

Atypical decay rates for atypical heights in RRT

Heng Ma 馬恒 (Technion)

Joint work (in progress) with

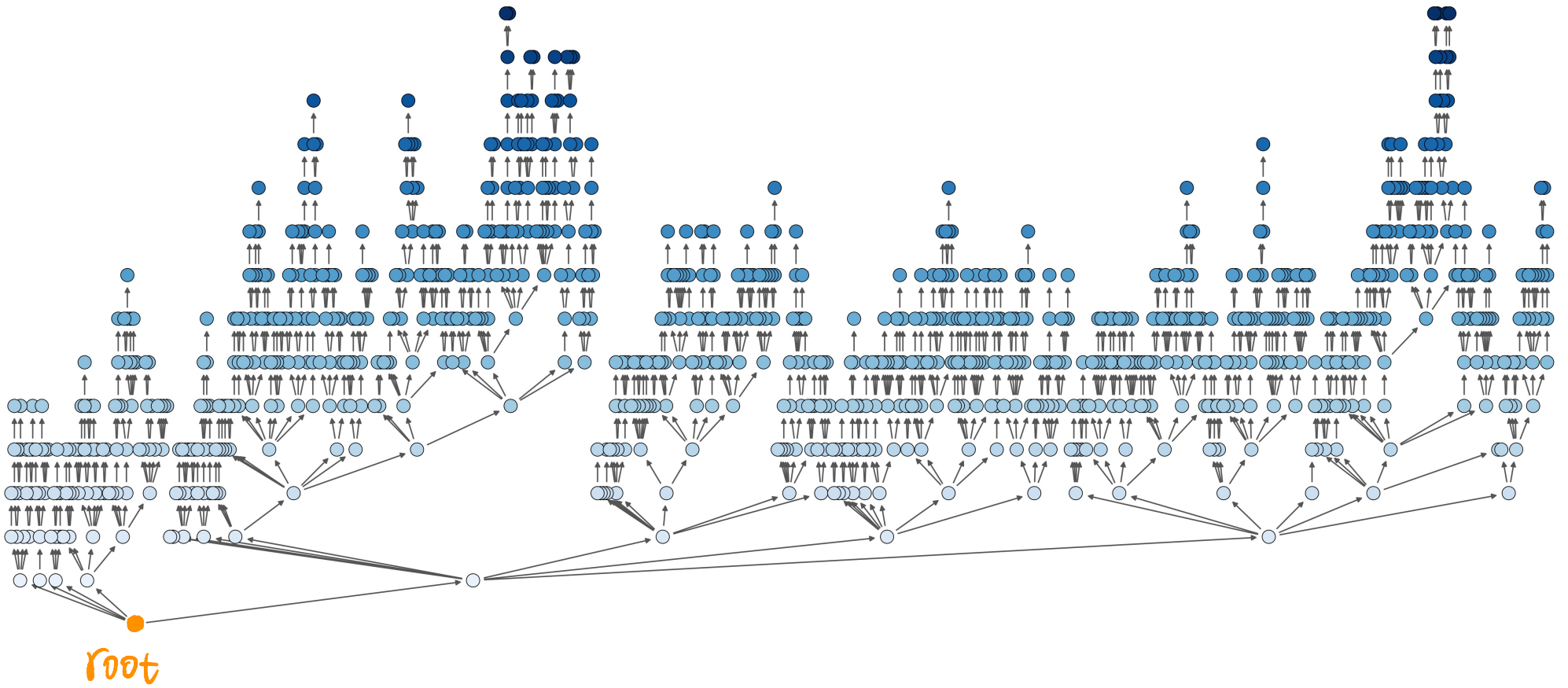
Xinxin Chen 陈昕昕

(Beijing Normal University)

Random recursive tree (RRT) T_n

- Start with a root v_1
- At each step $n \geq 2$, introduce a new vertex v_n .
- We attach v_n to an existing vertex, chosen **uniformly at random** from $\{v_1 \dots v_{n-1}\}$.

Also known as the "Uniform attachment tree".



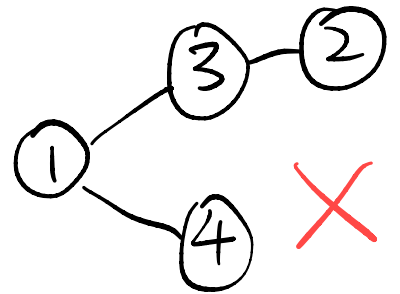
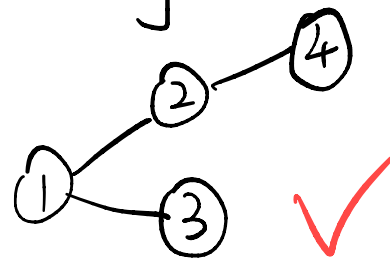
RRT on 1000 nodes

As uniform distribution on increasing trees

- A tree on $\{v_1, \dots, v_n\}$ is called an **increasing tree** if

1) The root has smallest label v_1

2) The label along any path starting from the root is increasing



- For any increasing tree τ ,

$$\mathbb{P}(\mathcal{T}_n = \tau) = \frac{1}{(n-1)!}$$

Remark (General attachment rules)

- Start with v_1 .
- At each step $n \geq 2$, introduce a new vertex v_n .
- Attach v_n to an existing vertex $v_{I_n} \in \{v_1, \dots, v_{n-1}\}$, where

$$P(I_n = j \mid \mathcal{T}_{n-1}) \propto \theta \cdot \deg_{\mathcal{T}_{n-1}}^+(v_j) + 1,$$

