The Shape of Unlabeled Rooted Trees

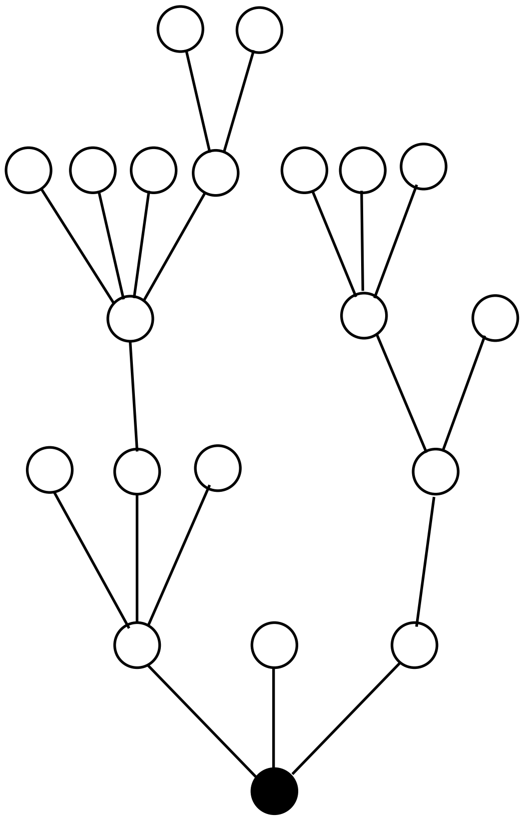
Bernhard Gittenberger
joint work with **Michael Drmota**

Institute of Discrete Mathematics and Geometry
TU Wien, Austria

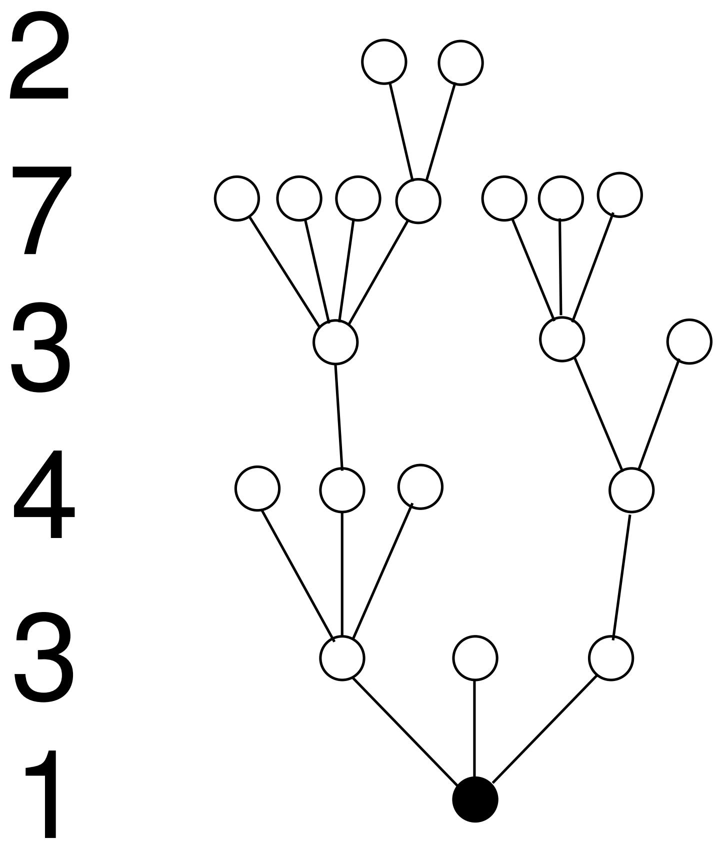
2008 International Conference on the
Analysis of Algorithms
Maresias, April 16, 2008

