Doktori értekezés

Katona Zsolt

2006

Random Graph Models

Doktori értekezés

Informatika Doktori Iskola, Az Informatika Alapjai program, vezető:

Demetrovics János

Katona Zsolt
Témavezető: Móri Tamás, docens

Eötvös Loránd Tudományegyetem
Valószínűségelméleti és Statisztika Tanszék

Contents

1 Introduction **2**

2 Preferential attachment model **6**

2.1 Results on width and level-wise degree distribution 8

2.2 The shape of the tree 11

2.2.1 Using martingales 11

2.2.2 Proof of Theorem 1 20

2.3 Degree distribution of levels 23

2.3.1 Generating functions 23

2.3.2 Proof of Theorem 2 33

3 An application: Directory trees **38**

4 A modified model: Preferential attachment with independent selection **41**

4.1 The maximum degree 42

4.2 Degree distribution 46

5 A Nash-equilibrium model of the World Wide Web **49**

6 Conclusion **51**

7 Acknowledgments **52**

1 Introduction

The dissertation studies random graph models used in describing complex real-world networks, focusing on random trees. The theory of random graphs was introduced by Erdős and Rényi in the early 1960's after Erdős employed random methods to solve extremal graph theory problems. These methods allow to prove the existence of some structures without constructing them. The first application of this method is by Szele [33], who showed in 1943, that there exists a tournament on n vertices, that contains at least $n!/2^{n-1}$ Hamilton cycles. Specifically, if we direct the edges independently with equal probabilities to the two directions, then the expected number of Hamilton cycles is $n!/2^{n-1}$. In 1947, Erdős proved similarly that the Ramsey-number $R(k)$ is at least $2^{k/2}$.

The first random graph models were introduced independently by different authors. The Erdős-Rényi model $\mathcal{G}(n, m)$ is a random selection from all the graphs with n vertices and m edges, with equal probabilities. The model $\mathcal{G}(n, p)$ described by [15] is very similar; edges are selected independently with probability p from all the possible edges between n vertices. (If, for example, $p = 1/2$ then this is equivalent to selecting among all the graphs with n node with equal probabilities.) Of course these models can be generalized in several ways.

As the Internet became an important way of communication, many researchers (physicists, computer scientists) began to study its structure. A secondary structure on the Internet is the World Wide Web (WWW), that is, the directed graph formed by the links connecting the web pages. We can find other large networks with complex structures in the focus of scientific interest, such as the genetic network consisting of proteins as nodes and the chemical reactions between them as links. The nervous system is a network of axons between the nerve cells, whereas social networks describe the interaction between people in an organization or the society. Since these networks are extremely large it is impossible to capture the exact structure. For example, in 2002, the WWW had around $1.4 \cdot 10^9$ pages. The naturally rising idea is to model these networks with random graphs.

One of the first papers that used random methods to study large real-world networks is by Watts and Strogatz [35]. They noticed that real networks have a small diameter, that is, the average distance between nodes is small, and these networks are highly clustered. The $C_1(G)$ clustering coefficient is the average of the node's individual clustering coefficients, that is, the proportion of pairs of the node's neighbors, that are themselves neighbors. Although the diameter in the original Erdős-Rényi random graph model is small, the clustering coefficient is also low, inconsistently with the empirical findings. The "small world" model by Watts and Strogatz [35] solves this problem. In the beginning, the n vertices of the graph lie on a line or a circle and each vertex is connected to its neighbors and second neighbors. In the next step every edge is relinked independently with probability p, that is, one end of the edge is moved to a randomly (uniformly) selected other vertex. If $p = 0$, the graph does not change, thus, the average path length is a linear function of n and each degree is four, whereas if $p \to 1$, then the average distance is a logarithmic function of n.

In 1999, Faloutsos, Faloutsos, and Faloutsos [12] published their paper about the Internet graph. They focused on the so called degree distribution, that is, the series $P(k) := P(d_i = k)$ for $k = 0, 1, 2, ...$, where d_i denotes the degree of node i. The results of empirical studies investigating the degree distributions show an interesting pattern. The Internet graph, for example, has the degree distribution $P(k) \sim c \cdot k^{-\gamma}$, with $\gamma \approx 2.3$. Some papers regarding the World Wide Web, such as the one studying the domain nd.edu [3], and [9] yield that this directed graph has a power-law degree distribution for both the in- and out-degrees with exponents 2.1 and 2.45 respectively. Figure 1 depicts the in-degree distribution of the Hungarian Web [4]. A much studied network is the collaboration graph of movie actors with almost 500,000 vertices. The degree distribution again has the form $P(k) \sim c \cdot k^{-\gamma}$, where $\gamma = 2,3 \pm 0.1$. This graph is also highly clustered, the coefficient is 100 times higher than in a random graph. Another example is the electric network of the western USA with $P(k) \sim c' \cdot k^{-\gamma'}$, where the exponent is around -4 and the coautorship network of scientific publications with an exponent around 3. Similar patterns can observed in the case of cellular networks, phone call networks, citation networks and neural networks. For a review of empirical results on large networks see [2]. Interestingly, power law distributions had been observed much earlier. In 1926, Lotka [24] claimed that citations in scientific publications follow a power-law distribution. Althoug Lotka did not talk about degree distributions, this is obviously the out-degree distribution of a citation network.

