WIDTH OF A SCALE-FREE TREE

ZSOLT KATONA,* *Eötvös Loránd University*

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 then this will be a tree. These graphs have been shown to have a power-law degree distribution, the same as that observed in some large real-world networks. We are interested in the width of the tree and we show that it is $W_n \sim n/\sqrt{\pi \log n}$ at the nth step; this also holds for a slight generalization of the model with another constant. We then see how this theoretical result can be applied to directory trees.

Keywords: Random graph; scale-free distribution; width of trees

2000 Mathematics Subject Classification: Primary 05C80; 60C05

1. Introduction

Consider the following random graph model of Barabási and Albert [1].

> 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.

Barabási and Albert pointed out that many complex real-world networks cannot be adequately described by the classical Erdös–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 World Wide Web), power-law degree distributions have been observed, with a parameter independent of the number of steps n. These are called scale-free degree distributions (or scale-free graphs).

The original definition may lead to different precise definitions; see, for example, Bollobás and Riordan [3] and for trees (i.e. $m = 1$), see Móri [7]. 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 had chosen an edge randomly, each with equal probability, then chosen one of the endpoints of that edge. This tree is also known as a *plane oriented recursive tree* or an *ordered recursive tree*. The reader might be interested in the survey [9] of 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 nth step. It was shown in [7] that the proportion of vertices of degree k converges almost surely (a.s.) 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 [4].

The following examples show the importance of these results and the sense of the generalization. Several graphs have been found with degree distributions $P(k) \sim ck^{-\gamma}$, where c is a constant; see [1]. One of these is the collaboration graph of movie actors (see [1, p. 510]), where $\gamma = 2.3 \pm 0.1$. Another example is the World Wide Web, 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 [2]; they found that, both for the in- and out-degrees, the distribution is $P(k) \sim ck^{-\gamma}$ with $\gamma_{\text{in}} = 2.29$ and $\gamma_{\text{out}} = 2.78$.

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, and so on. Let $X[n, k]$ denote the size of the kth level after the nth step (the first step is when we take the first edge). The following 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 [3] is in close connection with H_n. The results there yield the height of our original tree ($\beta = 0$) to be asymptotically $\mathcal{O}(\log n)$. On the other hand, Pittel proved in [8] that

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

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

Our goal is to determine the width of the tree. We use the method of [5], which was applied to binary search trees for the proof of $W_n \sim n/\sqrt{4\pi \log n}$. We now present the main results of this paper. Set $\alpha = (1 + \beta)/(2 + \beta)$.

Theorem 1. *With probability* 1, *we obtain*

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

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

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

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

In addition, our results show that the width of the tree is reached approximately at a level of $\alpha \log n$.

In Section 2, we introduce a way of using martingales for the proof of Theorem 1, which is postponed to Section 3. In Section 4, we present an application. V. Batagelj brought it to the author's 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

Firstly, we introduce the notation

$$Y[n, k] = X[n, k + 1] + (1 + \beta)X[n, k] \quad \text{for } k > 1,$$
$$Y[n, 0] = X[n, 1] + \beta,$$

for the sum of weights on 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

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

and, for $k > 0$,

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

These yield

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

thus, we obtain

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

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

The following lemma is about the asymptotics of the expectation.

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

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

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

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

Proof. As in the proof of Lemma 1, we obtain

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

$$= (1+\beta)(1+z)\prod_{j=1}^{n-1}\frac{j+\alpha(1+z)}{j+2\alpha-1}.$$

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 asymptotics can be determined as in [6] and [5], proving that

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

uniformly over any compact set. This yields

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

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

Next, we will study the convergence of the martingale $M_n(z)$. For this purpose, we need to determine the covariance function of $G_n(z)$.

