Fringe trees for random trees with given vertex degrees

University of Liverpool

Gabriel Hernan Berzunza Ojeda

Joint work with Cecilia Holmgren and Svante Janson (Uppsala University)

AofA 2024, University of Bath, June 17-21, 2024

Fringe Subtrees

Let $\mathcal{T}^{pl}$ be the set of all (finite) plane rooted trees (ordered rooted trees).

For $T \in \mathcal{T}^{pl}$, $|T|$ = size of T (number of vertices).

Fringe Subtrees

Let $\mathbb{T}^{pl}$ be the set of all (finite) plane rooted trees (ordered rooted trees).

For $T \in \mathbb{T}^{pl}$, $|T|$ = size of T (number of vertices)

Given $T \in \mathbb{T}^{pl}$ and a vertex $v \in T$, let T_v be the subtree rooted at v

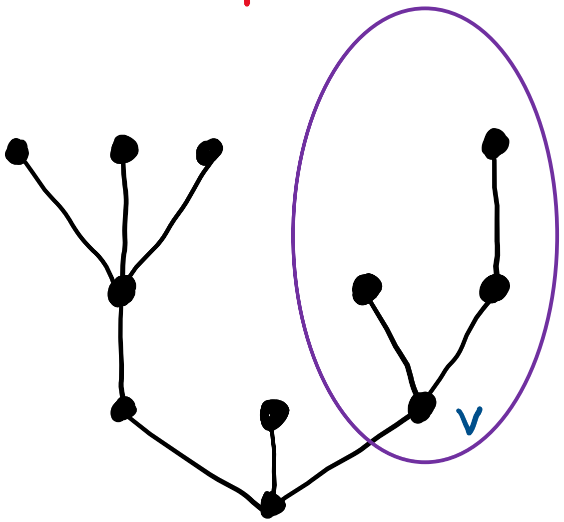
Fringe Subtrees

Let $\mathbb{T}^{pl}$ be the set of all (finite) plane rooted trees (ordered rooted trees).

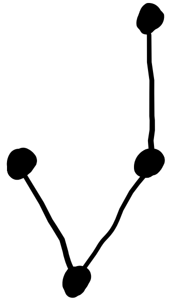
For $T \in \mathbb{T}^{pl}$, $|T| =$ size of T (number of vertices)

Given $T \in \mathbb{T}^{pl}$ and a vertex $v \in T$, let T_v be the subtree rooted at v

Example.



T

$T_v =$ 

These subtrees are called fringe subtrees.

We are interested in $\{T_v\}$ of a given T

Fringe Subtrees

For $T \in \mathbb{T}^{pl}$, consider the random fringe subtree T^*

$$T^* = T_v \quad \text{where } v \text{ is a uniform random vertex in } T$$

Fringe Subtrees

For $T \in \mathbb{T}^{pl}$, consider the random fringe subtree T^*

Aldous (1991). In general and for many important examples,

- Random recursive trees
- Binary search trees
- Conditioned Galton-Watson trees

$\vdots$

Subtree Counts

For $T, T' \in \mathbb{T}^{pl}$,

$$N_{T'}(T) = |\{v \in T : T_v = T'\}| = \sum_{v \in T} \mathbb{1}\{T_v = T'\},$$

i.e., the number of fringe subtrees of T that are equal (i.e., isomorphic) to T'.

Then the distribution of the random fringe subtree T^* is given by

$$\mathbb{P}(T^* = T') = \frac{N_{T'}(T)}{|T|}, \quad T' \in \mathbb{T}^{pl}.$$

Thus, the study the distribution of T^* is equivalent to study

$$N_{T'}(T).$$

Subtree Counts

In our case, T is a random tree.

- $N_{T'}(T)$ is a random variable for each $T' \in \Pi^{pl}$.
- The distribution of T^* is a random probability distribution on Π^{pl}, with

$$\mathbb{P}(T^* = T' \mid T) = \frac{N_{T'}(T)}{|T|}, \qquad T' \in \Pi^{pl}.$$

Galton - Watson Trees

Consider conditioned GW-trees to have size $m \in \mathbb{N}$ with offspring distribution ξ on $\mathbb{N}_0$ such that $\mathbb{E}[\xi] = 1$ and $\sigma^2 := \mathrm{Var}(\xi)$ (and non-zero)

Theorem. Aldous (1991). Let $\mathcal{T}_m$ be a conditioned (critical) G-W tree with $\sigma^2 < \infty$. For every fixed $T \in \mathbb{T}^{pl}$,

