Width of a scale-free tree

Zsolt Katona

Eötvös Loránd University
Department of Probability Theory and Statistics
zskatona@cs.elte.hu

Abstract

Consider the random graph model of Barabási and Albert, where we add a new vertex in every step and connect it to some old vertices with probabilities proportional to their degrees. If we connect it to only one of the old vertices this will be a tree. These graphs have been shown to have a power law degree distribution, the same as observed in some large real-world networks. We are interested in the width of the tree and we show that it is $W_n \sim n/\sqrt{\pi \log n}$ and this also holds for a slight generalization of the model with another constant. Then we see how this theoretical result can be applied to directory trees.

AMS classification: 05C80, 60C05

Keywords: random graphs, scale-free disribution, width of trees

1 Introduction

Consider the following random graph model of Barabási and Albert [?].

"Starting with a small number (m_0) of vertices, at every time step we add a new vertex with $m(\leq m_0)$ edges that link the new vertex to m different vertices already present in the system. To incorporate preferential attachment, we assume that the probability P that a new vertex will be connected to vertex i depends on the degree of that vertex."

They pointed out that many complex real world networks cannot be adequately described by the classical Erdős-Rényi random graph model, where the possible edges are included independently, with the same probability p. In this model the degree distribution is approximately Poisson with parameter np, while in real networks (for example the WWW) power law degree distributions have been observed, with a parameter independent of n. These are called scale-free degree distributions (or scale-free graphs).

The original definition may lead to different precise definitions. One is by Bollobás and Riordan in [?] and another, for trees ($m = 1$), by Móri in [?]. We use the latter model. Starting with a single point, at every step we add a new vertex and connect it to one of the old vertices by an edge. This old vertex is chosen randomly with probability proportional to its degree. This leads to the same model as if we chose an edge randomly, each with equal probability, then one of the endpoints of that edge. The tree is also known as *plane oriented recursive tree* or *ordered recursive tree.* The reader might be interested in the survey [**?**] about recursive trees.

A possible generalization of this model is where the probability of choosing an old vertex is $(k+\beta)/s_n$, instead of $k/2n$, with a given $\beta > -1$, where k is the degree of the vertex and $s_n = 2n + \beta(n+1) = (2+\beta)n + \beta$ is the sum of all weights in the n-th step. It was shown by Móri in [**?**] that the proportion of vertices of degree k converges almost surely to a limit c_k, which, as a function of k, decreases at the rate $k^{-(3+\beta)}$. Similar a.s. results were proved in a paper of Bollobás, Riordan, Spencer and Tusnády for their model (see [**?**]).

The following examples show the importance of these results and the sence of the generalization. Several graphs have been found with degree distributions $P(k) \sim c \cdot k^{-\gamma}$ with some constant c [**?**]. One of these is the collaboration graph of movie actors where $\gamma = 2.3 \pm 0.1$. Another example is the WWW, which is a directed graph, so it has an in-degree distribution and an out-degree distribution. The Hungarian Web was studied by the Websearch and Data Mining Group of the Hungarian Computer and Automation Research Institute in [**?**]; and they found that both for the in- and out-degrees the distribution is $P(k) \sim c \cdot k^{-\gamma}$ with $\gamma_{\text{in}} = 2.29$ and $\gamma_{out} = 2.78$.

In this paper we study the shape of the tree. Starting from the root (0th level) cut the tree into levels. The neighbours of the root will be on level 1, the neighbours of these will be on level 2, etc. Let $X[n,k]$ denote the size of the k-th level after the n-th step (the first step is when we take the first edge). These random variables determine the shape of the tree. Let $W_n := \max\{X[n,k] : 1 \leq k\}$ be its *width* and $H_n := \max\{k \geq 1 : X[n,k] \neq 0\}$ its *height*.

The diameter studied in [**?**] is in close connection with H_n. The results there yield that the height of our original tree ($\beta = 0$) is asymptotically $\mathcal{O}(\log n)$. On the other hand, Pittel proved in [**?**] that a.s.

$$\lim_{n\to\infty} \frac{H_n}{\log n} = \frac{1}{(2+\beta)y}$$

a.s., where y satisfies $(1+\beta)ye^{1+y} = 1$.

Our goal is to determine the width. We use the method of Chauvin, Drmota and Jabbour-Hattab [**?**], which they applied to binary search trees for the proof of $W_n \sim \frac{n}{\sqrt{4\pi \log n}}$. The main results of the present paper are the following. Set $\alpha = \frac{1+\beta}{2+\beta}$.

Theorem 1 *With probability* 1,

$$X[n,k] = \frac{n}{\sqrt{2\alpha\pi\log n}} \cdot \exp\left(-\frac{(k-\alpha\log n)^2}{2\alpha\log n}\right) + \mathcal{O}\left(\frac{n}{\log n}\right),$$

as $n\to\infty$, *where the error term is uniform for all* $k \geq 0$.