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

$$C_{n+1}^G(z_1, z_2) := \mathrm{E}(G_{n+1}(z_1)G_{n+1}(z_2))$$

$$= \sum_{j=1}^{n}\left(b_j(z_1, z_2)\prod_{k=j+1}^{n}a_k(z_1, z_2)\right)$$

$$+ (1+\beta)^2(1+z_1)(1+z_2)\prod_{j=1}^{n}a_j(z_1, z_2),$$

with

$$a_k(z_1, z_2) = 1 + \frac{2+(1+\beta)(z_1+z_2)}{s_k},$$

$$b_k(z_1, z_2) = \frac{(1+z_1+z_1\beta)(1+z_2+z_2\beta)}{s_k}\,\mathrm{E}\,G_k(z_1z_2).$$

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 $(n+1)$th step. With this notation, $G_{n+1}(z) - G_n(z) = z^{k_n}(1+(1+\beta)z)$. Thus, we obtain

$$C_{n+1}^G(z_1, z_2) = \mathrm{E}[\mathrm{E}((G_n(z_1) + z_1^{k_n}(1+z_1+z_1\beta))(G_n(z_2) + z_2^{k_n}(1+z_2+z_2\beta)) \mid \mathcal{F}_n)]$$

$$= C_n^G(z_1, z_2) + \mathrm{E}[\mathrm{E}(G_n(z_1)z_2^{k_n}(1+z_2+z_2\beta) + z_1^{k_n}(1+z_1+z_1\beta)G_n(z_2)$$

$$+ z_1^{k_n}z_2^{k_n}(1+z_1+z_1\beta)(1+z_2+z_2\beta) \mid \mathcal{F}_n)].$$

The conditional distribution of k_n with respect to $\mathcal{F}_n$ is

$$
P(k_n = k \mid \mathcal{F}_n) = \begin{cases} \dfrac{Y[n,k]}{s_n} & \text{if } k > 0, \\[2ex] \dfrac{Y[n,0]}{s_n} & \text{if } k = 0. \end{cases}
$$

Hence, the conditional expectation is

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

Similarly, we have

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

Finally, this yields

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

Hence,

$$
\begin{aligned}
C_{n+1}^G(z_1, z_2) &= \left(1 + \frac{2 + (1 + \beta)(z_1 + z_2)}{s_n} \right) C_n^G(z_1, z_2) \\
&\quad + \frac{(1 + z_1 + z_1 \beta)(1 + z_2 + z_2 \beta)}{s_n} \, E \, G_n(z_1 z_2).
\end{aligned}
$$

This proves the lemma, since $C_1^G(z_1, z_2) = (1 + \beta)^2 (1 + z_1)(1 + z_2)$.

Corollary 2. *The set of martingales* $\{M_n(z) \colon n \in \mathbb{N}\}$ *is bounded in* L^2 *for any fixed* $|z - 1| < \sqrt{1/\alpha}$, *where* L^p *is the space of p-power integrable functions. 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* $z \in \mathcal{H} := \{w \in \mathbb{C} \colon |w - 1| < \sqrt{1/\alpha}\}$.

Proof. Using the notation 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 $A_n \leq c B_n$ for every n. By Lemma 2,

$$
\begin{aligned}
C_n^G(z_1, z_2) &= \frac{(1 + z_1 + z_1 \beta)(1 + z_2 + z_2 \beta)}{2 + \beta} \sum_{j=1}^n \frac{E \, G_j(z_1, z_2)}{j + 2\alpha - 1} \prod_{j=k+1}^n a_k(z_1, z_2) \\
&\quad + (1 + \beta^2)(1 + z_1)(1 + z_2) \prod_{j=1}^n a_k(z_1, z_2) \\
&\ll \sum_{j=1}^n j^{\alpha \operatorname{Re}(z_1 z_2 - 1)} \left(\frac{n}{j} \right)^{2 + \alpha \operatorname{Re}(z_1 + z_2 - 2)} + n^{2 + \alpha \operatorname{Re}(z_1 + z_2 - 2)} \\
&\ll n^{2 + \alpha \operatorname{Re}(z_1 + z_2 - 2)} \sum_{j=1}^n j^{-(2 + \alpha \operatorname{Re}(z_1 + z_2 - z_1 z_2 - 1))}.
\end{aligned}
$$

Hence,

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

So, if

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

where $\overline{z}$ is the complex conjugate of z, 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 $2 + \alpha\,\mathrm{Re}(z_1 + z_2 - z_1z_2 - 1) > 1$; hence, $C_n^M(z_1, z_2)$ 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 arguments of [5]. The main idea is the following result, which can be proved in a similar way to Proposition 2 of [5].

