High degree vertices in random recursive trees

Laura Eslava joint work with Louigi Addario-Berry

McGill University

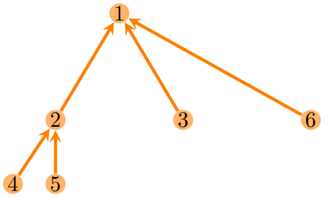
Random Structures and Algorithms, 2015

July 31st, 2015

Recursive trees: Standard construction

T_n is a rooted labelled tree on $[n] = \{1, 2, \ldots, n\}$.
Given T_{n-1}, construct T_n:

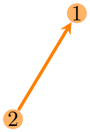
- Add a vertex n,
- Select $j \in [n-1]$ uniformly at random,
- Add the edge $n \to j$.



Recursive trees: Standard construction

T_n is a rooted labelled tree on $[n] = \{1, 2, \ldots, n\}$.
Given T_{n-1}, construct T_n:

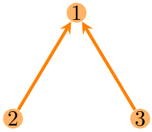
- Add a vertex n,
- Select $j \in [n-1]$ uniformly at random,
- Add the edge $n \to j$.



Recursive trees: Standard construction

T_n is a rooted labelled tree on $[n] = \{1, 2, \ldots, n\}$.
Given T_{n-1}, construct T_n:

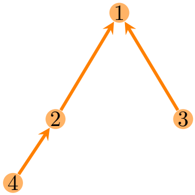
- Add a vertex n,
- Select $j \in [n-1]$ uniformly at random,
- Add the edge $n \to j$.



Recursive trees: Standard construction

T_n is a rooted labelled tree on $[n] = \{1, 2, \ldots, n\}$.
Given T_{n-1}, construct T_n:

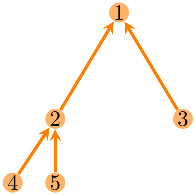
- Add a vertex n,
- Select $j \in [n-1]$ uniformly at random,
- Add the edge $n \to j$.



Recursive trees: Standard construction

T_n is a rooted labelled tree on $[n] = \{1, 2, \ldots, n\}$.
Given T_{n-1}, construct T_n:

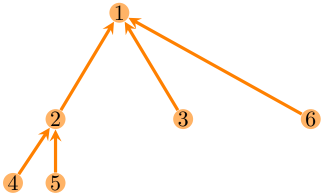
- Add a vertex n,
- Select $j \in [n-1]$ uniformly at random,
- Add the edge $n \to j$.



Recursive trees: Standard construction

T_n is a rooted labelled tree on $[n] = \{1, 2, \ldots, n\}$.
Given T_{n-1}, construct T_n:

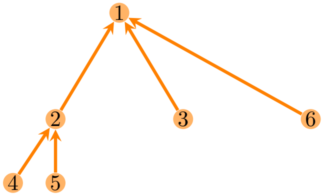
- Add a vertex n,
- Select $j \in [n-1]$ uniformly at random,
- Add the edge $n \to j$.



Recursive trees: Standard construction

T_n is a rooted labelled tree on $[n] = \{1, 2, \ldots, n\}$.
Given T_{n-1}, construct T_n:

- Add a vertex n,
- Select $j \in [n-1]$ uniformly at random,
- Add the edge $n \to j$.



Facts. - Labels are increasing along root-to-leaf paths.
- There are $n-1!$ possible outcomes for T_n.

Degree distribution

For $i \in [n]$ and $d \in \mathbb{N} \cup \{0\}$, let

$$\begin{aligned} \deg_n(i) &= \#\{j > i : \; j \to i \text{ in } \; T_n\}, \\ Z_d^{(n)} &= \#\{i \in [n] : \deg_n(i) = d\}. \end{aligned}$$

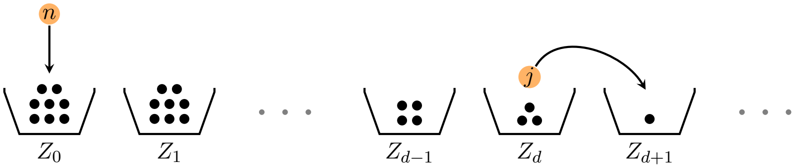
What can be said about $\{Z_d^{(n)}\}_{d \geq 0}$?

