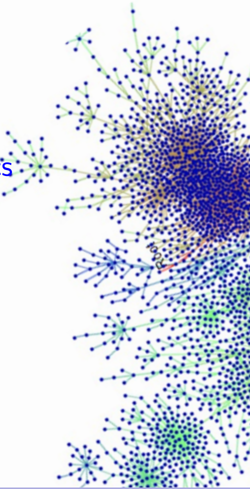
High degree vertices of (weighted) random recursive trees

Laura Eslava

IIMAS-UNAM

Analytic and Probabilistic Combinatorics
November 2022

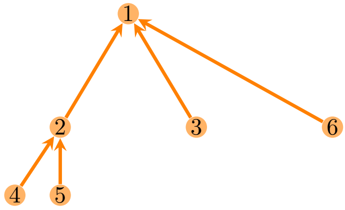


Joint work with: Louigi Addario-Berry; Bas Lodewijks and Marcel Ortgiese

Contents

Notation for rooted labelled trees: T

- Root / Leaves
- Children / Degree $\deg_T(\cdot)$
- Depth $\mathrm{ht}_T(\cdot)$ / Height
- Edges directed towards root
- Vertices are labeled with $[n] = \{1, \ldots, n\}$



$\deg_T(6) = 0, \mathrm{ht}_T(6) = 1$

Weighted random recursive trees

Tree growth process $(T_n,\ n \in \mathbb{N})$:

- ▷ Weights $(W_n)_{n\geq 1}$ i.i.d.
- ▷ T_1 is a single-vertex tree.
- ▷ For $n > 1$, build T_n from T_{n-1} adding: $\begin{cases} \text{vertex } n, \\ \text{edge } n \to j \end{cases}$

$$\mathbb{P}(n \to j \,|\, T_{n-1}) = \frac{W_j}{S_{n-1}}, \qquad S_{n-1} = \sum_{i=1}^{n-1} W_i.$$

Weighted random recursive trees

Tree growth process $(T_n,\ n \in \mathbb{N})$:

- ▷ Weights $(W_n)_{n\geq 1}$ i.i.d.
- ▷ T_1 is a single-vertex tree.
- ▷ For $n > 1$, build T_n from T_{n-1} adding: $\begin{cases} \text{vertex } n, \\ \text{edge } n \to j \end{cases}$

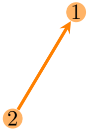
$$\mathbb{P}(n \to j \,|\, T_{n-1}) = \frac{W_j}{S_{n-1}}, \qquad S_{n-1} = \sum_{i=1}^{n-1} W_i.$$

Inhomogeneous probabilities,
independent from the current tree structure.

Random Recursive Trees (RRTs): $W_i = 1$ a.s.

At any step, a new vertex n attaches to a uniformly chosen vertex.

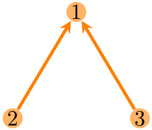
$$\mathbb{P}(n \to j \,|\, T_{n-1}) = \frac{1}{n-1}$$



Random Recursive Trees (RRTs): $W_i = 1$ a.s.

At any step, a new vertex n attaches to a uniformly chosen vertex.

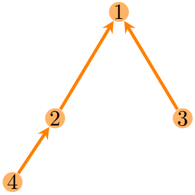
$$\mathbb{P}(n \to j | T_{n-1}) = \frac{1}{n-1}$$



Random Recursive Trees (RRTs): $W_i = 1$ a.s.

At any step, a new vertex n attaches to a uniformly chosen vertex.

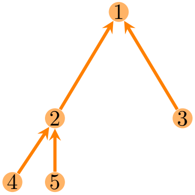
$$\mathbb{P}(n \to j | T_{n-1}) = \frac{1}{n-1}$$



Random Recursive Trees (RRTs): $W_i = 1$ a.s.

At any step, a new vertex n attaches to a uniformly chosen vertex.

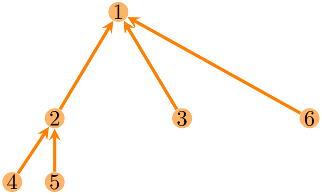
$$\mathbb{P}(n \to j | T_{n-1}) = \frac{1}{n-1}$$



Random Recursive Trees (RRTs): $W_i = 1$ a.s.