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 $C \subseteq I$,*

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

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 $C \subseteq \mathbb{R}$,

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

The uniform convergence of (M_n) comes from the following proposition. The proof, being essentially the same as the proof of Proposition 2 of [5], is omitted.

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

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

Generally, let $\gamma : \mathbb{R} \to \mathcal{H}$ be continuously 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. *The martingale $M_n(z)$ 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 $\gamma(t) = 1 + \rho \mathrm{e}^{\mathrm{i}t}$, for all $0 < \rho < \sqrt{2}$. Thus, for $|s - 1| < \rho$, we have

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

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 $G_n(z)$. Firstly, we approximate $\mathrm{E}\,|G_n(z)|^2$.

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

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

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

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

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

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

uniformly for $z \in C$.

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

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

For $|z - 1| \leq \sqrt{1/\alpha} - \delta$, the exponent of j is at most $-1 - \delta' < -1$, where δ' is an arbitrarily small number; hence,

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

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

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

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

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

$$\ll n^{2(1+\alpha(\mathrm{Re}\, z-1))} \frac{n^{-1-\alpha(2\,\mathrm{Re}\, z-|z|^2-1)}}{-1 - \alpha(2\,\mathrm{Re}\, z - |z|^2 - 1)}.$$

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

$$\mathrm{E}\, |G_n(z)|^2 \ll n^{2(1+\alpha(\mathrm{Re}\, z-1))} \frac{(n+1)^{-1-\alpha(2\,\mathrm{Re}\, z-|z|^2-1)}}{-1 - \alpha(2\,\mathrm{Re}\, z - |z|^2 - 1)}$$

$$\ll n^{1+\alpha(|z|^2-1)} \frac{1 - \mathrm{e}^{(-1-\alpha(2\,\mathrm{Re}\, z-|z|^2-1))\log(n+1)}}{-1 - \alpha(2\,\mathrm{Re}\, z - |z|^2 - 1)}$$

$$\ll n^{(1-\alpha)(1+(1+\beta)|z|^2)} \log n.$$

This completes the proof.

Now, we approximate $G_n'(z)$, where the prime denotes differentiation.

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

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

Proof. Obviously, $|G_n'(z)| \leq G_n'(|z|)$. By [8], we know that $H_n \ll \log n$, where H_n is the height of the tree. Hence, there exists an n_0, for each realization of the tree, such that $X[n, k] = 0$ a.s. for $n \geq n_0$, if $k > c \log n$. Hence, for sufficiently large n, with probability 1,

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

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 by arguments similar to those in the proof of Proposition 3 of [5], it will be omitted.

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

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

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

Remark 1. If $\beta = 0$ then Lemma 6 holds for the function $|G_n(z)|/|1 + z|$ on

$$\gamma(\delta) := \{z \colon |z| = 1, |z - 1| \geq \sqrt{2} - \delta, \operatorname{Re} z > -0.9\} \cup \{z \colon \operatorname{Re} z = -0.9, |z| \leq 1\}.$$

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 $n \to \infty$. Since $-1 \notin \gamma$, to see this it is enough to approximate $|G_n(z)|$ on $\gamma(\delta)$. From here, the argument follows in exactly the same way as that in the proof of Lemma 6.

3. Proof of Theorem 1

Finally, we can start to prove Theorem 1. 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 $z \neq -1/(1 + \beta)$. This exception does not matter if $\beta \neq 0$, since $|1/(1 + \beta)| \neq 1$, and the function can be extended 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 into two parts. Let $\varphi = \min(\pi, \arccos(1 - 1/2\alpha))$ and let

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

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

where δ is the same as in Lemma 6.

If $\beta = 0$, instead of $|z| = 1$ we integrate on

$$\gamma = \{\xi \colon |\xi| = 1, \, \mathrm{Re}\, \xi > -0.9\} \cup \{\xi \colon \mathrm{Re}\, \xi = -0.9, \, |\xi| \leq 1\}.$$

Let I_1 be the same as in the previous case and let

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

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:

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

where we integrate on

$$\{\xi \colon |\xi| = 1, \, |\xi - 1| \geq \sqrt{1/\alpha} - \delta\}$$

(where $\beta \neq 0$) and on $\gamma(\delta)$ (where $\beta = 0$).