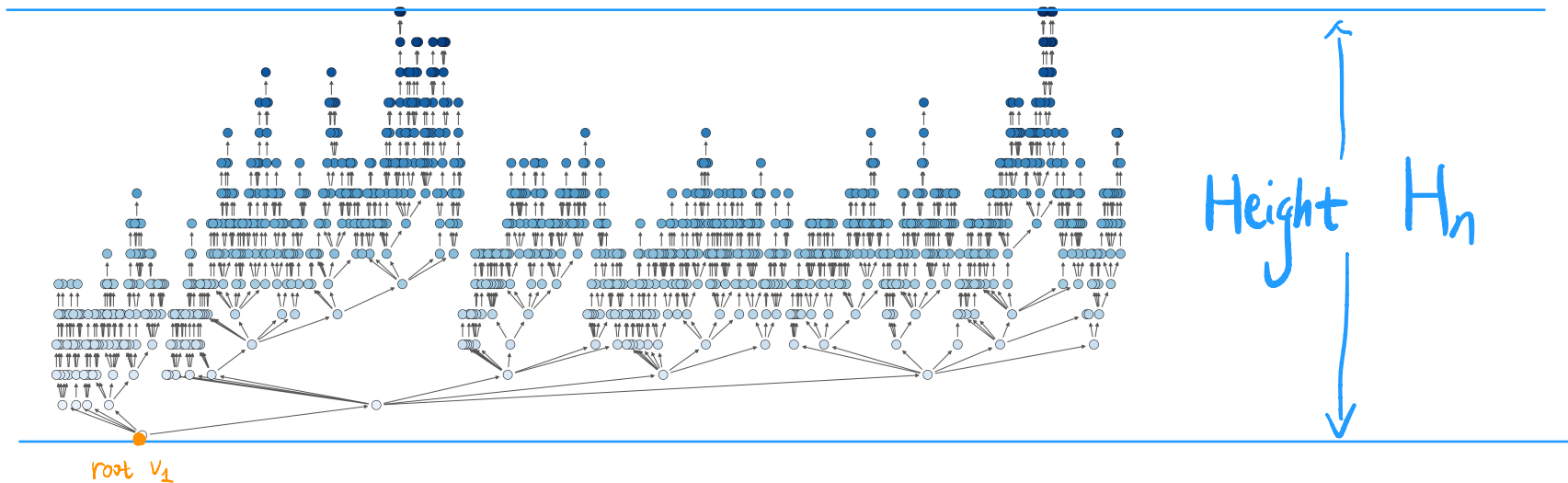
$\theta = 0$	RRT	Out degree of v_j Plane Oriented Recursive Tree
$\theta = 1$	preferential attachment tree	
$\theta = -\frac{1}{2}$	Binary Search tree	

Typical Behaviors
of RRT

The Height of RRT

- Denote by H_n the height of the tree \mathcal{T}_n :

$$H_n := \max_{1 \leq j \leq n} d(v_1, v_j)$$



- Szymanski '90 : with high prob. $1 \leq \frac{H_n}{\log n} \leq e$.
- Pittel '94 proved that $\lim_{n \rightarrow \infty} \frac{H_n}{\log n} = e$ almost surely,
by using a Continuous-time Embedding method
- Addario-Berry - Ford '13 :

$$\left\{ \begin{array}{l} \mathbb{E} H_n = e \log n - \frac{3}{2} \log \log n + O(1) \\ (H_n - \mathbb{E} H_n)_{n \geq 1} \text{ is tight : } \forall \lambda \in (0, \frac{1}{2e}), n \geq 1 \end{array} \right.$$

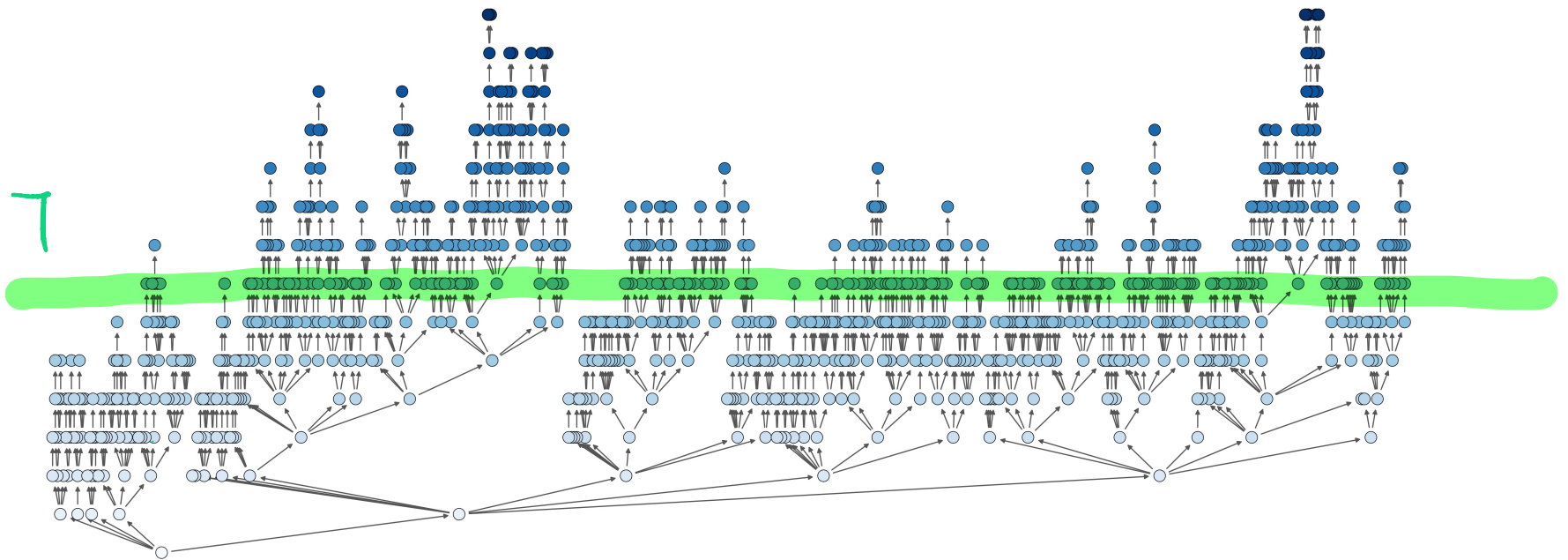
$$\mathbb{P}(|H_n - \mathbb{E} H_n| > k) \lesssim_{\lambda} e^{-\lambda k}$$

Level Sets of RRT

- Let $X_{n,k} :=$ number of nodes at level k in \mathcal{T}_n
$$= \sum_{j=1}^n \mathbb{1}_{\{d(v_1, v_j) = k\}}$$

$(X_{n,k})_{k \geq 1}$ is also called the profile of \mathcal{T}_n in the literature.