At any step, a new vertex n attaches to a uniformly chosen vertex.

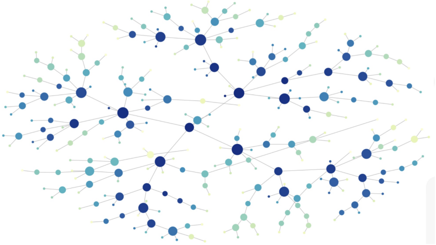
$$\mathbb{P}(n \to j | T_{n-1}) = \frac{1}{n-1}$$



Random Recursive Trees (RRTs): $W_i = 1$ a.s.

At any step, a new vertex n attaches to a uniformly chosen vertex.

$$\mathbb{P}(n \to j | T_{n-1}) = \frac{1}{n-1}$$



New edge-connection uniform, independent from evolution of the process.

Renewal theory for insertion depth

Theorem. (Devroye, 1988, Mahmoud 1991) For RRTs, as $n \to \infty$,

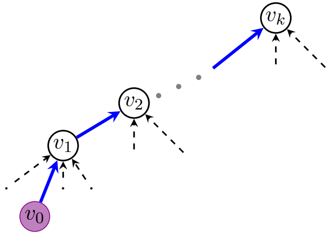
$$\frac{\mathrm{ht}_{T_n}(n) - \ln n}{\sqrt{\ln n}} \xrightarrow{\mathcal{L}} N(0,1).$$

Idea: For $(U_i)_{i \geq 0}$ i.i.d. $\mathrm{Unif}(0,1)$

$$v_0 = n, v_{i+1} = \lceil (v_i - 1) U_i \rceil,$$

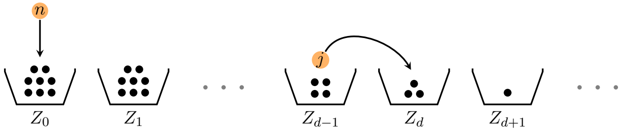
$$\mathrm{ht}_{T_n}(n) = \min\{k : v_k = 1\}$$

$$v_k \approx n e^{\sum_{i<k} \ln(U_i)}$$



Difference equations for degree sequence

Degree sequence: $Z_d^{(n)} = \#\{i \in [n] : \deg_{T_n}(i) = d\}$.



Theorem. [Na, Rapoport 1970] For RRTs, for each $d \geq 0$, as $n \to \infty$

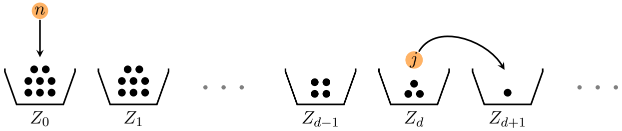
$$n^{-1}\mathbb{E}[Z_d^{(n)}] \to 2^{-(d+1)}.$$

Idea. For $d > 1$,

$$Z_d^{(n)} = Z_d^{(n-1)} + \mathbf{1}_{[\deg_{T_{n-1}}(j)=d-1]} - \mathbf{1}_{[\deg_{T_{n-1}}(j)=d]}.$$

Difference equations for degree sequence

Degree sequence: $Z_d^{(n)} = \#\{i \in [n] : \deg_{T_n}(i) = d\}$.



Theorem. [Na, Rapoport 1970] For RRTs, for each $d \geq 0$, as $n \to \infty$

$$n^{-1}\mathbb{E}[Z_d^{(n)}] \to 2^{-(d+1)}.$$

Idea. For $d > 1$,

$$Z_d^{(n)} = Z_d^{(n-1)} + \mathbf{1}_{[\deg_{T_{n-1}}(j)=d-1]} - \mathbf{1}_{[\deg_{T_{n-1}}(j)=d]}.$$

Pólya urn theory for degree sequence

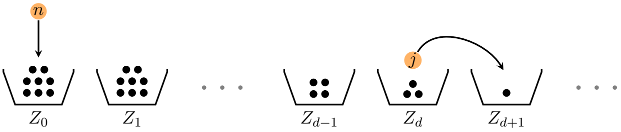
Theorem [Janson 2005] Jointly for all $d \geq 0$, as $n \to \infty$

$$n^{-1/2}(Z_d^{(n)} - 2^{-(d+1)}n) \xrightarrow{\text{dist}} V_d;$$

where V_d are gaussian r.v. with explicit covariance matrix.

