INCREASING TREE FAMILIES

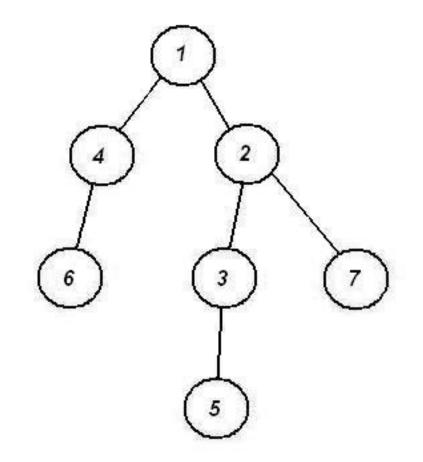
Michael Drmota

Inst. of Discrete Mathematics and Geometry Vienna University of Technology A 1040 Wien, Austria michael.drmota@tuwien.ac.at www.dmg.tuwien.ac.at/drmota/

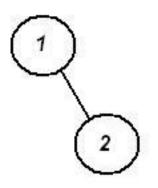
Complex Networks and Random Graphs Physikzentrum Bad Honnef, July 11–13, 2005

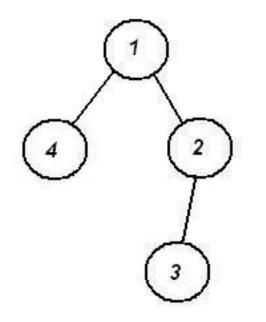
Outline of the Talk

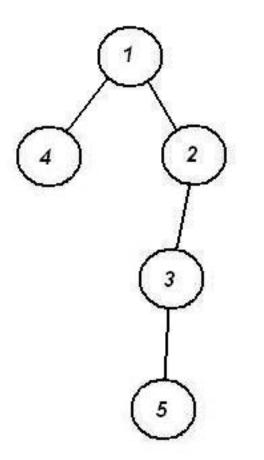
- Recursive Trees
- Plane Oriented Trees
- General Increasing Trees
- Degree Distribution
- Conclusion

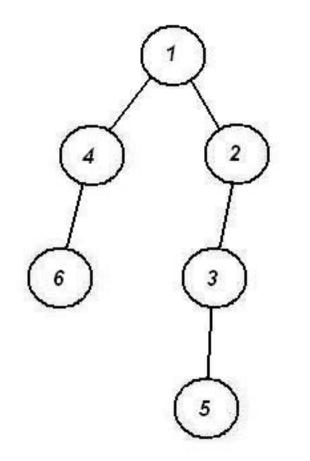


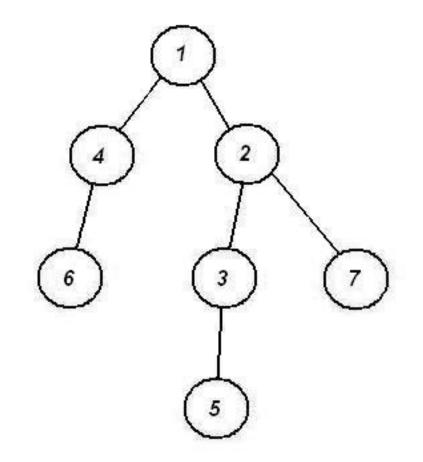












Combinatorial Description:

- labeled rooted tree
- labels are strictly increasing (starting at the root)
- no left-to-right order (non-planar)

Number of Recursive Trees:

$$y_n$$
 = number of recusive trees of size n
= $(n-1)!$

The node with label j has exactly j - 1 possibilities to be inserted $\implies y_n = 1 \cdot 2 \cdots (n - 1).$

Generating Functions:

$$y(x) = \sum_{n \ge 1} y_n \frac{x^n}{n!} = \sum_{n \ge 1} \frac{x^n}{n} = \log \frac{1}{1 - x}$$
$$y'(x) = 1 + y(x) + \frac{y(x)^2}{2!} + \frac{y(x)^3}{3!} + \dots = e^{y(x)}$$
$$R = 0 + \frac{9}{R} + \frac{9}{R} + \frac{9}{R} + \frac{9}{R} + \dots + \frac{9}{R} + \dots$$

