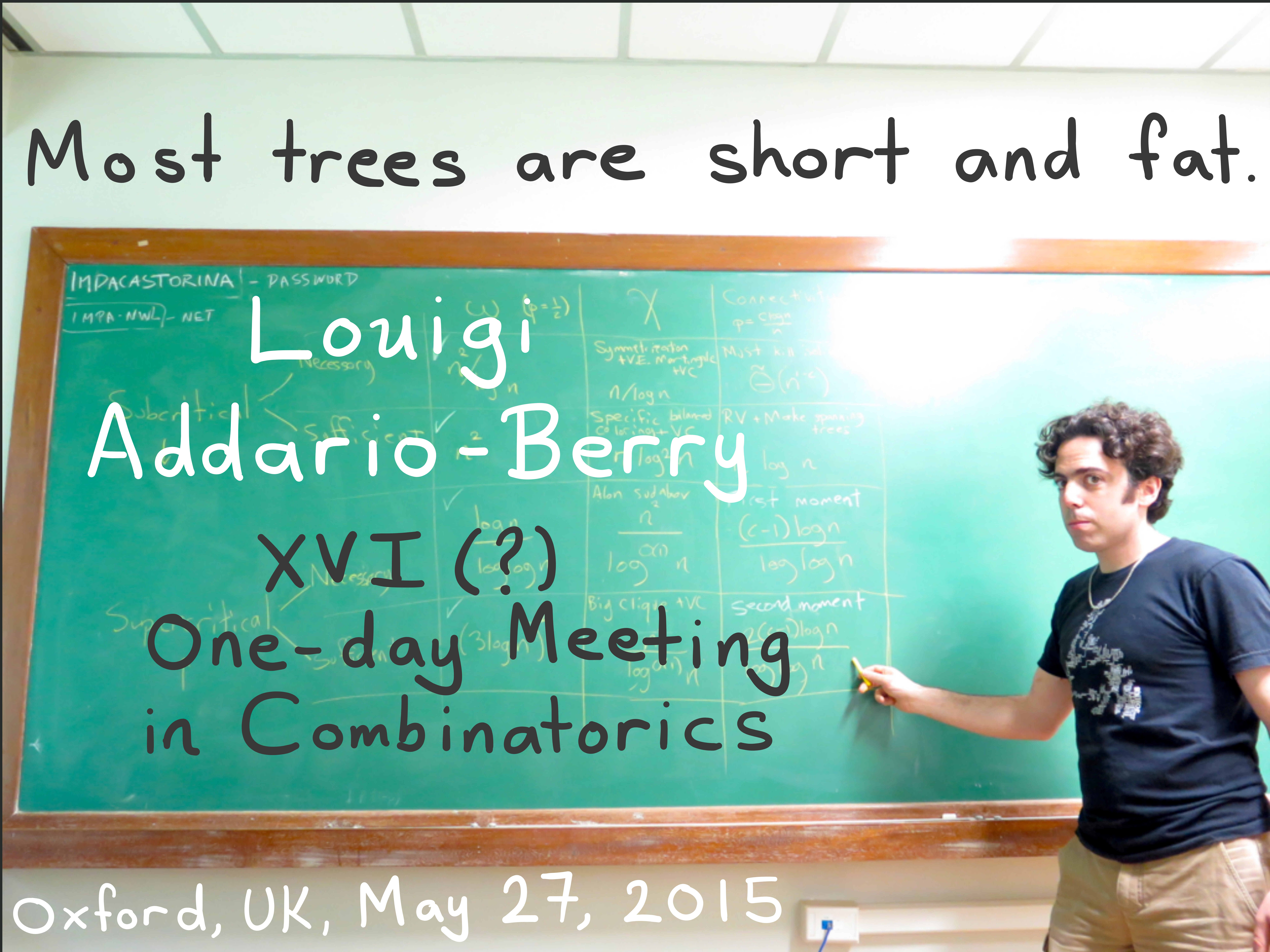
Most trees are short and fat.



