Levels of a scale-free tree

Zsolt Katona

Eötvös Loránd University
Department of Probability Theory and Statistics
zskatona@cs.bme.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 the graph will be a tree. These graphs have been shown to have power law degree distributions, the same as observed in some large real-world networks. We show that the degree distribution is the same on every sufficiently high level of the tree.

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 [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."

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).

In the case of $m = 1$, the resulting graph is a tree. These scale-free trees have been known since the 1980s as *nonuniform random recursive trees*. Two nearly identical classes of these trees are *random recursive trees with attraction of vertices proportional to the degrees* and *random plane-oriented recursive trees* (see [8] and [9]). In this model, starting with a single edge, 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 end points of that edge, both with equal probability.

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 [10] 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 $k^{-(3+\beta)}$. In the original case ($\beta = 0$), the formula of the expected proportion of vertices of degree k was determined by Szymański [12] and strong convergence of this proportion was shown by Lu and Feng [7]. Similar a.s. results were proved in a paper of Bollobás, Riordan, Spencer and Tusnády [4] for the general model ($m \geq 1$).

The following examples show the importance of these results and the sense of the generalization. Several graphs have been found with degree distributions $P(k) \sim c \cdot k^{-\gamma}$ with some constant c [1]. 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 a distinct in- and out-degree distribution. The Hungarain Web was studied by the Websearch and Data Mining Group of the Hungarian Computer and Automation Research Institute in [2]; and they found that both for the in- and out-degrees the distribution is $P(k) \sim c \cdot k^{-\gamma}$ with $\gamma_{in} = 2.29$ and $\gamma_{out} = 2.78$.

In [6], I studied 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*.

Set $\alpha = \frac{1+\beta}{2+\beta}$. In [6] the following result gives the sizes of the largest levels. With probability 1,

$$X[n,k] = \frac{n}{\sqrt{2\alpha\pi \log n}} \cdot \exp\left(-\frac{(k-\alpha\log n)^2}{2\alpha\log n}\right) + \mathcal{O}\left(\frac{n}{\log n}\right), \tag{1}$$

as $n \to \infty$, where the error term is uniform for all $k \geq 0$. Of course this also yields that the width is $W_n \sim \frac{n}{\sqrt{2\alpha\pi\log n}}$ and it is reached at about level $\alpha \log n$.

Knowing the degree distribution and the shape of the tree, Tamás Móri posed the problem whether the degree distribution is the same on all levels or not. He noticed that on lower levels it is different from that of the whole tree, for example on the first level the ratio of vertices with degree j

a.s. goes to $(1+\beta)\left(\frac{1}{j+\beta} - \frac{1}{j++\beta+1}\right)$. Hence the degree distribution on the lower level is still a power law distribution, but the exponent is -2, independently of the parameter β and the fixed level k. In this paper we show that the answer for Móri's question is yes for the middle levels (around $\alpha \log n$), that contain almost all vertices, hence determine the degree distribution of the whole tree.

Theorem 1 *Suppose $\beta = 0$. With any constants $0 < k_1 < k_2$, for $k_1\sqrt{\log n} < k - \frac{1}{2}\log n < k_2\sqrt{\log n}$ the ratio of vertices with degree j converges a.s. to a limit c_j on level k and c_j is equal to the limit of the ratio of j-degree vertices in the whole graph.*

Remark 1 *The theorem holds for any $\beta > -1$, that is, the limit of vertices with degree j on levels around $\alpha \log n$ a.s. converges to the same limit as in the whole tree.*

We will only prove the theorem. One can see that the proof of this generalization goes on the same lines, but needs longer and more complicated calculations.

2 Generating functions

The proof goes on similar lines as in [6] and [5], however, some of the recursive formulas had been known long before. The main idea is to consider the generating function

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

for any complex z. The sum is finite for a fixed n, hence $G_n(z)$ is holomorphic. To study degree distributions, we have to count the vertices with given degree. Instead of that let $X^{(\geq j)}[n, k]$ be the number of vertices with degree *at least* j on level k after step n. Let

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

be the corresponding generating function.

Obviously, $G_n^{(\geq j)}(1)$ is the number of vertices in the tree with degree at least j excluding the root. As shown in [12], the limit of the expected ratio of vertices with degree j is $\frac{4}{j(j+1)(j+2)}$. One can see that summing this quantity gives the ratio of vertices with degree at least j as $\frac{2}{j(j+1)}$.