Degree distribution

For $i \in [n]$ and $d \in \mathbb{N} \cup \{0\}$, let

$$\deg_n(i) = \#\{j > i : \ j \to i \text{ in } T_n\},$$
$$Z_d^{(n)} = \#\{i \in [n] : \deg_n(i) = d\}.$$

What can be said about $\{Z_d^{(n)}\}_{d \geq 0}$?



Asymptotic degree distribution

Theorem (Na, Rapoport 1970) For all $d \geq 0$, as $n \to \infty$

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

Asymptotic degree distribution

Theorem (Na, Rapoport 1970) For all $d \geq 0$, as $n \to \infty$

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

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 the V_d are jointly Gaussian r.v.'s with zero mean and explicit covariance.

Maximum degree: $\Delta_n = \max\{\deg_n(i) : i \in [n]\}$

Theorem (Devroye, Lu 1995) As $n \to \infty$, a.s.

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

Heuristic.

- Suppose $\mathbb{E}Z_m^{(n)} \approx n2^{-(m+1)}$ for $m(n) \to \infty$.
  Then $\mathbb{E}Z_{\lfloor \log_2 n \rfloor} \approx 1$.

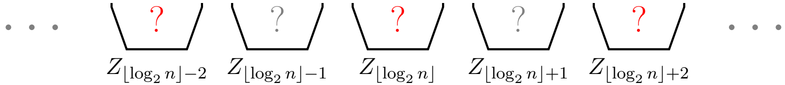
Maximum degree: $\Delta_n = \max\{\deg_n(i) : i \in [n]\}$

Theorem (Devroye, Lu 1995) As $n \to \infty$, a.s.

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

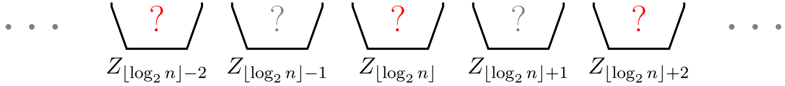
Heuristic.

- Suppose $\mathbb{E} Z_m^{(n)} \approx n 2^{-(m+1)}$ for $m(n) \to \infty$.
  Then $\mathbb{E} Z_{\lfloor \log_2 n \rfloor} \approx 1$.



Some new questions

- Which vertex attains the maximum degree?
- How many do?



For $d \in \mathbb{Z}$, let

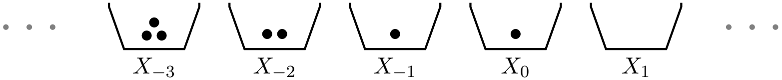
$$X_d^{(n)} = Z_{\lfloor \log_2 n \rfloor + d}^{(n)} = \#\{i \in [n] : \deg_n(i) = \lfloor \log_2 n \rfloor + d\}$$

Asymptotic High-degree distribution

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

$$X_d^{(n)} \xrightarrow{\text{dist}} W_d;$$

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



Furthermore: $\sum_{j \geq d} X_d^{(n)} \xrightarrow{\text{dist}} \sum_{j \geq d} W_j$.

Detour Maximum degree distribution

Theorem (Addario-Berry,E. 2015$^+$) Let $n = 2^k$. If $\liminf d(n) > -\infty$ and $\log_2 n + d(n) < 2 \ln n$, then

$$\mathbb{P}(\Delta_n \geq \log_2 n + d) = (1 - e^{-2^{-d}})(1 + o(1)).$$

Remarks.

- 'Discrete' version of the Gumbel's distribution.
- Goh and Schmutz (2002) proved the case $d \in \mathbb{Z}$ fixed.

Heuristic.

▷ $\{\Delta_n \geq \log_2 n + d\}$ iff $\{\sum_{j \geq d} X_d^{(n)} > 0\}$

▷ $\mathbb{P}(Poi(\lambda) > 0) = 1 - e^{-\lambda}$.

The crutial task

Let

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

Key Property. Let $n = 2^k$. For any $d \in \mathbb{Z}$, as $k \to \infty$

$$X_{\geq d}^{(n)} \xrightarrow{\text{dist}} \mathrm{Poi}(2^{-d}).$$

The crutial task

Let

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

Key Property. Let $n = 2^k$. For any $d \in \mathbb{Z}$, as $k \to \infty$

$$X_{\geq d}^{(n)} \xrightarrow{\text{dist}} \mathrm{Poi}(2^{-d}).$$

Proof's technique: Method of moments.
First step:

$$\mathbb{E}X_{\geq d}^{(n)} \to 2^{-d}.$$

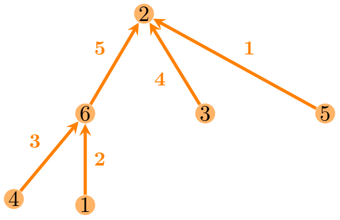
$$\mathbb{E}X_{\geq d}^{(n)} = \sum_{i \in [n]} \mathbb{P}(\deg_n(i) \geq \lfloor \log_2 n \rfloor + d)$$

A new approach: Coalescent

$F_1, \ldots, F_n$ are labelled forests on n vertices, with directed edges.
$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$, trees listed in $\nearrow$ order of least element.