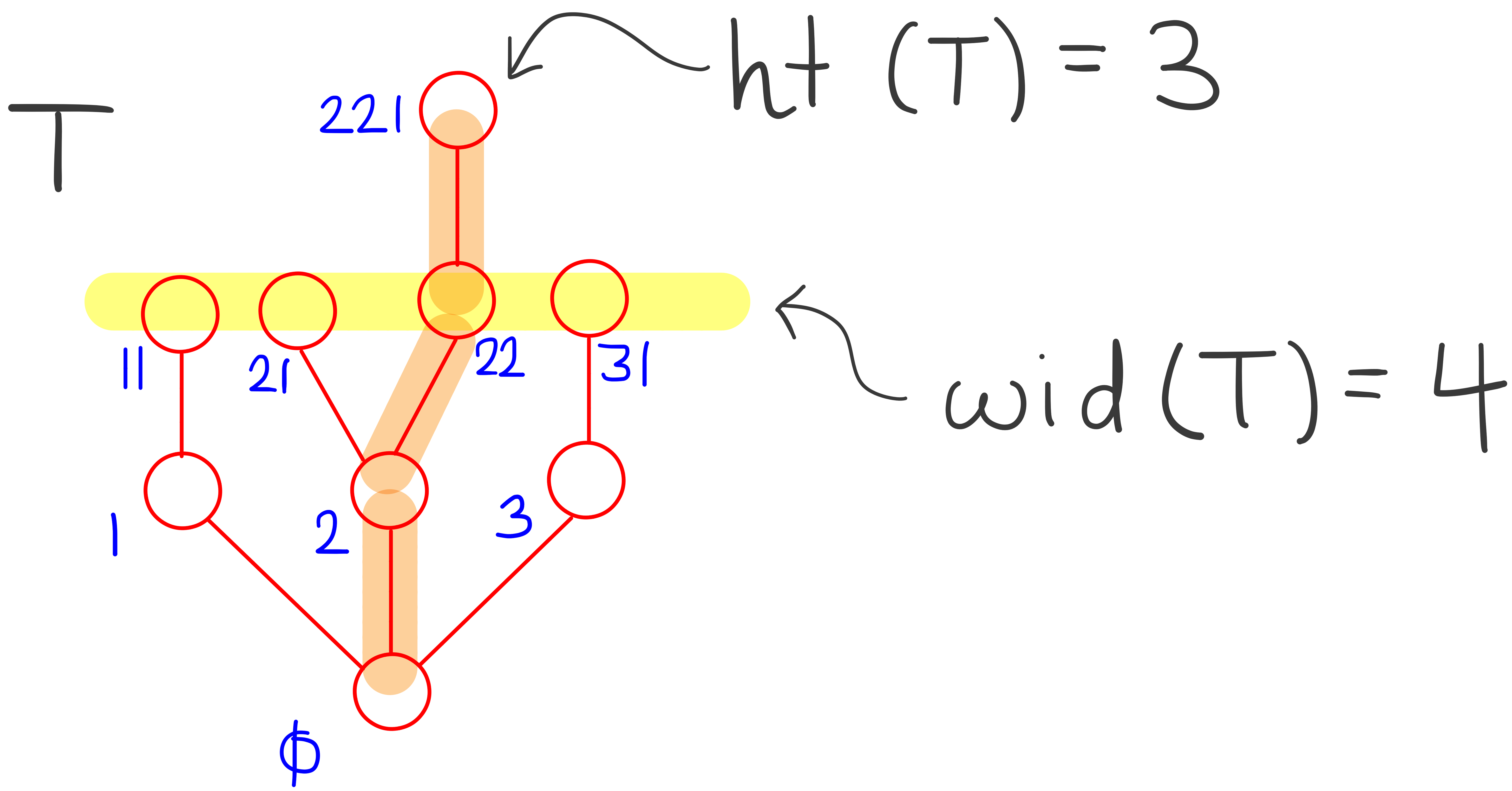
Louigi Addario-Berry

XVI (?) One-day Meeting in Combinatorics

Oxford, UK, May 27, 2015

Plane Trees:

Rooted; children of each node have left-to-right order.



Height: Greatest distance from any node to the root } ht(T)

Width: Greatest # nodes on a single level. } wid (T)

Random Plane Trees: Galton-Watson Trees

- Each node has random # of children
- Nodes reproduce independently
- Model Parameter: Reproduction Law/Offspring distribution.