To calculate $\mathbf{E}G_n^{(\geq j)}(z)$ we use conditional expectations. Let $\mathcal{F}_n$ denote the σ-field generated by the first n steps. The number of vertices with degree at least j on a given level either increases by one or does not change. For $j = 1$, the probability of an increase is

$$P(X[n+1, k] = X[n, k] + 1 | \mathcal{F}_n) = \begin{cases} \frac{X[n,k]+X[n,k-1]}{2n}, & \text{for } k > 1, \\ \frac{X[n,1]}{2n}, & \text{for } k = 1, \end{cases}$$

since the new vertex is connected to level $k-1$ with probability equal to the sum of the degrees on level $k-1$ over $2n$. Obviously, the sum of the degrees on level $k-1$ is $X[n,k]+X[n,k-1]$ for $k>1$ and $X[n,1]$ for $k=1$. For $j \geq 2$, probability of an increase is

$$P(X^{(\geq j)}[n+1,k] = X^{(\geq j)}[n,k]+1|\mathcal{F}_n) = (j-1)\frac{X^{(\geq j-1)}[n,k]-X^{(\geq j)}[n,k]}{2n},$$

since this event is equivalent to the event that the new vertex is connected to a vertex on level k with degree $j-1$. Thus, for $j=k=1$

$$\mathbf{E}(X[n+1,1]|\mathcal{F}_n) = (X[n,1]+1)\frac{X[n,1]}{2n} + X[n,1]\left(1-\frac{X[n,1]}{2n}\right) = \\ = \frac{X[n,1]}{2n} + X[n,1] = \frac{2n+1}{2n}X[n,1].$$

For $j=1$, $k>1$ we have

$$\mathbf{E}(X[n+1,k]|\mathcal{F}_n) = (X[n,k]+1)\frac{X[n,k]+X[n,k-1]}{2n} + X[n,k]\left(1-\frac{X[n,k]+X[n,k-1]}{2n}\right) = \\ = \frac{2n+1}{2n}X[n,k] + \frac{1}{2n}X[n,k-1],$$

and finally, for $j>1$

$$\mathbf{E}(X^{(\geq j)}[n+1,k]|\mathcal{F}_n) = (X^{(\geq j)}[n,k]+1)\frac{(j-1)(X^{(\geq j-1)}[n,k]-X^{(\geq j)}[n,k])}{2n} + \\ + X^{(\geq j)}[n,k]\left(1-\frac{(j-1)(X^{(\geq j-1)}[n,k]-X^{(\geq j)}[n,k])}{2n}\right) = \\ = \frac{2n-j+1}{2n}X^{(\geq j)}[n,k] + \frac{j-1}{2n}X^{(\geq j-1)}[n,k].$$

This gives the following recursive formula for the generating functions. For $j=1$,

$$\mathbf{E}(G_{n+1}(z)|\mathcal{F}_n) = \frac{2n+1}{2n}G_n(z) + \frac{z}{2n}G_n(z) = \frac{2n+1+z}{2n}G_n(z), \tag{2}$$

and for $j \geq 1$

$$\mathbf{E}(G_{n+1}^{(\geq j+1)}(z)|\mathcal{F}_n) = \frac{2n-j}{2n}G_n^{(\geq j+1)}(z) + \frac{j}{2n}G_n^{(\geq j)}(z). \tag{3}$$

In our calculations, just as in [6], we will use the fact that

$$\prod_{i=1}^{n}\frac{i+v}{i+w} = n^{\Re(v-w)}\left(\frac{\Gamma(1+w)}{\Gamma(1+v)} + O(1/n)\right),$$

for any complex v and $w \neq -1$.

Since $G_1(z)=1$,

$$\mathbf{E}G_n(z) = \mathbf{E}G_n^{(\geq 1)}(z) = \prod_{j=1}^{n-1}\frac{2j+1+z}{2j} = n^{(1+z)/2}\left(1/\Gamma\left(\frac{3+z}{2}\right) + \mathcal{O}(1/n)\right). \tag{4}$$

Remark 2 *For any fixed $z \in \mathbb{C}$ the sequence*

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

is a martingale with respect to the filtration $\mathcal{F}_n$.

PROOF: It follows from (4) and (2) that

$$\mathbf{E}(M_{n+1}(z)|\mathcal{F}_n) = \frac{\mathbf{E}(G_{n+1}(z)|\mathcal{F}_n)}{\mathbf{E}G_{n+1}(z)} = \frac{2n+1+z}{2n} G_n(z) \cdot \frac{1}{\mathbf{E}G_{n+1}(z)} = \frac{G_n(z)}{\mathbf{E}G_{n+1}(z)} = M_n(z).$$

□

In general, the following holds for the expectation of the generating functions.

