Asymptotic properties of weighted recursive and preferential attachment trees

ArXiv:1904.07115

Delphin Sénizergues

Retreat for Young Researchers in Probability and areas of Application, September 28, 2019

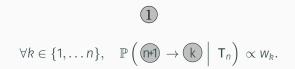
University of British Columbia

Introduction

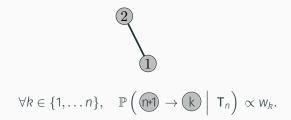
• Start with $(w_n)_{n\geq 1}$ a sequence of weights.

- Start with $(w_n)_{n\geq 1}$ a sequence of weights.
- Construct a sequence $(T_n)_{n\geq 1}$ recursively.

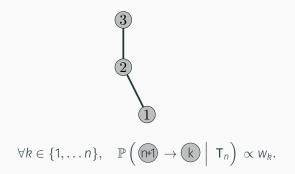
- Start with $(w_n)_{n\geq 1}$ a sequence of weights.
- Construct a sequence $(T_n)_{n\geq 1}$ recursively.



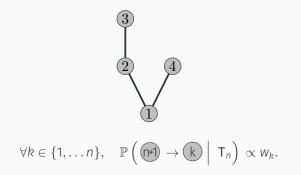
- Start with $(w_n)_{n\geq 1}$ a sequence of weights.
- Construct a sequence $(T_n)_{n\geq 1}$ recursively.



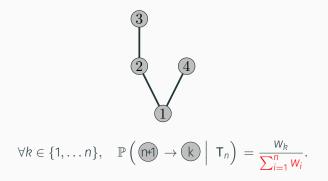
- Start with $(w_n)_{n\geq 1}$ a sequence of weights.
- Construct a sequence $(T_n)_{n\geq 1}$ recursively.



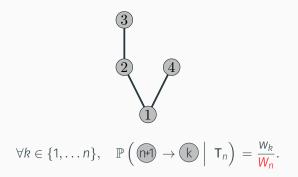
- Start with $(w_n)_{n\geq 1}$ a sequence of weights.
- Construct a sequence $(T_n)_{n\geq 1}$ recursively.



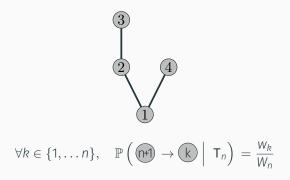
- Start with $(w_n)_{n\geq 1}$ a sequence of weights.
- Construct a sequence $(T_n)_{n\geq 1}$ recursively.



- Start with $(w_n)_{n\geq 1}$ a sequence of weights.
- Construct a sequence $(T_n)_{n\geq 1}$ recursively.

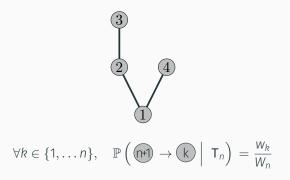


- Start with $(w_n)_{n\geq 1}$ a sequence of weights.
- Construct a sequence $(T_n)_{n\geq 1}$ recursively.



• We write $(T_n)_{n\geq 1} \sim WRT((w_n)_{n\geq 1})$.

- Start with $(w_n)_{n\geq 1}$ a sequence of weights.
- Construct a sequence $(T_n)_{n\geq 1}$ recursively.



- We write $(T_n)_{n\geq 1} \sim WRT((w_n)_{n\geq 1})$.
- We can also use *random* sequences of weights.

 If (w_n)_{n≥1} is constant, then (T_n)_{n≥1} follows the distribution of the Uniform Recursive Tree (URT).

- If $(w_n)_{n\geq 1}$ is constant, then $(T_n)_{n\geq 1}$ follows the distribution of the Uniform Recursive Tree (URT).
- We will work under the assumption that $W_n \simeq \operatorname{cst} \cdot n^{\gamma}$ for some constant $\gamma > 0$ + additional condition on the weights.