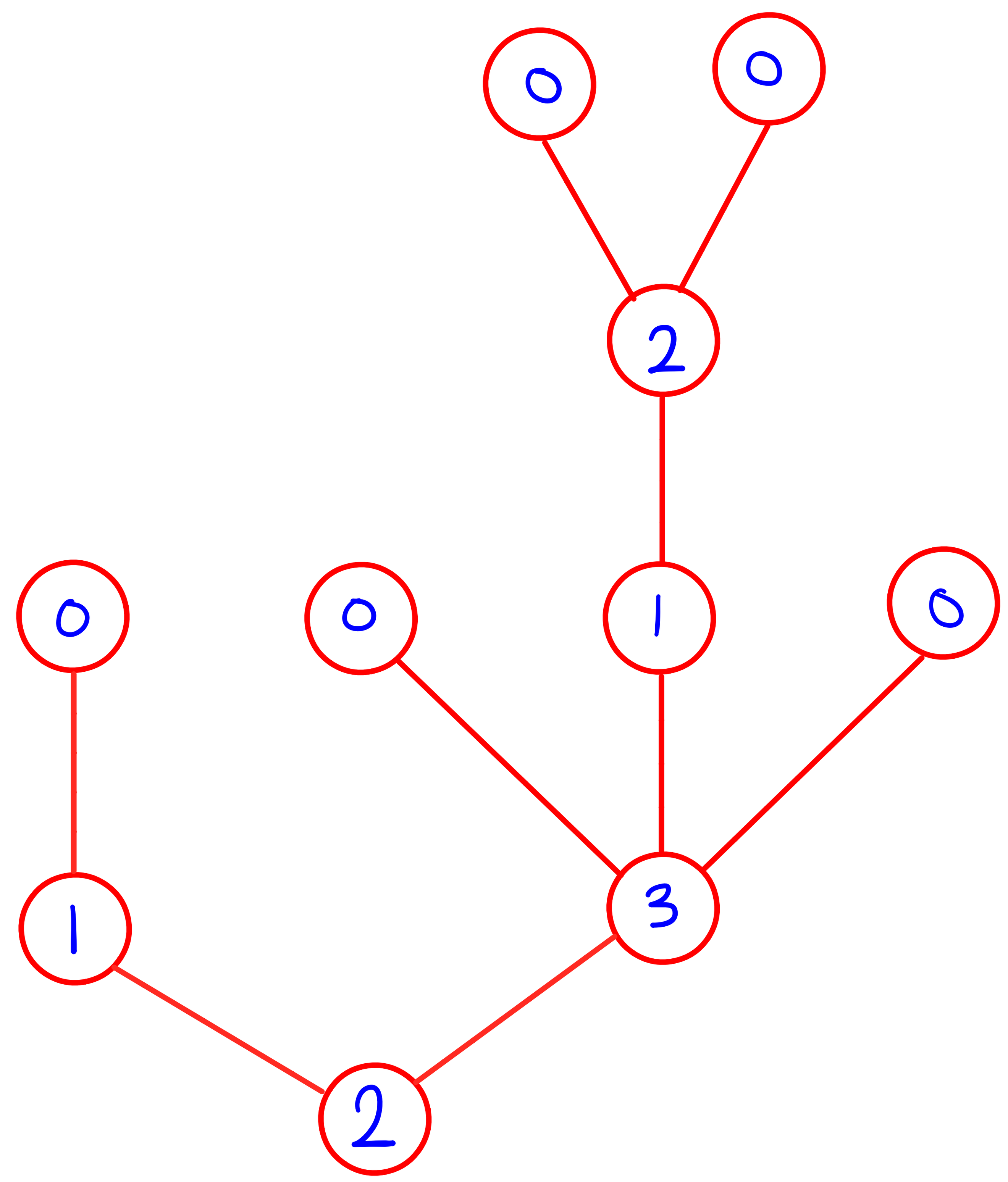
Construction

- $(C_i, i \geq 1)$ independent copies of a random variable C with $\sum_{k\geq 0} \mathbb{P}(C=k) = 1$.

Rule

The sequence $(C_i, i \geq 1)$ gives # children of nodes, in breadth-first search order

Example: $(2, 1, 3, 0, 0, 1, 0, 2, 0, 0, 1, 4, 0, \dots)$



Key definition: Such T is $GW(C)$-distributed

Size = # vertices =: σ

[possible that tree is infinite, if so then $\sigma = \infty$]

Main Results

Fix any r.v. C with $\sum_{k \geq 0} \mathbb{P}(C=k) = 1$

let T be $GW(C)$ distributed.

Write $p_k = \mathbb{P}(C=k)$.

Theorem ("Most trees are short & fat")

There is a universal constant $\delta > 0$ s.t.

$$\mathbb{P}\left(ht(T) \geq \frac{k}{1-p_1} \cdot wid(T)\right) \leq \exp(-\delta k).$$

Remarks

① If $\mathbb{E}C > 1$ then $\mathbb{P}(\sigma = \infty) > 0$, and $\mathbb{P}(ht(T) = wid(T) = \infty \mid \sigma = \infty) = 1$.

Also, given that $\sigma < \infty$, the cond. dist. of T is $GW(\hat{C})$

where $\mathbb{P}(\hat{C} = 1) = p_1$. So can assume $\mathbb{E}C \leq 1$.

② Write $T_k = \{\text{nodes at level } k\}$. Easy: $\mathbb{E}|T_k| = [\mathbb{E}C]^k$.