Lemma 1 *For any fixed $j \geq 2$*

$$\mathbf{E}G_n^{(\geq j)}(z) = n^{\frac{1+z}{2}} \left(c_j(z)/\Gamma\left(\frac{3+z}{2}\right) + \mathcal{O}\left(\frac{1}{n}\right)\right),$$

where

$$c_j(z) = \frac{(j-1)!}{(z+2)(z+3)\dots(z+j)}.$$

PROOF: We proceed by induction on j. For $j=1$ the equation is true with $c_1(z)=1$. Suppose that it is true for $j=l \geq 1$. By (3) we have

$$\mathbf{E}G_{n+1}^{(\geq l+1)}(z) = \frac{2n-l}{2n}\mathbf{E}G_n^{(\geq l+1)}(z) + \frac{l}{2n}\mathbf{E}G_n^{(\geq l)}(z).$$

Since $G_1^{(\geq l+1)}(z) = 0$, this recursive formula gives

$$\mathbf{E}G_{n+1}^{(\geq l+1)}(z) = \sum_{i=1}^{n} \frac{l}{2i}\mathbf{E}G_n^{(\geq i)}(z) \prod_{m=i+1}^{n} \frac{2m-l}{2m} = \sum_{i=1}^{n} \frac{c_l(z) l}{2i\Gamma(\frac{3+z}{2})} i^{\frac{1+z}{2}} (n/i)^{-l/2} \left(1 + \mathcal{O}\left(\frac{1}{i}\right)\right) =$$

$$= \frac{c_l(z) l}{2\Gamma(\frac{3+z}{2})} n^{-l/2} \frac{n^{\frac{1+z+l}{2}}}{(1+z+l)/2} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) = \frac{c_{l+1}(z)}{\Gamma(\frac{3+z}{2})} n^{\frac{1+z}{2}} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right).$$

□

From now on, we will study the functions

$$U_n^{(j)}(z) := G_n^{(\geq j)}(z) - c_j(z) G_n(z).$$

We will show that $U_n^{(j)}(z)$ is a.s. not far from 0. In order to be able to use martingales, we have to consider the following linear combination. For $j \geq 2$ let

$$W_n^{(j)}(z) := \sum_{i=2}^{j} (-1)^{j-i} \binom{j-1}{i-1} U_n^{(i)}(z). \tag{5}$$

The idea is based on a similar combination used in [10]. This combination will cancel out the different coefficients in the proof of Remark 4, originating from the recursive formulas for the generating functions in (3). Easy calculation shows that

$$W_n^{(j)}(z) = \sum_{i=1}^{j} b_i^{(j)}(z) G_n^{(\geq i)}(z), \tag{6}$$

where

$$b_1^{(j)}(z) = (-1)^{j-1}\frac{j-1}{j+z},$$

and for $2 \leq i \leq j$

$$b_i^{(j)}(z) = (-1)^{j-i}\binom{j-1}{i-1}.$$

Remark 3 *For $j \geq 2$,*

$$U_n^{(j)}(z) = \sum_{i=2}^{j}\binom{j-1}{i-1} W_n^{(i)}(z),$$

Proof: $W_n^{(j)}(z)$ is defined in (5). Plugging this definition into the right hand side of the equation yields

$$\sum_{i=2}^{j}\binom{j-1}{i-1} W_n^{(i)}(z) = \sum_{i=2}^{j}\sum_{k=2}^{i}(-1)^{i-k}\binom{j-1}{i-1}\binom{i-1}{k-1}U_n^{(k)}(z).$$

By changing the order of the sums, this is equal to

$$\sum_{k=2}^{j}\sum_{i=k}^{j}(-1)^{i-k}\binom{j-1}{i-1}\binom{i-1}{k-1}U_n^{(k)}(z) = U_n^{(j)}(z) + \sum_{k=2}^{j-1}U_n^{(k)}(z)\sum_{i=k}^{j}(-1)^{i-k}\binom{j-1}{i-1}\binom{i-1}{k-1}.$$

We only have to show that the last sum is equal to zero for any $2 \leq k \leq j-1$. By manipulating the binomial formulas, we have

$$\sum_{i=k}^{j}(-1)^{i-k}\binom{j-1}{i-1}\binom{i-1}{k-1} = k\binom{j}{k}\sum_{i=k}^{j}(-1)^{k-i}\binom{j-k}{j-i} = 0.$$

□

Remark 4 *For every fixed complex z and fixed $j \geq 2$,*

$$M_n^{(j)}(z) := W_n^{(j)}(z)\prod_{i=1}^{n}\frac{2i}{2i+1-j}$$

is a martingale with respect to $\mathcal{F}_n$.