- If $(w_n)_{n\geq 1}$ is constant, then $(T_n)_{n\geq 1}$ follows the distribution of the Uniform Recursive Tree (URT).
- We will work under the assumption that $W_n \simeq \operatorname{cst} \cdot n^{\gamma}$ for some constant $\gamma > 0$ + additional condition on the weights.
- One motivation is the connection with another model of growing trees: the preferential attachment trees.

- If (w_n)_{n≥1} is constant, then (T_n)_{n≥1} follows the distribution of the Uniform Recursive Tree (URT).
- We will work under the assumption that $W_n \simeq \operatorname{cst} \cdot n^{\gamma}$ for some constant $\gamma > 0$ + additional condition on the weights.
- One motivation is the connection with another model of growing trees: the preferential attachment trees.
- Also, connection to the monkey walk (Mailler-Uribe 2018+).

• Start with $(a_n)_{n\geq 1}$ a sequence of fitnesses.

- Start with $(a_n)_{n\geq 1}$ a sequence of fitnesses.
- Construct a sequence $(P_n)_{n\geq 1}$ recursively.

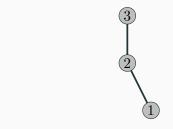
- Start with $(a_n)_{n\geq 1}$ a sequence of fitnesses.
- Construct a sequence $(P_n)_{n\geq 1}$ recursively.

$$\forall k \in \{1, \dots, n\}, \quad \mathbb{P}\left(\left[n+1\right] \to \left[k\right] \mid \mathsf{P}_n\right) \propto a_k + \mathsf{deg}_n^+(\left[k\right]).$$

- Start with $(a_n)_{n\geq 1}$ a sequence of fitnesses.
- Construct a sequence $(P_n)_{n\geq 1}$ recursively.

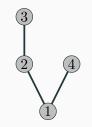
$$\forall k \in \{1, \dots, n\}, \quad \mathbb{P}\left(\left(n+1\right) \to (k) \mid \mathbf{P}_n\right) \propto a_k + \deg_n^+((k)).$$

- Start with $(a_n)_{n\geq 1}$ a sequence of fitnesses.
- Construct a sequence $(P_n)_{n\geq 1}$ recursively.



$$\forall k \in \{1, \dots, n\}, \quad \mathbb{P}\left(\left(n+1\right) \to k \mid \mathbf{P}_n\right) \propto a_k + \deg_n^+(k).$$

- Start with $(a_n)_{n\geq 1}$ a sequence of fitnesses.
- Construct a sequence $(P_n)_{n\geq 1}$ recursively.

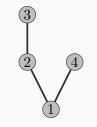


$$\forall k \in \{1, \dots n\}, \quad \mathbb{P}\left(\left(\mathbb{n}^{+}\right) \to \mathbb{k} \mid \mathsf{P}_n\right) \propto a_k + \mathsf{deg}_n^+(\mathbb{k}).$$

- Start with $(a_n)_{n\geq 1}$ a sequence of fitnesses.
- Construct a sequence $(P_n)_{n\geq 1}$ recursively.

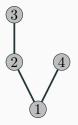
$$\forall k \in \{1, \dots, n\}, \quad \mathbb{P}\left(\texttt{P}^{+}\right) \to \texttt{k} \mid \mathsf{P}_{n}\right) = \frac{a_{k} + \mathsf{deg}_{n}^{+}(\texttt{k})}{\sum_{i=1}^{n} a_{i} + \sum_{i=1}^{n} \mathsf{deg}_{n}^{+}(\texttt{i})}.$$

- Start with $(a_n)_{n\geq 1}$ a sequence of fitnesses.
- Construct a sequence $(P_n)_{n\geq 1}$ recursively.



$$\forall k \in \{1, \dots, n\}, \quad \mathbb{P}\left(\left(n+1\right) \to \left(k\right) \mid \mathbf{P}_n\right) = \frac{a_k + \deg_n^+(k)}{A_n + n - 1}$$

- Start with $(a_n)_{n\geq 1}$ a sequence of fitnesses.
- Construct a sequence $(P_n)_{n\geq 1}$ recursively.



$$\forall k \in \{1, \dots, n\}, \quad \mathbb{P}\left(\text{net} \to \mathbb{k} \mid \mathbf{P}_n \right) = \frac{a_k + \deg_n^+(\mathbb{k})}{A_n + n - 1}$$

• We write $(P_n)_{n\geq 1} \sim PA((a_n)_{n\geq 1})$.