Supported by FWF, grant S9604, and the EU, Project NEMO, contract no.: 028875

Introduction

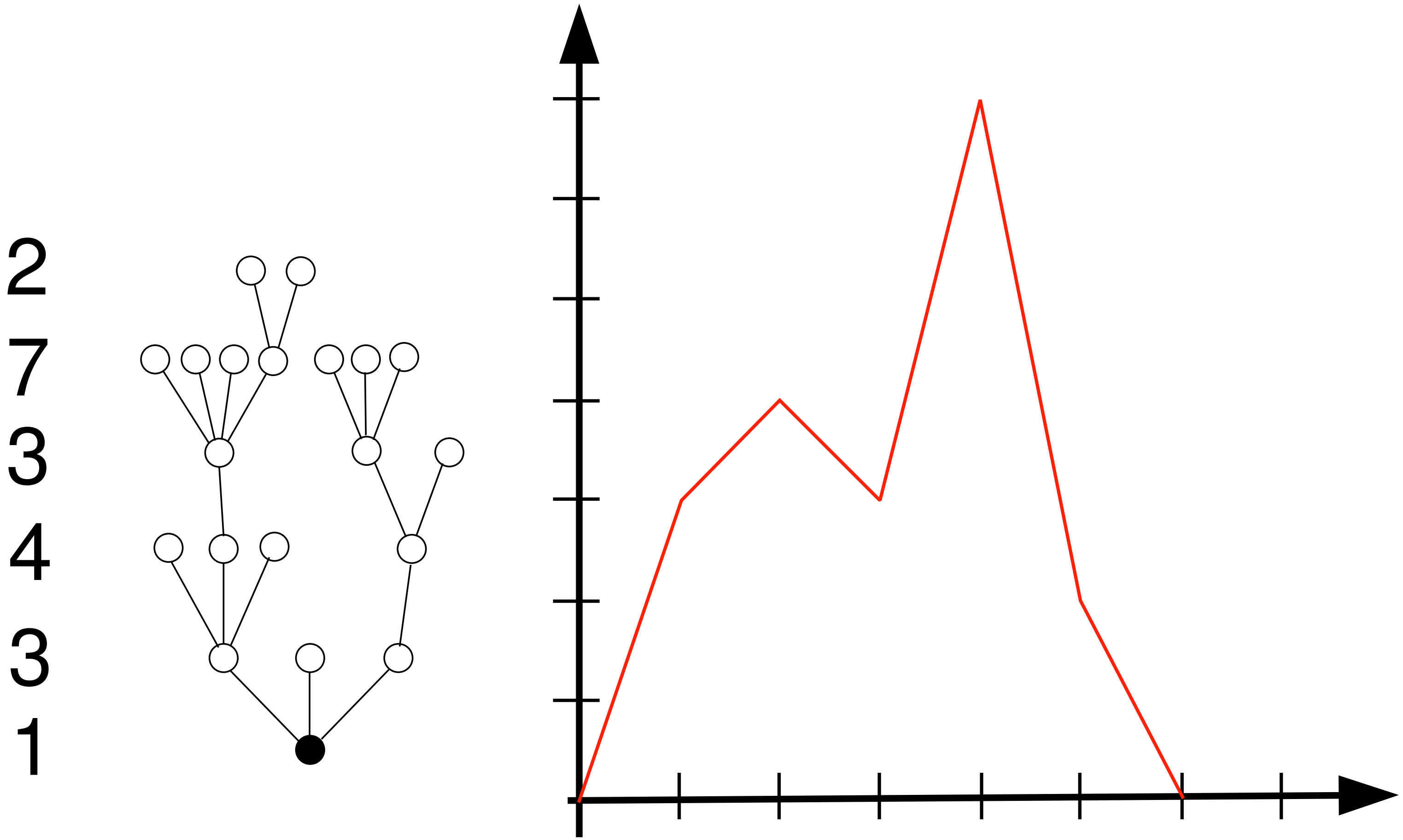


Introduction



Introduction

The profile:



Introduction

Two classes of trees

according to average height

1. $\log n$-trees: binary search trees, digital search trees, recursive trees, increasing trees

2. $\sqrt{n}$-trees: simply generated trees, Pólya trees

Historical Remarks

- Stepanov '69: Cayley trees, limiting distribution of level sizes, density and moments

- Kennedy '75, Kolchin '75, Takacs '91: simply generated trees, limiting density

- Aldous '91: simply generated trees, two conjectures for functional limit theorems, relation to Brownian excursion