A recursive tree can be interpreted as a root followed by an **unordered** sequence of recursive trees. $(y'(x) = \sum_{n \ge 0} y_{n+1}x^n/n!)$

Probability Model:

Process of growing trees

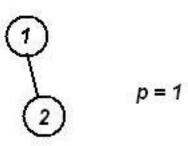
- The process starts with the root that is labeled with 1.
- At step j a new node (with label j) is attached to any previous node with equal probability 1/(j-1).

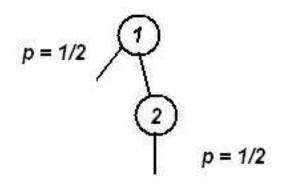
After *n* steps every tree (of size *n*) has equal probability $\frac{1}{(n-1)!}$.

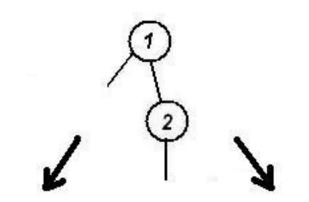
1

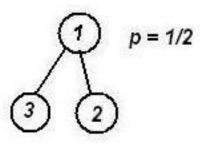


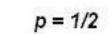
p = 1

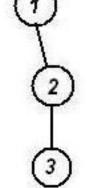




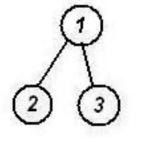




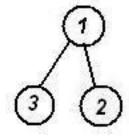




Remark



=



Depth D_n of the *n*-th node [Devroye, 1988; Mahmoud, 1991]

$$E D_n = H_{n-1} = \log n + O(1)$$

 $Var D_n = H_{n-1} - H_{n-1}^{(2)} = \log n + O(1)$

Central limit theorem:

$$\frac{D_n - \log n}{\sqrt{\log n}} \to \mathcal{N}(0, 1)$$

Harmonic numbers: $H_n = 1 + \frac{1}{2} + \dots + \frac{1}{n}$, $H_n^{(2)} = 1 + \frac{1}{2^2} + \dots + \frac{1}{n^2}$

Number L_n of leaves [Najock & Heyde, 1982]

$$E L_n = \frac{n}{2}$$
$$Var L_n = \frac{7}{12}n + \frac{1}{3}$$

Central limit theorem:

$$\frac{L_n - \frac{n}{2}}{\sqrt{\frac{7}{12}n}} \to \mathcal{N}(0, 1)$$

Distribution of out-degrees [Gastwirth, 1977]

$$\lambda_d = \lim_{n \to \infty} \text{probability that a random node of a trees of size } n$$

has out-degree d
$$= \lim_{n \to \infty} \frac{\text{expected number of nodes with out-degree } d}{n}$$

$$= \frac{1}{2^{d+1}}$$

E.g.: Number of leaves = number of nodes with out-degree $0 = \frac{n}{2}$.

Root degree $d_{0,n}$

$$E d_{0,n} = H_{n-1} = \log n + O(1)$$

Var $d_{0,n} = \log n + O(1)$

Central limit theorem:

$$\frac{d_{0,n} - \log n}{\sqrt{\log n}} \to \mathcal{N}(0,1)$$

This result follows from the correspondance to (random) permutations.

Height H_n

[Pittel 1994]

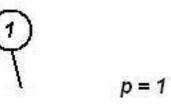
$$\frac{H_n}{\log n} \to e \quad (a.s.)$$

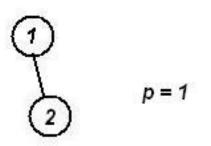
 $\mathbf{E} H_n \sim e \cdot \log n$

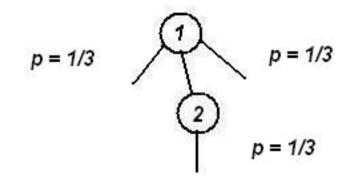
Exponential tails [Drmota 200?]