Level 7



Fuchs-Hwang-Neininger '06:

- For $0 \leq \alpha < e$, if $k = k(n)$ satisfies $\frac{k}{\log n} \rightarrow \alpha$ then

$$\mathbb{E}X_{n,k} = n^{\alpha(1-\log \alpha) + o(1)}, \quad \frac{X_{n,k}}{\mathbb{E}X_{n,k}} \xrightarrow{d} X_\alpha$$

where X_α satisfies

$$X \stackrel{d}{=} \alpha U^\alpha X + (1-U)^\alpha \tilde{X}$$

\tilde{X} : an i.i.d.

copy of X

$U \sim \text{Unif}[0,1]$

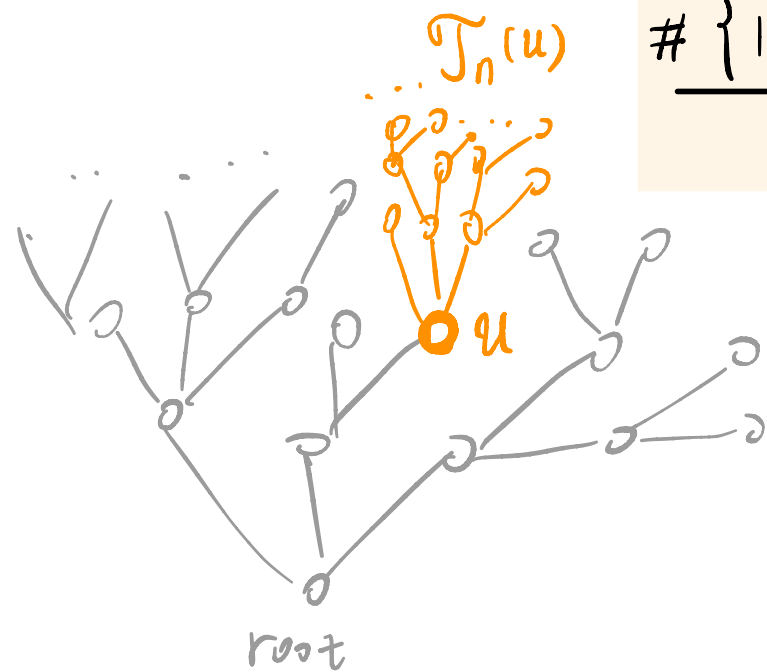
- CLT when $\alpha = 0$: If $k = k(n)$ satisfies $\frac{k}{\log n} \rightarrow 0$ then

$$\frac{(k-1)! \sqrt{2k-1}}{(\log n)^{k-\frac{1}{2}}} \left[X_{n,k} - \frac{(\log n)^k}{k!} \right] \xrightarrow[\eta \rightarrow +\infty]{d} \text{Normal}(0, 1)$$

Limiting Fringe Tree of RRT

Aldous '91: Uniformly choose a node u

in \mathcal{T}_n . What does the subtree of u look like?



$$\frac{\#\{1 \leq j \leq n : \mathcal{T}_n(u) \sqsubseteq \tau\}}{n} \xrightarrow{\text{a.s.}} P(\mathcal{T}_M \sqsubseteq \tau)$$

where M is a R.V. independent of (\mathcal{T}_n) , with $P(M \geq n) = \frac{1}{n}$.

Degree Profile of RRT.

As a consequence,

$$\frac{\#\{1 \leq j \leq n : \deg_{T_n}(v_j) = k\}}{n} \xrightarrow[n \rightarrow +\infty]{\text{a.s.}} 2^{-k}$$

In particular,

proportion of leaves \longrightarrow 50%

Our focus.

Rare events concerning
RRT height

?

RRT being atypically short

Fix $\alpha \in (0, 1)$, what is the decay rate of

$$\mathbb{P}(H_n \leq \alpha e \log n)$$

?

RRT being atypically tall

Fix $\beta > 1$, what is the decay rate of

$$\mathbb{P}(H_n \geq \beta e \log n)$$

Theorem 1 (Chen-M. '26+)

Define $m_\alpha(n) := n^{1-\alpha} (\log n)^{-\frac{3}{2e}}$ and $w_\alpha(n)$ as follows

$$\mathbb{P}(H_n \geq \alpha e \ln n) = e^{-w_\alpha(n) m_\alpha(n)}$$

Then for $\alpha < 1$, $\liminf_{n \rightarrow +\infty} w_\alpha(n) = +\infty$

Moreover, for any integer $k \in \mathbb{N}$,

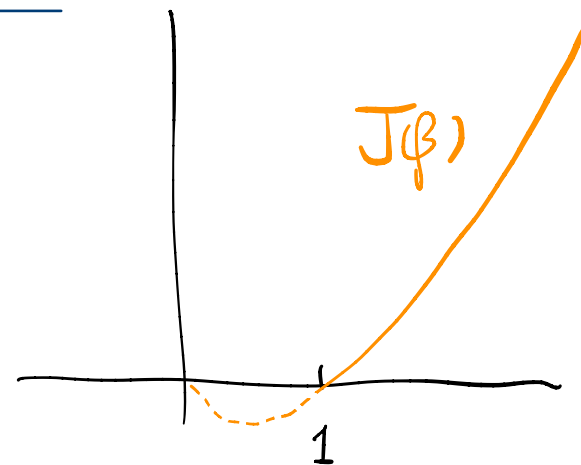
$$\limsup_{n \rightarrow +\infty} \frac{w_\alpha(n)}{\log^{(k)}(n)} < +\infty.$$

$$\begin{aligned} & \log^{(k)}(x) \\ & := \log^{(k-1)}(\log x) \end{aligned}$$

Theorem 2 (Chen-M. '26+)

- Define $J(\beta) := e \cdot \beta \ln \beta$

Then for $\beta > 1$, we have



$$\mathbb{P}(H_n > \beta e \log(n))$$

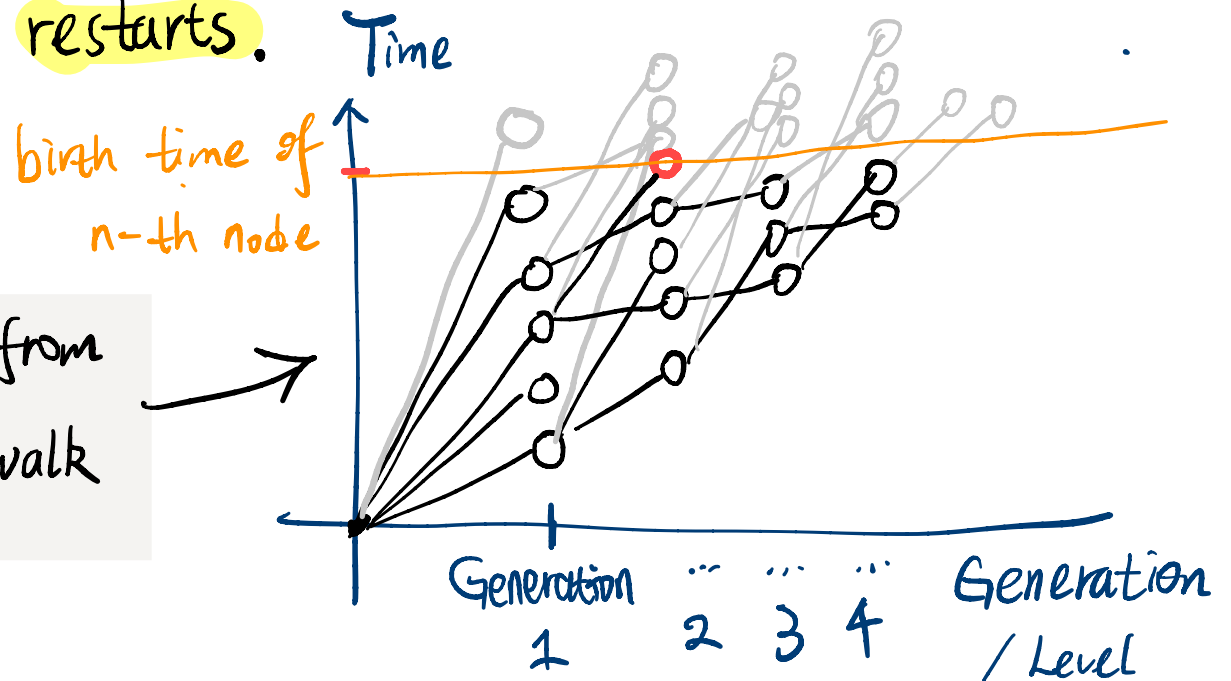
$$= n^{-J(\beta) + o(1)} = e^{-[J(\beta) + o(1)] \log n}$$