- Drmota and G. '97, G. '99, Pitman '99: Proof of Aldous' conjectures

Historical Remarks

Representations of the Brownian excursion local time

- Getoor and Sharpe ’79: double Laplace transform approach directly via Brownian functionals

- Cohen and Hooghiemstra ’82: M/M/1 queues

- Knight ’82: convolution formula

- Louchard ’84: Laplace transform, Kac formula

Historical Remarks

Extensions:

- Pavlov 80ies, 90ies: random forests
- Drmota '96: nodes of fixed degree, functional limit theorems
- G. '02: functional limit theorem for simply generated forests

Historical Remarks

Extensions:

- Pavlov 80ies, 90ies: random forests
- Drmota '96: nodes of fixed degree, functional limit theorems
- G. '02: functional limit theorem for simply generated forests

Other tree classes:

- Chauvin, Drmota, Jabbour-Hattab 01, Drmota 04, Drmota and Hwang 2005: binary search trees
- Hwang et al. 2006, 2007: recursive trees, increasing trees

Brownian Excursion and Local Time

$W(t)$ ... Brownian motion (Wiener process)

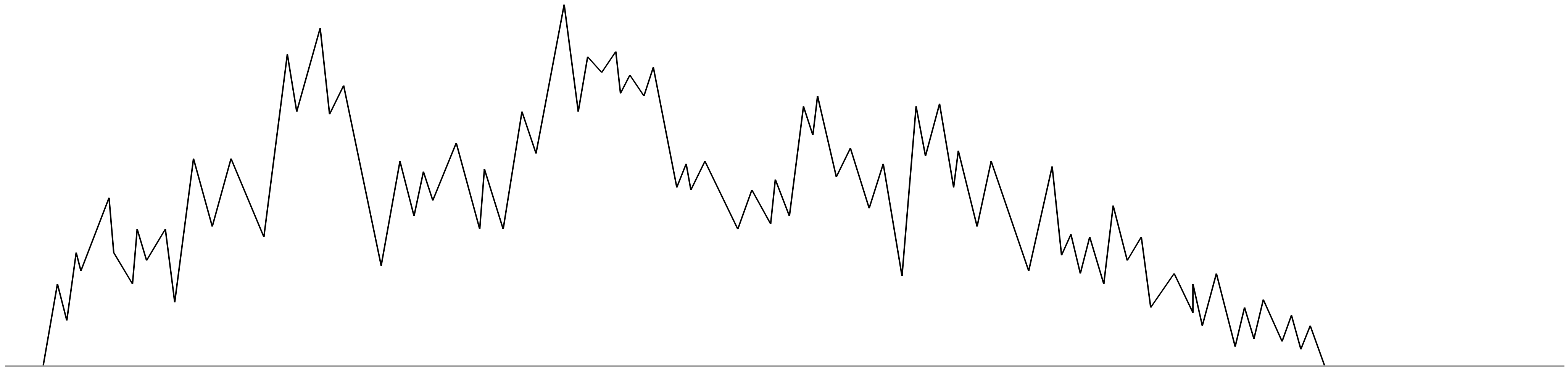
- $P[W(0) = 0] = 1$

- for $0 \le t_0 < t_1 < \cdots < t_k$ the increments $W(t_i) - W(t_{i-1})$ are independent