Idea.

▷ Vertex color is given by its degree.



Pólya urn theory for degree sequence

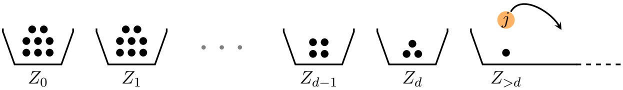
Theorem [Janson 2005] Jointly for all $d \geq 0$, as $n \to \infty$

$$n^{-1/2}(Z_d^{(n)} - 2^{-(d+1)}n) \xrightarrow{\text{dist}} V_d;$$

where V_d are gaussian r.v. with explicit covariance matrix.

Idea.

- Vertex color is given by its degree.
- Requires finite number of colors.



Degree and depth of a random vertex

Select a uniformly random vertex u in T_n,

- Classical methods:

$$\frac{\mathrm{ht}_{T_n}(u) - \ln n}{\sqrt{\ln n}} \approx N(0,1)$$

$$\mathbb{P}(\deg_{T_n}(u) = k | T_n) = \frac{Z_k^{(n)}}{n} \approx 2^{-(k+1)}$$

Degree and depth of a random vertex

Select a uniformly random vertex u in T_n,

- Classical methods:

$$\frac{\mathrm{ht}_{T_n}(u) - \ln n}{\sqrt{\ln n}} \approx N(0,1)$$

$$\mathbb{P}(\deg_{T_n}(u) = k | T_n) = \frac{Z_k^{(n)}}{n} \approx 2^{-(k+1)}$$

- Kingman's coalescent (spoiler):

$$\mathrm{ht}_{T_n}(u) \stackrel{\mathcal{L}}{=} \mathrm{Bin}(|\mathcal{S}|, 1/2)$$

$$\deg_{T_n}(u) \stackrel{\mathcal{L}}{=} \min\{\mathrm{Geo}(1/2), |\mathcal{S}|\}$$

$|\mathcal{S}| \stackrel{\mathcal{L}}{=} \sum_{i=2}^{n} \mathrm{Ber}(2/i)$, which is concentrated around $2 \ln n$.

Maximum degree

$$\Delta_n = \max\{\deg_{T_n}(i) : i \in [n]\}$$

Theorem [Devroye, Lu 1995] If T_n is a recursive tree. As $n \to \infty$, a.s.

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

Heuristic:

- ▷ Classical methods:

$$\mathbb{E}[\#\{i \in [n] : \deg_{T_n}(i) = d\}] \approx 2^{-(d+1)} n \approx 1 \qquad \text{if } d = \log_2 n.$$

Maximum degree

$$\Delta_n = \max\{\deg_{T_n}(i) : i \in [n]\}$$

Theorem [Devroye, Lu 1995] If T_n is a recursive tree. As $n \to \infty$, a.s.

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

Heuristic:

▷ Classical methods:

$$\mathbb{E}[\#\{i \in [n] : \deg_{T_n}(i) = d\}] \approx 2^{-(d+1)} n \approx 1 \qquad \text{if } d = \log_2 n.$$

▷ Kingman's coalescent (spoiler):

$$\mathbb{E}[\#\{i \in [n] : \deg_{T_n}(i) \geq d\}] = 2^{-d} n(1 + o(1)) \quad \text{for } d < 2 \ln n.$$

Generating Functions for tails of maximum degree

Theorem [Goh, Schmutz 2002] For $i \in \mathbb{N}$ fixed, and $n = 2^m$

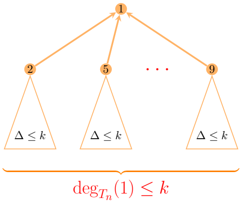
$$\mathbb{P}(\Delta_n - \log_2 n < i) = \exp\{-2^{-i}\} + o(1).$$

Idea.