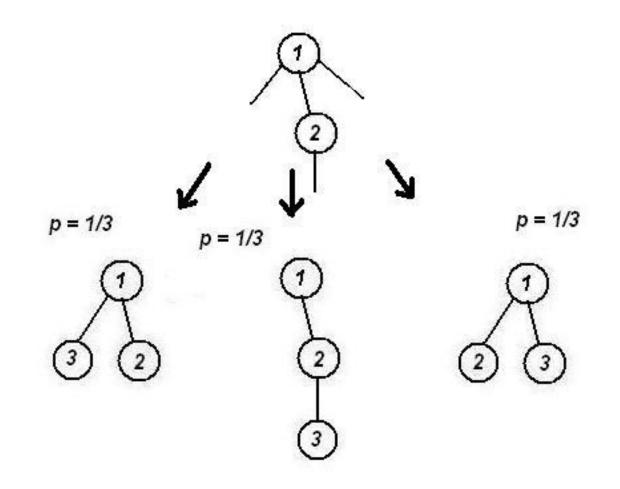
$$\Pr\left\{|H_n - \operatorname{E} H_n| \ge z\right\} = O(e^{-\eta z})$$

for some $\eta > 0$. ($\implies \operatorname{Var} H_n = O(1)$ etc.)

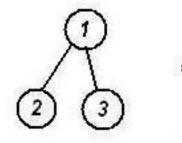




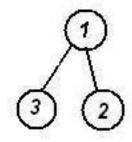




Remark







Number of Plane Oriented Trees:

$$y_n = \text{number of plane oriented trees of size } n$$
$$= 1 \cdot 3 \cdot 5 \cdots (2n - 3) = (2n - 3)!!$$
$$= \frac{(2n - 2)!}{2^{n-1}(n-1)!}$$

The node with label j has exactly 2j - 3 possibilities to be inserted $\implies y_n = 1 \cdot 3 \cdots (2n - 3).$

Generating Functions:

$$y(x) = \sum_{n \ge 1} y_n \frac{x^n}{n!} = \sum_{n \ge 1} \frac{1}{2^{n-1}} {\binom{2(n-1)}{n-1}} \frac{x^n}{n} = 1 - \sqrt{1-2x}$$
$$y'(x) = 1 + y(x) + y(x)^2 + y(x)^3 + \dots = \frac{1}{1-y(x)}$$

A plane oriented tree can be interpreted as a root followed by an **ordered** sequence of plane oriented trees. $(y'(x) = \sum_{n \ge 0} y_{n+1}x^n/n!)$

Probability Model:

Process of growing trees

- The process starts with the root that is labeled with 1.
- At step j a new node (with label j) is attached to any previous node of outdegree d with probability (d + 1)/(2j 3).
 (Barabási-Albert model)

After *n* steps every tree (of size *n*) has equal probability $\frac{1}{(2n-3)!!}$.

Results for Plane Oriented Trees

Depth D_n of the *n*-th node [Mahmoud, 1992]

$$E D_n = H_{2n-1} - \frac{1}{2} H_{n-1} = \frac{1}{2} \log n + O(1)$$

$$Var D_n = H_{2n-1} - \frac{1}{2} H_{n-1} - H_{2n-1}^{(2)} + \frac{1}{4} H_{n-1}^{(2)}$$

$$= \frac{1}{2} \log n + O(1)$$

Central limit theorem:

$$\frac{D_n - \frac{1}{2}\log n}{\sqrt{\frac{1}{2}\log n}} \to \mathcal{N}(0, 1)$$

Results for Plane Oriented Trees

Number L_n of leaves [Mahmoud, Smythe & Szymanski, 1993]

$$\mathbf{E} L_n = \frac{2n-1}{3}$$
$$\mathbf{Var} L_n = \frac{n}{9} - \frac{1}{18} - \frac{1}{6(2n-1)}$$