- $\mathbb{P}(\mathcal{T}_m^\bullet = T \mid \mathcal{T}_m) = \dfrac{N_T(\mathcal{T}_m)}{m} \xrightarrow[m \to \infty]{p} \mathbb{P}(\mathcal{T} = T)$ (Quenched Version)
- $\mathbb{P}(\mathcal{T}_m^\bullet = T) = \dfrac{\mathbb{E}\, N_T(\mathcal{T}_m)}{m} \xrightarrow[m \to \infty]{} \mathbb{P}(\mathcal{T} = T)$ (Annealed Version)

$\mathcal{T}$ is the corresponding unconditioned G-W tree

$$\mathbb{P}(\mathcal{T} = T) = \prod_{i \geq 0} p_i^{n_T(i)}, \quad n_T(i) = \#\text{ vertices in } T \text{ with out-degree } i$$

$$p_i = \mathbb{P}(\xi = i).$$

Galton - Watson Trees

Let $\mathcal{T}_m$ be a conditioned (critical) G-W tree with $\sigma^2 < \infty$.

Let $\mathcal{T}$ be the corresponding unconditioned G-W tree. Let $\pi(T) := \mathbb{P}(\mathcal{T} = T)$, $T \in \mathbb{T}^{\mathrm{pl}}$.

Theorem.

- $\mathbb{E}[N_T(\mathcal{T}_m)] = m\pi(T) + o(\sqrt{m})$
- $\mathrm{Var}(N_T(\mathcal{T}_m)) = m\left(\pi(T) - (2|T| - 1 + \sigma^{-2})\pi(T)^2\right) + o(m)$
- $$\frac{N_T(\mathcal{T}_m) - m\pi(T)}{\sqrt{m}} \xrightarrow[m \to \infty]{d} N\left(0, \pi(T) - (2|T| - 1 + \sigma^{-2})\pi(T)^2\right).$$

- Janson (2001) (assuming a third moment),
- Minami (2005) and Drmota (2009) (assuming an exponential moment)
- Janson (2016) (in general, but $\sigma^2 < \infty$)

Galton - Watson Trees

Let $\mathcal{T}_n$ be a conditioned (critical) G-W tree with $\sigma^2 < \infty$.

Let $\mathcal{T}$ be the corresponding unconditioned G-W tree. Let $\pi(T) := \mathbb{P}(\mathcal{T} = T)$, $T \in \mathbb{T}^{PI}$.

Theorem.

- $\mathbb{E}[N_T(\mathcal{T}_n)] = n\pi(T) + o(\sqrt{n})$
- $\mathrm{Var}(N_T(\mathcal{T}_n)) = n\left(\pi(T) - (2|T| - 1 + \sigma^{-2})\pi(T)^2\right) + o(n)$
- $$\frac{N_T(\mathcal{T}_n) - n\pi(T)}{\sqrt{n}} \xrightarrow[n\to\infty]{d} N\left(0, \pi(T) - (2|T| - 1 + \sigma^{-2})\pi(T)^2\right).$$

Remark. Janson (2016) ($\sigma^2 < \infty$). Extended it to a Multivariate Version, i.e.

$(N_{T_1}(\mathcal{T}_n), \dots, N_{T_k}(\mathcal{T}_n))$ is asymptotically normal (up to renormalization)

Trees with given vertex degrees

For $T \in \mathbb{T}^{\mathrm{Pl}}$ and $v \in T$, Let $d_T(v)$ be the out-degree of v (number of children)

The degree statistic of T is the sequence $n_T = (n_T(i))_{i \geq 0}$ where

$$n_T(i) = |\{v \in T : d_T(v) = i\}|$$

is the number of vertices of T with i children. We have

$$|T| = \sum_{i \geq 0} n_T(i) = 1 + \sum_{i \geq 0} i\, n_T(i).$$

Trees with given vertex degrees

A sequence of non-negative integers $n = (n(i))_{i \geq 0}$ is the degree statistic of some tree if and only if $\sum_{i \geq 0} n(i) = 1 + \sum_{i \geq 0} i\, n(i)$.

Trees with given vertex degrees

A sequence of non-negative integers $n = (n(i))_{i \geq 0}$ is the degree statistic of some tree if and only if $\sum_{i \geq 0} n(i) = 1 + \sum_{i \geq 0} i\, n(i)$.

Let $|n| = \sum_{i \geq 0} n(i)$ be the size of $n = (n(i))_{i \geq 0}$.

Trees with given vertex degrees

A sequence of non-negative integers $\mathbf{n} = (n(i))_{i \geq 0}$ is the degree statistic of some tree if and only if $\sum_{i \geq 0} n(i) = 1 + \sum_{i \geq 0} i\, n(i)$.