Since the empirical degree distributions show the same pattern, the natural question is whether this is consistent with the Erdős-Rényi random graph model or not. Considering the model $\mathcal{G}(n,p)$, that is, when edges are selected independently with probability p from all the possible edges between n vertices, the degree distribution is asymptotically Poisson with parameter np as $n \to \infty$ [7]. That is,

$$P(k) \sim e^{-np} \frac{(np)^k}{k!}.$$

In the $\mathcal{G}(n,m)$ model, the degree distribution is also asymptotically Poisson with expectation $\frac{2m}{n}$

In both cases the number of vertices with degree k is exponentially decreasing and the exponent changes with the number of vertices. Therefore these models are not consistent with the empirical results. Another problem is that the networks under consideration are expanding, whereas the above models consider the number of vertices fixed.

In order to overcome the problem of power-law degree distributions and the natural growth of real networks Barabási and Albert suggested the preferential attachment model [1].

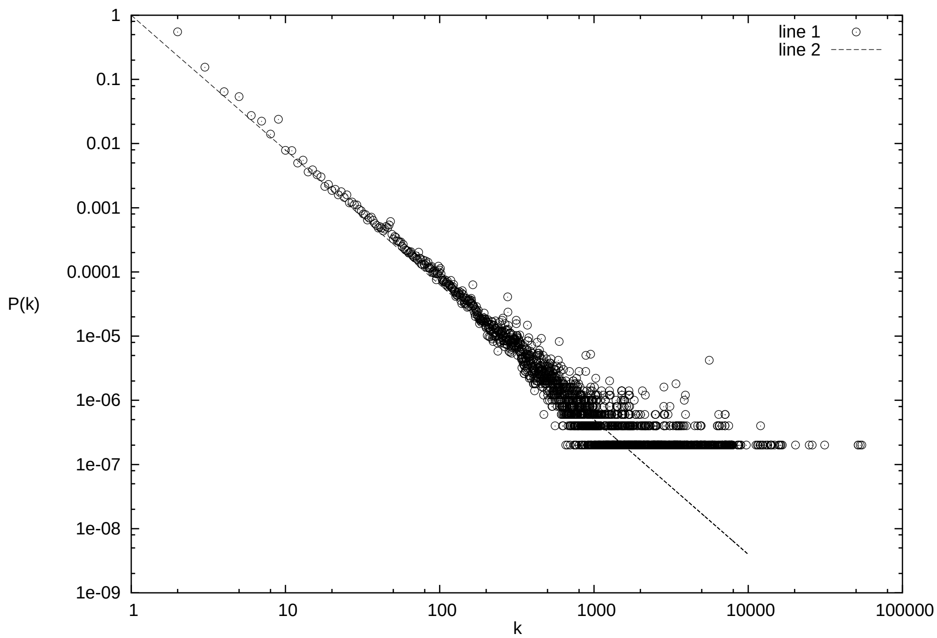

Figure 1: The in-degree distribution of the Hungarian Web [4]. The horizontal axis represents in-degree (k), while the vertical axis measures the number of nodes with a given in-degree $(P(k))$, both on logarithmic scales. The empirical distribution is $P(k) \sim c \cdot k^{-2.29}$.

2 Preferential attachment model

The original description of Barabási and Albert [1] is the following.

"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 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 [27] and [26]). 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. Figure 2 shows a simulated random tree with 1000 nodes.

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 [28] 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 [34] and strong convergence

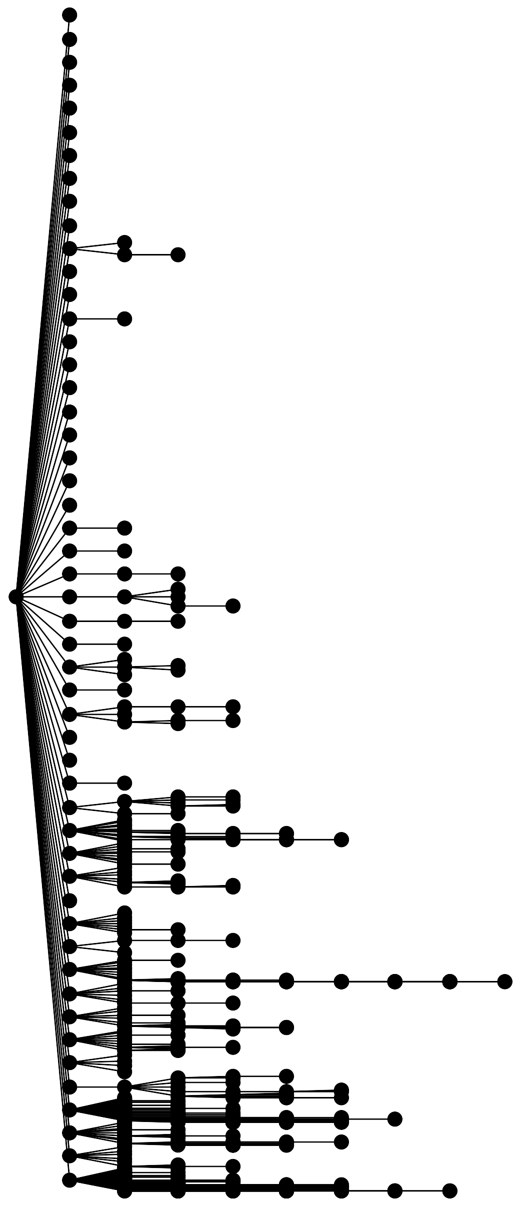

Figure 2: Simulation of the random tree process, $n = 1000$.

of this proportion was shown by Lu and Feng [25]. Rudas, Tóth and Valkó [32] studied similar trees where general weight functions were used to build the tree.