Central limit theorem:

$$\frac{L_n-\frac{2}{3}n}{\sqrt{\frac{n}{9}}} \to \mathcal{N}(0,1)$$

Distribution of out-degrees [Bergeron, Flajolet & Salvy, 1992]

$$\lambda_{d} = \lim_{n \to \infty} \text{probability that a random node in } \mathcal{P}_{n} \text{ has out-degree } d$$
$$= \lim_{n \to \infty} \frac{\text{expected number of nodes with out-degree } d}{n}$$
$$= \frac{4}{(d+1)(d+2)(s+3)}$$

Remark. $\lambda_d \sim 4 d^{-3}$ as $d \to \infty$.

Root degree $d_{0,n}$ [Bergeron, Flajolet & Salvy, 1992]

$$\Pr\left\{d_{0,n}=k\right\} = \frac{(2n-3-k)!}{2^{n-1-k}(n-1-k)!} \sim \sqrt{\frac{2}{\pi n}} e^{-k^2/(4n)}$$

$$\operatorname{E} d_{0,n} = \sqrt{\pi n} + O(1)$$

Height H_n

[Pittel 1994]

$$\frac{H_n}{\log n} \to \frac{1}{2s} = 1.79556\dots \quad (a.s.),$$

where s = 0.27846... is the positive solution of $se^{s+1} = 1$.

$$\mathbf{E} H_n \sim \frac{1}{2s} \log n$$

Exponential tails [Drmota 200?]

$$\Pr\left\{|H_n - \operatorname{E} H_n| \ge z\right\} = O(e^{-\eta z})$$

for some $\eta > 0$. (\implies Var $H_n = O(1)$ etc.)

Distance E_n between 2 random points

[Bollobas & Riordan, 200?; Morris, Panholzer & Prodinger, 2004]

 $\mathbf{E} E_n = \log n + O(1)$

 $\operatorname{Var} E_n = \log n + O(1)$

Central limit theorem:

$$\frac{E_n - \log n}{\sqrt{\log n}} \to \mathcal{N}(0, 1)$$

[Bergeron, Flajolet & Salvy, 1992]

 \mathcal{P}_n : set of all *plane oriented trees* of size n

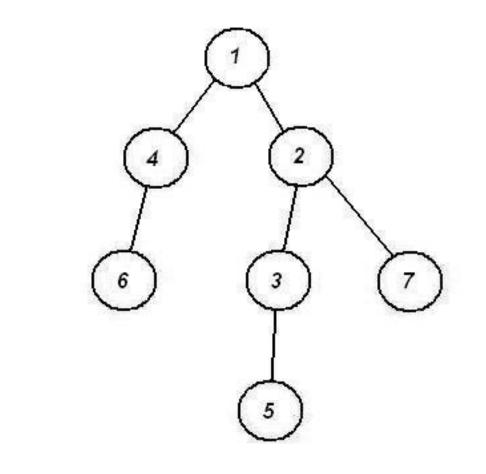
 ϕ_0, ϕ_1, \ldots : weight sequence $(\phi_0 > 0, \phi_j > 0$ for some $j \ge 2)$ $\phi(t) = \phi_0 + \phi_1 t + \phi_2 t^2 + \cdots$

Weight of a tree $T \in \mathcal{P}_n$:

$$\omega(T) = \prod_{j \ge 0} \phi_j^{N_j(T)},$$

where $N_j(T)$ = the number of nodes in T with **outdegree** j.

Т



 $\omega(T) = \phi_0^3 \phi_1^2 \phi_2^2$

Generating Functions:

$$y_n = \sum_{T \in \mathcal{P}_n} \omega(T)$$

$$y(x) = \sum_{n \ge 1} y_n \frac{x^n}{n!}$$

$$y'(x) = \phi_0 + \phi_1 y(x) + \phi_2 y(x)^2 + \dots = \phi(y(x))$$

$$R = 0 + 9 + 8 + 4 + \dots$$
$$R = 0 + 9 + 8 + 8 + \dots$$

Probability distribution on \mathcal{P}_n

For $T \in \mathcal{P}_n$ set:

$$\pi_n(T) := \frac{\omega(T)}{y_n}$$

Remark. In general it is **not** clear whether π_n is induced by **a tree** evolution process. It is just a sequence of probability measures.