Proof: In order to prove that $M_n^{(j)}(z)$ is a martingale with respect to $\mathcal{F}_n$, we have to show that $\mathbf{E}(M_{n+1}^{(j)}(z)|\mathcal{F}_n) = M_n^{(j)}(z)$, or equivalently, that $\mathbf{E}(W_{n+1}^{(j)}(z)|\mathcal{F}_n) = \frac{2n+1-j}{2n}W_n^{(j)}(z)$. Using the recursive formulas (2) and (3), we have

$$\mathbf{E}(W_{n+1}^{(j)}(z)|\mathcal{F}_n) = \sum_{i=1}^{j} b_i^{(j)}(z)\mathbf{E}(G_{n+1}^{(\geq i)}(z)|\mathcal{F}_n) =$$

$$= b_1^{(j)}(z)\frac{2n+1+z}{2n}G_n(z) + \sum_{i=2}^{j} b_i^{(j)}(z)\left(\frac{2n-j+1}{2n}G_n^{(\geq i)}(z) + \frac{i-1}{2n}G_n^{(\geq i-1)}(z)\right)$$

The coefficient of $G_n(z)$ here is

$$(-1)^{j-1}\left(\frac{j-1}{j+z}\frac{2n+1+z}{2n} - \frac{j-1}{2n}\right) = (-1)^{j-1}\frac{j-1}{j+z}\frac{2n+1-j}{2n} = b_1^{(j)}(z)\frac{2n+1-j}{2n}.$$

For $j \geq i \geq 2$, we have $\frac{b_{i+1}^{(j)}(z)}{b_i^{(j)}(z)} = \frac{j-i}{i}$, thus the coefficient of $G_n^{(\geq i)}(z)$ is

$$b_i^{(j)}(z)\left(\frac{2n-i+1}{2n} - \frac{i}{2n}\cdot\frac{j-i}{i}\right) = b_i^{(j)}(z)\frac{2n-j+1}{2n}.$$

Thus, $\mathbf{E}(W_{n+1}^{(j)}(z)|\mathcal{F}_n) = \frac{2n+1-j}{2n}W_n^{(j)}(z)$, completing the proof. □

The key to the approximation of $U_n^{(j)}(z)$ is to find an upper bound for its variance. The following lemma gives an upper bound for this variance for an arbitrary z, however, it will only be needed for $|z| = 1$ in the proof of the theorem.

Lemma 2 *For any complex z and fixed $k \geq 2$,*

$$E|W_n^{(k)}(z)|^2 = \mathcal{O}\left(n^{\frac{1+|z|^2}{2}}\right),$$

which yields

$$E|U_n^{(k)}(z)|^2 = \mathcal{O}\left(n^{\frac{1+|z|^2}{2}}\right).$$

Proof: Let

$$C_n^{(i,j)}(z_1,z_2) := \mathbf{E}(G_n^{(\geq i)}(z_1)G_n^{(\geq j)}(z_2)),$$

$$D_n^{(k)}(z_1,z_2) := \mathbf{E}(W_n^{(k)}(z_1)W_n^{(k)}(z_2)).$$

Using equation (6), this can be rewritten as

$$D_n^{(k)}(z_1,z_2) = \sum_{l=1}^{k}\sum_{m=1}^{k} b_l^k(z_1)b_m^k(z_2)C_n^{(l,m)}(z_1,z_2). \tag{7}$$

The objective is to give a recursive formula for $D_n^{(k)}(z_1,z_2)$. Note that $G_{n+1}^{(\geq j)}(z) = G_n^{(\geq j)}(z) + K_n^{(j)}(z)$ where the distribution of $K^{(j)}$ is given by

$$P(K_n^{(1)}(z) = z^{k-1}|\mathcal{F}_n) = \begin{cases} \frac{X[n,k]+X[n,k-1]}{2n}, & \text{for } k>1, \\ \frac{X[n,1]}{2n}, & \text{for } k=1, \end{cases},$$

and for $j \geq 2$

$$P(K_n^{(j)}(z) = z^{k-1}|\mathcal{F}_n) = (j-1)\frac{X^{(\geq j-1)}[n,k] - X^{(\geq j)}[n,k]}{2n}$$

These yield

$$\mathbf{E}(K_n^{(1)}(z)|\mathcal{F}_n) = \frac{1+z}{2n}G_n(z) \quad \text{and} \quad \mathbf{E}(K_n^{(j)}(z)|\mathcal{F}_n) = \frac{j-1}{2n}(G_n^{(\geq j-1)}(z) - G_n^{(\geq j)}(z)).$$