To construct F_t, given F_{t-1}:

- ▷ Select two trees with indices $\{a_t, b_t\}$ uniformly at random.
- ▷ According to a fair coin flip ξ_t, select the direction of the edge that will connect the root of such trees.
- * All r.v.'s are independent.



A new approach: Coalescent

$F_1, \ldots, F_n$ are labelled forests on n vertices, with directed edges.
$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$, trees listed in $\nearrow$ order of least element.

To construct F_t, given F_{t-1}:

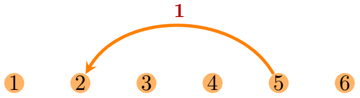
- ▷ Select two trees with indices $\{a_t, b_t\}$ uniformly at random.
- ▷ According to a fair coin flip ξ_t, select the direction of the edge that will connect the root of such trees.
- * All r.v.'s are independent.

A new approach: Coalescent

$F_1, \ldots, F_n$ are labelled forests on n vertices, with directed edges.
$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$, trees listed in $\nearrow$ order of least element.

To construct F_t, given F_{t-1}:

- ▷ Select two trees with indices $\{a_t, b_t\}$ uniformly at random.
- ▷ According to a fair coin flip ξ_t, select the direction of the edge that will connect the root of such trees.
- * All r.v.'s are independent.

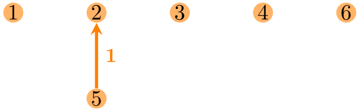


A new approach: Coalescent

$F_1, \ldots, F_n$ are labelled forests on n vertices, with directed edges.
$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$, trees listed in $\nearrow$ order of least element.

To construct F_t, given F_{t-1}:

- ▷ Select two trees with indices $\{a_t, b_t\}$ uniformly at random.
- ▷ According to a fair coin flip ξ_t, select the direction of the edge that will connect the root of such trees.
- * All r.v.'s are independent.



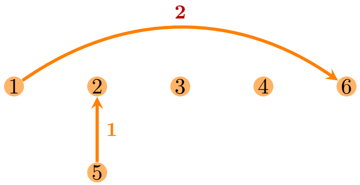
A new approach: Coalescent

$F_1, \ldots, F_n$ are labelled forests on n vertices, with directed edges.
$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$, trees listed in $\nearrow$ order of least element.

To construct F_t, given F_{t-1}:

- Select two trees with indices $\{a_t, b_t\}$ uniformly at random.
- According to a fair coin flip ξ_t, select the direction of the edge that will connect the root of such trees.

* All r.v.'s are independent.

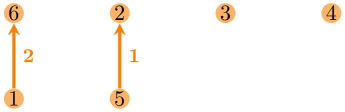


A new approach: Coalescent

$F_1, \ldots, F_n$ are labelled forests on n vertices, with directed edges.
$F_t = \{ T_1^{(t)}, \ldots, T_{n-t+1}^{(t)} \}$, trees listed in $\nearrow$ order of least element.

To construct F_t, given F_{t-1}:

- ▷ Select two trees with indices $\{a_t, b_t\}$ uniformly at random.
- ▷ According to a fair coin flip ξ_t, select the direction of the edge that will connect the root of such trees.
- * All r.v.'s are independent.

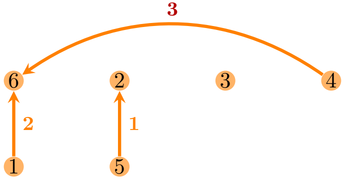


A new approach: Coalescent

$F_1, \ldots, F_n$ are labelled forests on n vertices, with directed edges.
$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$, trees listed in $\nearrow$ order of least element.

To construct F_t, given F_{t-1}:

- ▷ Select two trees with indices $\{a_t, b_t\}$ uniformly at random.
- ▷ According to a fair coin flip ξ_t, select the direction of the edge that will connect the root of such trees.
- * All r.v.'s are independent.