Let $|\mathbf{n}| = \sum_{i \geq 0} n(i)$ be the size of $\mathbf{n} = (n(i))_{i \geq 0}$

Let $\Pi_{\mathbf{n}}^{d}$ be the set of plane rooted trees with degree statistic $\mathbf{n} = (n(i))_{i \geq 0}$

Trees with given vertex degrees

A sequence of non-negative integers $\mathbf{n}=(n(i))_{i\geq 0}$ is the degree statistic of some tree if and only if $\sum_{i\geq 0} n(i) = 1 + \sum_{i\geq 0} i\, n(i)$.

Let $|\mathbf{n}| = \sum_{i\geq 0} n(i)$ be the size of $\mathbf{n}=(n(i))_{i\geq 0}$

Let $\mathbb{T}_{\mathbf{n}}^{d}$ be the set of plane rooted trees with degree statistic $\mathbf{n}=(n(i))_{i\geq 0}$

Let $T_{\mathbf{n}} \sim \mathrm{Unif}(\mathbb{T}_{\mathbf{n}}^{d})$ (i.e. a uniform tree with given degree statistic $\mathbf{n}=(n(i))_{i\geq 0}$)

Pitman (2002): $|\mathbb{T}_{\mathbf{n}}^{d}| = \frac{1}{|\mathbf{n}|} \frac{|\mathbf{n}|!}{\prod_{i\geq 0} n(i)!}$.

Trees with given vertex degrees

A sequence of non-negative integers $n = (n(i))_{i \geq 0}$ is the degree statistic of some tree if and only if $\sum_{i \geq 0} n(i) = 1 + \sum_{i \geq 0} i\, n(i)$.

Let $|n| = \sum_{i \geq 0} n(i)$ be the size of $n = (n(i))_{i \geq 0}$

Let $\mathbb{T}_n^d$ be the set of plane rooted trees with degree statistic $n = (n(i))_{i \geq 0}$

Let $T_n \sim \mathrm{Unif}(\mathbb{T}_n^d)$ (i.e. a uniform tree with given degree statistic)

Remark. A G-W tree, conditioned on having degree statistic $n = (n(i))_{i \geq 0}$, is uniformly distributed on $\mathbb{T}_n^d$.

Trees with given vertex degrees

For a degree statistic $n = (n(i))_{i \geq 0}$, denote by $p(n) = (p_i(n))_{i \geq 0}$ its (empirical) degree distribution, i.e.,

$$p_i(n) = \frac{n(i)}{|n|}, \quad i \geq 0$$

Condition. $n_k = (n_k(i))_{i \geq 0}$, $k \geq 1$, are degree statistics s.t. as $k \to \infty$,

$|n_k| \to \infty$ and $p_i(n_k) \to p_i$, where $p = (p_i)_{i \geq 0}$ is a probility distribution on $\mathbb{N}_0$.

Trees with given vertex degrees

Condition. $n_k = (n_k(i))_{i\geq 0}$, $k \geq 1$, are degree statistics s.t. as $k \to \infty$,

$|n_k| \to \infty$ and $p_i(n_k) \to p_i$, where $p = (p_i)_{i\geq 0}$ is a probility distribution on $\mathbb{N}_0$.

Theorem 1. Let $T_{n_k} \sim \text{Uniform}(\Pi^d_{n_k})$

- $\mathbb{P}(T^\bullet_{n_k} = T \mid T_{n_k}) = \dfrac{N_T(T_{n_k})}{|n_k|} \xrightarrow[k \to \infty]{P} \pi_p(T).$
- $\mathbb{P}(T^\bullet_{n_k} = T) = \dfrac{\mathbb{E}\, N_T(T_{n_k})}{|n_k|} \xrightarrow[k\to\infty]{} \pi_p(T).$

where $\pi_p(T) = \prod_{i\geq 0} p_i^{n_T(i)}$, $T \in \Pi$. (with $0^0 = 1$)

Remark. $\pi_p(\cdot)$ is the distribution of a (uncondioned) G-W tree

with offspring distribution $p = (p_i)_{i\geq 0}$.

Trees with given vertex degrees

Condition. $\mathbf{n}_k = (n_k(i))_{i\geq 0}$, $k\geq 1$, are degree statistics s.t. as $k\to\infty$,

$|\mathbf{n}_k|\to\infty$ and $p_i(\mathbf{n}_k)\to p_i$, where $\mathbf{p}=(p_i)_{i\geq 0}$ is a probability distribution on $\mathbb{N}_0$.