Theorem (S. 2019+)

Preferential attachment trees are weighted recursive trees i.e. for any sequence of fitnesses $\mathbf{a} = (a_n)_{n \ge 1}$, there exists a random sequence $(\mathbf{w}_n^{\mathbf{a}})_{n \ge 1}$ such that the distributions $\mathsf{PA}((a_n)_{n \ge 1})$ and $\mathsf{WRT}((\mathbf{w}_n^{\mathbf{a}})_{n \ge 1})$ coincide.

Theorem (S. 2019+)

Preferential attachment trees are weighted recursive trees i.e. for any sequence of fitnesses $\mathbf{a} = (a_n)_{n \ge 1}$, there exists a random sequence $(\mathbf{w}_n^{\mathbf{a}})_{n \ge 1}$ such that the distributions $\mathsf{PA}((a_n)_{n \ge 1})$ and $\mathsf{WRT}((\mathbf{w}_n^{\mathbf{a}})_{n \ge 1})$ coincide.

• If $A_n \simeq c \cdot n$ as $n \to \infty$ then the sequence $(w_n^a)_{n \ge 1}$ satisfies $W_n^a \simeq \operatorname{cst} \cdot n^{\gamma}$ a.s. with $\gamma = \frac{c}{c+1}$.

Theorem (S. 2019+)

Preferential attachment trees are weighted recursive trees i.e. for any sequence of fitnesses $\mathbf{a} = (a_n)_{n \ge 1}$, there exists a random sequence $(\mathbf{w}_n^{\mathbf{a}})_{n \ge 1}$ such that the distributions $\mathsf{PA}((a_n)_{n \ge 1})$ and $\mathsf{WRT}((\mathbf{w}_n^{\mathbf{a}})_{n \ge 1})$ coincide.

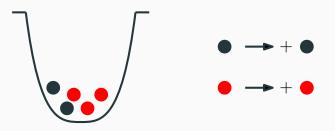
- If $A_n \simeq \mathbf{c} \cdot n$ as $n \to \infty$ then the sequence $(\mathbf{w}_n^{\mathbf{a}})_{n \ge 1}$ satisfies $W_n^{\mathbf{a}} \simeq \operatorname{cst} \cdot n^{\gamma}$ a.s. with $\gamma = \frac{c}{c+1}$.
- Results for weighted recursive trees automatically apply to preferential attachment trees!

For any sequence $\mathbf{a} = (a_n)_{n \ge 1}$ of fitnesses, the random sequence $(\mathbf{w}_n^{\mathbf{a}})_{n > 1}$ is defined as

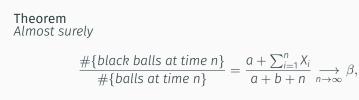
$$W_1^a = W_1^a = 1$$
 and $\forall n \ge 2$, $W_n^a = \prod_{k=1}^{n-1} \beta_k^{-1}$,

where the $(\beta_k)_{k\geq 1}$ are independent with respective distribution Beta $(A_k + k, a_{k+1})$, with $A_k = \sum_{i=1}^k a_i$.

The classical Pólya urn



- At time 0, the urn contains *a* black balls and *b* red balls.
- At each time $n \ge 1$, a ball is drawn from the urn and returned to the urn together with a new ball of the same colour.
- For all $n \ge 1$, we let $X_n := \mathbf{1}_{\{\text{the ball drawn at time } n \text{ is black}\}}$.



where β is a random variable with distribution Beta(a, b).