A powerful approach

Pittel's Continuous-time embedding

- Each node has an **independent** alarm clock, **exponentially** distributed with rate 1
- When it rings, this node give birth to a new child and its clock **restarts**.

Extracting a RRT from a branching random walk



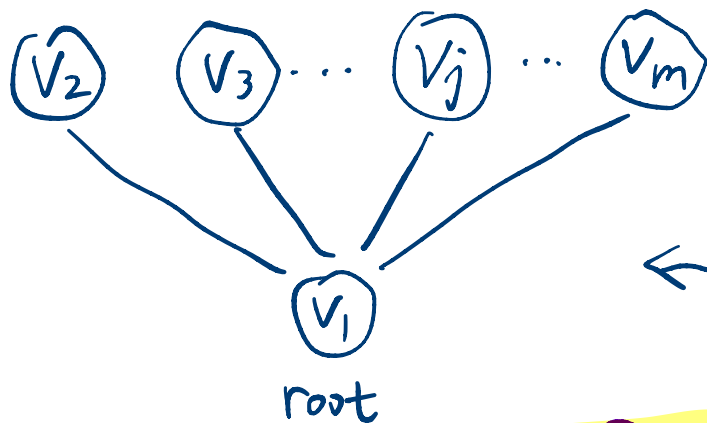
Proof ideas of Thm 1

Strategy to make

the tree atypically short

A crude strategy:

Make the tree as short as possible at an early stage, then let it grow "freely".

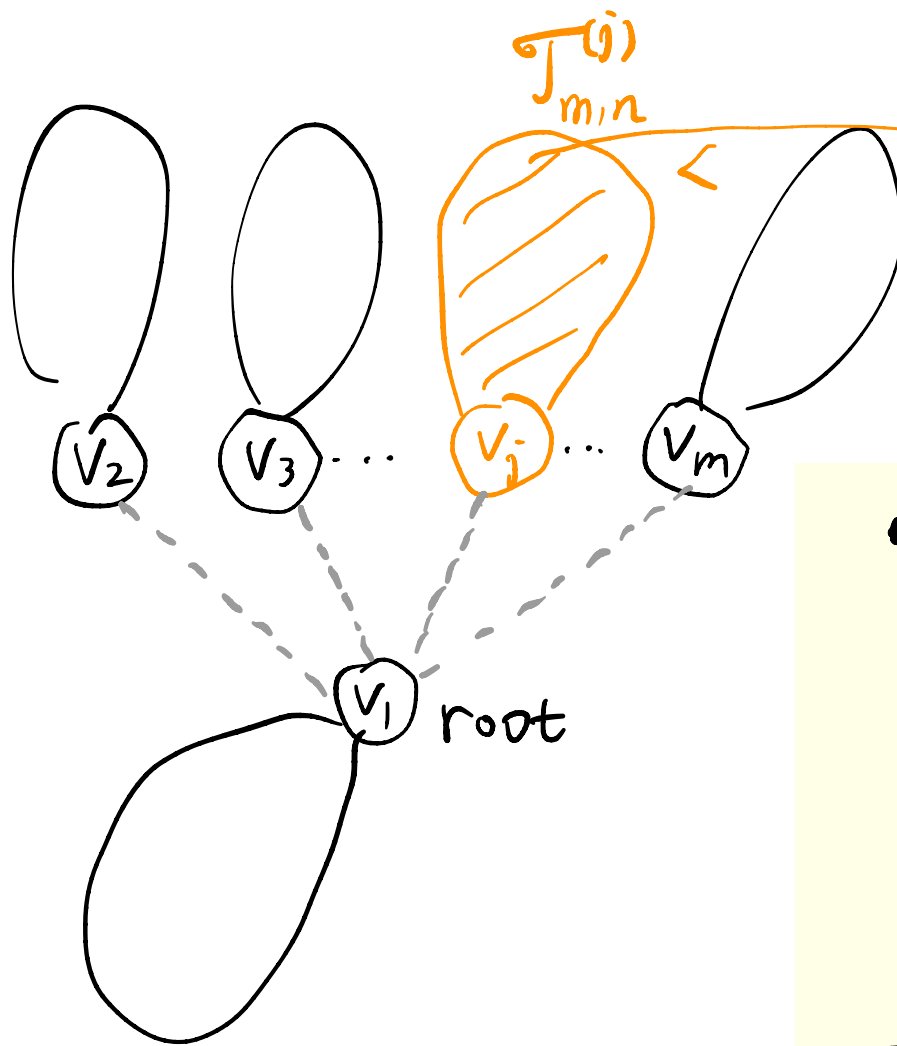


That is, fix $m \geq 2$,
we ask $H(\mathcal{T}_m) = 1$

Cost: $\mathbb{P}(H(\mathcal{T}_m) = 1)$

$$= \frac{1}{(m-1)!} = e^{-m(\ln m + O(m))}$$

Make the tree as short as possible at an early stage, then let it grow "freely".



- For each $z \leq j \leq m$, let $T_{m,n}^{(j)}$ denote the tree containing V_j after removing edges in T_m .

• Then

$$\begin{cases} H(T_m) = 1 \\ \max_{1 \leq j \leq m} H(T_{m,n}^{(j)}) \leq 2e \log n - 1 \end{cases}$$

$$\Rightarrow H_n := H(T_n) \leq 2e \log n$$

Observation 1: Conditionally on $\{\mathcal{T}_m = \tau, |\mathcal{T}_{m,n}^{(j)}| = n_j, 1 \leq j \leq m\}$

$\mathcal{T}_{m,n}^{(j)}$ are independent RRT on n_j nodes.