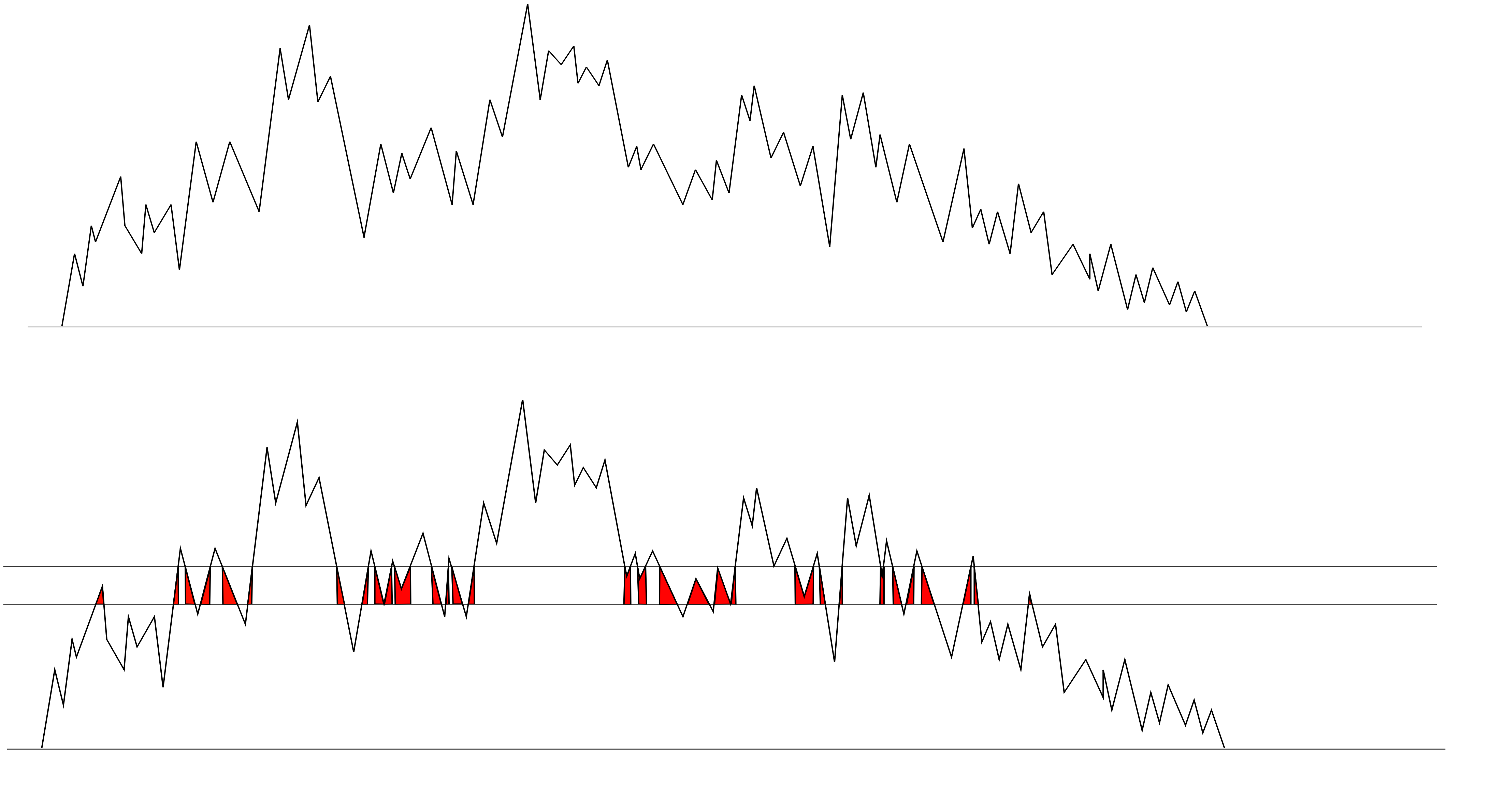
- for $0 \le s < t$ the random variable $W_t - W_s \sim \mathcal{N}(0, t-s)$

$e(t) \stackrel{d}{=} (W(t) | W(1) = 0)$ Brownian excursion

Brownian Excursion and Local Time



Brownian Excursion and Local Time



Brownian Excursion and Local Time

$\int_0^1 I_{[a,b]}(e(s))\,ds$ ... occupation time of $[a,b]$

local time:

$$\ell(\kappa) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_0^1 I_{[\kappa,\kappa+\varepsilon]}(e(s))\,ds$$

Brownian Excursion and Local Time

$\int_0^1 I_{[a,b]}(e(s))\,ds$ ... occupation time of $[a,b]$

local time:

$$\ell(\kappa) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^1 I_{[\kappa,\kappa+\varepsilon]}(e(s))\,ds$$