Similar a.s. results were proved in a paper of Bollobás, Riordan, Spencer and Tusnády [6] for general graphs, not only trees. Bollobás and Riordan [8] also determined the diameter of these graphs. Cooper and Frieze [11] obtained results for the maximum degree and degree distribution for a another generalization of the model; they allow new edges to be inserted between existing nodes, a variable number of edges to be added in each step, and endpoints for new vertices are chosen by a mixture of uniform selection and copying or preferential attachment. The recent mathematical results on scale-free graphs are summarized in [5].

The next section summarizes the results obtained in [19] and [20] on the Barabási-Albert type tree obtained with the generalization of Móri, that is, a new edge is connected to an existing vertex with degree k with probability $(k+\beta)/s_n$.

2.1 Results on width and level-wise degree distribution

First we study the shape of the tree. Starting from the root (0th level), we divide the tree to levels. The neighbors of the root will be on level 1, the neighbors 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 $W_n := \max\{X[n,k] : 1 \le k\}$ be its *width* and $H_n := \max\{k \ge 1 : X[n,k] \ne 0\}$ its *height*.

The diameter studied in [8] 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 [31] 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 [10], which they applied to binary search trees for the proof of $W_n \sim \frac{n}{\sqrt{4\pi \log n}}$. The main results regarding the shape 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 $\alpha \log n$. The first image in Figure 3 shows the level sizes of a simulated random tree process with 1000 nodes. The curve shows the function $\frac{n}{\sqrt{2\alpha\pi \log n}} \cdot \exp\left(-\frac{(k-\alpha\log n)^2}{2\alpha\log n}\right)$, which estimates these level sizes. The second image in Figure 3 shows the width of the same tree as n increases from and the function $\frac{n}{\sqrt{2\alpha\pi\log n}}$.

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

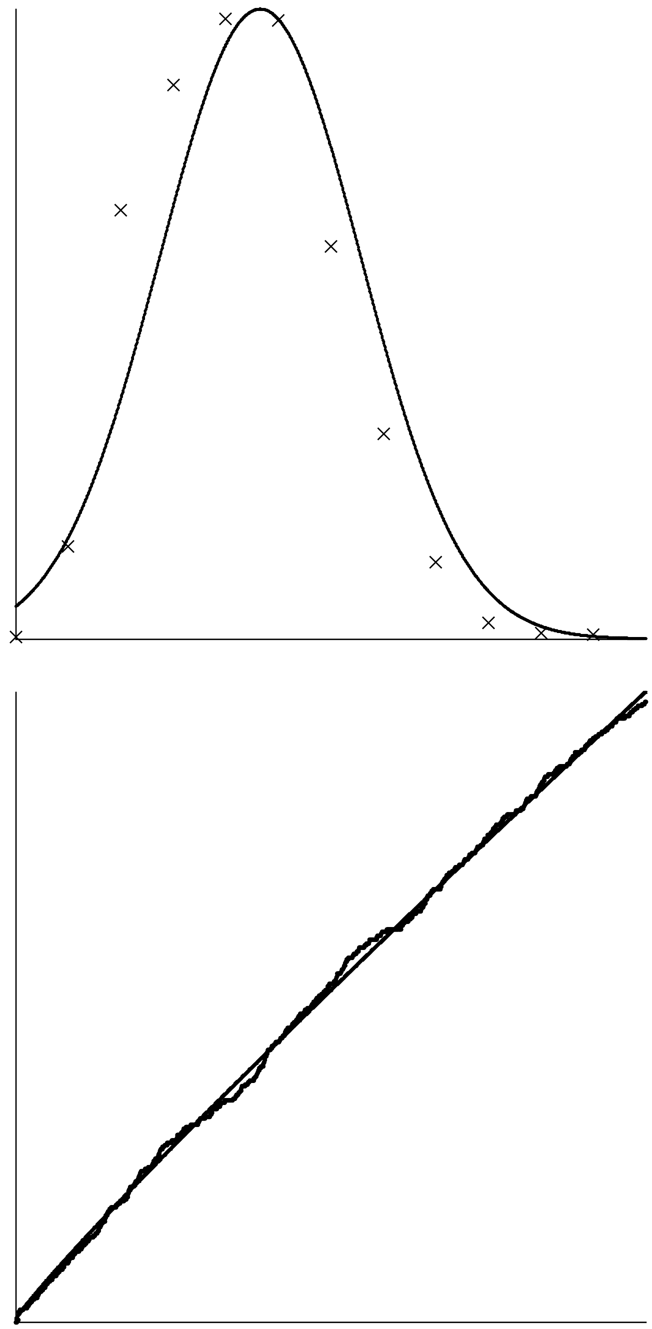

Figure 3: Simulated and theoretical level sizes of a random tree process, ($n = 1500$) and simulated and theoretical width of the same realization of the process, as n increases from 1 to 1500.