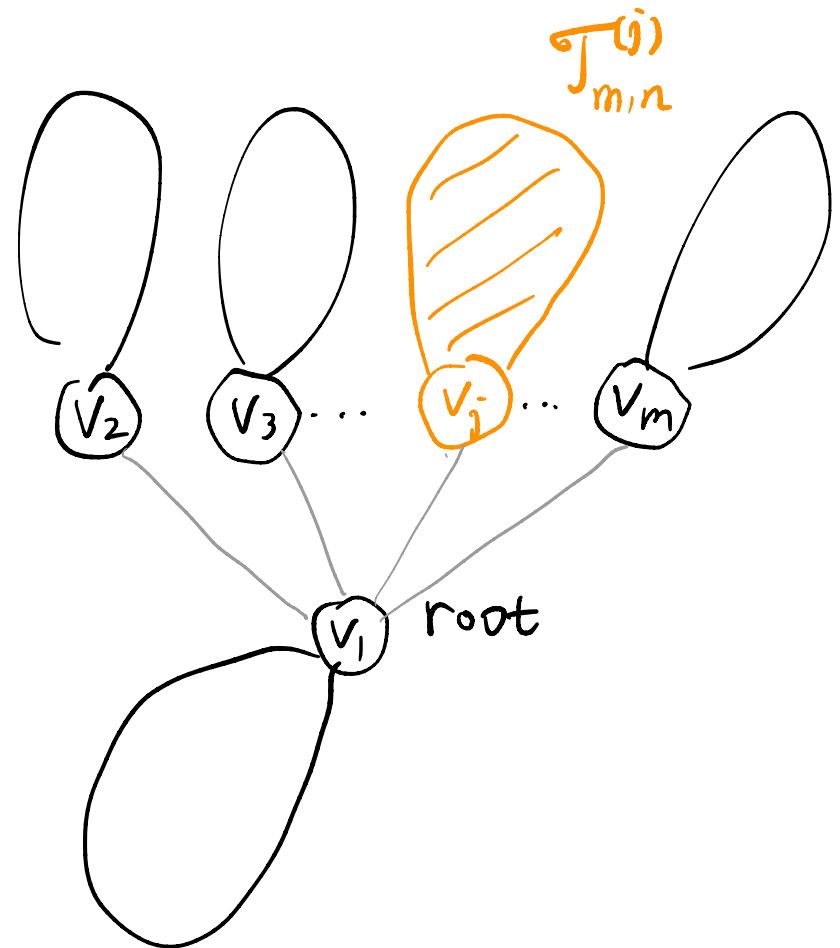
Observation 2:

$(|\mathcal{T}_{m,n}^{(j)}| : 1 \leq j \leq m)$

has the Uniform distribution on

$\left\{ (n_j)_{j=1}^m : n_j \in \mathbb{N}; \sum_{j=1}^m n_j = n \right\}$.

In particular, $\mathbb{E} |\mathcal{T}_{m,n}^{(j)}| = \frac{n}{m}$.

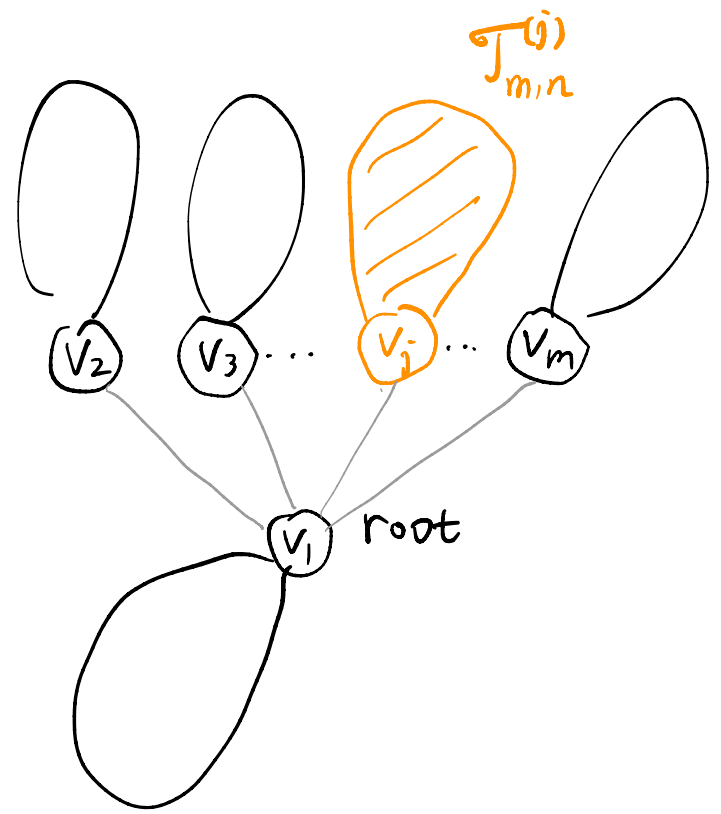


- Need a lower bound for

$$\mathbb{P} \left(\max_{1 \leq j \leq m} H(\mathcal{T}_{m,n}^{(j)}) \leq 2e \log n - 1 \right)$$

- Suppose $|\mathcal{T}_{m,n}^{(j)}| \leq 100 \frac{n}{m}$, then

$$H(\mathcal{T}_{m,n}^{(j)}) \stackrel{\text{st}}{\leq} H_{100 \frac{n}{m}} = \mathbb{E}[H_{\frac{n}{m}}] + O_{\mathbb{P}}(1)$$



- Choose m carefully so that

$$\mathbb{E}[H_{\frac{n}{m}}] = 2e \log n - \text{large constant}$$

and hence

$$\mathbb{P} \left(H_{100 \frac{n}{m}} \leq 2e \log n - 1 \right) \geq \frac{1}{2}$$

$$m = C_{\text{st}} \times m_2(n)$$



All together,

$$\mathbb{P}(H_n \leq 2e \ln n) \geq \underbrace{\mathbb{P}(H(J_m) = 1)}_{\frac{1}{(m-1)!}} = e^{-m/\ln m + O(m)}$$

$$\times \underbrace{\mathbb{P}\left(|J_{m,n}^{(j)}| \leq 100 \frac{n}{m} \quad \forall 1 \leq j \leq m\right)}$$