If $\mathbb{E}C < 1$ then $\mathbb{P}(ht(T) \geq k) = \mathbb{P}(|T_k| > 0) \leq \mathbb{E}|T_k| = [\mathbb{E}C]^k$

so $\mathbb{P}(ht(T) \geq k \cdot wid(T)) \leq \exp\left(-k \cdot \log\left(\frac{1}{\mathbb{E}C}\right)\right)$.

Exp. tails; but $\log\left(\frac{1}{\mathbb{E}C}\right)$ isn't $\approx \frac{1}{1-p_1}$.

③ If $\mathbb{E}C = 1$ & $\mathbb{E}[C^2] < \infty$ then $\mathbb{P}(wid(T) \geq x) = \Theta\left(\frac{1}{x}\right)$,

$\mathbb{P}(ht(T) \geq x) = \Theta\left(\frac{1}{x}\right)$, so not trivially true.

[Actually, when $\mathbb{E}C = 1$ & $\mathbb{E}[C^2] < \infty$, $\mathbb{P}\left(ht(T) \geq \frac{k}{1-p_1} wid(T)\right) \geq \exp(-\gamma k)$; thm is tight.]

Random Plane Trees: Galton-Watson Trees

- Each node has random # of children
- Nodes reproduce independently
- Model Parameter: Reproduction Law/Offspring distribution.