Also we have

$$\mathbf{E}(K_n^{(1)}(z_1)K_n^{(1)}(z_2)|\mathcal{F}_n) = \frac{1+z_1z_2}{2n}G_n(z),$$

$$\mathbf{E}(K_n^{(j)}(z_1)K_n^{(1)}(z_2)|\mathcal{F}_n) = z_2\frac{j-1}{2n}(G_n^{(\geq j-1)}(z_1z_2) - G_n^{(\geq j)}(z_1z_2)) \ \text{ for } \ j \geq 2,$$

and for $i \geq j \geq 2$,

$$\mathbf{E}(K_n^{(i)}(z_1)K_n^{(j)}(z_2)|\mathcal{F}_n) = \frac{i-1}{2n}(G_n^{(\geq i-1)}(z_1z_2) - G_n^{(\geq i)}(z_1z_2)).$$

It follows for $i \geq j \geq 2$ that

$$\begin{aligned} C_{n+1}^{(i,j)}(z_1,z_2) &= \mathbf{E}\left[\mathbf{E}\left((G_n^{(\geq i)}(z_1) + K_n^{(i)}(z_1))(G_n^{(\geq j)}(z_2) + K_n^{(j)}(z))|F_n\right)\right] = \\ &= \mathbf{E}\Bigg[G_n^{(\geq i)}(z_1)G_n^{(\geq j)}(z_2) + G_n^{(\geq i)}(z_1)\frac{j-1}{2n}\left(G_n^{(\geq j-1)}(z_2) - G_n^{(\geq j)}(z_2)\right) + \\ &+\frac{i-1}{2n}\left(G_n^{(\geq i-1)}(z_1) - G_n^{(\geq i)}(z_1)\right)G_n^{(\geq j)}(z_2) + \frac{i-1}{2n}\left(G_n^{(\geq i-1)}(z_1z_2) - G_n^{(\geq i)}(z_1z_2)\right)\Bigg] = \\ &= \frac{2n+2-i-j}{2n}\mathbf{E}(G_n^{(\geq i)}(z_1)G_n^{(\geq j)}(z_2)) + \frac{i-1}{2n}\mathbf{E}(G_n^{(\geq i-1)}(z_1)G_n^{(\geq j)}(z_2)) + \\ &+ \frac{j-1}{2n}\mathbf{E}(G_n^{(\geq i)}(z_1)G_n^{(\geq j-1)}(z_2)) + \frac{i-1}{2n}\mathbf{E}\left(G_n^{(\geq i-1)}(z_1z_2) - G_n^{(\geq i)}(z_1z_2)\right). \end{aligned}$$

This gives the formula

$$\begin{aligned} C_{n+1}^{(i,j)}(z_1,z_2) = \frac{2n+2-i-j}{2n}C_n^{(i,j)}(z_1,z_2) + \frac{i-1}{2n}C_n^{(i-1,j)}(z_1,z_2) + \frac{j-1}{2n}C_n^{(i,j-1)}(z_1,z_2) + \\ + \frac{i-1}{2n}\mathbf{E}\left(G_n^{(\geq i-1)}(z_1z_2) - G_n^{(\geq i)}(z_1z_2)\right). \end{aligned}$$

For $j \geq 2$, similar calculations lead to

$$\begin{aligned} C_{n+1}^{(j,1)}(z_1,z_2) = \frac{2n+2-j+z_2}{2n}C_n^{(j,1)}(z_1,z_2) + \frac{j-1}{2n}C_n^{(j-1,1)}(z_1,z_2) + \\ + z_2\frac{j-1}{2n}\mathbf{E}\left(G_n^{(\geq j-1)}(z_1z_2) - G_n^{(\geq j)}(z_1z_2)\right), \end{aligned}$$

and

$$C_{n+1}^{(1,1)}(z_1, z_2) = \frac{2n + 2 + z_1 + z_2}{2n} C_n^{(1,1)}(z_1, z_2) + \frac{1 + z_1 z_2}{2n} \mathbf{E} G_n(z_1 z_2).$$