One can show it's greater than $\frac{e^{-cm}}{m}$

$$\times \underbrace{\mathbb{P}\left(H_{100 \frac{n}{m}} \leq 2e \log n - 1\right)}$$

by our choice of m , it is greater than 2^{-m}

$$\geq e^{-cm/\ln m} \geq e^{-cm_2(n) \log n}$$

□

All together,

$$\mathbb{P}(H_n \leq 2e \ln n) \geq \underbrace{\mathbb{P}(H(T_m) = 1)}_{\text{circled in red}} \frac{1}{(m-1)!} = e^{-m/\ln m + O(m)}$$

$$\times \mathbb{P}\left(\left|T_{m,n}^{(j)}\right| \leq \log \frac{n}{m} \quad \forall 1 \leq j \leq m\right)$$

One can show it's greater than $\frac{e}{m}$

$$\times \mathbb{P}\left(H_{\log \frac{n}{m}} \leq 2e \log n - 1\right)$$

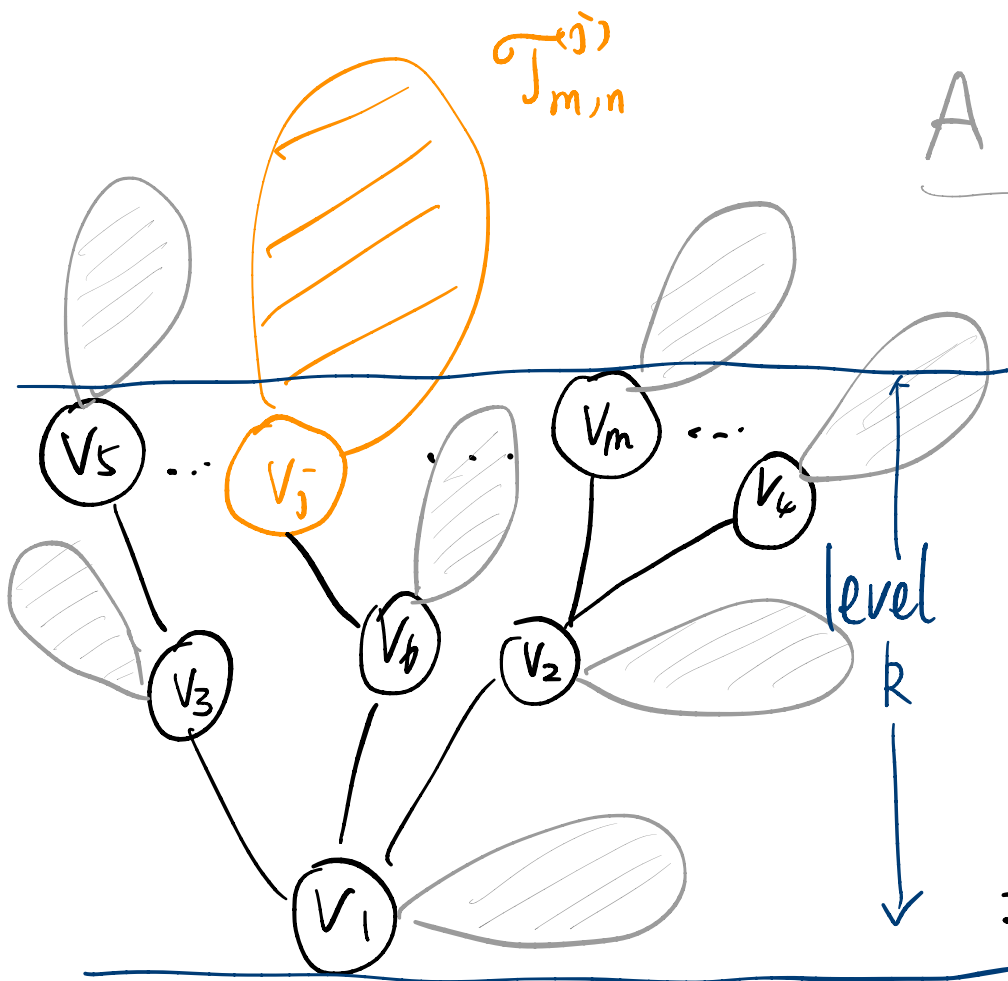
by our choice of m , it is greater than $\frac{e^{-m}}{m} H(T_m) = 1$

$$\geq e^{-c m/\ln m} \geq e^{-c m_2(n) \log n}$$

□

Just need
to improve
the crude
requirement

A slight improvement:



Fix any $k \in \mathbb{N}$. Then

$$\begin{cases} H(\mathcal{T}_m) = k \\ \max_{1 \leq j \leq m} H(\mathcal{T}_{m,n}^{(j)}) \leq 2e \log n - k \end{cases}$$

$$\Rightarrow H_n := H(\mathcal{T}_n) \leq 2e \log n$$

$$\begin{aligned} \bullet \mathbb{P}(H(\mathcal{T}_m) = k) &= \frac{A_k(m)}{(m-1)!} \\ &\geq e^{-m \log^{(k)} m} = e^{-m \log^{(k)} n} \end{aligned}$$

Increasing trees with height k on m nodes

□

Remarks

- We have a intuitive enumeration showing that

$$A_k(n) \approx \exp(n \log n - n \log^{(k)} n + O(n)) \quad (*)$$

- It is well known that, the generating function

$$F_k(z) = \sum_{n=1}^{\infty} \frac{A_k(n)}{n!} z^n = \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{P}(H_n \leq k) z^n$$

Satisfies

$$\begin{cases} F_k'(z) = \exp(F_{k-1}(z)) \\ F_1(z) = e^z - 1; \quad F_k(0) = 0 \end{cases}$$

In principle, (*) should be proved by analyzing $F_k(z)$

But it seems hard (for us) to do this ...

Proof ideas of Thm 1

Constraints on
many sub-structures

(1) Reselect m so that

$$\mathbb{E}[H_{\frac{1}{4}\frac{n}{m}}] = 2\epsilon \log n + \text{large constant}$$