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 2 *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. In the next section we will introduce the way of using martingales and prove Theorem 1. The proof of 2 is postponed to Section 2.3.

In Section 3 we will see an application. V. Batagelj called my attention to directory trees, which might follow the preferential attachment model. We will study some of them and see how their widths can be approximated by applying Theorem 1.

2.2 The shape of the tree

2.2.1 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

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

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 [13] and [10], 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*

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

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

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

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

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

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

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 $A_n \leq cB_n$ for every n. By Lemma 2,

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

Hence,

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

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 $2 + \alpha\Re(z_1 + z_2 - z_1 z_2 - 1) > 1$, 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 lines of [10]. The main idea is the following result, which can be proved similarly to Proposition 2 of [10].

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 [16] 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 $|z-1| \le \sqrt{1/\alpha} - \delta$ 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 $\sqrt{1/\alpha} - \delta \le |z-1| \le \sqrt{1/\alpha}$ 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

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

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

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 [31] 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 $n \geq n_0$, if $k > c \log n$. Hence, for sufficiently large n, with probability 1

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

□

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 [10], 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 2 *If $\beta = 0$, the same is true for the function $\frac{|G_n(z)|}{|1+z|}$ on*

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

For any $K > 0$ 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. □

2.2.2 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 $\gamma = \{\xi \,|\, |\xi| = 1, \Re\xi > -0.9\} \cup \{\xi \,|\, \Re\xi = -0.9,\ |\xi| \leq 1\}$. 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 2.

By Lemma 6 and Remark 2, 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

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

The remaining part of the integral is

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

Again we are going to use

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

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

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

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

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

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

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

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

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

Integration gives

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

Hence we have

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

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

□

2.3 Degree distribution of levels

2.3.1 Generating functions

The proof goes on similar lines as the previous one, however, we use generating functions in a slightly different way. 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. Note that $G_n(z)$ denotes a generating functions which is slightly different from the previous section. 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 [34], 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 $j = 1$,

$$\mathbf{E}(G_{n+1}(z)|\mathcal{F}_n) = \frac{2n+1}{2n}G_n(z) + \frac{z}{2n}G_n(z) = \frac{2n+1+z}{2n}G_n(z), \tag{6}$$

and for $j \geq 1$

$$\mathbf{E}(G_{n+1}^{(\geq j+1)}(z)|\mathcal{F}_n) = \frac{2n-j}{2n}G_n^{(\geq j+1)}(z) + \frac{j}{2n}G_n^{(\geq j)}(z). \tag{7}$$

In our calculations, just as in Section 2.2, 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 $G_1(z) = 1$,

$$\mathbf{E}G_n(z) = \mathbf{E}G_n^{(\geq 1)}(z) = \prod_{j=1}^{n-1} \frac{2j+1+z}{2j} = n^{(1+z)/2}\left(1/\Gamma\left(\frac{3+z}{2}\right) + \mathcal{O}(1/n)\right). \tag{8}$$

Remark 3 *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 (8) and (6) 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 7 *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

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

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 (7) 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

$$W_n^{(j)}(z) := \sum_{i=2}^{j} (-1)^{j-i} \binom{j-1}{i-1} U_n^{(i)}(z). \tag{9}$$

The idea is based on a similar combination used in [28]. This combination will cancel out the different coefficients in the proof of Remark 5, originating from the recursive formulas for the generating functions in (7). Easy calculation shows that

$$W_n^{(j)}(z) = \sum_{i=1}^{j} b_i^{(j)}(z) G_n^{(\geq i)}(z), \tag{10}$$

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 4 *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 (9). 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 $2 \le k \le j-1$. 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 5 *For every fixed complex z and fixed $j \ge 2$,*

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

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 (6) and (7), we have

$$\mathbf{E}(W_{n+1}^{(j)}(z)|\mathcal{F}_n) = \sum_{i=1}^{j} b_i^{(j)}(z)\mathbf{E}(G_{n+1}^{(\ge i)}(z)|\mathcal{F}_n) =$$
$$= b_1^{(j)}(z)\frac{2n+1+z}{2n}G_n(z) + \sum_{i=2}^{j} b_i^{(j)}(z)\left(\frac{2n-j+1}{2n}G_n^{(\ge i)}(z) + \frac{i-1}{2n}G_n^{(\ge i-1)}(z)\right)$$

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 8 *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 (10), this can be rewritten as

$$D_n^{(k)}(z_1,z_2)=\sum_{l=1}^{k}\sum_{m=1}^{k}b_l^k(z_1)b_m^k(z_2)C_n^{(l,m)}(z_1,z_2). \tag{11}$$

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

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

and

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

Also we have

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

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

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

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

This gives the formula

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

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

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

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 (8) and Lemma 7. 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 (11), 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 4,

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

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 9 *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 [31] 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.

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

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

□

2.3.2 Proof of Theorem 2

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

Lemma 10 *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 8, 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

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

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 9. Suppose that $|z| = 1$ and $\frac{2\pi l}{K} < \arg z < \frac{2\pi(l+1)}{K}$. Then we have uniformly

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

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 10. We have a.s.