$y_{n,k} = \#$ increasing trees with $\Delta_n \leq k$

$$f_k(z) = \sum_{k \geq 1} \frac{y_{n,k} z^n}{n!}$$

$$f_k'(z) = \sum_{j=0}^{k} \frac{(f_k(z))^j}{j!}$$



▷ Deleting the root is equivalent to taking the derivative of $f_k(z)$.

High degree vertices: Motivation

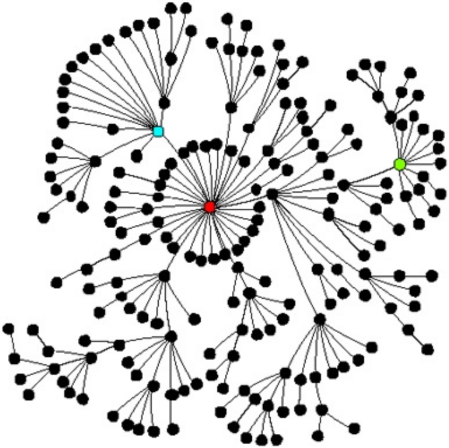


Image from scalefreenetworks, Flickr

High degree vertices [Addario-Berry, E. 2017, E. 2020, E. 2021]

- **Poisson Point Process for near-maximum degree vertices:**

  Number and their depth

- Central Limit Theorems (critical value $1 < c < \log e$):

$$Z^{(n)}_{\geq c \ln n} = \{v \in [n], \deg_{T_n}(v) \geq c \ln n\}$$

- Gumbel Distribution:

  Tighten tails for Δ_n

High degree vertices [Addario-Berry, E. 2017, E. 2020, E. 2021]

▷ **Poisson Point Process for near-maximum degree vertices:**

Number and their depth

▷ Central Limit Theorems (critical value $1 < c < \log e$):

$$Z^{(n)}_{\geq c \ln n} = \{v \in [n], \deg_{T_n}(v) \geq c \ln n\}$$

▷ Gumbel Distribution:

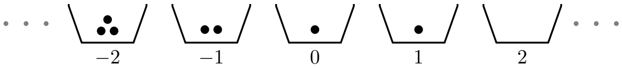
Tighten tails for Δ_n

Recent Advances: [E., Lodewijks, Ortgiese, 2022^+, Lodewijks 2022^+]
Same qualitative properties for WRRT with weight distribution $W \in (0,1]$ satisfying $\mathbb{P}(W = 1) > 0$.

A Poisson point process

For each vertex in T_n, place a point on $\mathbb{Z} \cup \{\infty\}$; $n = 2^m$.

$$\bullet = \left(\deg_{T_n}(v) - \log_2 n, \frac{\mathrm{ht}_{T_n}(v) - (1-\alpha)\ln n}{\sqrt{(1-\alpha/2)\ln n}} \right)$$



$$X_d = \#\{v \in [n], \deg_{T_n}(v) = d + \log_2 n\}$$

- Good news:
  - $(X_d)_{d\in\mathbb{Z}}$ have independent Poisson distribution
  - depth marks have Gaussian fluctuations,
  - independent from $(X_d)_{d\in\mathbb{Z}}$.

- **Surprising: Never-ending race** of vertices to become max-degree.

Kingman's coalescent

Kingman's Coalescent or Union-Find tree

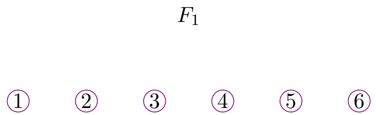
Fix $n \in \mathbb{N}$, for each $1 \leq t \leq n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

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

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.

Kingman's Coalescent or Union-Find tree

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

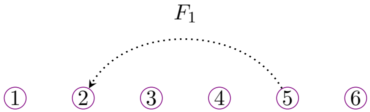
$$F_t = \{ T_1^{(t)}, \ldots, T_{n-t+1}^{(t)} \}$$

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.