$\phi_\kappa(t)$ ... characteristic function

Then

$$\phi_\kappa(t) = 1 + \frac{\sqrt{2}}{\sqrt{\pi}} \int_\gamma \frac{t\sqrt{-s}e^{-\kappa\sqrt{-2s}}}{\sqrt{-s}e^{\kappa\sqrt{-2s}} - it\sqrt{2}\sinh(\kappa\sqrt{-2s}\,)}\, e^{-s}\,ds$$

where γ is a straight line parallel to $\mathbb{R} \cdot i$

Generating Functions

$\mathcal{Y}_n$ class of trees of size n

y_n ... number of trees of size n

$y(x)$ ... generating function of the sequence y_n

$$y(x) = \sum_{n \geq 1} y_n x^n$$

$\mathcal{Y}_n$ simply generated trees

$$y(x) = x\varphi(y(x))$$

Pólya trees:

$$y(x) = xe^{y(x)} \exp\left(\sum_{i \geq 2} \frac{y(x^i)}{i}\right)$$

Generating Functions

Pólya trees:

$$y(x) = xe^{y(x)} \exp\left(\sum_{i\geq 2} \frac{y(x^i)}{i}\right)$$

Pólya '37: radius of convergence ρ, $0 < \rho < 1$,
$x = \rho$ only singularity on circle of convergence

Otter '48: $y(\rho) = 1$

$$y(x) = 1 - b(\rho - x)^{1/2} + c(\rho - x) + d(\rho - x)^{3/2} + \cdots ,$$

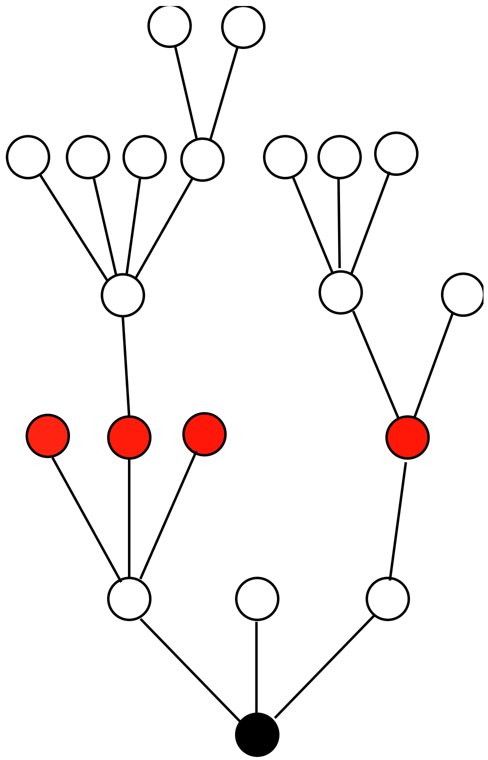
with $b \approx 2.681$, $c = b^2/3 \approx 2.396$, $\rho \approx 0.338$.

Hence

$$y_n \sim \frac{b\sqrt{\rho}}{2\sqrt{\pi}} n^{-3/2} \rho^{-n}$$

The Profile

$L_n(k)$ number of nodes at distance k from the root of an unlabeled rooted tree of size n



The Profile – Results

Theorem 1 *Let*

$$l_n(\kappa) = \frac{1}{\sqrt{n}} L_n\left(\kappa\sqrt{n}\right)$$

Then we have in $C[0,\infty)$

$$(l_n(t))_{t\geq 0} \xrightarrow{w} \frac{b\sqrt{\rho}}{2\sqrt{2}}\left(l\left(\frac{b\sqrt{\rho}}{2\sqrt{2}}\,t\right)\right)_{t\geq 0},$$

as $n\to\infty$*, where* b *and* ρ *are the constants in*

$$y(x) \sim 1 - b(\rho - x)^{1/2}, \qquad x \to \rho.$$

The Profile – Results

Theorem 2 *Let* $l_n(t)$ *as before.*

Then, as $n \to \infty$*, for any choice of fixed* $t_1, \ldots, t_d$

$$(l_n(\kappa_1), \ldots, l_n(\kappa_d)) \xrightarrow{w} \frac{b\sqrt{\rho}}{2\sqrt{2}} \left(l\left(\frac{b\sqrt{\rho}}{2\sqrt{2}}\kappa_1\right), \ldots, l\left(\frac{b\sqrt{\rho}}{2\sqrt{2}}\kappa_d\right)\right),$$