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

$$M_n(\mathrm{e}^{\mathrm{i}t}) = \frac{G_n(\mathrm{e}^{\mathrm{i}t})}{E_n(\mathrm{e}^{\mathrm{i}t})}$$

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

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

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

$$\frac{1}{2\pi} \int_{(\log n)^{-(1-\vartheta)/2} \leq |t| \leq \phi - \delta} |G_n(\mathrm{e}^{\mathrm{i}t})| \, \mathrm{d}t \ll n \int_{(\log n)^{-(1-\vartheta)/2}}^{\infty} \mathrm{e}^{-c't^2 \log n} \, \mathrm{d}t \ll n \mathrm{e}^{-c'(\log n)^{\vartheta}}. \quad (1)$$

The remaining part of the integral is

$$I_0 := \frac{1}{2\pi} \int_{|t| \leq (\log n)^{-(1-\vartheta)/2}} \frac{G_n(\mathrm{e}^{\mathrm{i}t})}{1 + (1 + \beta)\mathrm{e}^{\mathrm{i}t}} \mathrm{e}^{-kit} \, \mathrm{d}t.$$

Again, we are going to use

$$G_n(z) = E_n(z)M_n(z) = \mathrm{E}\, G_n(z)\frac{M_n(z)}{(1+\beta)(1+z)} \tag{2}$$

and Lemma 2, which can be written in the form

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

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

$$\begin{aligned}
\frac{\mathrm{E}\, G_n(\mathrm{e}^{\mathrm{i}t})}{1+(1+\beta)z} &= n\mathrm{e}^{(\log n)(\mathrm{e}^{\mathrm{i}t}-1)\alpha}\left(\frac{(1+\beta)(1+\mathrm{e}^{\mathrm{i}t})\Gamma(2\alpha)}{(1+(1+\beta)\mathrm{e}^{\mathrm{i}t})\Gamma(1+\alpha(1+\mathrm{e}^{\mathrm{i}t}))} + \mathcal{O}\left(\frac{1}{n}\right)\right) \\
&= n\mathrm{e}^{-(\alpha t^2/2)\log n+(\mathrm{i}t\alpha)\log n} \\
&\quad \times \left(1 - \mathrm{i}t\left(\alpha - \frac{1}{2} + 2\alpha^2\Gamma'(1+2\alpha)\right) - \frac{\alpha t^3}{6}\mathrm{i}\log n + \mathcal{O}(t^2+t^4\log n)\right).
\end{aligned} \tag{3}$$

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

$$\frac{M_n(\mathrm{e}^{\mathrm{i}t})}{(1+\beta)(1+\mathrm{e}^{\mathrm{i}t})} = 1 + \mathrm{i}t\frac{M_n'(1)-(1+\beta)}{2(1+\beta)} + \mathcal{O}(t^2). \tag{4}$$

Then, by (2)–(4), we conclude that, with probability 1,

$$\begin{aligned}
\frac{G_n(\mathrm{e}^{\mathrm{i}t})\mathrm{e}^{-k\mathrm{i}t}}{1+(1+\beta)\mathrm{e}^{\mathrm{i}t}} &= n\mathrm{e}^{-(\alpha t^2/2)\log n+\mathrm{i}t(\alpha\log n-k)} \\
&\quad \times \left(1 - \mathrm{i}t\left(\alpha - \frac{1}{2} + 2\alpha^2\Gamma'(1+2\alpha) - \frac{M_n'(1)-(1+\beta)}{2(1+\beta)}\right)\right. \\
&\quad \left. - \frac{\alpha t^3}{6}\mathrm{i}\log n + \mathcal{O}(t^2+t^4\log n)\right)
\end{aligned}$$

uniformly with respect to k. For the same reason as in (1), here we also have

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

Hence,

$$\begin{aligned}
\frac{I_0}{n} &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-(\alpha t^2/2)\log n+\mathrm{i}t(\alpha\log n-k)} \\
&\quad \times \left(1 - \mathrm{i}t\left(\alpha - \frac{1}{2} + 2\alpha^2\Gamma'(1+2\alpha) - \frac{M_n'(1)-(1+\beta)}{2(1+\beta)}\right) - \frac{\alpha t^3}{6}\mathrm{i}\log n\right)\mathrm{d}t \\
&\quad + \mathcal{O}((\log n)^{-3/2}).
\end{aligned}$$