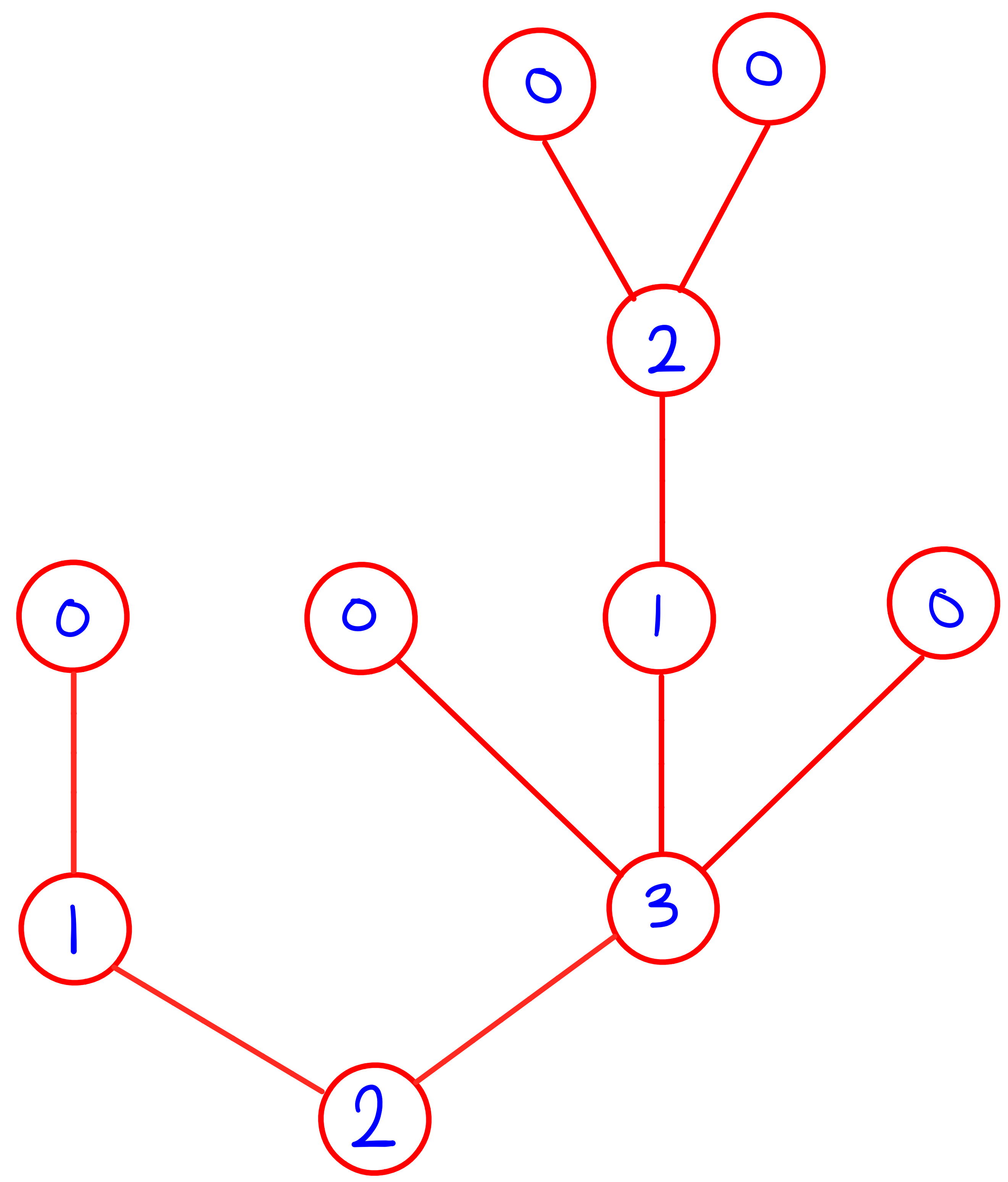
Construction

- $(C_i, i \geq 1)$ independent copies of a random variable C with $\sum_{k \geq 0} \mathbb{P}(C = k) = 1$.

Rule

The sequence $(C_i, i \geq 1)$ gives # children of nodes, in breadth-first search order

Example: $(2, 1, 3, 0, 0, 1, 0, 2, 0, 0, 1, 4, 0, \ldots)$



Key definition: Such T is $GW(C)$-distributed

Halting Condition

Trees with n nodes have $n-1$ edges.

For $n \geq 0$

$$\#\text{nodes discovered by } \overbrace{\text{time } n}^{n \text{ steps of BFS}}. = 1 + \sum_{i=1}^{n} C_i$$

$$\#\text{ vertices explored by time } n = n$$

$$\text{Size} = \#\text{ vertices} =: \sigma$$

$$= \inf\{n : 1 + \sum_{i=1}^{n} C_i = n\}.$$

Setup

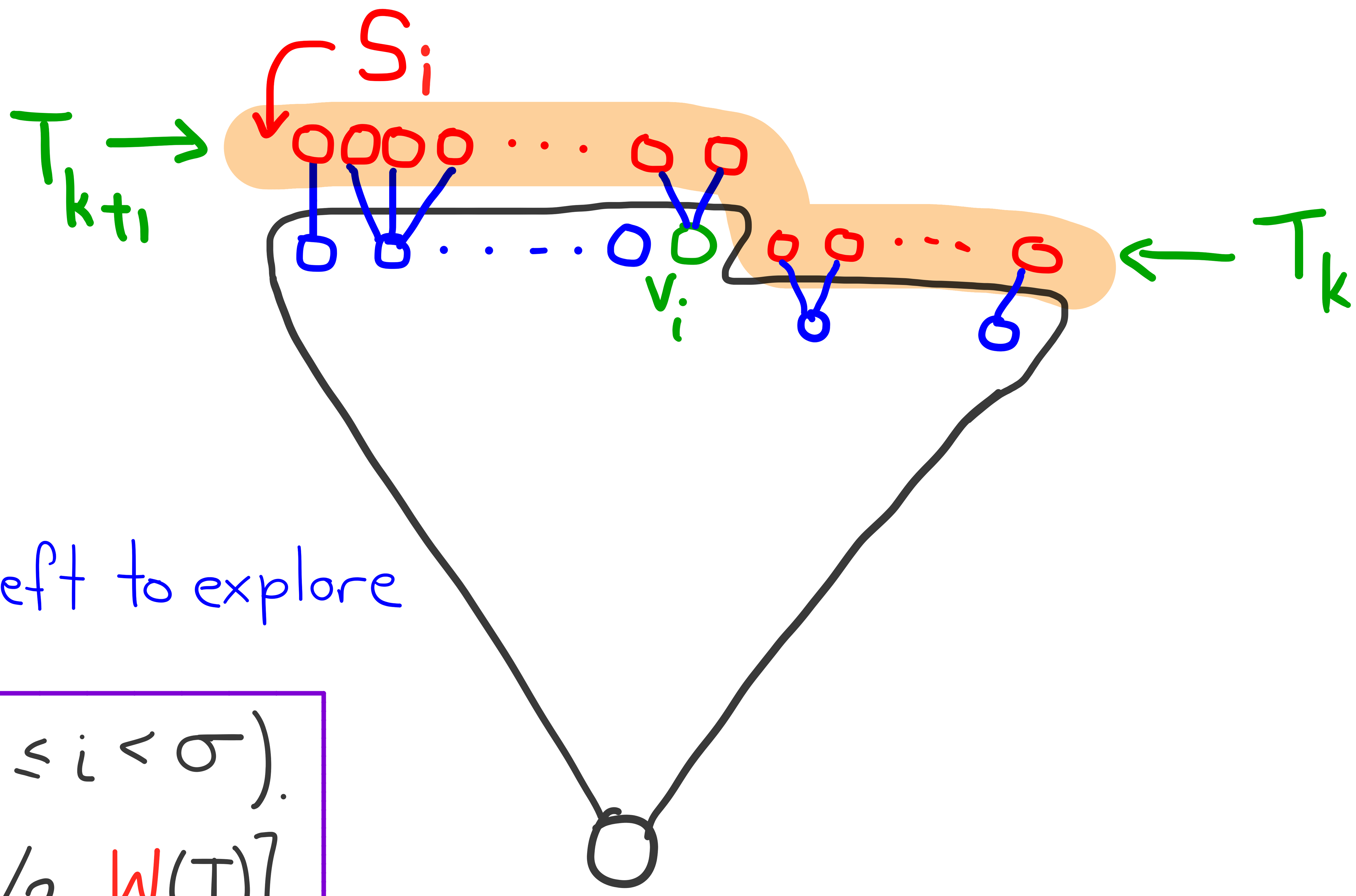
$1+\sum_{j=1}^{i} C_j$ = # nodes discovered by time i.