Theorem Almost surely

$$\frac{\#\{\text{black balls at time } n\}}{\#\{\text{balls at time } n\}} = \frac{a + \sum_{i=1}^{n} X_i}{a + b + n} \xrightarrow[n \to \infty]{} \beta,$$

where β is a random variable with distribution Beta(a, b). Furthermore, conditionally on β , the sequence $(X_n)_{n\geq 1}$ is a sequence of i.i.d. Bernoulli random variables with parameter β .

Convergence results for weighted recursive trees

Proposition (S. 2019+) If $W_n \sim C \cdot n^{\gamma}$ as $n \to \infty$ for $0 < \gamma <$ 1, then we almost surely have

$$n^{-(1-\gamma)} \cdot (\deg_n^+(1)), \deg_n^+(2)), \ldots) \xrightarrow[n \to \infty]{} \frac{1}{C(1-\gamma)} \cdot (W_1, W_2, \ldots),$$

for the product topology.

Proposition (S. 2019+) If $W_n \sim C \cdot n^{\gamma}$ as $n \to \infty$ for $0 < \gamma < 1$, then we almost surely have

$$n^{-(1-\gamma)} \cdot (\deg_n^+(1), \deg_n^+(2), \dots) \xrightarrow[n \to \infty]{} \frac{1}{C(1-\gamma)} \cdot (W_1, W_2, \dots),$$

for the product topology.

• Under some additional condition on $(w_n)_{n\geq 1}$, the last convergence can be proved to hold in ℓ^p for $p > \frac{1}{1-\gamma}$.

Proposition (S. 2019+) If $W_n \sim C \cdot n^{\gamma}$ as $n \to \infty$ for $0 < \gamma < 1$, then we almost surely have

$$n^{-(1-\gamma)} \cdot (\deg_n^+(1), \deg_n^+(2), \dots) \xrightarrow[n \to \infty]{} \frac{1}{C(1-\gamma)} \cdot (W_1, W_2, \dots),$$

for the product topology.

- Under some additional condition on $(w_n)_{n\geq 1}$, the last convergence can be proved to hold in ℓ^p for $p > \frac{1}{1-\gamma}$.
- Thanks to the representation theorem, this also holds for preferential attachment trees.

CONVERGENCE OF DEGREES: ELEMENTS OF PROOFS

For any $k \ge 1$ we can write:

$$\operatorname{deg}_{n}^{+}(\overline{\mathbb{k}}) = \sum_{i=k+1}^{n} \mathbf{1}_{\left\{ i \to k \right\}}.$$

For any $k \ge 1$ we can write:

$$\operatorname{deg}_{n}^{+}(\overline{\mathbb{k}}) = \sum_{i=k+1}^{n} \mathbf{1}_{\left\{ i \to k \right\}}.$$

The events $\{(i) \rightarrow (k)\}$ for $i \in \{k + 1, k + 2, ..., n\}$ are *independent* and have respective probability $\frac{W_k}{W_{i-1}}$.

For any $k \ge 1$ we can write:

$$\deg_n^+(\mathbf{k}) = \sum_{i=k+1}^n \mathbf{1}_{\left\{\mathbf{i}\}\to\mathbf{k}\right\}}.$$

The events $\{i \to k\}$ for $i \in \{k + 1, k + 2, ..., n\}$ are independent and have respective probability $\frac{w_k}{W_{i-1}}$. Hence by a law of large numbers, with probability 1 we have

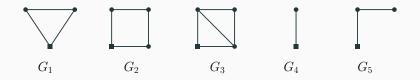
$$\deg_n^+(\mathbb{k}) \underset{n \to \infty}{\sim} \sum_{i=k+1}^n \mathbb{P}\left((\mathbb{i}) \to \mathbb{k}\right) \underset{n \to \infty}{\sim} \sum_{i=k+1}^n \frac{W_k}{W_{i-1}} \underset{n \to \infty}{\sim} W_k \cdot \sum_{i=k+1}^n \frac{1}{C \cdot i^{\gamma}} \underset{n \to \infty}{\overset{\sim}{\sim}} W_k \cdot \frac{n^{1-\gamma}}{C(1-\gamma)}.$$