Integration gives

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

Hence, we obtain

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

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 them are trees. V. Batagelj brought it to the author's attention that directory trees should be studied. The following examples all have power-law degree distributions $P(k) \sim ck^{-\gamma}$, with $2 < \gamma < 3$. This allows us to compare the width of the tree with the result of Theorem 1.
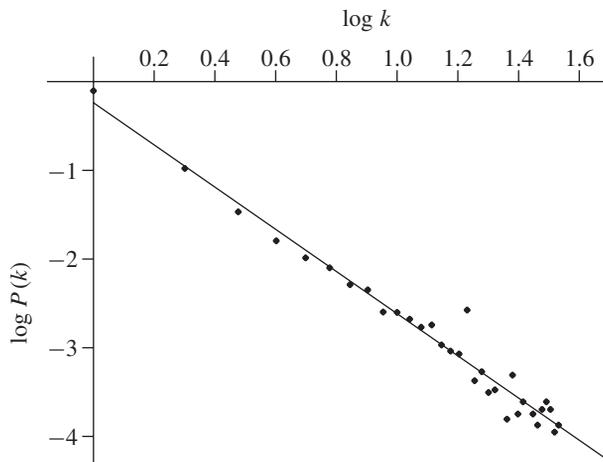

FIGURE 1.

TABLE 1.

Directory tree	Vertices	β	Real width	Theoretical width	β'
Server of CS Dept., Budapest Univ. of Technology	39 182	-0.62	10 159	9 162	-0.71
Server of CS Dept., Eötvös Loránd Univ.	18 609	-0.96	10 916	12 519	-0.95
Server of Fazekas High School	48 898	-0.25	9 721	9 071	-0.40
Home Linux	22 797	-0.27	4 026	4 415	$+0.04$
Home Windows	6 999	-0.53	2 097	1 662	-0.75

The first example is the directory tree of the main server of the Department of Computer Science and Information Theory, Budapest University of Technology and Economics. Figure 1 shows the degree distribution with logarithmic scales.

Linear regression gives $\gamma \approx 2.38$. Substituting $\beta = \gamma - 3 = -0.62$ and $n = 39\,182$ into Theorem 1 gives a width of 9162. We can compare this with the real width of the tree, which is 10 159. We can also calculate the β' that would give the same theoretical width as the real width. From Theorem 1, we have $\beta' \approx -0.71$

Table 1 shows the results of studying several directory trees. The servers are all Unix systems with a lot of users who are free to create their own directories in the home directory. This is a reason why we should consider these graphs to be random. The ratio of theoretical to real widths is between 0.85 and 1.15 in Table 1; hence, we can approximate the width of directory trees using Theorem 1.

Acknowledgement

References

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

[2] BENCZÚR, A. *et al.* (2003). Searching a small national domain—Preliminary report. Preprint. Available at http://www.ilab.sztaki.hu/.

[3] BOLLOBÁS, B. AND RIORDAN, O. (2004). The diameter of a scale-free random graph. *Combinatorica* **24,** 5–34.

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

[5] CHAUVIN, B., DRMOTA, M. AND JABBOUR-HATTAB, J. (2001). The profile of binary search trees. *Ann. Appl. Prob.* **11,** 1042–1062.

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

[7] MÓRI, T. F. (2002). On random trees. *Studia Sci. Math. Hung.* **39,** 143–155.

[8] PITTEL, B. (1994). Note on the heights of random recursive trees and random *m*-ary search trees. *Random Structures Algorithms* **5,** 337–347.

[9] SMYTHE, R. T. AND MAHMOUD, H. M. (1994). A survey of recursive trees. *Teor. Ǐmovīr. Mat. Stat.* **51,** 1–29 (in Ukrainian). English translation: *Theory Prob. Math. Statist.* **51** (1995), 1–27.