Corollary 1 *As* $n\to\infty$, *we have a.s.*

$$W_n = \frac{n}{\sqrt{2\alpha\pi\log n}}\left(1+\mathcal{O}\left(\frac{1}{\sqrt{\log n}}\right)\right).$$

In addition our results also yield that the width is reached at about level $\alpha\log n$.

In the next section we will introduce the way of using martingales for the proof, which is postponed to Section 3.

In Section **??** we will see an application. V. Batagelj called my attention that directory trees may have the power law degree distribution property. We will study some of them and see how their widths can be approximated by applying Theorem 1.

2 Using martingales

First, introduce the notation

$$Y[n,k] = X[n,k+1] + (1+\beta)X[n,k] \text{ for } k > 1,$$

$$Y[n,0] = X[n,1] + \beta,$$

for the sum of weights on the level k. Our basic tool is the study of the following series of complex generating functions

$$G_n(z) = \sum_{k=0}^{\infty} Y[n,k]z^k.$$

Let $\mathcal{F}_n$ denote the natural σ-field generated by the first n steps.

Lemma 1 *For any fixed* $z\in\mathbb{C}$ *the sequence*

$$M_n(z) := \frac{G_n(z)}{E_n(z)}$$

is a martingale with respect to the filtration $\mathcal{F}_n$, *where*

$$E_n(z) = \prod_{j=1}^{n-1}\frac{S_j+1+(1+\beta)z}{S_j}.$$

PROOF: Easy calculation gives that

$$\mathbf{E}(Y[n+1,0]|\mathcal{F}_n) = Y[n,0]\frac{s_n+1}{s_n},$$

and for $k>0$

$$\mathbf{E}(Y[n+1,k]|\mathcal{F}_n) = Y[n,k]\frac{s_n+1}{s_n} + Y[n,k-1]\frac{1+\beta}{s_n}.$$

These yield

$$\mathbf{E}(G_{n+1}(z)|\mathcal{F}_n) = \frac{s_n+1}{s_n}G_n(z) + \frac{1+\beta}{s_n}zG_n(z) = \frac{s_n+1+(1+\beta)z}{s_n}G_n(z),$$

thus the expectation

$$\mathbf{E}G_n(z) = (1+\beta)(1+z)\prod_{j=1}^{n-1}\frac{S_j+1+(1+\beta)z}{S_j} = (1+\beta)(1+z)E_n(z),$$

since $G_1(z) = (1+\beta)(1+z)$. Hence $M_n(z)$ is a martingale. □

The next lemma is about the asymptotics of the expectation.

Lemma 2 *For any compact set of complex numbers $C \subset \mathbb{C}$ we have*

$$\mathbf{E}G_n(z) = \frac{n^{1+\alpha(z-1)}(1+\beta)(1+z)\Gamma(2\alpha)}{\Gamma(1+\alpha(1+z))} + \mathcal{O}\left(n^{\alpha\Re(z-1)}\right),$$

$$E_n(z) = \frac{n^{1+\alpha(z-1)}\Gamma(2\alpha)}{\Gamma(1+\alpha(1+z))} + \mathcal{O}\left(n^{\alpha\Re(z-1)}\right)$$

uniformly for $z \in C$, as $n \to \infty$.

PROOF: As it is in the previous lemma's proof,

$$\mathbf{E}G_n(z) = (1+\beta)(1+z)\prod_{j=1}^{n-1}\frac{S_j+1+(1+\beta)z}{S_j} = (1+\beta)(1+z)\prod_{j=1}^{n-1}\frac{j+\alpha(1+z)}{j+2\alpha-1},$$

The product is equal to

$$\frac{\Gamma(n+\alpha(1+z))}{\Gamma(1+\alpha(1+z))}\frac{\Gamma(2\alpha)}{\Gamma(n+2\alpha-1)}$$

Its asymptotitcs can be determined as in [**?**] and [**?**], proving that

$$\frac{\Gamma(n+z)}{\Gamma(n)} = n^z + \mathcal{O}(n^{\Re z-1})$$

uniformly over any compact set. It yields, that

$$E_n(z) = \frac{n^{1+\alpha(z-1)}\Gamma(2\alpha)}{\Gamma(1+\alpha(1+z))} + \mathcal{O}\left(n^{\alpha\Re(z-1)}\right)$$

uniformly in any compact set, as $n \to \infty$. □

Next, we are going to study the convergence of the martingale $M_n(z)$. On this purpose, we determine the covariance function of $G_n(z)$.