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

The second integral can be approximated just as in Section 2.2, we will use two Lemmas to do so.

Lemma 11 *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 3 we have already seen that $M_n(z)$ is a martingale. Corollary 3 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 12 *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 2 is equal to the $G_n(z)$ of the present section.

□

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 12. By the lemma, for any $L > 0$ we can approximate the second integral as follows.

$$|I_2| \leq \sup_{\gamma(\delta)} |c_j(z) - c_j(1)| \frac{1}{2\pi} \int_{\gamma(\delta)} |G_n(e^{it})| dt \ll \frac{n}{(\log n)^L}. \tag{12}$$

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 11. On the other hand, (8) provides us the asymptotics of the denominator, hence

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

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

$$\frac{1}{2\pi} \int_{(\log n)^{-(1-\vartheta)/2} \leq |t| \leq \pi/2 - \delta} |G_n(e^{it})| dt \ll \\ \ll n \int_{(\log n)^{-(1-\vartheta)/2}}^{\infty} e^{-c' t^2 \log n} dt \ll n e^{-c' (logn)^{\vartheta}} \ll \frac{n}{(\log n)^L}. \tag{13}$$

The remaining part of the integral is

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

Again, we are going to use

$$G_n(z) = \mathbf{E}G_n(z) M_n(z) \tag{14}$$

and (8), 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

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

$$= ne^{-(t^2/4)\log n + (it/2)\log n}\left(1 - \frac{it}{2}\Gamma'(2) - \frac{t^3}{12}i\log n + \mathcal{O}(t^4\log n)\right). \quad (15)$$

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

$$M_n(e^{it}) = 1 + itM_n'(1) + \mathcal{O}(t^2). \quad (16)$$

and trivially

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

with $c = c_j'(e^{it})|_{t=0}$. Then, by (14), (15) and (16) 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 (13), 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 Theorem 1 completes the proof. □

3 An application: 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. Figure 4 shows the subdirectory structure and Figure 5 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	theor. 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 many users who create their directories to store their own files. 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.