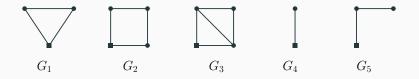
If $(\mathbf{P}_n)_{n\geq 1}$ is a sequence of trees with distribution $\mathsf{PA}((a_n)_{n\geq 1})$ with $A_n \simeq \mathbf{c} \cdot n$ as $n \to \infty$ then we have the almost sure convergence in some ℓ^p space,

$$n^{-\frac{1}{c+1}} \cdot (\operatorname{deg}_n^+(1), \operatorname{deg}_n^+(2), \dots) \xrightarrow[n \to \infty]{} (m_1, m_2, \dots),$$

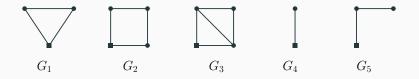
where $(m_n)_{n\geq 1}$ is a constant times the sequence $(w_n^a)_{n\geq 1}$.

Scaling limits for generalisation of Rémy's algorithm

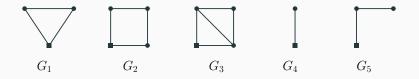


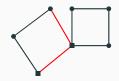


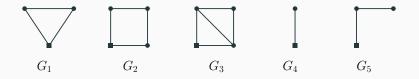


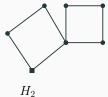




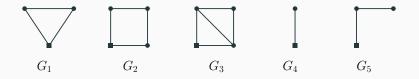


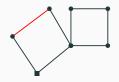


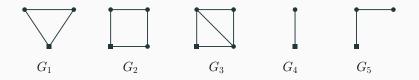


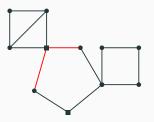


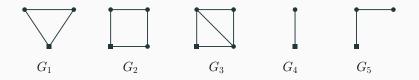


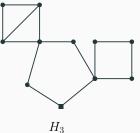




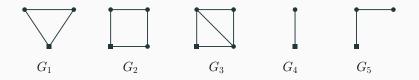


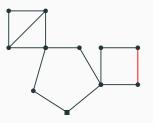


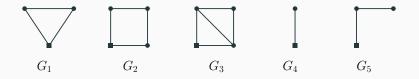




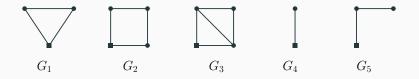


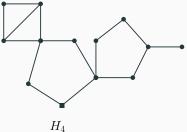


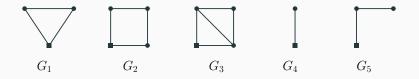




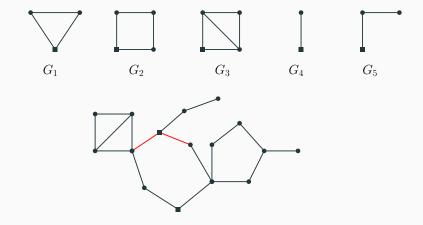


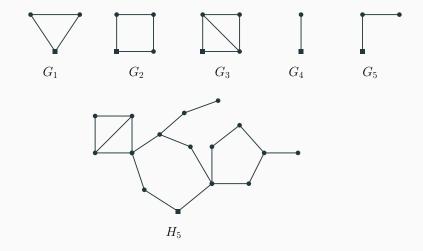












$$(H_n, n^{-1/2} \cdot \mathsf{d}_{gr}) \xrightarrow[n \to \infty]{} \mathcal{T},$$

where ${\cal T}$ is Aldous' CRT.

$$(H_n, n^{-1/2} \cdot \mathsf{d}_{gr}) \underset{n \to \infty}{\to} \mathcal{T},$$

where ${\cal T}$ is Aldous' CRT.