Lemma 3 *For every pair $z_1, z_2 \in \mathbb{C}$, we have*

$$C_{n+1}^G(z_1,z_2) := \mathbf{E}(G_{n+1}(z_1)G_{n+1}(z_2)) =$$
$$= \sum_{j=1}^{n}\left(b_j(z_1,z_2)\prod_{k=j+1}^{n}a_k(z_1,z_2)\right) + (1+\beta)^2(1+z_1)(1+z_2)\prod_{j=1}^{n}a_j(z_1,z_2),$$

with

$$a_k(z_1,z_2) = 1 + \frac{2+(1+\beta)(z_1+z_2)}{S_k},\ b_k(z_1,z_2) = \frac{(1+z_1+z_1\beta)(1+z_2+z_2\beta)}{S_k}\mathbf{E}G_k(z_1z_2).$$

PROOF:

We give a linear recursion for $C_n^G(z_1,z_2)$. Let k_n+1 denote the level of the vertex added in the $(n+1)$-st step. With this notation $G_{n+1}(z) - G_n(z) = z^{k_n}(1+(1+\beta)z)$. Thus

$$C_{n+1}^G(z_1,z_2) = \mathbf{E}\left[\mathbf{E}\left((G_n(z_1)+z_1^{k_n}(1+z_1+z_1\beta))(G_n(z_2)+z_2^{k_n}(1+z_2+z_2\beta))|\mathcal{F}_n\right)\right] =$$
$$= C_n^G(z_1,z_2) + \mathbf{E}\left[\mathbf{E}\left(G_n(z_1)z_2^{k_n}(1+z_2+z_2\beta) + z_1^{k_n}(1+z_1+z_1\beta)G_n(z_2) + \right.\right.$$
$$\left.\left. + z_1^{k_n}z_2^{k_n}(1+z_1+z_1\beta)(1+z_2+z_2\beta)|\mathcal{F}_n\right)\right].$$

The conditional distribution of k_n w.r.t. $\mathcal{F}_n$ is the following

$$P(k_n = k|\mathcal{F}_n) = \begin{cases} \frac{Y[n,k]}{s_n}, & \text{if } k>0,\\ \frac{Y[n,0]}{s_n}, & \text{if } k=0.\end{cases}$$

Hence, the conditional expectation is

$$\mathbf{E}(G_n(z_1)z_2^{k_n}(1+z_2+z_2\beta)|\mathcal{F}_n) = \frac{1+z_2+z_2\beta}{s_n}G_n(z_1)G_n(z_2).$$

Similarly we have

$$\mathbf{E}(G_n(z_2)z_1^{k_n}(1+z_1+z_1\beta)|\mathcal{F}_n) = \frac{1+z_1+z_1\beta}{s_n}G_n(z_1)G_n(z_2),$$

finally, this yields

$$\mathbf{E}(z_1^{k_n}z_2^{k_n}(1+z_1+z_1\beta)(1+z_2+z_2\beta)|\mathcal{F}_n) = \frac{(1+z_1+z_1\beta)(1+z_2+z_2\beta)}{s_n}G_n(z_1z_2).$$

Hence

$$C^G_{n+1}(z_1,z_2) = \left(1+\frac{2+(1+\beta)(z_1+z_2)}{s_n}\right)C^G_n(z_1,z_2)+ \\ +\frac{(1+z_1+z_1\beta)(1+z_2+z_2\beta)}{s_n}\mathbf{E}G_n(z_1z_2).$$

This proves the lemma since $C^G_1(z_1,z_2)=(1+\beta)^2(1+z_1)(1+z_2)$. □

Corollary 2 *$\{M_n(z): n\in\mathbb{N}\}$ is bounded in L^2 for any fixed $|z-1|<\sqrt{1/\alpha}$. Thus, there exists a random variable $M(z)\in L^2$ such that $M_n(z)\to M(z)$ a.s. in L^2 as $n\to\infty$ for $z\in\mathcal{H}:=\{w\in\mathbb{C}: |w-1|<\sqrt{1/\alpha}\}$.*

PROOF: Using the notations of Lemma 3, we have

$$\prod_{k=j+1}^{n} a_k(z_1,z_2) = \left(\frac{n}{j}\right)^{2+\alpha(z_1+z_2-2)}\left(1+\mathcal{O}\left(\frac{1}{j}\right)\right).$$

We write $A_n \ll B_n$ if there is a constant $c>0$ such that $A_n\le cB_n$ for every n. By Lemma 2,

$$C^G_n(z_1,z_2) = \frac{(1+z_1+z_1\beta)(1+z_2+z_2\beta)}{2+\beta}\sum_{j=1}^{n}\frac{\mathbf{E}G_j(z_1,z_2)}{j+2\alpha-1}\prod_{j=k+1}^{n}a_k(z_1,z_2)+ \\ +(1+\beta^2)(1+z_1)(1+z_2)\prod_{j=1}^{n}a_k(z_1,z_2)\ll \\ \ll\sum_{j=1}^{n} j^{\alpha(\Re z_1z_2-1)}\left(\frac{n}{j}\right)^{2+\alpha\Re(z_1+z_2-2)}+n^{2+\alpha\Re(z_1+z_2-2)}\ll \\ \ll n^{2+\alpha\Re(z_1+z_2-2)}\sum_{j=1}^{n} j^{-(2+\alpha\Re(z_1+z_2-z_1z_2-1)}.$$