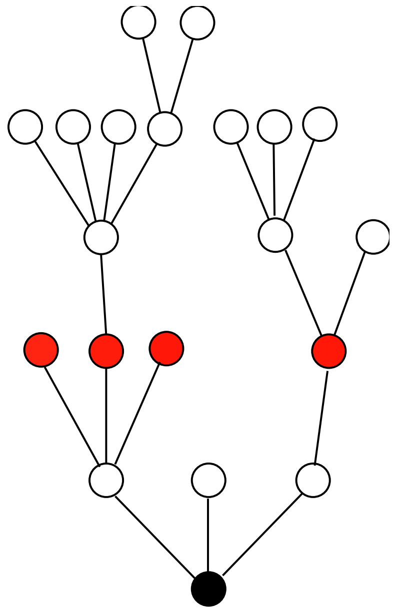
Theorem 3 *For all* $n, r, h \geq 0$ *and some* $C > 0$ *we have*

$$\mathbf{E}(L_n(r) - L_n(r+h))^4 \leq Ch^2 n$$

and therefore the processes $(l_n(t))_{t\geq 0}$ *are tight.*

The Profile – Combinatorial Setup

$L_n(k)$ number of nodes at distance k from the root of an unlabeled rooted tree of size n



y_{nm} ... number of trees in $\mathcal{Y}_n$ with m nodes in level k

$$y_k(x,u) = \sum_{n\geq 0}\sum_{m\geq 0} y_{nm} z^n u^m$$

The Profile – Combinatorial Setup

$\Longrightarrow$ the generating function $y_k(x,u)$ is given by

$$y_0(x,u) = uy(x)$$
$$y_{k+1}(x,u) = z \exp\left(y_k(x,u) + \sum_{i\ge 2} \frac{y_k(x^i,u^i)}{i}\right),\; k \ge 0$$

char. function of $\frac{1}{\sqrt{n}} L_n(k)$:

$$\phi_{kn}(t) = \frac{1}{y_n}[x^n] y_k\left(x, e^{it/\sqrt{n}}\right)$$

The Profile – Combinatorial Setup

the sequence $y_k(x,u)$ satisfies

$$\lim_{k\to\infty} y_k(x,u) = y(x)$$

hence define

$$w_k(x,u) = y_k(x,u) - y(x)$$

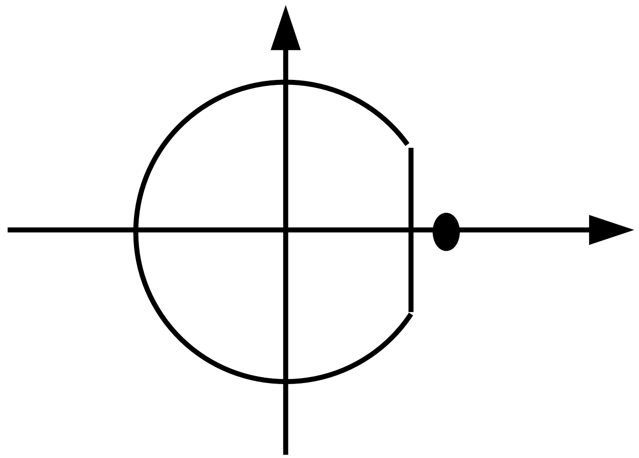
Then

$$\lim_{k\to\infty} w_k(x,u) = 0, \quad w_k(x,1) = 0$$

The Profile

Get more precise information and use

singularity analysis to determine



$$\phi_{kn}(t) = \frac{1}{y_n}[x^n] y_k\left(x, e^{it/\sqrt{n}}\right)$$

$$= \frac{1}{2\pi i y_n} \oint \frac{y_k\left(x, e^{it/\sqrt{n}}\right)}{x^{n+1}}\, dx$$

Notice:

$$y_k(x,u) = y(x) + w_k(x,u)$$

The Profile – Local Behaviour of w_k

Theorem 3 *Set* $x = \rho\left(1 + \frac{s}{n}\right)$, $u = e^{it/\sqrt{n}}$, $k = \kappa\sqrt{n}$. *Assume* $s = \mathcal{O}\left(\log^n\right)$, $t = \mathcal{O}\left(1\right)$, $n \to \infty$.