Examples

• Recursive Trees:
$$\phi(t) = \sum_{j \ge 0} \frac{t^j}{j!} = e^t$$
, $\phi_j = \frac{1}{j!}$

The factor 1/j! "reduces" planar trees to non-planar ones.

- Plane Oriented Trees: $\phi(t) = 1 + t + t^2 + \dots = \frac{1}{1-t}$, $\phi_j = 1$
- Binary Search Trees: $\phi(t) = (1+t)^2$, $\phi_0 = 1$, $\phi_1 = 2$, $\phi_2 = 1$.

For all these three examples, π_n is induced by a tree evolution process.

Theorem [Panholzer & Prodinger, 200?]

The sequence π_n of probability measures on \mathcal{P}_n is induced by a tree evolution process if and only if $\phi(t)$ has one of the three forms:

•
$$\phi(t) = \phi_0 \left(1 + \frac{\phi_1}{D\phi_0} t \right)^D$$
 for some $D \in \{2, 3, ...\}$ and $\phi_0 > 0, \phi_1 > 0.$

•
$$\phi(t) = \phi_0 e^{\frac{\phi_1}{\phi_0}t}$$
 with $\phi_0 > 0$, $\phi_1 > 0$.

•
$$\phi(t) = \frac{\phi_0}{\left(1 - \frac{\phi_1}{r\phi_0}t\right)^r}$$
 for some $r > 0$ and $\phi_0 > 0$, $\phi_1 > 0$.

Probabilistic tree evolution model

- The process starts with the root that is labeled with 1.
- At step j a new node (with label j) is attached to any previous node (with out-degree d) with probability **proportional** to $\frac{(d+1)\phi_{d+1}\phi_0}{d}$

$$\phi_d$$

In order to obtain all possible π_n it is sufficient to work with $\phi_0 = \phi_1 = 1$:

 $\phi(t) = (1+t)^D, \quad \phi(t) = e^t, \quad \phi(t) = 1/(1-t)^r$

Recursive Trees: $\phi(t) = e^t$

$$\phi_d = \frac{1}{d!} \implies \frac{(d+1)\phi_{d+1}\phi_0}{\phi_d} = 1$$

A new node is attached to previous nodes with equal probability.

Generalized Plane Oriented Trees: $\phi(t) = 1/(1-t)^r$ for some r > 0

$$\phi_d = \binom{r+d-1}{d} \implies \frac{(d+1)\phi_{d+1}\phi_0}{\phi_d} = d+r$$

A new node is attached to a previous nodes with probability **propor**tional to d + r, where d is the **out-degree**.

For r = 1 this these are (usual) plane oriented trees.

Theorem

Let $\phi(t) = 1/(1-t)^r$ for some r > 0 and set

 $\lambda_d = \lim_{n \to \infty} \text{probability that a random node in } \mathcal{P}_n \text{ has out-degree } d$ $= \lim_{n \to \infty} \frac{\text{expected number of nodes with out-degree } d}{n}$

Then

$$\lambda_d = \frac{(r+1)\Gamma(2r+1)\Gamma(r+d)}{\Gamma(r)\Gamma(2r+d+2)}.$$

In particular

$$\lambda_d \sim \frac{(r+1)\Gamma(2r+1)}{\Gamma(r)} \cdot d^{-2-r}$$

Remark 1

There is also a **central limit theorem** for each d.