Hence,

$$C^M_n(z_1,z_2) := \mathbf{E}(M_n(z_1)M_n(z_2)) = \frac{\mathbf{E}(G_n(z_1)G_n(z_2)}{\mathbf{E}G_n(z_1)\mathbf{E}G_n(z_2)}\ll\sum_{j=1}^{n} j^{-(2+\alpha\Re(z_1+z_2-z_1z_2-1))}.$$

So, if

$$2+\alpha\Re(z+\overline{z}-z\overline{z}-1)>1,$$

then the sum is bounded. The inequality above is true exactly in $\mathcal{H}$, hence $M_n(z)$ is bounded in L^2 for $z\in\mathcal{H}$. □

Also, if $z_1,z_2\in\mathcal{H}$ then $2+\alpha\Re(z_1+z_2-z_1z_2-1)>1$, hence $C^M_n(z_1,z_2)$ converges to some $C^M(z_1,z_2)$ uniformly over the compact subsets of $\mathcal{H}^2$ and $C^M(z_1,z_2)$ is holomorphic over $\mathcal{H}^2$.

To prove the uniform convergence of $M_n(z)$ we follow the lines of [**?**]. The main idea is the following result, which can be proved similarly to Proposition 2 of [**?**].

Proposition 1 *Let $I = (1-\sqrt{1/\alpha}, 1+\sqrt{1/\alpha})$. Then $(M(t))_{t\in I}$ has a continuous modification $\widetilde{M}$ such that for any compact $C \subseteq I$,*

$$\mathbf{E}\left(\sup_{t\in C} |\widetilde{M}|^2\right) < \infty.$$

Generally, if $\gamma : \mathbb{R} \to \mathcal{H}$ is continuously differentiable, then $(M_n(\gamma(t)))_{t\in R}$ has a modification $\widetilde{M}_\gamma$ such that for any compact set $C \subseteq \mathbb{R}$,

$$\mathbf{E}\left(\sup_{t\in C} |\widetilde{M}_\gamma(t)|^2\right) < \infty.$$

□

The uniform convergence of (M_n) comes from the following proposition. The proof being essentially the same as in [**?**] will be omitted.

Proposition 2 *For any compact set $C \subseteq I$, we have $M_n \to M$ uniformly over C and*

$$\mathbf{E}\left(\sup_{t\in C} |M_n(t) - M(t)|^2\right) \to 0,$$

Generally, let $\gamma : \mathbb{R} \to \mathcal{H}$ be continously differentiable and let $M_{n,\gamma}(t) = M_n(\gamma(t))$ and $M_\gamma(t) = M(\gamma(t))$. Then the same result holds for $(M_{n,\gamma})$.

□

Corollary 3 *$M_n(z)$ and all its derivatives converge uniformly over the compact subsets of $\mathcal{H}$.*

PROOF: By Proposition 2 M_n is uniformly convergent over the arc $\gamma(t) = 1 + \rho e^{it}$ for all $0 < \rho < \sqrt{2}$, thus for $|s-1| < \rho$ we have

$$M_n(s) = \frac{1}{2\pi i}\oint_\gamma \frac{M_n(z)}{z-s} dz,$$

by Cauchy's formula. Thus M_n and its derivatives converge uniformly over the compact subsets of $\mathcal{H}$. □

In order to prove theorem 1, we will need two more lemmas on the asymptotics of $G_n(z)$. First, we approximate $\mathbf{E}|G_n(z)|^2$.

Lemma 4 *For every $\delta > 0$ and $|z-1| \leq \sqrt{1/\alpha} - \delta$,*

$$\mathbf{E}|G_n(z)|^2 = \mathcal{O}\left(n^{2(1+\alpha(\Re z-1))}\right).$$

For any z such that $\sqrt{1/\alpha} - \delta \leq |z-1| \leq \sqrt{1/\alpha}$,

$$\mathbf{E}|G_n(z)|^2 = \mathcal{O}\left(n^{2(1+\alpha(\Re z-1))}\log n\right),$$

with uniform error terms as $n \to \infty$. Furthermore, for any compact $C \subseteq \mathbb{C} - \mathcal{H}$,

$$\mathbf{E}|G_n(z)|^2 = \mathcal{O}\left(n^{1+\alpha(|z|^2-1)}\log n\right)$$

uniformly for $z \in C$.