Let $S_i = 1+\sum_{j=1}^{i}(C_j-1)$

= # nodes in "BFS queue" at time i

$\mathbb{E}C \le 1 \Rightarrow \mathbb{E}S_n = 1 + n(\mathbb{E}C-1) \le 1.$

$\sigma = \inf\{t : S_t = 0\}$ = first time no nodes left to explore



Prop: Let $W(T) = \max(S_i, 0 \le i < \sigma)$.
Then $\mathrm{wid}(T) \in (W(T)/2, W(T)]$

Proof: During BFS on level k, "exploration queue" $\subset T_k \cup T_{k+1}$; and $= T_k$ at start of level k. ∎

Idea: $ht(T) = \sum_{k=1}^{ht(T)} 1 = \sum_{k=1}^{ht(T)} \sum_{v \in T_k} \frac{1}{|T_k|}.$

When $v_i \in T_k$ then $S_i \approx |T_k|$ so perhaps

$$ht(T) \approx \sum_{k=1}^{ht(T)} \sum_{v_i \in T_k} \frac{1}{S_i} = \sum_{i=1}^{\sigma} \frac{1}{S_i} =: H(T)\ ?$$

[False; consider a star with n leaves. But...]

Prop: $ht(T) \le 3H(T).$

Corollary Suffices to prove

$$\mathbb{P}\left(H(T) \ge \frac{k}{1-p_1} W(T)\right) \le e^{-\delta k},$$

thm. follows.

$W(\sigma)=\max(S_i,\ 0\le i<\sigma)$ $\quad H(\sigma)=\sum_{i=1}^{\sigma}\frac{1}{S_i}$ $\quad$ **Aim**: $\mathbb{P}\left(H(\sigma)\ge\frac{k}{1-p_1}W(\sigma)\right)\le e^{-\delta k}$

Decomposition into scales.

When $S_i\approx 2^k$, takes about 2^k steps for $H(i)$ to increase by 1

$$H(i+2^k)-H(i)=\sum_{j=i+1}^{i+2^k}\frac{1}{S_j}\simeq 2^k\cdot\frac{1}{S_i}\simeq 1.$$