Remark 2

This result is in accordance to [Dorogovtsev, Mendes & Samukhin, 2000] and [Buckley & Osthus, 2004]. (r = A, m = 1)

Generating Functions

 $N_d(T)$ number of nodes of T with out-degree d $N_{d,n}$ (random) number of nodes in \mathcal{P}_n with out-degree d.

$$y_{n,k} = \sum_{T \in \mathcal{P}_n, N_d(T) = k} \omega(T) = y_n \cdot \Pr(N_{d,n} = k)$$

$$y(x,u) = \sum_{n,k} y_{n,k} u^k \frac{x^n}{n!} = \sum_T \omega(T) \frac{x^{|T|}}{|T|!} u^{N_d(T)}$$
$$= \sum_{n \ge 1} y_n \cdot \mathbf{E} u^{N_{d,n}} \cdot \frac{x^n}{n!}$$

Generating Functions

$$\frac{\partial y(x,u)}{\partial x} = \phi(y(x,u)) + \phi_d \cdot (u-1) \cdot y(x,u)^d$$

d = 3:

$$\phi(y) + \phi_d \cdot (u-1) \cdot y^d = \phi_0 + \phi_1 y + \phi_2 y^2 + u \phi_3 y^3 + \phi_4 y^4 + \cdots$$

Expected value of nodes of degree d

 $N_{d,n}$ denotes the random variable that counts the number of nodes in \mathcal{P}_n with out-degree d.

$$\implies y(x,u) = \sum_{n,k} y_{n,k} \frac{x^n}{n!} u^k = \sum_{n \ge 1} y_n \cdot \mathbf{E} \, u^{N_{d,n}} \cdot \frac{x^n}{n!}$$

$$\implies \frac{\partial y(x,u)}{\partial u}\Big|_{u=1} = \sum_{n\geq 1} y_n \cdot \mathbf{E} N_{d,n} \cdot \frac{x^n}{n!}$$

Expected value of nodes of degree \boldsymbol{d}

Set
$$S(x) = \frac{\partial y(x,u)}{\partial u}\Big|_{u=1} = \sum_{n\geq 1} y_n \cdot \mathbf{E} N_{d,n} \cdot \frac{x^n}{n!}$$
.

Recall that
$$\frac{\partial y(x,u)}{\partial x} = \phi(y(x,u)) + \phi_d(u-1)y(x,u)^d$$

$$\implies S'(x) = \phi'(y(x))S(x) + \phi_d y(x)^d$$

$$\implies \qquad S(x) = \phi_d y'(x) \int_0^x \frac{y(t)^d}{y'(t)} dt$$

Singular behaviour of y(x)

 $\phi(t) = 1/(1-t)^r$, $y'(x) = \phi(y(x))$

$$\implies \qquad y(x) = 1 - (1 - (r+1)x)^{\frac{1}{r+1}}$$

 $x_0 = \frac{1}{r+1}$ is the (only) singularity of y(x).

$$\implies \frac{y_n}{n!} = (-1)^{n-1} (r+1)^n {\binom{\frac{1}{r+1}}{n}} \sim \frac{-1}{\Gamma(-\frac{1}{r+1})} (r+1)^n n^{-\frac{r+2}{r+1}}$$

Lemma

Suppose that

$$y(x) = (1-x)^{-\alpha}$$

Then

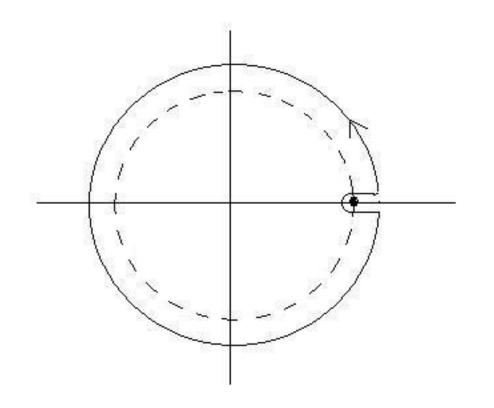
$$y_n = (-1)^n \binom{-\alpha}{n} = \frac{n^{\alpha-1}}{\Gamma(\alpha)} + \mathcal{O}\left(n^{\alpha-2}\right).$$

Remark. This lemma applies in the previous situation for $\alpha = -\frac{1}{r+1}$.