• How does the diameter of H_n grow as $n \to \infty$ for more general sequences $(G_n)_{n \ge 1}$?

$$(H_n, n^{-1/2} \cdot \mathsf{d}_{gr}) \underset{n \to \infty}{\to} \mathcal{T},$$

where ${\cal T}$ is Aldous' CRT.

- How does the diameter of H_n grow as $n \to \infty$ for more general sequences $(G_n)_{n \ge 1}$?
- Do we get a scaling limit?

$$(H_n, n^{-1/2} \cdot \mathsf{d}_{gr}) \underset{n \to \infty}{\to} \mathcal{T},$$

where ${\cal T}$ is Aldous' CRT.

- How does the diameter of H_n grow as $n \to \infty$ for more general sequences $(G_n)_{n \ge 1}$?
- Do we get a scaling limit?
- If yes, what does it look like?

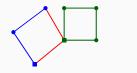








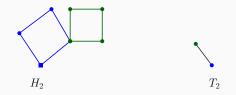




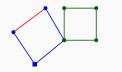


13

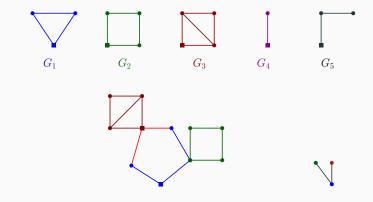


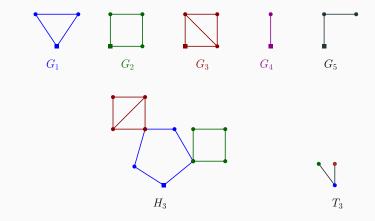


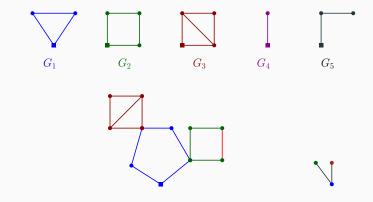


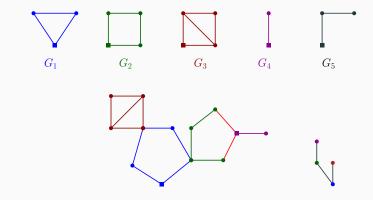


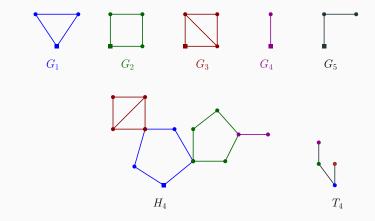


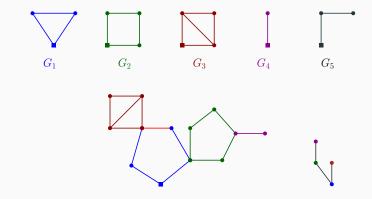


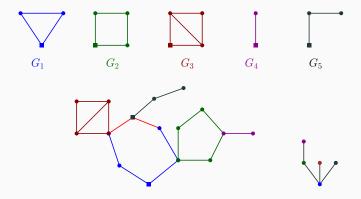




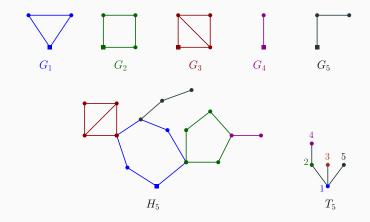




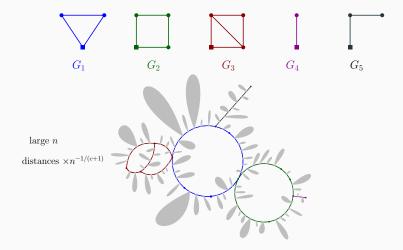




A GENERALISED VERSION OF RÉMY'S ALGORITHM



A GENERALISED VERSION OF RÉMY'S ALGORITHM



Theorem (S. 2019+) Suppose that $\sum_{i=1}^{n} a_i = c \cdot n + O(n^{1-\epsilon})$ for some constant c > 0 and $a_n \le n^{\frac{1}{c+1}-\epsilon+o(1)}$. Then we have the following convergence

$$\left(H_n, n^{-\frac{1}{c+1}} \cdot \mathrm{d}_{\mathrm{gr}}, \mu_{n,\mathrm{unif}}\right) \underset{n \to \infty}{\longrightarrow} \left(\mathcal{H}, \mathsf{d}, \mu\right),$$

almost surely in Gromov-Hausdorff-Prokhorov topology.