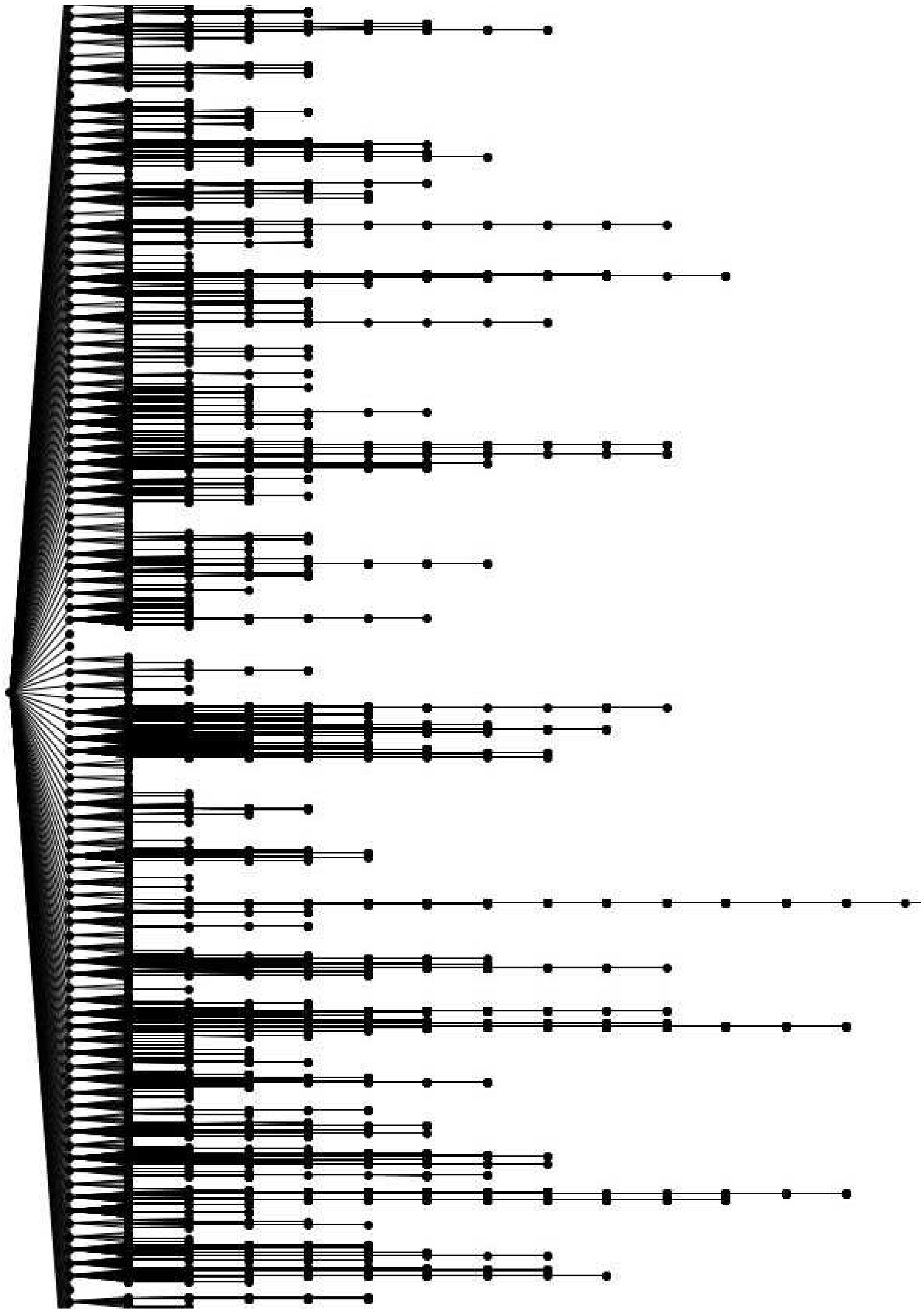

Figure 4: Directory Tree, Department of Computer Science, Budapest University of Technology.

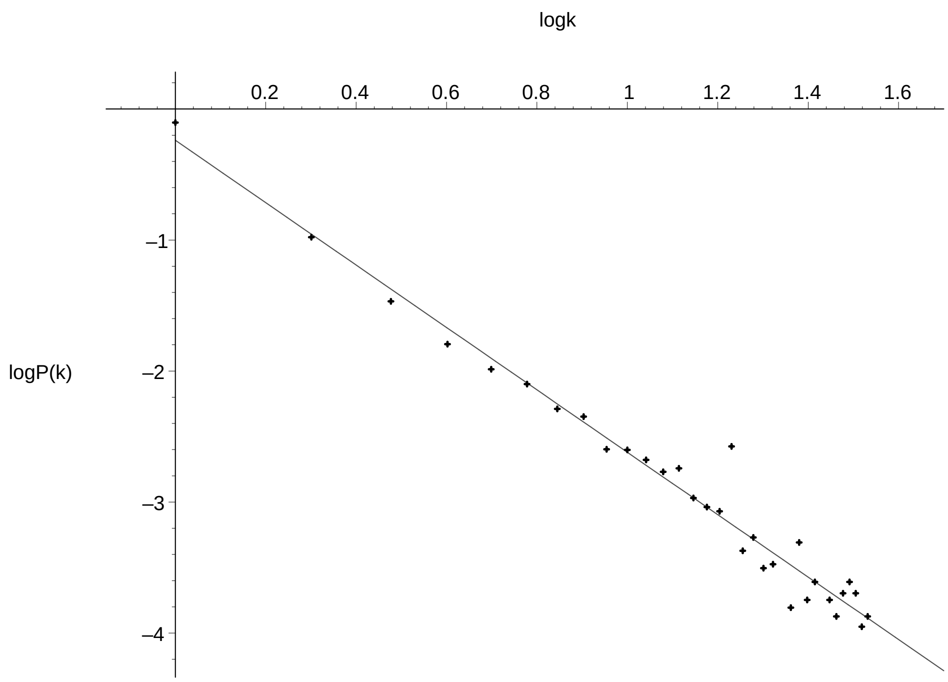

Figure 5: Degree distribution, Department of Computer Science, Budapest University of Technology.

4 A modified model: Preferential attachment with independent selection

Consider the following modification of the Barabási–Albert random graph [21]. At every step a new vertex is added to the graph. It is connected to the old vertices randomly, with probabilities proportional to the degree of the other vertex, and independently of each other. Since we are interested in asymptotic analysis, the initial configuration can be arbitrary, but, for the sake of simplicity, let us start from the very simple graph consisting of two points and the edge between them.

Let us number the vertices in the order of their creation; thus the vertex set of the graph after n steps is $\{0, 1, \dots, n\}$. Let $U[n, k]$ and $V[n, k]$ denote the number of vertices of degree exactly k and at least k, resp., after n steps. Thus, $U[n, 0] + U[n, 1] + \cdots = n+1$. Let $S_n = \sum_{k\geq 1} kU[n, k] = \sum_{k\geq 1} V[n, k]$, the sum of degrees, or equivalently, the double of the number of edges. At the nth step an old vertex of degree k is connected to the new one with probability $\lambda k/S_{n-1}$. This quantity remains below 1, provided the proportionality coefficient λ is less than 2, which will therefore be assumed in the sequel.

Let $\mathcal{F}_n$ denote the σ-field generated by the first n steps. Let $\Delta[n, k]$ be the number of new edges into the set of old vertices of degree k at the nth step, and let $\Delta_n = \sum_{k\geq 1} \Delta[n, k]$ be the total number of new edges. Obviously, the conditional distribution of $\Delta[n+1, k]$ with respect to $\mathcal{F}_n$ is binomial with parameters $U[n, k]$ and $\lambda \dfrac{k}{S_n}$, hence $E\left(\Delta_{n+1} | \mathcal{F}_n\right) = \lambda$.

The aim of the present section is to study some asymptotic properties of this random graph as the number of vertices tends to infinity. In Section 4.1 we prove a strong law of large numbers for the maximum degree, and in Section 4.2 it will be shown that the proportion of vertices of degree k converges a.s. to a constant, which, as a function of k, decreases in the order of k^{-3} as $k \to \infty$.

4.1 The maximum degree

First we deal with the asymptotics of S_n.