Proof: Recall the proof of Corollary 2. It follows that

$$\mathbf{E}|G_n(z)|^2 \ll n^{2(1+\alpha(\Re z-1))}\sum_{j=1}^{n} j^{(-2-\alpha(2\Re z-|z|^2-1))}.$$

For $|z-1| \leq \sqrt{1/\alpha} - \delta$ the exponent of j is at most $-1-\delta' < -1$, hence

$$\mathbf{E}|G_n(z)|^2 \ll n^{2(1+\alpha(\Re z-1))}\sum_{j=1}^{n} j^{-1-\delta'} \ll n^{2(1+\alpha(\Re z-1))}.$$

On the other hand, for $\sqrt{1/\alpha} - \delta \leq |z-1| \leq \sqrt{1/\alpha}$ we can write

$$\mathbf{E}|G_n(z)|^2 \ll n^{2(1+\alpha(\Re z-1))}\sum_{j=1}^{n} j^{-1-\delta'} \ll n^{2(1+\alpha(\Re z-1))}\log n.$$

In the third case, for $|z-1| > \sqrt{1/\alpha}$ we have

$$\mathbf{E}|G_n(z)|^2 \ll n^{2(1+\alpha(\Re z-1))}\sum_{j=1}^{n} j^{(-2-\alpha(2\Re z-|z|^2-1))} \ll n^{2(1+\alpha(\Re z-1))}\frac{n^{-1-\alpha(2\Re z-|z|^2-1)}}{-1-\alpha(2\Re z-|z|^2-1)}.$$

For the uniform equality we need more. The numerator might tend to 0, so

$$\mathbf{E}|G_n(z)|^2 \ll n^{2(1+\alpha(\Re z-1))}\cdot\frac{(n+1)^{(-1-\alpha(2\Re z-|z|^2-1))}}{-1-\alpha(2\Re z-|z|^2-1)} \ll$$
$$\ll n^{1+\alpha(|z|^2-1)}\cdot\frac{1-e^{(-1-\alpha(2\Re z-|z|^2-1)\log(n+1)}}{-1-\alpha(2\Re z-|z|^2-1)} \ll n^{(1-\alpha)(1+(1+\beta)|z|^2)}\log n$$

This completes the proof.

□

Now we approximate $G_n'(z)$.

Lemma 5 *For every $0 < |z| < 2$, we have a.s.*

$$|G_n'(z)| \ll |z|^{-1} \log n \cdot n^{(1-\alpha)(1+|z|+|z|\beta)}.$$

Proof: Obviously, $|G_n'(z)| \le G_n'(|z|)$. By [?] we now that the height of the tree $H_n \ll \log n$. Hence there exists an n_0, for each realization of the tree, such that $X[n,k] = 0$ a.s. for $n \ge n_0$, if $k > c\log n$. Hence, for sufficiently large n, with probability 1

$$G_n'(|z|) = \sum_{k=1}^{\infty} kY[n,k]|z|^{k-1} \le c\log n \sum_{k=1}^{\infty} Y[n,k]|z|^{k-1} \le c\log n \cdot \frac{G_n(|z|)}{|z|}.$$

□

We need the following lemma to approximate $G_n(z)$ outside $\mathcal{H}$. Since the proof of this lemma follows from Lemma 4 and Lemma 5 the same way as the proof of Proposition 3 in [?], it will be omitted.

Lemma 6 *For any $K > 0$ there exists a $\delta > 0$ such that*

$$\sup_{|z|=1, |z-1| \ge \sqrt{1/\alpha} - \delta} |G_n(z)| = \mathcal{O}\left(\frac{n}{(\log n)^K}\right),$$

a.s., as $n \to \infty$.

Remark 1 *If $\beta = 0$, the same is true for the function $\frac{|G_n(z)|}{|1+z|}$ on*

$$\gamma(\delta) := \{z \,|\, |z| = 1,\, |z-1| \ge \sqrt{2} - \delta,\, \Re z > -0.9\} \cup \{z \,|\, \Re z = -0.9,\, |z| \le 1\}.$$

For any $K > 0$ there exists a $\delta > 0$ such that

$$\sup_{\gamma(\delta)} \left|\frac{G_n(z)}{1+z}\right| = \mathcal{O}\left(\frac{n}{(\log n)^K}\right),$$

a.s., as $n \to \infty$.

Proof: Since γ evades -1 it is enough to approximate $|G_n(z)|$ on $\gamma(\delta)$. From here the proof goes exactly the same as that of the previous lemma. □

3 Proof of Theorem 1

Finally, we can start to prove the theorem. By definition,

$$G_n(z) = \sum_{k=0}^{\infty} Y[n,k] z^k,$$

$$\frac{G_n(z) - \beta}{1 + (1+\beta)z} = \sum_{k=0}^{\infty} X[n,k+1] z^k,$$

if $z \neq -\frac{1}{1+\beta}$. This exception does not matter if $\beta \neq 0$, since $\left|\frac{1}{1+\beta}\right| \neq 1$ and the function can be expanded to this point regularly. We can extract $X[n,k]$ from the generating function by using Cauchy's formula.

If $\beta \neq 0$, then

$$X[n,k+1] = \frac{1}{2\pi i} \int_{|z|=1} \frac{G_n(\xi) - \beta}{(1 + (1+\beta)\xi)\xi^{k+1}} d\xi = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{G_n(e^{it}) - \beta}{1 + (1+\beta)e^{it}} e^{-kit} dt.$$

We split the integral to two parts. Let $\varphi = \min(\pi, \arccos(1 - 1/2\alpha))$, and

$$I_1 := \frac{1}{2\pi} \int_{|t| \leq \varphi - \delta} \frac{G_n(e^{it}) - \beta}{1 + (1+\beta)e^{it}} e^{-kit} dt,$$

$$I_2 := \frac{1}{2\pi} \int_{\pi \geq |t| \geq \varphi - \delta} \frac{G_n(e^{it}) - \beta}{1 + (1+\beta)e^{it}} e^{-kit} dt,$$

where δ is the same as in Lemma 6.