Then $w_k(x,u)$ *admits the local representation*

$$w_k(x,u) = \frac{y(x)^k w_0}{1 - w_0\left(\frac{1-y(x)^k}{2(1-y(x))} + f_k(x,u)\right)(1 + \mathcal{O}\left(w_0\right))}$$

where

$$f_k(x,u) = \sum_{\ell=0}^{k} \frac{\sum_{i \geq 2} \frac{w_\ell(x^i,u^i)}{i}}{w_\ell(x,u)^2}\, y(x)^\ell$$

$$w_0 = (u-1)y(x), \quad y(x) \sim 1 - b\sqrt{\rho}\sqrt{\rho - x}$$

The Profile – Idea of the Proof

<u>Idea of the Proof</u>

$$w_{k+1}(x,u) = y(x)\left(\exp\left(w_k(z,u) + \sum_{i\ge 2}\frac{w_k(z^i,u^i)}{i}\right) - 1\right)$$

Lemma 1 $|x| \le \rho^2 + \varepsilon$, $|u| = 1 \Longrightarrow$ *there exists* $0 < L < 1$ *with*

$$|w_k(x,u)| \le C|u-1| \cdot |x| \cdot L^k$$

Consequence: sums like

$$\sum_{i\ge 2}\frac{w_\ell(x^i,u^i)}{i}$$

behave relatively nicely.

The Profile – Idea of the Proof

Lemma 2 *Under the assumption of Theorem 2 we have*

$$|w_k(z,u)| \le C_1 |w_0| |y(z)|^k$$

and

$$|w_k(z,u)| \ge C_2 |w_0| |y(z)|^k (1 - C_3\varepsilon)^k$$

$\Longrightarrow w_0(x,u) f_k(x,u)$ is bounded

The Profile – Idea of the Proof

Lemma 2 *Under the assumption of Theorem 2 we have*

$$|w_k(z,u)| \le C_1 |w_0| |y(z)|^k$$

and

$$|w_k(z,u)| \ge C_2 |w_0| |y(z)|^k (1 - C_3 \varepsilon)^k$$

$\Longrightarrow w_0(x,u) f_k(x,u)$ is bounded

and

w_k and f_k have square root type singularities w.r.t. x, in particular,

$$w_0 f_k(x,u) = (u-1) y(x) f_k(x,u) = C_1(x,u) + C_2(x,u) \sqrt{1 - \frac{x}{\rho}}$$

and

$$\lim_{k \to \infty} f_k(x,u) = c_0$$

The Profile – Idea of the Proof

Lemma 3 *For* $k \sim \kappa\sqrt{n}$ *we have* $L_n(k) \xrightarrow{w} A\ell(B\kappa)$.

$$A = \frac{1}{(1-c_0)b\sqrt{2\rho}}, \qquad B = \frac{b\sqrt{\rho}}{2\sqrt{2}}$$

The Profile – Normalisation

We have

$$\mathbf{E}L_n(k) = [x^n]\gamma_k(x)$$

with

$$\gamma_k(x) = \frac{\partial}{\partial u} y_k(x,u)\Big|_{u=1}$$

It can be shown that $\gamma_k(x) \sim C(x)y(x)^k$.

But

$$\mathbf{E}\sum_k L_n(k) = n \qquad \Longrightarrow \qquad C(x) \sim xy'(x)(1-y(x))$$

The Profile – Normalisation

We have

$$\mathbf{E}L_n(k) = [x^n]\gamma_k(x)$$

with

$$\gamma_k(x) = \frac{\partial}{\partial u} y_k(x,u)\Big|_{u=1}$$

It can be shown that $\gamma_k(x) \sim C(x)y(x)^k$.

But

$$\mathbf{E}\sum_k L_n(k) = n \qquad \Longrightarrow \qquad C(x) \sim xy'(x)(1-y(x))$$

$\Longrightarrow$ $\lim_{n\to\infty} \mathbf{E}L_n(k) = \mathbf{E}(A\ell(Bt))$ with $A = B = \frac{b\sqrt{\rho}}{2\sqrt{2}}$

Tightness

Tightness follows from

$$\mathbf{E}\,(L_n(r) - L_n(r+h))^\beta \le C\,(h\sqrt{n})^\alpha$$

for all non-negative integers n, r, h and some $\alpha > 1$, $\beta > 0$

We also know that showing

$$\mathbf{E}\,(L_n(r) - L_n(r+h))^4 \le C\,h^2 n$$

doable for simply generated trees.