and hence

$$\mathbb{P}(H_{\frac{1}{4}\frac{n}{m}} \leq 2\epsilon \log n) < \frac{1}{2}$$

$$\Rightarrow m = \frac{M_2(n)}{\text{large Cst}}$$

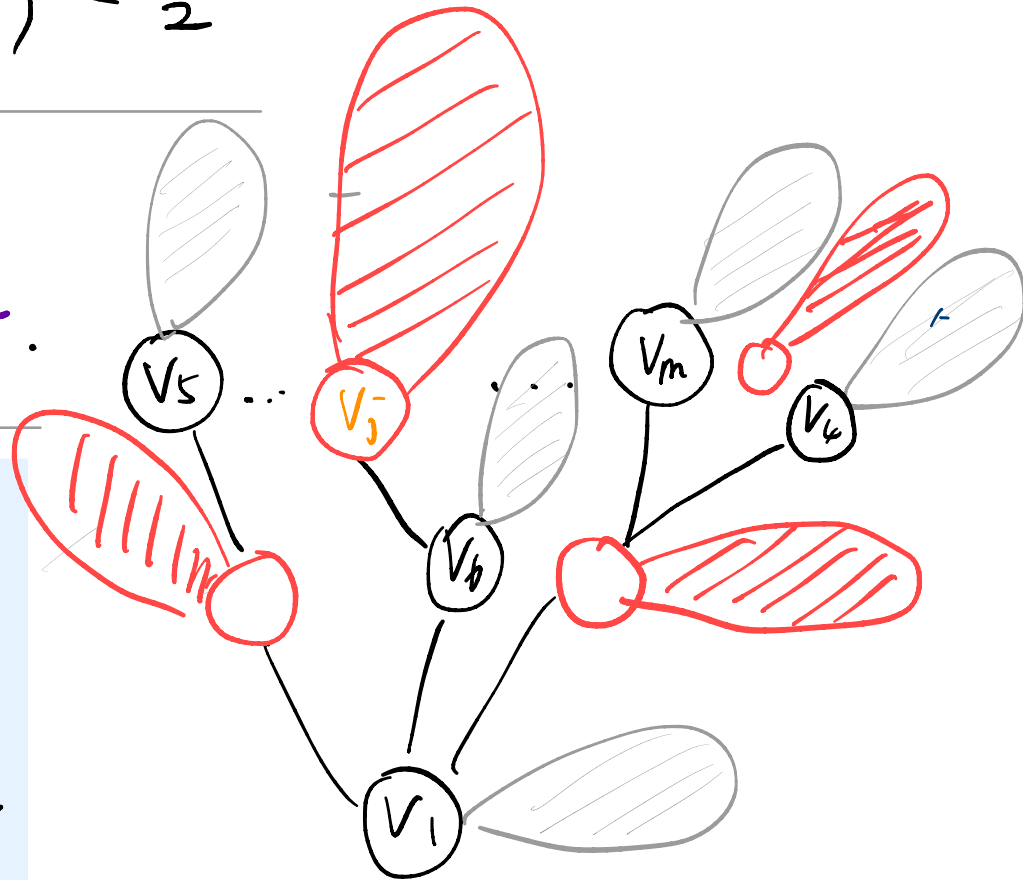
(2) Let $BIG_{m,n}$ be the collection of

all $T_{m,n}^{(j)}$'s with $|T_{m,n}^{(j)}| \geq \frac{1}{4}\frac{n}{m}$.

Combining (2) with (1):

$$\mathbb{P}(H_n \leq 2\epsilon \log n)$$

$$\leq \mathbb{E}[2^{-|BIG_{m,n}|}].$$



$$\bullet \mathbb{P}(H_n \leq 2 \log n) \leq \mathbb{E} [2^{-|BIG_{m,n}|}]$$

$$\text{CLAIM: } \mathbb{P}(|BIG|_{m,n} < \gamma m) \leq e^{-C_\gamma^* m}$$

$$\leq 2^{-cm} + e^{-c'm} \leq e^{-c''m} \quad \square$$

Recall that $(|J_{m,n}^{(j)}|)_{j=1}^m \stackrel{\text{Unif}}{\sim} \{(\pi_j)_{j=1}^m : \pi_j \in \mathbb{N}, \sum_1^m \pi_j = n\}$

$\xi_j := \mathbb{1}\{|J_{m,n}^{(j)}| \text{ is small}\}$ are negative correlated

$$\mathbb{P}\left(\sum_1^m \xi_j \geq (1-\gamma)m\right) \leq \mathbb{P}\left(\sum_1^m \xi_j^{\text{IID}} \geq (1-\gamma)m\right) \\ \leq \exp\left\{-\underline{D(1-\gamma \| P_0)} m\right\}$$

[Rmk: $\lim_{\gamma \downarrow 0} C_\gamma^* = +\infty$] \longleftarrow

Where can be improved?

- Fix $k \in \mathbb{N}_+$

- Let $BIG_{m,n}^{(k)} = \left\{ 1 \leq j \leq m : \boxed{d(v_1, v_j) \geq k} \right.$
 $\left. \begin{array}{l} \sigma_{m,n}^{(j)} \text{ is big, i.e., } |\sigma_{m,n}^{(j)}| \geq \frac{1}{4} \frac{n}{m} \end{array} \right\}$