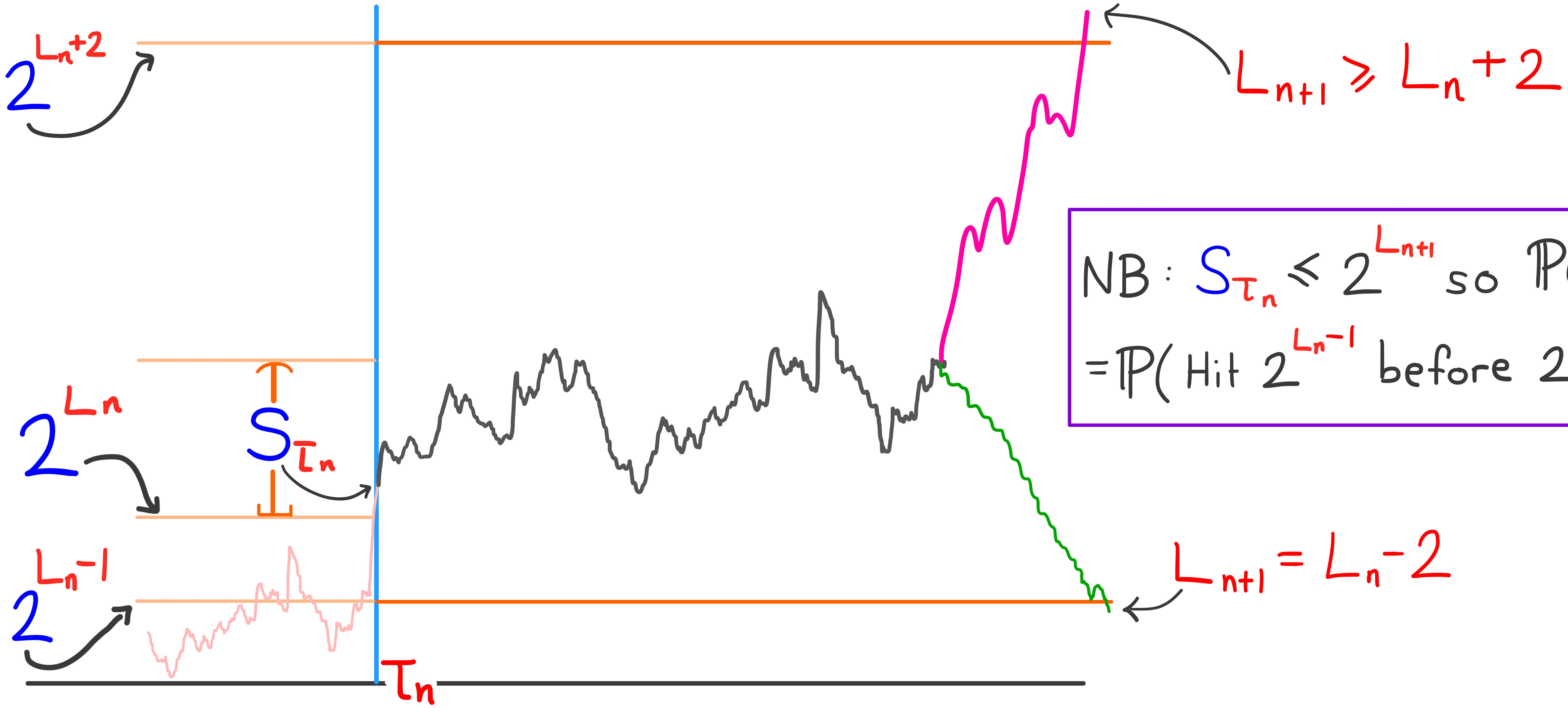
Def

$\tau_0=0=$ initial time

$L_0=0=\log_2 S_0=$ initial scale

$\tau_{n+1}=\min\{t\ge\tau_n : S_t\notin[2^{L_n-1},2^{L_n+2})\}$

$L_{n+1}=\sup\{k : 2^k\le S_{\tau_{n+1}}\}$



NB: $S_{\tau_n}\le 2^{L_{n+1}}$ so $\mathbb{P}(L_{n+1}<L_n)$
$=\mathbb{P}(\text{Hit } 2^{L_n-1} \text{ before } 2^{L_n+2})>\frac{1}{2}$.

$W(\sigma) = \max(S_i, 0 \le i < \sigma)$ $H(\sigma) = \sum_{i=1}^{\sigma} \frac{1}{S_i}$ **Aim**: $\mathbb{P}\left(H(\sigma) \geq \frac{k}{1-p_1} W(\sigma)\right) \leq e^{-\delta k}$

Number of visits to a scale

Let $\Lambda(t)$ = current scale at time $t = L_n$ (with n s.t. $\tau_n \leq t < \tau_{n+1}$)

Let $N(\ell) = \#\{t < \sigma : \Lambda(t) = \ell\}$

Then $\boxed{H(\sigma) = \sum_{\ell \geq 0} \sum_{\{t < \sigma : \Lambda(t) = \ell\}} \frac{1}{S_t} \leq \sum_{\ell \geq 0} N(\ell) \cdot 2^{-(\ell-1)}}$

$$N(\ell) = N(\ell,1) + N(\ell,2) + \cdots + N(\ell, M(\ell))$$

$N(\ell,i)$ = Duration of i'th visit to scale ℓ

$M(\ell)$ = # visits to scale ℓ $= \#\{i : L_i = \ell\}$

Fact: Given that $M(\ell) \neq 0$, $M(\ell)$ dominated by sum of 2 Geom$(\frac{1}{2})$ r.v.s; $\Longrightarrow \mathbb{P}(M(\ell) > k \mid M(\ell) > 0) \leq 2^{-k/2}$.