Proof.

Cauchy's formula:

$$(-1)^n \binom{-\alpha}{n} = \frac{1}{2\pi i} \int_{\gamma} (1-x)^{-\alpha} x^{-n-1} dx.$$



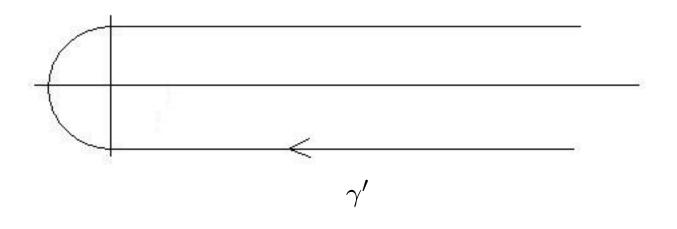
More precisely ...

$$\begin{split} \gamma &= \gamma_1 \cup \gamma_2 \cup \gamma_3 \cup \gamma_4 \\ \gamma_1 &= \left\{ x = 1 + \frac{t}{n} \middle| |t| = 1, \Re t \le 0 \right\} \\ \gamma_2 &= \left\{ x = 1 + \frac{t}{n} \middle| 0 < \Re t \le \log^2 n, \Im t = 1 \right\} \\ \gamma_3 &= \overline{\gamma_2} \\ \gamma_4 &= \left\{ x \middle| |x| = \left| 1 + \frac{\log^2 n + i}{n} \right|, \arg(1 + \frac{\log^2 n + i}{n}) \le |\arg(x)| \le \pi \right\}. \end{split}$$

Substitution for $x \in \gamma_1 \cup \gamma_2 \cup \gamma_3$:

$$x = 1 + \frac{t}{n} \Longrightarrow x^{-n-1} = e^{-t} \left(1 + \mathcal{O}\left(\frac{t^2}{n}\right) \right)$$

 $\implies t \in \gamma' = \{t \mid |t| = 1, \Re t \le 0\} \cup \{t \mid 0 < \Re t \le \log^2 n, \Im t = \pm 1\}:$



With Hankel's integral representation for $1/\Gamma(\alpha)$

$$\frac{1}{2\pi i} \int_{\gamma_1 \cup \gamma_2 \cup \gamma_3} (1-x)^{-\alpha} x^{-n-1} dx = \frac{n^{\alpha-1}}{2\pi i} \int_{\gamma'} (-t)^{-\alpha} e^{-t} dt$$
$$+ \frac{n^{\alpha-2}}{2\pi i} \int_{\gamma'} (-t)^{-\alpha} e^{-t} \cdot \mathcal{O}\left(t^2\right) dt$$
$$= n^{\alpha-1} \frac{1}{\Gamma(\alpha)} + \mathcal{O}\left(n^{\alpha-2}\right).$$

Lemma [Flajolet and Odlyzko, 1992]

Let

$$y(x) = \sum_{n \ge 0} y_n x^n$$

be analytic in a region

$$\Delta = \{x : |x| < x_0 + \eta, |\arg(x - x_0)| > \delta\},\$$

 $x_0 > 0, \ \eta > 0, \ 0 < \delta < \pi/2.$

Suppose that for some real α

$$y(x) = \mathcal{O}\left((1 - x/x_0)^{-\alpha}\right) \qquad (x \in \Delta).$$

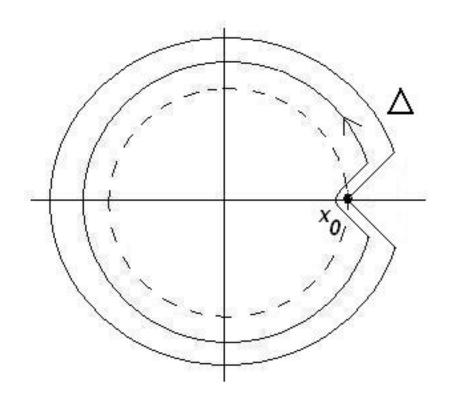
Then

$$y_n = \mathcal{O}\left(x_0^{-n} n^{\alpha-1}\right).$$

Proof

Cauchy's formula:

$$y_{\boldsymbol{n}} = \frac{1}{2\pi i} \int_{\gamma} y(x) \, x^{-\boldsymbol{n}-1} \, dx,$$