Theorem 3 $S_n = 2\lambda n + o\left(n^{1/2+\varepsilon}\right)$, $\forall \varepsilon > 0$.

PROOF: With $\Delta_1 = 1$ define $\zeta_n = \sum_{j=1}^{n} (\Delta_j - \lambda) = S_n/2 - n\lambda$. Then $(\zeta_n, \mathcal{F}_n)$ is a square integrable martingale, and the increasing process associated with ζ_n^2 by the Doob decomposition is

$$A_n = \sum_{j=2}^{n} \mathrm{Var}\left(\Delta_j | \mathcal{F}_{j-1}\right) = \sum_{j=2}^{n-1} \sum_{k \geq 1} U[j,k] \frac{k\lambda}{S_j} \left(1 - \frac{k\lambda}{S_j}\right) \leq n\lambda.$$

It is well known ([30], Proposition VII-**2**-4) that $\zeta_n = o\left(A_n^{1/2+\varepsilon}\right)$ a.e. on the event $A_n \to \infty$. This completes the proof. $\square$

Let us turn to $M_n = \max\{k : U[n,k] > 0\}$, the maximum degree after n steps.

Theorem 4 *We have* $\lim_{n\to\infty} M_n/\sqrt{n} = \mu$ *with probability* 1, *where the limit* μ *differs from zero with positive probability.*

PROOF: We follow the same lines as in the proof of Theorem 3.1 of [29].

Let $W[n,j]$ denote the degree of vertex j after the nth step, with the initial values $W[n,j] = 0$ for $n < j$, $W[1,0] = W[1,1] = 1$, $W[j,j] = \Delta_j$. Then $M_n = \max\{W[n,j] : j \geq 0\}$. Let us introduce the normalizing terms

$$c[n,k] = \prod_{i=1}^{n-1} \frac{S_i}{S_i + k\lambda}, \quad n \geq 1,\ k \geq 1.$$

For $n \to \infty$, with probability 1 we have

$$c[n,k] = \exp\left(-k\lambda \sum_{i=1}^{n-1} \frac{1}{S_i} + \frac{k^2\lambda^2}{2} \sum_{i=1}^{n-1} \frac{1+o(1)}{S_i^2}\right).$$

Since $\frac{1}{S_i} = \frac{1}{2\lambda i}(1 + o\left(i^{-1/2+\varepsilon}\right))$, we obtain that the exponent in 4.1 differs from $-\frac{k}{2}\log n$ only by a term converging with probability 1. Thus $c[n,k] \sim \gamma_k\, n^{-k/2}$, with an appropriate positive random variable γ_k.

We clearly have that

$$E\left(W[n+1,j]|\mathcal{F}_n\right) = W[n,j] + \lambda\,\frac{W[n,j]}{S_n} = W[n,j]\,\frac{S_n+\lambda}{S_n}.$$

Hence

$$\left(Z[n;j,1] = c[n,1]\,W[n,j], \mathcal{F}_n\right), \quad n \geq \max\{j,1\}$$

is either a positive martingale or constant zero, thus it converges a.s. to some ζ_j. To estimate the moments of ζ_j, consider

$$Z[n;j,k] = c[n,k]\binom{W[n,j]+k-1}{k}.$$

Since $W[n+1,j] - W[n,j]$ is equal to either 1 or 0, we can write

$$\binom{W[n+1,j]+k-1}{k} =$$
$$= \binom{W[n,j]+k-1}{k} + (W[n+1,j] - W[n,j])\binom{W[n,j]+k-1}{k-1} =$$
$$= \binom{W[n,j]+k-1}{k}\left(1 + \frac{k(W[n+1,j] - W[n,j])}{W[n,j]}\right).$$

By taking conditional expectation with respect to $\mathcal{F}_n$ we can see that

$$E\left(\binom{W[n+1,j]+k-1}{k}|\mathcal{F}_n\right) = \binom{W[n,j]+k-1}{k}\left(1 + \frac{k\lambda}{S_n}\right),$$

hence

$$(Z[n;j,k],\,\mathcal{F}_n)\,,\quad n\geq\max\{j,1\}$$

is also a (convergent) martingale. Since $c[n,1]^k \leq c[n,k]$, we can majorize $Z[n;j,1]^k$ by $k!\,Z[n;j,k]$.

Now, $c[n,1]M_n = \max\{Z[n;j,1] : 0\leq j\leq n\}$, being the maximum of an increasing number of nonnegative martingales, is a submartingale. The proof can be completed by showing that this submartingale is bounded in L_k, for some $k\geq 1$.

