\sqrt{2}} \sqrt{\ln n}. \end{aligned} \qquad \begin{aligned} & B_m \stackrel{\mathcal{L}}{\longrightarrow} N_1 \stackrel{\mathcal{L}}{=} N(0, 1) \\ & S_n \stackrel{\mathcal{L}}{\longrightarrow} N_2 \stackrel{\mathcal{L}}{=} N(0, 1) \\ & \frac{N_1 + N_2}{\sqrt{2}} \stackrel{\mathcal{L}}{=} N(0, 1). \end{aligned}$$

Conditional asymptotic normality of $ht_{F_n}(1)$

Proposition. Fix $m \in \mathbb{Z}$, conditional on deg_{*F_n*}(1) = $\lfloor \log_2 n \rfloor + m$,

$$\frac{\operatorname{ht}_{F_n}(1) - (1 - \alpha) \ln n}{\sqrt{(1 - \alpha + \alpha^2/2) \ln n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

where $2\alpha = \log_2 e$.

Conditional asymptotic normality of $ht_{F_n}(1)$

Proposition. Fix $m \in \mathbb{Z}$, conditional on $\deg_{F_n}(1) = \lfloor \log_2 n \rfloor + m$,

$$\frac{\operatorname{ht}_{F_n}(1) - (1 - \alpha) \ln n}{\sqrt{(1 - \alpha + \alpha^2/2) \ln n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

where $2\alpha = \log_2 e$.

Proof's sketch: Given that $\deg_{F_n}(1) = \beta \ln n$, $\beta < 2$, then

$$|\mathcal{S}| - \deg_{F_n}(1) = (2 - \beta) \ln n + S_n \sqrt{2 \ln n} \qquad S_n \stackrel{\mathcal{L}}{\longrightarrow} N(0, 1).$$

Previous heuristic holds, now for $Bin(|\mathcal{S}| - \deg_{F_n}(1), 1/2)$.

Extending the study to many vertices

$$\bullet = \left(\deg_n(v) - \lfloor \log_2 n \rfloor, \frac{\operatorname{ht}_n(v) - (1-\alpha) \ln n}{\sqrt{(1-\alpha-\alpha^2/2) \ln n}} \right)$$
$$\mathcal{DH}^{(n)} : \cdots \underbrace{\bigvee}_{-2} / \underbrace{\bigvee}_{-1} / \underbrace{\bigvee}_{0} / \underbrace{\bigvee}_{1} / \underbrace{\bigvee}_{2} / \cdots$$

Tools.

- Method of moments.
- Exchangeability of vertices in F_n ,
- weak correlation between selection sets of distinct vertices.

Thanks!



Imagen from scalefreenetworks, Flickr