If $\beta = 0$, instead of $|z| = 1$ we integrate on $\gamma = \{\xi \mid |\xi| = 1, \Re\xi > -0.9\} \cup \{\xi \mid \Re\xi = -0.9,\ |\xi| \leq 1\}$. Let I_1 be the same as in the latter case and

$$I_2 := \frac{1}{2\pi i} \int_{\gamma(\delta)} \frac{G_n(\xi)}{(1+\xi)\xi^{k+1}} d\xi,$$

where δ is the same as in Remark 1.

By Lemma 6 and Remark 1, for any $K > 0$ we can approximate the second integral in both cases as follows.

$$|I_2| \leq \frac{1}{2\pi} \int \left| \frac{G_n(\xi) - \beta}{1 + (1+\beta)\xi} \right| d\xi \ll \frac{n}{(\log n)^K}, \tag{1}$$

where we integrate on $\{|\xi| = 1,\ |\xi - 1| \geq \sqrt{1/\alpha} - \delta\}$ in case of $\beta \neq 0$ and on $\gamma(\delta)$ if $\beta = 0$.

For $|t| \leq \varphi - \delta$,

$$M_n(e^{it}) = \frac{G_n(e^{it})}{E_n(e^{it})}$$

is a.s. uniformly bounded by Corollary 3. On the other hand, Lemma 2 provides us the asymptotics of the denominator, hence

$$|G_n(e^{it})| \ll n^{(1-\alpha)(1+(1+\beta)\Re e^{it})} = n \cdot n^{\alpha(\cos t-1)} = n \cdot e^{(\log n)(\cos t-1)\alpha} \ll n e^{-c' t^2 (\log n)}$$

for some constant $c' > 0$. By fixing a sufficiently small positive ϑ we have

$$\frac{1}{2\pi}\int_{(\log n)^{-(1-\vartheta)/2}\le |t|\le \phi-\delta} |G_n(e^{it})|dt \ll n\int_{(\log n)^{-(1-\vartheta)/2}}^{\infty} e^{-c' t^2 \log n}dt \ll n e^{-c'(\log n)^{\vartheta}}. \qquad (2)$$

The remaining part of the integral is

$$I_0 := \frac{1}{2\pi}\int_{|t|\le (\log n)^{-(1-\vartheta)/2}} \frac{G_n(e^{it})}{1+(1+\beta)e^{it}} e^{-kit}dt.$$

Again we are going to use

$$G_n(z) = E_n(z)M_n(z) = \mathbf{E}G_n(z)\frac{M_n(z)}{(1+\beta)(1+z)} \qquad (3)$$

and Lemma 2, which can be written in the form

$$\mathbf{E}G_n(z) = n^{(1-\alpha)(1+z+z\beta)} \cdot \frac{(1+\beta)(1+z)\Gamma(2\alpha)}{\Gamma(1+\alpha(1+z))} + \mathcal{O}\left(n^{(\Re z-1)\alpha}\right) = \\ = n\cdot n^{(z-1)\alpha}\left(\frac{(1+\beta)(1+z)\Gamma(2\alpha)}{\Gamma(1+\alpha(1+z))} + \mathcal{O}\left(\frac{1}{n}\right)\right)$$

uniformly. If $t \to 0$ in such a way, that $|t| \le (\log n)^{-(1-\vartheta)/2}$, then

$$\frac{\mathbf{E}G_n(e^{it})}{1+(1+\beta)z} = n e^{(\log n)(e^{it}-1)\alpha}\left(\frac{(1+\beta)(1+e^{it})\Gamma(2\alpha)}{(1+(1+\beta)e^{it})\Gamma(1+\alpha(1+e^{it}))} + \mathcal{O}\left(\frac{1}{n}\right)\right) = \\ = n e^{-(\alpha t^2/2)\log n + (it\alpha)\log n}\left(1 - it\left(\alpha - \frac{1}{2} + 2\alpha^2\Gamma'(1+2\alpha)\right) - \frac{\alpha t^3}{6} i\log n + \mathcal{O}(t^2 + t^4\log n)\right). \qquad (4)$$

On the other hand, $M_n(1) = 2(1+\beta)$, hence

$$\frac{M_n(e^{it})}{(1+\beta)(1+e^{it})} = 1 + it\frac{M_n'(1)-(1+\beta)}{2(1+\beta)} + \mathcal{O}(t^2). \qquad (5)$$