Theorem (S. 2019+) Suppose that $\sum_{i=1}^{n} a_i = c \cdot n + O(n^{1-\epsilon})$ for some constant c > 0 and $a_n \le n^{\frac{1}{c+1}-\epsilon+o(1)}$. Then we have the following convergence

$$\left(H_n, n^{-\frac{1}{c+1}} \cdot \mathrm{d}_{\mathrm{gr}}, \mu_{n,\mathrm{unif}}\right) \xrightarrow[n \to \infty]{} (\mathcal{H}, \mathsf{d}, \mu),$$

almost surely in Gromov-Hausdorff-Prokhorov topology.

• Convergence of degrees in the tree \rightarrow convergence of each coloured portion of the graph.

Theorem (S. 2019+) Suppose that $\sum_{i=1}^{n} a_i = c \cdot n + O(n^{1-\epsilon})$ for some constant c > 0 and $a_n \le n^{\frac{1}{c+1}-\epsilon+o(1)}$. Then we have the following convergence

$$\left(H_n, n^{-\frac{1}{c+1}} \cdot \mathrm{d}_{\mathrm{gr}}, \mu_{n,\mathrm{unif}}\right) \underset{n \to \infty}{\longrightarrow} (\mathcal{H}, \mathsf{d}, \mu),$$

almost surely in Gromov-Hausdorff-Prokhorov topology.

- Convergence of degrees in the tree \rightarrow convergence of each coloured portion of the graph.
- Controlling the degrees in the tree \rightarrow controlling distances in H_n .

Theorem (S. 2019+) Suppose that $\sum_{i=1}^{n} a_i = c \cdot n + O(n^{1-\epsilon})$ for some constant c > 0 and $a_n \le n^{\frac{1}{c+1}-\epsilon+o(1)}$. Then we have the following convergence

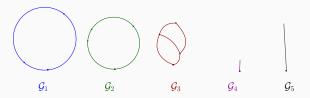
$$\left(H_n, n^{-\frac{1}{c+1}} \cdot \mathrm{d}_{\mathrm{gr}}, \mu_{n,\mathrm{unif}}\right) \underset{n \to \infty}{\longrightarrow} (\mathcal{H}, \mathsf{d}, \mu),$$

almost surely in Gromov-Hausdorff-Prokhorov topology.

- Convergence of degrees in the tree \rightarrow convergence of each coloured portion of the graph.
- Controlling the degrees in the tree \rightarrow controlling distances in H_n .
- $\cdot\,$ Description of the tree as a WRT \rightarrow iterative gluing construction.



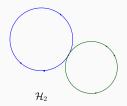




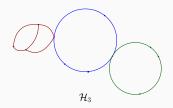


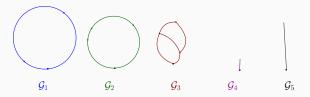
 \mathcal{H}_1

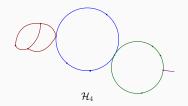


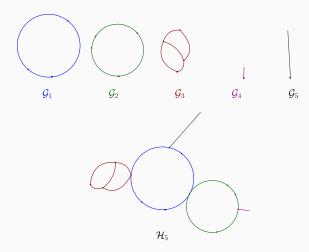




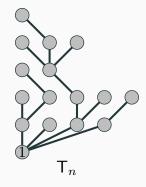


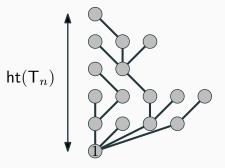


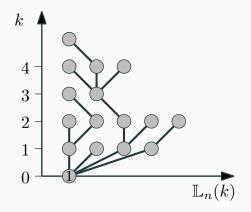




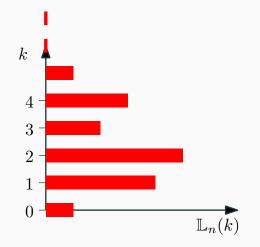
Thank you for your attention!