Notice, that in these recursive formulas, all the $\mathbf{E}G_n^{(\geq .)}(z_1 z_2)$ type expressions are $\mathcal{O}\left(n^{\frac{1+z_1 z_2}{2}}\right)$, according to (4) and Lemma 1. Since all of them are divided by n, all those terms are $\mathcal{O}\left(n^{\frac{z_1 z_2 - 1}{2}}\right)$. If we plug the above recursive formulas into (7), easy but tedious calculation gives that

$$D_{n+1}^{(k)}(z_1, z_2) = \frac{n - k + 1}{n} \left[D_n^{(k)}(z_1, z_2) + \mathcal{O}\left(n^{\frac{z_1 z_2 - 1}{2}}\right) \right].$$

Since after the first step there is no vertex with degree at least two, $D_1^{(k)}(z_1, z_2) = 0$. Thus, the recursive formula for $D_n^{(k)}$ yields

$$D_{n+1}^{(k)}(z_1, z_2) = \sum_{i=1}^{n} \mathcal{O}\left(i^{\frac{z_1 z_2 - 1}{2}}\right) \prod_{m=i+1}^{n} \frac{m - k + 1}{m} = \sum_{i=1}^{n} \mathcal{O}\left(i^{\frac{z_1 z_2 - 1}{2}}\right) \mathcal{O}\left(\left(\frac{n}{i}\right)^{\frac{1-k}{2}}\right) = \mathcal{O}\left(n^{\frac{1+z_1 z_2}{2}}\right).$$

Obviously, $|W_n^{(k)}(z)|^2 = W_n^{(k)}(z)\overline{W_n^{(k)}(z)} = W_n^{(k)}(z) W_n^{(k)}(\overline{z})$, hence

$$\mathbf{E}|W_n^{(k)}(z)|^2 = D_n^{(k)}(z, \overline{z}) = \mathcal{O}\left(n^{\frac{1+z\overline{z}}{2}}\right) = \mathcal{O}\left(n^{\frac{1+|z|^2}{2}}\right).$$

For the second part, recall that according to Remark 3,

$$U_n^{(k)}(z) = \sum_{i=2}^{k} \binom{k-1}{i-1} W_n^{(i)}(z),$$

hence,

$$\mathbf{E}|U_n^{(k)}(z)|^2 \leq c \cdot \mathbf{E}|W_n^{(k)}(z)|^2 = \mathcal{O}\left(n^{\frac{1+|z|^2}{2}}\right)$$

with some constant c. □

Now we approximate $(U_n^{(j)})'(z)$. Denote by $A \ll B$ if there is a c constant such that $A \leq cB$.

Lemma 3 *For every $z \neq 0$, and fixed $j \geq 2$ we have a.s.*

$$|(U_n^{(j)})'(z)| \ll \frac{\log n}{|z|} U_n^{(j)}(|z|).$$

PROOF: Trivially, $|(U_n^{(j)})'(z)| \leq (U_n^{(j)})'(|z|)$. In [11] it was shown that the height of the tree $H_n \sim c \log n$ a.s., where $c \approx 4.31$ is the (greater than 2) solution of $c \log(2e/c) = 1$. Hence, a.s.

there exists n_0 such that for $n \geq n_0$, we have $X^{(\geq j)}[n,k] = X[n,k] = 0$ if $k > (c+1)\log n$. Thus, a.s.

$$(U_n^{(j)})'(|z|) = \sum_{k=0}^{\infty} k\left(X^{(\geq j)}[n,k] - c_j(|z|)X[n,k]\right)|z|^{k-1} \ll$$

$$\ll \log n \sum_{k=0}^{\infty}\left(X^{(\geq j)}[n,k] - c_j(|z|)X[n,k]\right)|z|^{k-1} = \frac{\log n}{|z|}U_n^{(j)}(|z|).$$

□

3 Proof of the Theorem

Before directly entering the proof we study $U_n^{(j)}(z)$ for $|z| = 1$.

Lemma 4 *For every $\varepsilon > 0$ we have a.s.*

$$\sup_{z=1} |U_n^{(j)}(z)| = \mathcal{O}\left(n^{3/4+\varepsilon}\right)$$

as $n \to \infty$.

PROOF: By Markov's inequality and Lemma 2, we have

$$P(|U_n^{(j)}(z)| \geq n^{3/4+\varepsilon}) \leq \frac{\mathbf{E}|U_n^{(j)}(z)|^2}{n^{3/2+2\varepsilon}} \ll n^{-1/2-2\varepsilon}$$

Let $z(n,l) = \exp(i\frac{2\pi l}{K})$ for $l = 1,\ldots,K$, where $K = \lfloor \log n \rfloor$. These points split the circle $|z| = 1$ into K equal arcs. We have

$$P(|U_n^{(j)}(z(n,l))| \geq n^{3/4+\varepsilon} \text{ for any } l) \ll n^{-1/2-2\varepsilon}K \ll n^{-1/2-\varepsilon}$$