Then, by (3), (4) and (5) we conclude that, with probability 1,

$$\frac{G_n(e^{it})e^{-kit}}{1+(1+\beta)e^{it}} = n e^{-(\alpha t^2/2)\log n + it(\alpha\log n - k)} \cdot \\ \cdot\left(1 - it\left(\alpha - \frac{1}{2} + 2\alpha^2\Gamma'(1+2\alpha) - \frac{M_n'(1)-(1+\beta)}{2(1+\beta)}\right) - \frac{\alpha t^3}{6} i\log n + \mathcal{O}(t^2+t^4\log n)\right).$$

uniformly with respect to k. For the same reason as in (2), here we also have

$$\int_{|t|\ge (\log n)^{-(1-\vartheta)/2}} e^{-t^2\log n}(1+t+t^3\log n) \ll e^{-(\log n)^{\vartheta}}.$$

Hence

$$\frac{I_0}{n} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-(\alpha t^2/2)\log n + it(\alpha\log n - k)}\cdot$$
$$\cdot\left(1 - it\left(\alpha - \frac{1}{2} + 2\alpha^2\Gamma'(1+2\alpha) - \frac{M_n'(1)-(1+\beta)}{2(1+\beta)}\right) - \frac{\alpha t^3}{6} i\log n\right) dt + \mathcal{O}((\log n)^{-3/2}).$$

Integration gives

$$\frac{I_0}{n} = \frac{1}{\sqrt{2\alpha\pi\log n}}\exp\left(-\frac{((\log n)\alpha - k)^2}{2\alpha\log n}\right)\cdot\left(1 + \frac{((\log n)\alpha - k)}{2\alpha\log n} - \frac{((\log n)\alpha - k)^3}{6\alpha^2(\log n)^2} + \right.$$
$$\left. + \frac{(\log n)\alpha - k}{\alpha\log n}(\alpha - 1/2 + 2\alpha^2\Gamma'(1+2\alpha) - \frac{M_n'(1)-(1+\beta)}{2(1+\beta)})\right) + \mathcal{O}((\log n)^{-3/2}).$$

Hence we have

$$\frac{X[n,k]}{n/\sqrt{2\alpha\pi\log n}} = \exp\left(-\frac{((\log n)\alpha - k)^2}{2\alpha\log n}\right)\cdot\left(1 + \frac{((\log n)\alpha - k)}{2\alpha\log n} - \frac{((\log n)\alpha - k)^3}{6(\alpha\log n)^2} + \right.$$
$$\left. + \frac{(\log n)\alpha - k}{\alpha\log n}(\alpha - 1/2 + 2\alpha^2\Gamma'(1+2\alpha) - \frac{M_n'(1)-(1+\beta)}{2(1+\beta)})\right) + \mathcal{O}\left(\frac{1}{\log n}\right)$$

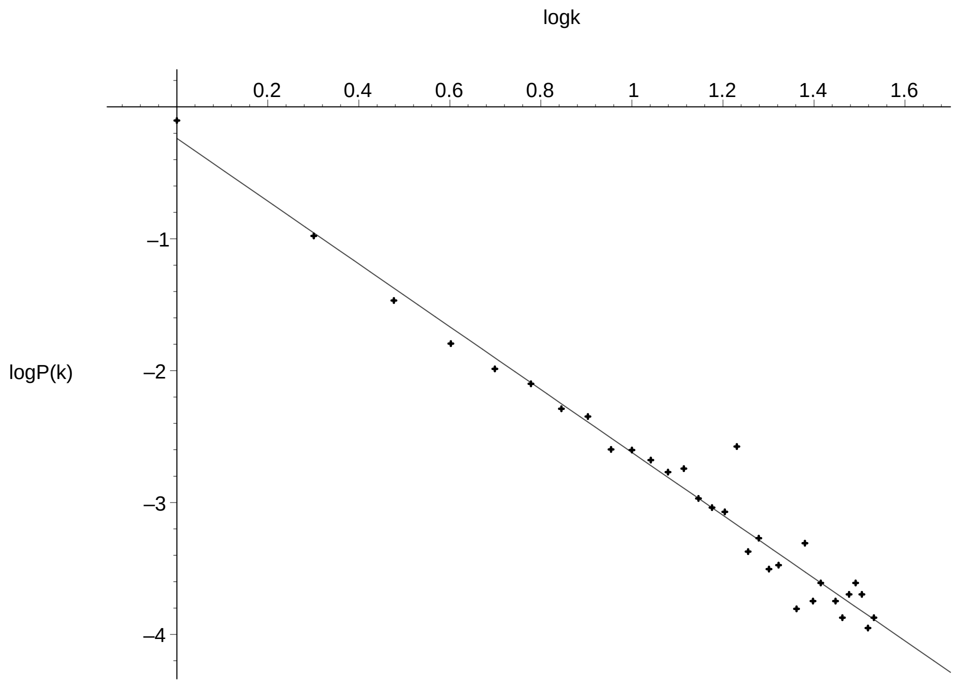
a.s., with an error term uniform in k. This completes the proof.