Kingman's Coalescent or Union-Find tree

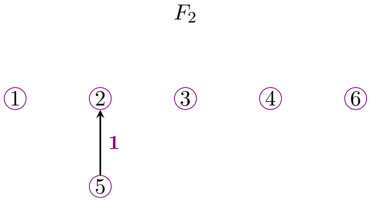
Fix $n \in \mathbb{N}$, for each $1 \leq t \leq n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

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

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.

Kingman's Coalescent or Union-Find tree

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

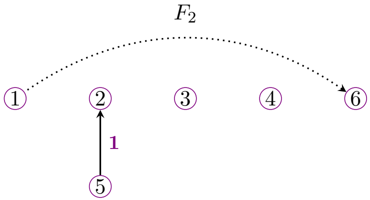
$$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$$

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.



Kingman's Coalescent or Union-Find tree

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

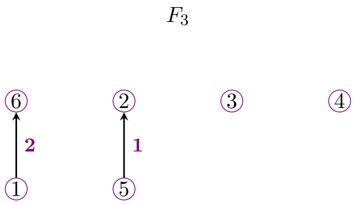
$$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$$

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.



Kingman's Coalescent or Union-Find tree

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

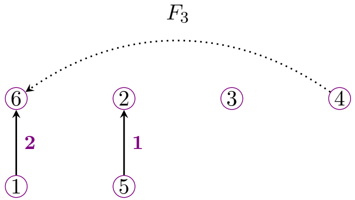
$$F_t = \{T_1^{(t)}, \dots, T_{n-t+1}^{(t)}\}$$

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.

All choices are independent.



Kingman's Coalescent or Union-Find tree

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

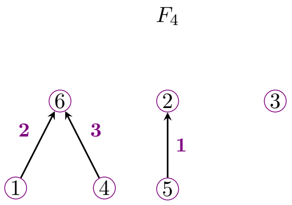
$$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$$

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.



Kingman's Coalescent or Union-Find tree

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

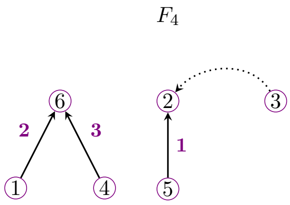
$$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$$

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.



Kingman's Coalescent or Union-Find tree

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

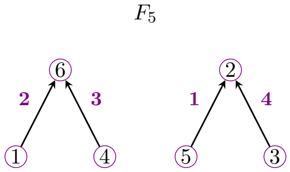
$$F_t = \{ T_1^{(t)}, \ldots, T_{n-t+1}^{(t)} \}$$

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.



Kingman's Coalescent or Union-Find tree

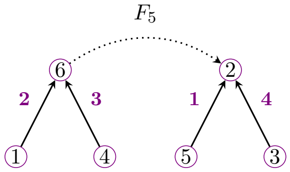
Fix $n \in \mathbb{N}$, for each $1 \leq t \leq n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

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

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.

Kingman's Coalescent or Union-Find tree

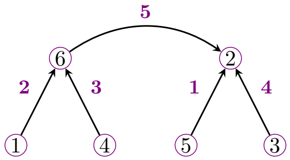
Fix $n \in \mathbb{N}$, for each $1 \leq t \leq n$ construct a forest of rooted labelled trees on $V(F_t) = \{1, \ldots, n\}$.

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

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.

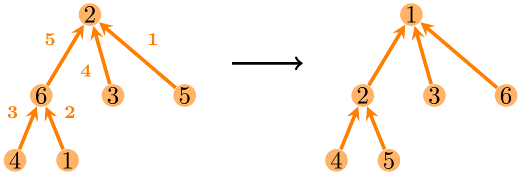


F_n

All choices are independent.

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.