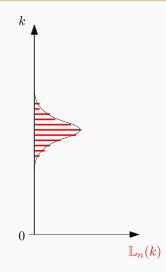


The profile of the tree T_n is the function $\mathbb{L}_n : \mathbb{N} \to \mathbb{N}$ defined as $\forall k \ge 0, \quad \mathbb{L}_n(k) = \# \{ \text{vertices at height } k \text{ in the tree } T_n \}.$

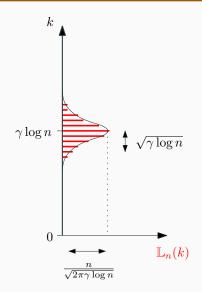


The profile of the tree T_n is the function $\mathbb{L}_n : \mathbb{N} \to \mathbb{N}$ defined as $\forall k \ge 0, \quad \mathbb{L}_n(k) = \# \{ \text{vertices at height } k \text{ in the tree } T_n \}.$

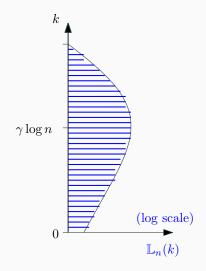
THE PROFILE IS ALMOST SURELY ASYMPTOTICALLY GAUSSIAN



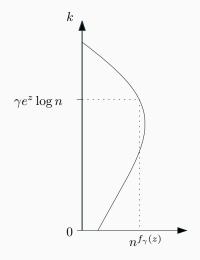
THE PROFILE IS ALMOST SURELY ASYMPTOTICALLY GAUSSIAN



THE LOG OF THE PROFILE CONVERGES TO A FUNCTION

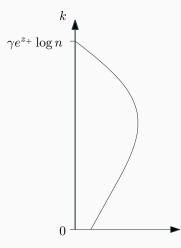


THE LOG OF THE PROFILE CONVERGES TO A FUNCTION



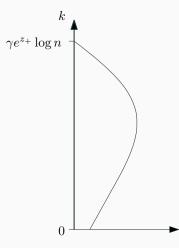
$$\cdot f_{\gamma}(z) := 1 + \gamma(e^z - 1 - ze^z).$$

THE LOG OF THE PROFILE CONVERGES TO A FUNCTION



- $\cdot f_{\gamma}(z) := 1 + \gamma(e^z 1 ze^z).$
- z_+ is the unique positive zero of f_{γ} .

The log of the profile converges to a function



- $f_{\gamma}(z) := 1 + \gamma(e^z 1 ze^z).$
- z_+ is the unique positive zero of f_{γ} .
- Almost surely $ht(T_n) \underset{n \to \infty}{\sim} \gamma \cdot e^{z_+} \log n.$

Theorem (S. 2019+) Under the assumption $W_n \simeq \operatorname{cst} \cdot n^{\gamma}$ (+ additional condition) we almost surely have

$$\cdot \frac{\operatorname{ht}(\mathsf{T}_n)}{\log n} \xrightarrow[n \to \infty]{} \gamma \cdot e^{Z_+}.$$

$$\cdot \mathbb{L}_n(k) \underset{n \to \infty}{=} \frac{n}{\sqrt{2\pi\gamma \log n}} \exp\left\{-\frac{1}{2} \cdot \left(\frac{k - \gamma \log n}{\sqrt{\gamma \log n}}\right)^2\right\} + O\left(\frac{n}{\log n}\right),$$

$$\cdot \text{ for all } z \in (z_-, z_+), \quad \mathbb{L}_n(|\gamma e^{Z} \log n|) = n^{f_\gamma(Z) + O(1)}.$$