Since

$$\sum_{n=1}^{\infty}(n^2)^{-1/2-\varepsilon} < \infty,$$

we can apply the Borel-Cantelli Lemma. Hence for all but finitely many n we have a.s.

$$\sup_l |U_{n^2}^{(j)}(z(n^2,l))| \leq (n^2)^{3/4+\varepsilon}.$$

Between the points $z(n^2,l)$ we can use Lemma 3. Suppose that $|z| = 1$ and $\frac{2\pi l}{K} < \arg z < \frac{2\pi(l+1)}{K}$. Then we have uniformly

$$|U_{n^2}^{(j)}(z)| = |U_{n^2}^{(j)}(z(n^2,l)) + \mathcal{O}(U_{n^2}'(z(n^2,l))(1/K))| \leq (n^2)^{3/4+\varepsilon} + \mathcal{O}\left((n^2)^{3/4+\varepsilon}\right) \ll (n^2)^{3/4+\varepsilon}$$

Hence for all but finitely many n we have a.s.

$$\sup_{|z|=1} |U_{n^2}^{(j)}(z)| \leq (n^2)^{3/4+\varepsilon}.$$

Finally, recall that for $|z|=1$ we have $|G_{n+1}^{(\geq j)}(z) - G_n^{(\geq j)}(z)| = |K_n^{(j)}(z)| \leq 1$, and $|G_{n+1}(z) - G_n(z)| = |K_n^{(1)}(z)| \leq 1$. Hence $|U_{n+1}^{(j)}(z) - U_n^{(j)}(z)| \leq |G_{n+1}^{(\geq j)}(z) - G_n^{(\geq j)}(z)| - |c_j(z)| \cdot |(G_{n+1}(z) - G_n(z))| \leq 2$. For $1 \leq k \leq 2n$, we have uniformly

$$|U_{n^2+k}^{(j)}(z)| = |U_{n^2}^{(j)}(z) + \mathcal{O}(k)| \ll (n^2)^{3/4+\varepsilon} + \mathcal{O}(n) \ll (n^2+k)^{3/4+\varepsilon}.$$

This completes the proof, as it yields for all but finitely many n a.s.

$$\sup_{|z|=1} |U_n^{(j)}(z)| \leq (n)^{3/4+\varepsilon}.$$

□

Now we directly start the proof of the Theorem.

PROOF: We can extract $X[n,k]$ from the generating function by using Cauchy's formula.

$$X^{(\geq j)}[n,k+1] - c_j(1)X[n,k] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{G_n^{(\geq j)}(e^{it}) - c_j(1)G_n(e^{it})}{e^{kit}} dt =$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{G_n^{(\geq j)}(e^{it}) - c_j(e^{it})G_n(e^{it})}{e^{kit}} dt + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{(c_j(e^{it}) - c_j(1))G_n(e^{it})}{e^{kit}} dt = J + I.$$

First, we can approximate J with Lemma 4. We have a.s.

$$|J| \ll n^{3/4+\varepsilon}.$$

The second integral can be approximated just as in [6], we will use two Lemmas to do so.

Lemma 5 *The martingale* $M_n(z) = \frac{G_n(z)}{\mathbf{E}G_n(z)}$ *and all its derivatives converge uniformly over the compact subsets of* $\mathcal{H} := \{z \in \mathbb{C} \,||z-1| < \sqrt{2}\}$.

PROOF: In Remark 2 we have already seen that $M_n(z)$ is a martingale. Corollary 3 of [6] with $\beta = 0$ says that $(1+z)M_n(z)$ and all its derivatives converge uniformly over the compact subsets of $\mathcal{H} := \{z \in \mathbb{C} \,||z-1| < \sqrt{2}\}$. Since $-1 \notin \overline{\mathcal{H}}$, this proves the lemma. □

Lemma 6 *Let*

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

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

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

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

PROOF: The function $\frac{G_n(z)}{1+z}$ of Remark 1 in [6] is equal to the $G_n(z)$ of the present paper.

□

Return to the proof of the Theorem. Since $(c_j(e^{it}) - c_j(1))G_n(e^{it})$ is regular for $|z| < 2$,

$$I = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{(c_j(e^{it}) - c_j(1))G_n(e^{it})}{e^{kit}} dt = \frac{1}{2\pi i}\int_{\gamma} \frac{(c_j(\xi) - c_j(1))G_n(\xi)}{\xi^{k+1}} d\xi$$

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

We split the integral I into two parts.