□

4 Directory trees

Although there are several examples of networks that have power-law degree distributions none of these is a tree. V. Batagelj called my attention to directory trees that should be studied. The following examples all have power-law degree distributions $P(k) \sim c\cdot k^{-\gamma}$ with $2 < \gamma < 3$. This allows to compare the width of the tree with the result of Theorem 1.

The first example is the directory tree of the main server of the Department of Computer Science, Budapest University of Technology. The figure shows the degree distribution with logarithmic scales.



Linear regression gives that $\gamma \approx 2.38$. Substituting $\beta = \gamma - 3 = -0.62$ and $n = 39182$ to Theorem 1 gives 9162 to the width. We can compare this with the real width of the tree which is 10159. We can also calculate the β' that would give the same theoretical width as the real. From Theorem 1 it is $\beta' \approx -0.71$

This table shows the results of studying several directory trees.

Directory tree of	vertices	β	real width	theoretical width	β'
Server of CS Dep., Tech. Univ.	39182	-0.62	10159	9162	-0.71
Server of CS Dep., Eotvos Univ.	18609	-0.96	10916	12519	-0.95
Server of Fazekas High School	48898	-0.25	9721	9071	-0.4
Home Linux	22797	-0.27	4026	4415	+0.04
Home Windows	6999	-0.53	2097	1662	-0.75

The servers are all unix systems with a lot of users who make their directories by their own decisions. This can be a reason why we should consider these graphs random. The ratio of theoretical and real width is between 0.85 and 1.15 in the examples, hence we can approximate the width of directory trees with Theorem 1.

I am indebted to Tamás Móri for his valuable suggestions as my PhD supervisor.

References

[1] BARABÁSI, A.-L., ALBERT, R., Emergence of scaling in random networks, *Science* **286** (1999), 509-512.

[2] BARABÁSI, A.-L., ALBERT, R., HAWOONG JEONG, Mean-field theory for scale-free random networks, *Physica A* **272** (1999) 173-187.

[3] BARABÁSI, A.-L., ALBERT, R., HAWOONG JEONG, Scale-free characteristics of random networks: the topology of the world-wide web, *Physica A* **281** (2000) 69-77.

[4] BENCZÚR, A., CSALOGÁNY, K., FOGARAS, D., FRIEDMAN, E., SARLÓS, T., UHER, M., WINDHAGER, E., Searching a small national domain – Preliminary report *Proceedings of WWW 2003.*

[5] BOLLOBÁS, B. Random Graphs, *Academic Press,* London (1985)

[6] BOLLOBÁS, B., RIORDAN, O., The diameter of a scale-free random graph, *Combinatorica* (2003)

[7] BOLLOBÁS, B., RIORDAN, O., SPENCER, J., TUSNÁDY, G., The degree sequence of a scale-free random graph process, *Random Structures Algorithms* **18** (2001), 279-290.

[8] CHAUVIN, B., DRMOTA, M., JABBOUR-HATTAB, J., The profile of binary search trees, *The Annals of Applied Probability* **11** (2001), 1042-1062.

[9] ERDŐS, P., Some remarks on the theory of graphs, *Bull. Res. Council Israel, Section F* **7** (1947) 292-294.

[10] FLAJOLET, P., ODLYZKO, A.M. Singularity analysis of generating functions, *SIAM J. Discrete Math.* **3** (1990), 216-240

[11] GILBERT, E.N., Random graphs, *Ann. Math. Statist.* **30** (1959), 1141-1144.

[12] JOFFE, A. LE CAM, L., NEVEU, J. Sur la loi des grands nombres pour des variables aléatoires attachées à un arbre dyadique, *C. R. Acad. Sci. Paris. Ser. A* **277** (1973), 963-964.

[13] MÓRI, T.F.,On random trees, *Studia Sci. Math. Hungar.* **39** (2002), 143-155.

[14] MÓRI, T.F.,The maximum degree of the Barabási random tree, *Combinatorics, Probability & Computing* (megjelenés alatt)

[15] NEVEU, J., Martingales à temps discret, *Masson and Cie, Paris* (1972)

[16] RÉNYI, A., SZEKERES, G., On the height of trees, *J. Austral. Math. Soc.* **7** (1967) 497-507.

[17] PITTEL, B. Note on the heights of random recursive trees and random m-ary search trees, *Random Structures Algorithms* **5** (1994), no.2, 337-347.

[18] REVUZ, D., YOR, M., Continuous Martingales and Brownian Motion, *Springer, New York* (1990)

[19] SMYTHE, R. T., MAHMOUD, H. M., A survey of recursive trees, *Teor.Ĭmovīr. Mat. Stat.,* **51** (1994), 1-29.

[20] SZELE, T., Combinatorial investigations concerning complete directed graphs, *Matematikai és Fizikai Lapok* **50** (1943), 287-293.

[21] WATTS, D.J., STROGATZ, S.H., Collective dynamics of 'small-worlds' networks, *Nature* **393** (1998), 440-442.