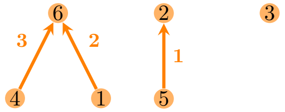


A new approach: Coalescent

$F_1, \ldots, F_n$ are labelled forests on n vertices, with directed edges.
$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$, trees listed in $\nearrow$ order of least element.

To construct F_t, given F_{t-1}:

- ▷ Select two trees with indices $\{a_t, b_t\}$ uniformly at random.
- ▷ According to a fair coin flip ξ_t, select the direction of the edge that will connect the root of such trees.
- * All r.v.'s are independent.

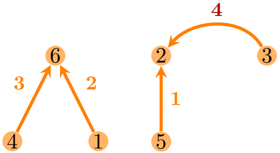


A new approach: Coalescent

$F_1, \ldots, F_n$ are labelled forests on n vertices, with directed edges.
$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$, trees listed in $\nearrow$ order of least element.

To construct F_t, given F_{t-1}:

- ▷ Select two trees with indices $\{a_t, b_t\}$ uniformly at random.
- ▷ According to a fair coin flip ξ_t, select the direction of the edge that will connect the root of such trees.
- * All r.v.'s are independent.

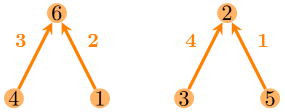


A new approach: Coalescent

$F_1, \dots, F_n$ are labelled forests on n vertices, with directed edges.
$F_t = \{ T_1^{(t)}, \dots, T_{n-t+1}^{(t)} \}$, trees listed in $\nearrow$ order of least element.

To construct F_t, given F_{t-1}:

- ▷ Select two trees with indices $\{a_t, b_t\}$ uniformly at random.
- ▷ According to a fair coin flip ξ_t, select the direction of the edge that will connect the root of such trees.
- * All r.v.'s are independent.

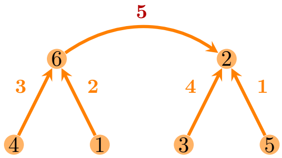


A new approach: Coalescent

$F_1, \ldots, F_n$ are labelled forests on n vertices, with directed edges.
$F_t = \{T_1^{(t)}, \ldots, T_{n-t+1}^{(t)}\}$, trees listed in $\nearrow$ order of least element.

To construct F_t, given F_{t-1}:

- ▷ Select two trees with indices $\{a_t, b_t\}$ uniformly at random.
- ▷ According to a fair coin flip ξ_t, select the direction of the edge that will connect the root of such trees.
- * All r.v.'s are independent.

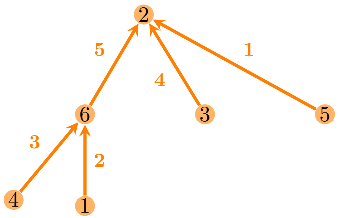


A new approach: Coalescent

$F_1, \dots, F_n$ are labelled forests on n vertices, with directed edges.
$F_t = \{T_1^{(t)}, \dots, T_{n-t+1}^{(t)}\}$, trees listed in $\nearrow$ order of least element.

To construct F_t, given F_{t-1}:

- ▷ Select two trees with indices $\{a_t, b_t\}$ uniformly at random.
- ▷ According to a fair coin flip ξ_t, select the direction of the edge that will connect the root of such trees.
- * All r.v.'s are independent.



Exchangeable construction

Fact.

- Vertex labels in $T_1^{(n)}$ are exchangeable.

Claim. $T_1^{(n)}$ has same degree distribution as a RRT tree.

Exchangeable construction

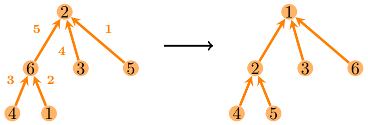
Fact.

- Vertex labels in $T_1^{(n)}$ are exchangeable.

Claim. $T_1^{(n)}$ has same degree distribution as a RRT tree.

Proof sketch.

- 'Edge labels' are decreasing along root-to-leaf paths.
- There are $n!(n-1)!$ possible outcomes for $T_1^{(n)}$.



□

Degree of vertex 1 in $T_1^{(n)} :=$ deg$^*(1)$

Let $S = \{t \in [n-1] : 1 \in \{a_t, b_t\}\}$.