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

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

with the δ we get from Lemma 6. By the lemma, for any $L > 0$ we can approximate the second integral as follows.

$$|I_2| \leq \sup_{\gamma(\delta)} |c_j(z) - c_j(1)| \frac{1}{2\pi}\int_{\gamma(\delta)} |G_n(e^{it})| dt \ll \frac{n}{(\log n)^L}. \tag{8}$$

For $|t| \leq \pi/2 - \delta$

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

is a.s. uniformly bounded by Lemma 5. On the other hand, (4) provides us the asymptotics of the denominator, hence

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

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

$$\frac{1}{2\pi}\int_{(\log n)^{-(1-\vartheta)/2} \leq |t| \leq \pi/2 - \delta} |G_n(e^{it})| dt \ll n\int_{(\log n)^{-(1-\vartheta)/2}}^{\infty} e^{-c' t^2 \log n} dt \ll ne^{-c'(logn)^{\vartheta}} \ll \frac{n}{(\log n)^L}. \tag{9}$$

The remaining part of the integral is

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

Again, we are going to use

$$G_n(z) = \mathbf{E}G_n(z)M_n(z) \tag{10}$$

and (4), which can be written in the form

$$\mathbf{E}G_n(z) = \frac{n^{(1+z)/2}}{\Gamma(\frac{3+z}{2})} + \mathcal{O}\left(n^{(\Re z-1)/2}\right) = n\cdot n^{(z-1)/2}\left(\frac{1}{\Gamma(\frac{3+z}{2})}+\mathcal{O}\left(\frac{1}{n}\right)\right)$$

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

$$\mathbf{E}G_n(e^{it}) = ne^{\frac{1}{2}(e^{it}-1)(\log n)}\left(\frac{1}{\Gamma(\frac{3+e^{it}}{2})}+\mathcal{O}\left(\frac{1}{n}\right)\right) =$$

$$= ne^{-(t^2/4)\log n+(it/2)\log n}\left(1-\frac{it}{2}\Gamma'(2)-\frac{t^3}{12}i\log n+\mathcal{O}(t^4\log n)\right). \tag{11}$$

On the other hand, $M_n(1)=1$, hence

$$M_n(e^{it}) = 1 + itM_n'(1)+\mathcal{O}(t^2). \tag{12}$$

and trivially

$$c_j(e^{it})-c_j(1) = ct+\mathcal{O}(t)$$

with $c=c_j'(e^{it})|_{t=0}$. Then, by (10), (11) and (12) we conclude that, with probability 1,

$$G_n(e^{it})e^{-kit} = ne^{it((\log n)/2-k)-(t^2/4)\log n}\cdot(ct+\mathcal{O}(t^2+t^4\log n)).$$

uniformly with respect to k. Partial integration gives

$$\int_{-\infty}^{\infty} e^{-(t^2/4)\log n}(t^2+t^4\log n)dt = 72\sqrt{\pi}(\log n)^{-3/2}.$$

For the same reason as in (9), here we also have

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

Hence

$$\frac{I_0}{n} = \frac{1}{2\pi}\int_{-\infty}^{\infty} cte^{it((1/2)\log n-k)-(t^2/4)\log n}dt+\mathcal{O}((\log n)^{-3/2}).$$

Integration gives

$$\frac{I_0}{n} = \mathcal{O}\left(\frac{1}{\log n}\right).$$

We can summarize the results in

$$X^{(\ge j)}[n,k+1]-c_j(1)X[n,k] = \mathcal{O}\left(\frac{n}{\log n}\right).$$

Comparing this with (1) (the result of [6]) completes the proof. □

4 Acknowledgements

The author is greatful to the referees for pointing out minor errors and for their valuable bibliographic suggestions.

References

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

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

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

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

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

[6] KATONA, ZS.,The width of a scale-free tree, *Journal of Applied Probability*, **42.3**, (2005), forthcoming

[7] LU, J.L., Q. FENG, Strong consistency of the number of vertices of given degrees in nonuniform random recursive tree, *Yokohama Math J.* 45 (1998), 61-69.

[8] MAHMOUD, H., R. SMYTHE, A survey of recursive trees, *Theory of Probability and Mathematical Statistics* **51** (1996), 1-29.

[9] MAHMOUD, H., R. SMYTHE, J. SZYMAŃSKI, On the structure of plane-oriented recursive trees and their their branches, *Random Structures & Algorithms* **4** (1993), 151-176.

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

[11] 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.

[12] SZYMAŃSKI, J., On the nonuniform random recursive tree,
Annals of Discrete Math. **33** (1987), 297-306.