Theorem 2. Let $T_{\mathbf{n}_k}\sim \text{Uniform}(\mathbb{T}^d_{\mathbf{n}_k})$

- $\mathbb{E}\, N_T(T_{\mathbf{n}_k}) = \pi_{\mathbf{p}}(T)\,|\mathbf{n}_k| + o(|\mathbf{n}_k|)$
- $\text{Var}(N_T(T_{\mathbf{n}_k})) = \left(\pi_{\mathbf{p}}(T) + \eta_{\mathbf{p}}(T)\,\pi_{\mathbf{p}}(T)^2\right)|\mathbf{n}_k| + o(|\mathbf{n}_k|)$
- $$\frac{N_T(T_{\mathbf{n}_k}) - \pi_{\mathbf{p}(\mathbf{n}_k)}(T)\,|\mathbf{n}_k|}{\sqrt{|\mathbf{n}_k|}} \xrightarrow[k\to\infty]{d} N\left(0,\ \pi_{\mathbf{p}}(T) + \eta_{\mathbf{p}}(T)\,\pi_{\mathbf{p}}(T)^2\right).$$

where $\eta_{\mathbf{p}}(T) = (|T|-1)^2 - \sum_{i\geq 0}\frac{n_T(i)^2}{p_i}$ (here $0/0=0$).

Trees with given vertex degrees

Theorem 2. Let $T_{n_k} \sim \mathrm{Uniform}(\Pi^{d}_{n_k})$

- $\mathbb{E}\, N_T(T_{n_k}) = \pi_p(T)\,|n_k| + o(|n_k|)$
- $\mathrm{Var}\,(N_T(T_{n_k})) = \left(\pi_p(T) + \eta_p(T)\,\pi_p(T)^2\right)|n_k| + o(|n_k|)$
- $\dfrac{N_T(T_{n_k}) - \pi_{p(n_k)}(T)\,|n_k|}{\sqrt{|n_k|}} \xrightarrow[k\to\infty]{d} N\left(0, \pi_p(T) + \eta_p(T)\,\pi_p(T)^2\right)$

where $\eta_p(T) = (|T|-1)^2 - \sum_{i\geq 0} \frac{n_T(i)^2}{p_i}$ (here $0/0 = 0$)

Remark. We also proved a Multivariate Version, i.e.

$(N_{T_1}(T_{n_k}), \dots, N_{T_q}(T_{n_k}))$ is asymptotically normal (up to renormalization)

Galton - Watson Trees

Let $\mathcal{T}_m$ be a conditioned (critical) G-W tree with offspring distribution in the domain of attraction of a stable law of index $\alpha \in (1,2)$. $(\sigma^2 = \infty)$

Theorem.

$$\frac{N_T(\mathcal{T}_m) - m\pi(T)}{\sqrt{m}} \xrightarrow[m\to\infty]{d} N\left(0,\ \pi(T) - (2|T|-1)\,\pi(T)^2\right)$$

where $\mathcal{T}$ is the corresponding unconditioned G-W tree and $\pi(T) := \mathbb{P}(\mathcal{T} = T)$, $T \in \mathbb{T}^{Pl}$.

Sketch of Proofs (Theorem 2)

For $x \in \mathbb{R}$ and $q \in \mathbb{N}_0$, $(x)_q = x(x-1)\cdots(x-q+1)$ the q-th falling factorial of x

(Here $(x)_0 = 1$. Note that $(x)_q = 0$ whenever $x \in \mathbb{N}_0$ and $x-q+1 \leq 0$).

Sketch of Proofs (Main ingredient)

Theorem (Gao and Wormald (2004)). Let $(X_m)_{m\geq 1}$ be a sequence of non-negative r.v. Supose that μ_m and σ_m are positive real numbers s.t., as $m\to\infty$

$$\sigma_m \ll \mu_m \ll \sigma_m^3$$

Let $\gamma > 0$ be a fixed real number. Let $c > 0$ be a constant, and suppose further that, as $m\to\infty$, uniformly for all integer sequences $(k_m)_{m\geq 1}$ with $0 \leq k_m \leq c\,\mu_m/\sigma_m$,

$$\mathbb{E}(X_m)_{k_m} = \mu_m^{k_m} \cdot \exp\left(\frac{1}{2}\,\frac{\gamma\sigma_m^2 - \mu_m}{\mu_m^2}\,k_m^2 + o(1)\right).$$

Then,

$$\frac{X_m - \mu_m}{\sigma_m} \xrightarrow[m\to\infty]{d} N(0,\gamma).$$

Sketch of Proofs (Main ingredient)