Tightness

GF encoding the bivariate distribution of $L_n(k), L_n(k+m)$ is $y_{km}(x,u,v)$ with

$$\tilde{y}_{0,m}(x,u,v) = u y_m(x,v)$$
$$\tilde{y}_{k+1,m}(x,u,v) = x \exp\left(\sum_{i\ge 1} \frac{\tilde{y}_{k,m}(x^i,u^i,v^i)}{i}\right), \qquad k \ge 0.$$

Hence tightness follows from

$$[x^n]\left[\left(\frac{\partial}{\partial u} + 7\frac{\partial^2}{\partial u^2} + 6\frac{\partial^3}{\partial u^3} + \frac{\partial^4}{\partial u^4}\right)\tilde{y}_{r,h}\left(x,u,\frac{1}{u}\right)\right]_{u=1} \le C\,\frac{h^2}{\sqrt{n}}\rho^{-n}$$

and therefore from

$$\left[\left(\frac{\partial}{\partial u} + 7\frac{\partial^2}{\partial u^2} + 6\frac{\partial^3}{\partial u^3} + \frac{\partial^4}{\partial u^4}\right)\tilde{y}_{r,h}\left(x,u,\frac{1}{u}\right)\right]_{u=1} = \mathcal{O}\left(\frac{h^2}{1-|y(x)|}\right)$$

for $x \in \Delta$.

The Height

$y_n^{(k)}$ ... number of trees with n nodes and height at most k

$y_k(x) = \sum_{n\ge1} y_n^{(k)} x^n$

Then

$$y_0(x) = 0$$
$$y_{k+1}(x) = x \exp\left(\sum_{i\ge1} \frac{y_k(x^i)}{i} \right), \qquad k \ge 0.$$

$\Longrightarrow y_k(x) = y_k(x, 0)$

$$w_k(x) := w_k(x, 0) = y(x) - y_k(x) = \sum_{n\ge0} \mathbf{P}\{H_n > k\}\, y_n x^n$$

The Height

Theorem 4 *Let $k = \kappa\sqrt{n}$. Then, as $n \to \infty$, we have*

$$w_k(x) = \frac{-y(x)^k}{\frac{1}{2}\frac{1-y(x)^k}{1-y(x)} + \mathcal{O}\left(\sqrt{k}\right)}$$

uniformly for $|x - \rho| = \mathcal{O}\left(1/\sqrt{n}\right)$ such that $x \in \Delta$.

Thanks to Flajolet, Odlyzko '82
and Flajolet, Gao, Odlyzko, Richmond '95 this is enough.

The Height

Theorem 5 *Let* H_n *denote the height of a tree of size* n*. Then*

$$\mathbf{E}H_n \sim \frac{2\sqrt{\pi}}{b\sqrt{\rho}}\sqrt{n}.$$

Moreover, if we set $\beta = \frac{2\sqrt{n}}{hb\sqrt{\rho}}$*, then*

$$\mathbf{P}\{H_n = h\} = \frac{y_n^{(h)}}{y_n} \sim 4b\sqrt{\frac{\rho\pi^5}{n}}\beta^4 \sum_{m\geq 1} m^2(2(m^2\pi^2\beta^2 - 3)e^{-m^2\pi^2\beta^2}$$

The Height

Theorem 5 *Let* H_n *denote the height of a tree of size* n. *Then*

$$\mathbf{E}H_n \sim \frac{2\sqrt{\pi}}{b\sqrt{\rho}}\sqrt{n}.$$

Moreover, if we set $\beta = \frac{2\sqrt{n}}{hb\sqrt{\rho}}$, *then*

$$\mathbf{P}\left\{H_n = h\right\} = \frac{y_n^{(h)}}{y_n} \sim 4b\sqrt{\frac{\rho\pi^5}{n}}\beta^4 \sum_{m\geq 1} m^2(2(m^2\pi^2\beta^2 - 3)e^{-m^2\pi^2\beta^2}$$

Large deviation results by Broutin & Flajolet 2008

THANK YOU!