Claim.

$$\mathbb{P}(\deg^*(1) \geq m) = 2^{-m}\mathbb{P}(|S| \geq m).$$

Degree of vertex 1 in $T_1^{(n)} := \deg^*(1)$

Let $S = \{t \in [n-1] : 1 \in \{a_t, b_t\}\}$.

Claim.

$$\mathbb{P}(\deg^*(1) \geq m) = 2^{-m}\mathbb{P}(|S| \geq m).$$

Proof.

Vertex 1 is in $T_1^{(t)}$ for all $t \in [n]$. Then, its degree increases

- only at steps $t \in S$,
- only if vertex 1 is a root in F_t.

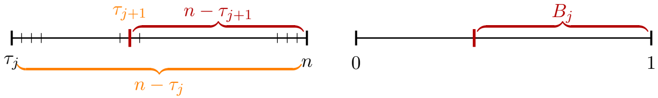
□

Stick-breaking process: $S = \{\tau_1, \tau_2, \ldots, \tau_{|S|}\}$

Lemma. For any $m < 2\ln n$, as $n \to \infty$

$$\mathbb{P}(|S| \geq m) \to 1.$$

Heuristic.
Let τ_j be the j-th time $1 \in \{a_t, b_t\}$ then $n - \tau_{j+1}$ is close to $B_j(n - \tau_j)$, where B_j are independent $\mathrm{Beta}(2, 1)$ r.v.'s.



□

Note. $\log_2 n < 2\ln n$.

Perks of the new contruction

Key Property. Let $n = 2^k$. For any $d \in \mathbb{Z}$, as $k \to \infty$

$$X_{\geq d}^{(n)} \xrightarrow{\text{dist}} \mathrm{Poi}(2^{-d}).$$

Perks of the new contruction

Key Property. Let $n = 2^k$. For any $d \in \mathbb{Z}$, as $k \to \infty$

$$X_{\geq d}^{(n)} \xrightarrow{\text{dist}} \mathrm{Poi}(2^{-d}).$$

Proof sketch.

- Alternative construction + Exchangeability

$$\mathbb{E}X_{\geq d}{}^{(n)} = n\mathbb{P}(\deg^*(1) \geq \lfloor \log_2 n \rfloor + d),$$

Perks of the new contruction

Key Property. Let $n = 2^k$. For any $d \in \mathbb{Z}$, as $k \to \infty$

$$X_{\geq d}^{(n)} \xrightarrow{\text{dist}} \mathrm{Poi}(2^{-d}).$$

Proof sketch.

- ▷ Alternative construction + Exchangeability
- ▷ Decoupling randomness + 'Stick-breaking process'

$$\begin{aligned} \mathbb{E}X_{\geq d}{}^{(n)} &= n\mathbb{P}(\deg^*(1) \geq \lfloor \log_2 n \rfloor + d), \\ &= n \cdot 2^{-(\lfloor \log_2 n \rfloor + d)}\mathbb{P}(|S| \geq \lfloor \log_2 n \rfloor + d), \\ &= 2^{-d}(1 + o(1)). \end{aligned}$$

Perks of the new contruction

Key Property. Let $n = 2^k$. For any $d \in \mathbb{Z}$, as $k \to \infty$

$$X_{\geq d}^{(n)} \xrightarrow{\text{dist}} \mathrm{Poi}(2^{-d}).$$

Proof sketch.

- Alternative construction + Exchangeability
- Decoupling randomness + 'Stick-breaking process'

$$\begin{aligned}\mathbb{E}X_{\geq d}{}^{(n)} &= n\mathbb{P}(\deg^*(1) \geq \lfloor \log_2 n \rfloor + d), \\ &= n \cdot 2^{-(\lfloor \log_2 n \rfloor + d)} \mathbb{P}(|S| \geq \lfloor \log_2 n \rfloor + d), \\ &= 2^{-d}(1 + o(1)).\end{aligned}$$

Apply analogous arguments for higher moments of $X_{\geq d}^{(n)}$. □

Thank you!

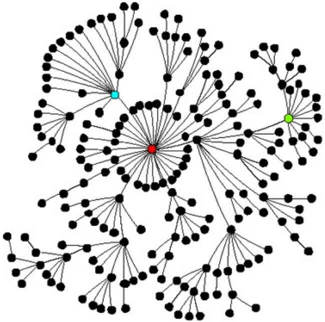


Image from scalefreenetworks, Flickr