Proof: visits to scale ℓ entail upcrossings of $[2^{\ell-1}, 2^{-\ell})$ or of $[2^{\ell+1}, 2^{\ell+2})$.
Both are hard since walk has non-positive drift. ∎

$W(\sigma) = \max(S_i, 0 \le i < \sigma)$ $H(\sigma) = \sum_{i=1}^{\sigma} \frac{1}{S_i}$ **Aim**: $\mathbb{P}\left(H(\sigma) \geq \frac{k}{1-p_1} W(\sigma)\right) \leq e^{-\delta k}$

$H(\sigma) \leq \sum_{\ell \geq 0} N(\ell) \cdot 2^{-(\ell-1)}$ $N(\ell)$ = time spent at scale ℓ

$= \sum_{\ell \geq 0} \sum_{i=1}^{M(\ell)} N(\ell, i) \cdot 2^{-(\ell-1)}$ $N(\ell, i)$ = Duration of i'th visit to scale ℓ

Thm (Lévy; Doeblin; Kolmogorov; Rogozin; Le Cam; Esséen; Kesten):

With $p = \max p_i$, have $\boxed{\max_k \mathbb{P}(S_n = k) \leq \frac{Cp}{\sqrt{n(1-p)}}}$ $C > 0$ universal.

"Any random walk spreads out over $\geq \sqrt{n}$ values by time n". Here $\sqrt{n} \simeq 2^\ell$.

Corollary: $\mathbb{P}\left(N(\ell,i) > \frac{64C^2 p^2}{(1-p)} \cdot m \cdot 4^\ell\right) \leq 2^{-m}$

Recall that $N(\ell) = \sum_{i=1}^{M(\ell)} N(\ell,i)$ and that $\mathbb{P}(M(\ell) \geq k) \leq 2^{-k/2}$

It then follows that $\mathbb{P}\left(\frac{N(\ell)}{2^{\ell-1}} \geq C' m \cdot 2^\ell\right) \leq 2^{-m}$

$W(\sigma) = \max(S_i, 0 \le i < \sigma)$ $H(\sigma) = \sum_{i=1}^{\sigma} \frac{1}{S_i}$ **Aim**: $\mathbb{P}\left(H(\sigma) \ge \frac{k}{1-p_1} W(\sigma)\right) \le e^{-\delta k}$

$$H(\sigma) \le \sum_{\ell \ge 0} N(\ell) \cdot 2^{-(\ell-1)} \qquad \mathbb{P}\left(\frac{N(\ell)}{2^{\ell-1}} \ge C' m \cdot 2^{\ell}\right) \le 2^{-m}$$

Wrapping up

Note that if $N(\ell) > 0$ then $W(\sigma) \ge 2^{\ell-1}$

So with $\ell^* = \max\{\ell : N(\ell) > 0\}$,

$$\mathbb{P}(H(\sigma) > C'' m \cdot W(\sigma))$$

$$\le \mathbb{P}\left(\sum_{\ell=0}^{\ell^*} \frac{N(\ell)}{2^{\ell-1}} > C'' m \cdot 2^{\ell^*}\right)$$

$$\le \sum_{\ell=0}^{\ell^*} \mathbb{P}\left(\frac{N(\ell)}{2^{\ell-1}} > C' m \cdot 2^{(\ell^*+\ell)/2}\right) \qquad \left[\sum_{\ell=0}^{\ell^*} 2^{(\ell^*+\ell)/2} \simeq 2^{\ell^*} \cdot \frac{1}{\sqrt{2}-1}\right]$$

$$\le \sum_{\ell=0}^{\ell^*} 2^{-m \cdot 2^{(\ell^*-\ell)/2}} < 2^{1-m} \quad \blacksquare$$

Behind the curtain: Use Markov property and non-positive drift to "fix" bogus $\le$ above.

Remarks

- Claimed theorem with dependence only on p_1 proved it with dependence on $p = \max p_i$.

Fix: non-trivial; requires modifying "disperson" bound for our setting.

(Idea: If subcritical then $p_{max} = p_0$ or p_1; if p_0 close to 1 then either very subcritical or make large jumps.)

- Stronger results if add info. about tails of degrees. Ex: If $\mathbb{P}(C \geq k) = \Theta(t^{-\alpha})$, $\alpha \in (1,2)$,

$$\text{then } \mathbb{P}(H(\sigma) > C \cdot m \cdot W(\sigma)^{\alpha-1}) \leq 2^{-\delta m}$$

- Conjecture: All this works even conditional on size of tree: $\mathbb{P}(H(\sigma) > C \cdot m \cdot W(\sigma) \mid \sigma = n) \leq 2^{-\delta m}$.
- Conjecture: Binary trees are the tallest & slimmest.

Let $GW_n(C) = [GW(C) \mid \sigma = n]$, $ht_n(C) = ht(GW_n(C))$, $wid_n(C) = wid(GW_n(C))$

Assuming $p_1 = 0$, to stochastically maximize $ht_n(C)$ to ______ minimize $wid_n(C)$

Take $p_0 = p_2 = \frac{1}{2}$.


IMPACASTORINA - PASSWORD
IMPA-NWL - NET
Thank you!