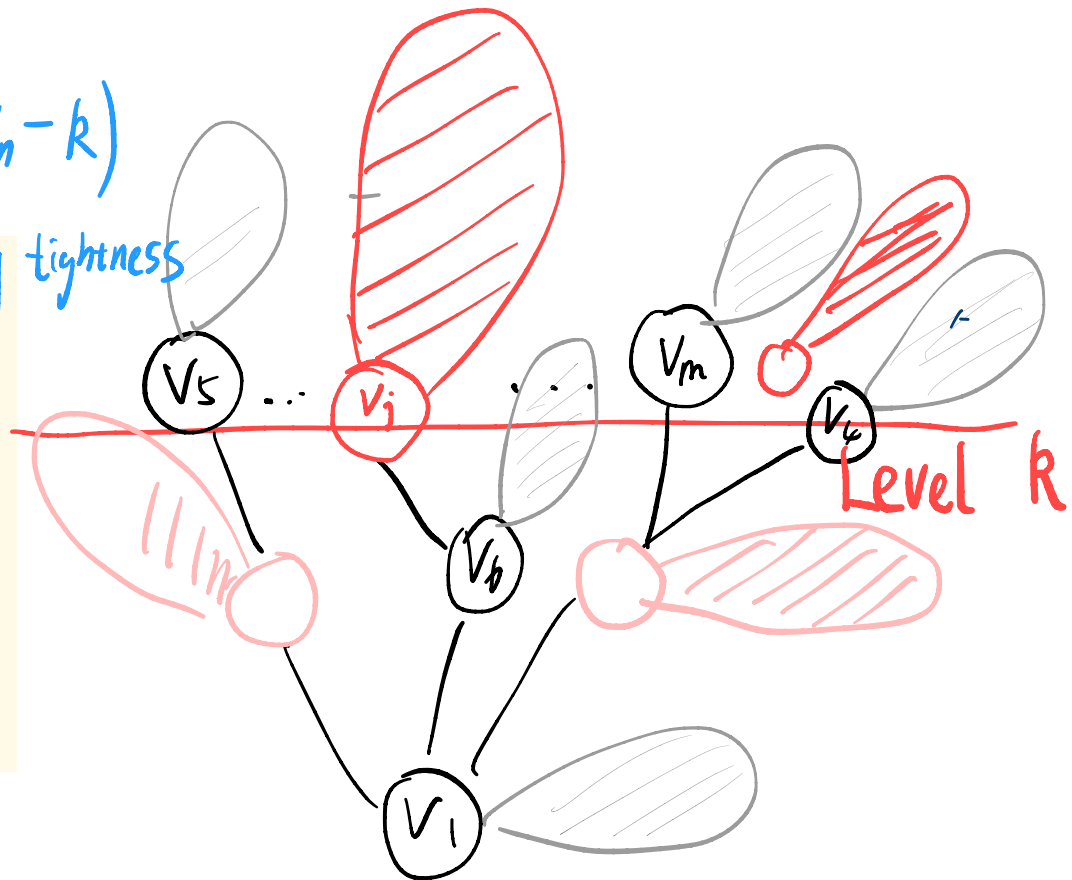
Then

$$\delta(k) = \sup_{n \geq 1} \mathbb{P}(H_n \leq \mathbb{E}H_n - k)$$

$\xrightarrow{k \rightarrow +\infty} 0$ by tightness

$$\mathbb{P}(H_n \leq 2e \ln n)$$

$$\leq \mathbb{E} \left[\delta(k) |BIG_{m,n}^{(k)}| \right]$$



★ Then for any $\varepsilon \in (0, 1]$, $\exists C_\varepsilon > 0$ s.t. for $n \geq 1$

$$\mathbb{P}\left(\sum_{l=0}^k X_{n,l} > \varepsilon n\right) \leq e^{-C_\varepsilon n \log^{(k)}(n)}$$

↖ size of level l in \mathcal{T}_n

- Consequently, $\mathbb{P}(|\text{BIG}_{m,n}^{(k)}| < \frac{\gamma}{2} m)$

$$\leq \mathbb{P}(|\text{BIG}|_{m,n} < \gamma m) + \mathbb{P}\left(\sum_{l=0}^k X_{m,l} > \frac{\gamma}{2}\right)$$

$$\lesssim e^{-C_\gamma^* m} + e^{-\Theta(m \log^{(k)} m)}$$

$$BIG_{m,n}^{(k)} = \left\{ \begin{array}{l} 1 \leq j \leq m: d(v_1, v_j) \geq k \\ \bigcup_{m,n}^{(j)} \text{ is big, } \left| \bigcup_{m,n}^{(j)} \right| \geq \frac{1}{4} \frac{n}{m} \end{array} \right\}$$

Then

$$\delta(k) = \sup_{n \geq 1} \mathbb{P} (H_n \leq \mathbb{E} H_n - k)$$

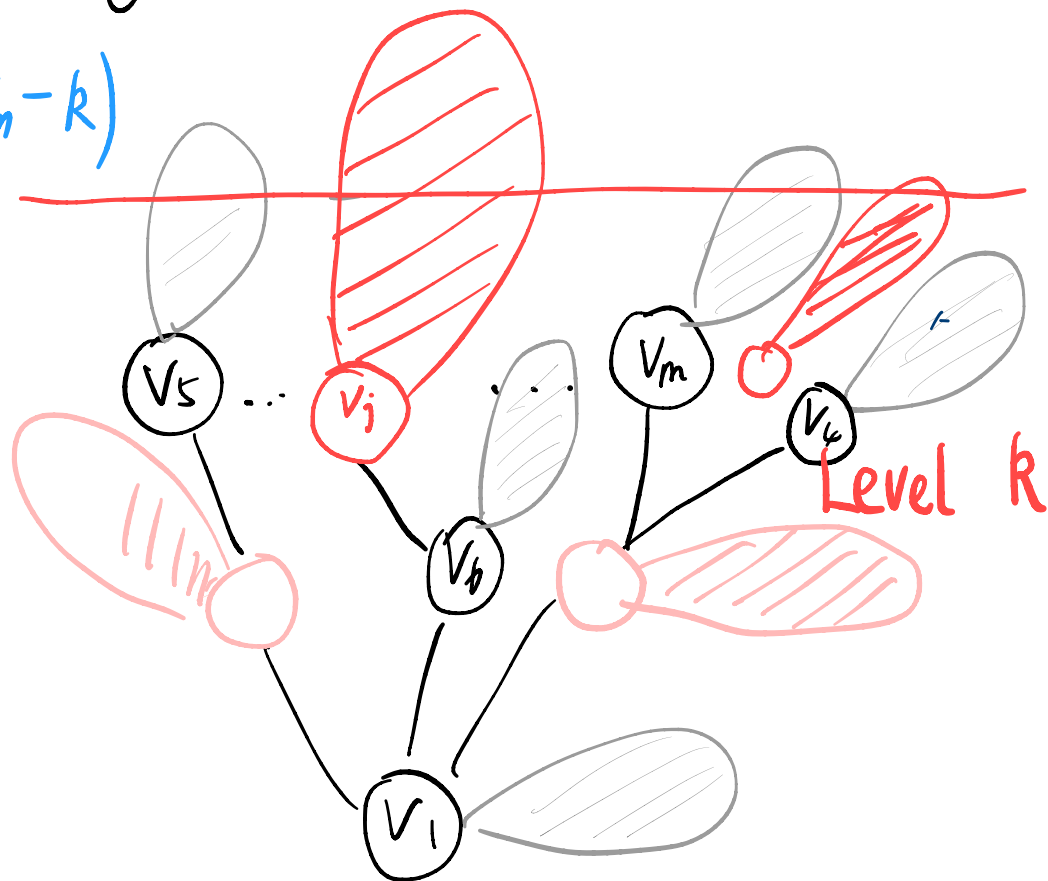
$$\mathbb{P}(H_n \leq \alpha \ln n)$$

$$\leq \mathbb{E} \left[\delta(k) \left| BIG_{m,n}^{(k)} \right| \right]$$

$$\leq \delta(k) \gamma^m + e^{-C \gamma^* m}$$

$$\geq e^{-Km}$$

for any $k > 0$, provided γ small then k large enough. \square



Future directions

- Identifying the decay rate of $\omega_2(n)$
- LDP on the height of Preferential attachment tree
binary search tree and other random tree models .

•

Thanks for

your attention!

As a shortest-path tree

- K_n = Complete graph on $\{v_1 \dots v_n\}$
- assign **IID exponential** weights (mean 1) to each edge

The tree \mathbb{T}_n , formed by the shortest-path from v_1 to all other nodes $\{v_2 \dots v_n\}$, is a RRT.