Selection set $\mathcal{S}^{(n)}$

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

At step $t \in \mathcal{S}$, two trees are selected:

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

Selection set $\mathcal{S}^{(n)}$

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

At step $t \in \mathcal{S}$, two trees are selected:

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

Selection set $\mathcal{S}^{(n)}$

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

At step $t \in \mathcal{S}$, two trees are selected:

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

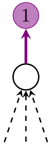
Selection set $\mathcal{S}^{(n)}$

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

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

At step $t \in \mathcal{S}$, two trees are selected:

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



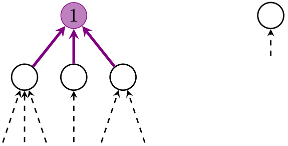
Selection set $\mathcal{S}^{(n)}$

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

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

At step $t \in \mathcal{S}$, two trees are selected:

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



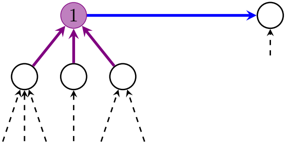
Selection set $\mathcal{S}^{(n)}$

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

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

At step $t \in \mathcal{S}$, two trees are selected:

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



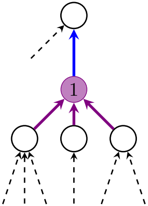
Selection set $\mathcal{S}^{(n)}$

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

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

At step $t \in \mathcal{S}$, two trees are selected:

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



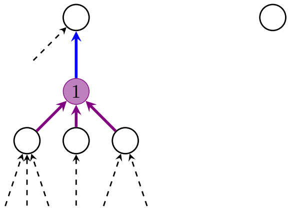
Selection set $\mathcal{S}^{(n)}$

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

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

At step $t \in \mathcal{S}$, two trees are selected:

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



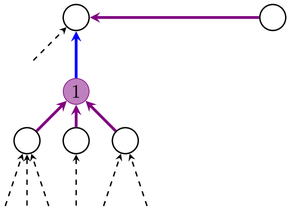
Selection set $\mathcal{S}^{(n)}$

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

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

At step $t \in \mathcal{S}$, two trees are selected:

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



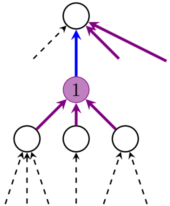
Selection set $\mathcal{S}^{(n)}$

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

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

At step $t \in \mathcal{S}$, two trees are selected:

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



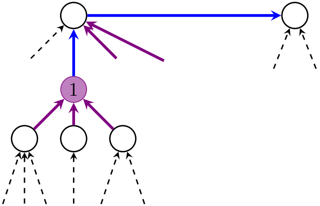
Selection set $\mathcal{S}^{(n)}$

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

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

At step $t \in \mathcal{S}$, two trees are selected:

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



Degree and depth of vertex 1 in F_n

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

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

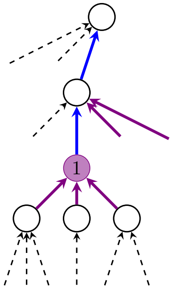
Proposition.

Total # non-favourable merges.

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

First streak favourable merges.

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



Degree and depth of vertex 1 in F_n

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

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

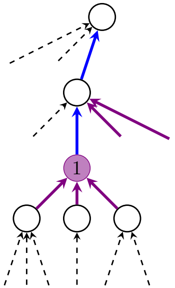
Proposition.

Total # non-favourable merges.

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

First streak favourable merges.

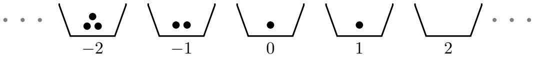
$$\deg_{F_n}(1) \stackrel{\mathcal{L}}{=} \min\{Geo(1/2), |\mathcal{S}|\}.$$



Recent Advances: [Lodewijks 2022$^+$]
Analysis can include the label of vertex 1 (in the RRT mapping).

Summary

- ▷ No persistency of vertex centrality
  - Never-ending race of vertices to become max-degree



- ▷ Advantages of Kingman's coalescent
  - **Combinatorial** foundation of known heuristics
  - Degree and depth of uniformly random vertices in T_n
  - Degree and depth of high-degree vertices in T_n

- ▷ Recent advances
  - **Labels** of high-degree vertices in T_n
  - High-degree results for **Weighted** random recursive trees