Asymptotic Transfer

Suppose that a function y(x) is analytic in a region of the form Δ and that it has an expansion of the form

$$y(x) = C\left(1 - \frac{x}{x_0}\right)^{-\alpha} + O\left(\left(1 - \frac{x}{x_0}\right)^{-\beta}\right) \qquad (x \in \Delta),$$

where $\beta < \alpha$. Then we have (as $n \to \infty$)

$$y_n = [x^n]y(x) = C\frac{n^{\alpha-1}}{\Gamma(\alpha)}x_0^{-n} + \mathcal{O}\left(x_0^{-n}n^{\max\{\alpha-2,\beta-1\}}\right).$$

The Degree Distribution (cont.)

Singular behaviour of $S(x) = \sum_{n \ge 1} y_n \mathbf{E} N_{d,n} \frac{x^n}{n!} = \phi_d y'(x) \int_0^x \frac{y(t)^d}{y'(t)} dt$

•
$$y'(x) = \frac{1}{(1 - (r+1)x)^{1 - \frac{1}{r+1}}}$$

•
$$\int_0^x \frac{y(t)^d}{y'(t)} dt = \int_0^{1/(r+1)} \frac{y(t)^d}{y'(t)} dt + O\left((1 - (r+1)x)^{\frac{1}{r+1}}\right)$$

$$\implies S(x) = \phi_d \int_0^{1/(r+1)} \frac{y(t)^d}{y'(t)} dt \cdot \frac{1}{(1-(r+1)x)^{1-\frac{1}{r+1}}} + O\left((1-(r+1)x)^{\frac{2}{r+1}-1}\right)$$

The Degree Distribution (cont.)

$$\implies \frac{y_n}{n!} \cdot \mathbf{E} N_{d,n} = \phi_d \int_0^{1/(r+1)} \frac{y(t)^d}{y'(t)} dt \cdot (-1)^n (r+1)^n {\binom{\frac{1}{r+1} - 1}{n}} + O\left((r+1)^n n^{-\frac{2}{r+1}}\right)$$

Recall:
$$\frac{y_n}{n!} \sim \frac{-1}{\Gamma(-\frac{1}{r+1})} (r+1)^n n^{-\frac{r+2}{r+1}}$$

$$\implies \mathbf{E} N_{d,n} = \phi_d \cdot (r+1) \int_0^{1/(r+1)} \frac{y(t)^d}{y'(t)} dt \cdot \mathbf{n} + O\left(n^{1-\frac{1}{r+1}}\right)$$

The Degree Distribution (cont.)

$$\implies \lambda_d = \lim_{n \to \infty} \frac{\mathbb{E} N_{d,n}}{n}$$
$$= \phi_d \cdot (r+1) \int_0^{1/(r+1)} \frac{y(t)^d}{y'(t)} dt$$
$$= \frac{(r+1)\Gamma(2r+1)\Gamma(r+d)}{\Gamma(r)\Gamma(2r+d+2)}$$
$$\sim \frac{(r+1)\Gamma(2r+1)}{\Gamma(r)} \cdot d^{-2-r}$$

Conclusion

General Plane Oriented Trees (defined by $\phi(t) = \frac{1}{(1-t)^r}$) have the following properties:

- Average distance between 2 nodes is of order log n (+ central limit theorem)
- Height is order log n (+ exponential tails)
- Degree distribution is of the form $\lambda_d \sim cd^{-2-r}$ (scale free)

There is **no clustering** (it is a tree !!!) but these kind of trees can be used as a **prototype for scale free random graphs** (trees) where several asymptotic properties can be proved rigorously. Thank You!