Theorem (Multivariate G-W Theorem). For $m, n \in \mathbb{N}$, let $(X_{1n}, \dots, X_{mn})$ be vectors of nonnegative r.v. Suppose that μ_{in} and σ_{in} are positive real numbers s.t for each $1 \le i \le m$, as $n \to \infty$,

$$\sigma_{in} \ll \mu_{in} \ll \sigma_{in}^3$$

Let $\Gamma = (\gamma_{ij})_{i,j=1}^{m}$ be a fixed matrix. Let $c > 0$ be a constant, and suppose further that, as $m \to \infty$, uniformly for all integer sequences $(k_{in})_{i=1}^{m}$ with $0 \le k_{in} \le c\mu_{in}/\sigma_{in}$

$$\mathbb{E} \prod_{i=1}^{m} (X_{in})_{k_{in}} = \prod_{i=1}^{m} \mu_{in}^{k_{in}} \cdot \exp\left(\frac{1}{2} \sum_{i,j=1}^{m} \frac{\gamma_{ij}\sigma_{in}\sigma_{jn} - \delta_{ij}\mu_{in}}{\mu_{in}\mu_{jn}} k_{in}k_{jn} + o(1) \right)$$

Then,

$$\left(\frac{X_{1n} - \mu_{1n}}{\sigma_{1n}}, \dots \frac{X_{mn} - \mu_{mn}}{\sigma_{mn}} \right) \xrightarrow[n \to \infty]{d} N(0, \Gamma)$$

* δ_{ij} is Kronecker's delta

Sketch of Proofs

Lemma 1. Let $n=(n(i))_{i\geq 0}$ be a degree statistic and let $T_n \sim \mathrm{Uniform}(\Pi_n^d)$

For $q \in \mathbb{N}$ and $T \in \Pi^{pl}$ s.t. $|n| \geq q|T|-q+1$,

$$\mathbb{E}(N_T(T_n))_q = \frac{|n|}{(|n|)_{q|T|-q+1}} \prod_{i\geq 0} (n(i))_{q n_T(i)}$$

Recall $|n| = \sum_{i\geq 0} n(i)$ and $n_T(i) = |\{v \in T : d_T(v) = i\}|$, $i \geq 0$.

Recall:

Condition. $n_k = (n_k(i))_{i\geq 0}$, $k \geq 1$, are degree statistics s.t. as $k \to \infty$,

$|n_k| \to \infty$ and $p_i(n_k) \to p_i$, where $p = (p_i)_{i\geq 0}$ is a probability distribution on $\mathbb{N}_0$

Theorem 2. Let $T_{n_k} \sim \text{Uniform}(\Pi^d_{n_k})$

- $\mathbb{E}\, N_T(T_{n_k}) = \Pi_p(T)\,|n_k| + o(|n_k|)$
- $\mathrm{Var}\,(N_T(T_{n_k})) = \left(\Pi_p(T) + \eta_p(T)\,\Pi_p(T)^2\right)|n_k| + o(|n_k|)$
- $\dfrac{N_T(T_{n_k}) - \Pi_{p(n_k)}(T)\,|n_k|}{\sqrt{|n_k|}} \xrightarrow[k\to\infty]{d} N\left(0, \Pi_p(T) + \eta_p(T)\,\Pi_p(T)^2\right)$

where $\eta_p(T) = (|T|-1)^2 - \sum_{i\geq 0} \frac{n_T(i)^2}{p_i}$ (here $0/0 = 0$)

Sketch of Proofs (Theorem 2)

By using the estimation, $(x)_k = x^k \exp\left(-\frac{k(k-1)}{2x} + O\left(\frac{k^3}{x^2}\right)\right)$ $x \geq 1$ a real number, $0 \leq k \leq x/2$ an integer,

We see that, for $q = q_k = O(|n_k|^{1/2})$,

$$\mathbb{E}(N_T(\Upsilon_{n_k}))_{q_k} = \frac{|n_k|}{(|n_k|)_{q|T|-q+1}} \prod_{i\geq 0} (n_k(i))_{q n_T(i)}$$

$$= \mu_{n_k}(T)^{q_k} \exp\left(\frac{\gamma_p(T)|n_k| - \mu_{n_k}(T)}{2\mu_{n_k}(T)} q_k^2 + o(1)\right),$$

where

$\mu_{n_k}(T) = |n_k| \pi_{p(n_k)}(T) = |n_k| \prod_{i\geq 0} p_i(n_k)^{n_T(i)}$ and $\gamma_p(T) = \pi_p(T) + \eta_p(T)\pi_p(T)^2$

Finally, we apply Gao-Wormald Theorem.

Thanks