Let us start from the estimation

$$E\left(c[n,1]M_n\right)^k = E\left(\max\{Z[n;j,1]^k : 0\leq j\leq n\}\right)\leq k!\sum_{j=0}^{n} EZ[n;j,k] =$$

$$= k!EZ[1;0,k] + k!\sum_{j=1}^{\infty} EZ[j;j,k] = k! + k!\sum_{j=1}^{n} E(c[j,k]\binom{\Delta_j+k-1}{k}).$$

Here

$$E(c[j,k]\binom{\Delta_j+k-1}{k}) = E[E\left(c[j,k]\binom{\Delta_j+k-1}{k}|\mathcal{F}_{|-\infty}\right)] =$$

$$= E[c[j,k]E\left(\binom{\Delta_j+k-1}{k}|\mathcal{F}_{|-\infty}\right)].$$

Next we show that, independently of j,

$$E\left(\binom{\Delta_j+k-1}{k}|\mathcal{F}_{|-\infty}\right)\leq E\binom{\pi+k-1}{k},$$

where π stands for a Poisson(λ) random variable. Remembering that $\Delta_j = \Delta[j,0]+\cdots+\Delta[j,j-1]$, we can write

$$\binom{\Delta_j+k-1}{k} = \sum_{\ell_0+\cdots+\ell_j=k}\binom{\Delta[j,0]}{\ell_0}\cdots\binom{\Delta[j,j-1]}{\ell_{j-1}}\binom{k-1}{\ell_j}.$$

The binomial coefficients on the right-hand side are conditionally independent. The conditional distribution of each $\Delta[j,i]$ is binomial. Let ξ be a

Binomial(n,p) random variable and η a Poisson one, with the same expectation. Then

$$E\binom{\xi}{\ell} = \binom{n}{\ell}p^\ell \le \frac{(np)^\ell}{\ell\,!} = E\binom{\eta}{\ell}.$$

Thus, if all random variables $\Delta[j,i]$ on the right-hand side of 4.1 are replaced by conditionally independent Poisson variables, the conditional expectation cannot decrease. Hence 4.1 follows.

By 4.1 and 4.1, for the L_k-boundedness of the submartingale $c[n,1]M_n$ it is sufficient to check that $\sum_{j=1}^{\infty} Ec[j,k] < \infty$. The convergence of $\sum_{j=1}^{\infty} c[j,k]$ when $k > 2$ is clear from the asymptotics obtained for $c[n,k]$, but the integrability does not follow immediately.

Let $k = 8$ and $N = \max\{n : S_n > 4\lambda n\}$. Then for $j > N$ we have

$$c[j,8] = \prod_{i=1}^{j-1}\left(1 - \frac{8\lambda}{S_i + 8\lambda}\right) \le \prod_{i=N+1}^{j-1}\left(1 - \frac{2}{n+2}\right) = \frac{(N+1)(N+2)}{j(j+1)},$$

but this is obviously true even for $j \le N$. Thus, for the finiteness of $\sum_{j=1}^{\infty} Ec[j,8]$ it is sufficient to prove that $EN^2 < \infty$.

By the usual large deviation arguments we have

$$P(N=n) \le P\left(S_n > 4\lambda n\right) = P\left(2^{S_n/2} > 4^{\lambda n}\right) \le E\left(2^{S_n/2}\right)4^{-\lambda n}.$$

Thus, we have to estimate the moment generating function of S_n. With $\Delta_1 = 1$ we can write $S_n/2 = \sum_{i=1}^{n} \Delta_i$, and

$$\begin{aligned} E\left(2^{S_n/2}\right) &= E(E\left(2^{S_n/2}|\mathcal{F}_{\setminus-\infty}\right)) = E(2^{S_{n-1}/2}E\left(2^{\Delta_n}|\mathcal{F}_{\setminus-\infty}\right)) = \\ &= E\left(2^{S_{n-1}/2}\prod_{j=1}^{n-1}\left(1+\frac{\lambda j}{S_{n-1}}\right)^{U[n-1,j]}\right) \le \\ &\le E\left(2^{S_{n-1}/2}\exp\left\{\sum_{j=1}^{n-1}\frac{\lambda j U[n-1,j]}{S_{n-1}}\right\}\right) = e^\lambda E\left(2^{S_{n-1}/2}\right). \end{aligned}$$

Therefore $E\left(2^{S_n/2}\right) \leq e^{\lambda n}$, which, combined with 4.1, implies that $P(N = n) \leq (e/4)^{\lambda n}$. Thus $EN^2 < \infty$, indeed. □

4.2 Degree distribution

In this section we prove that the degree distribution of our graph stabilizes almost surely, as $n \to \infty$, around a power law with exponent -3.

First we show that stochastic process of new vertices approaches a stationary regime. More precisely, the random variables Δ_n are asymptotically independent and asymptotically Poisson(λ) distributed. By LeCam's theorem on Poisson approximation [23] we have

$$\sum_{k=0}^{\infty} \left| P\left(\Delta_{n+1} = k | \mathcal{F}_n\right) - \frac{\lambda^k}{k!}\, e^{-\lambda} \right| \leq 2 \sum_{k=1}^{n} U[n,k] \left(\frac{k\lambda}{S_n}\right)^2 \leq 2\lambda^2 \frac{M_n}{S_n} = o(1).$$

Similarly, the conditional distribution of $\Delta[n+1, k]$ with respect to $\mathcal{F}_n$ is also asymptotically Poisson with parameter $\lambda k\, \dfrac{U[n,k]}{S_n}$.

Theorem 5 *For every $k = 0, 1, \ldots$ The proportion of vertices of degree k converges a.s. as $n \to \infty$:*

$$P\left(\lim_{n\to\infty} \frac{U[n,k]}{n+1} = x_k\right) = 1,$$

where

$$x_0 = p_0, \quad x_k = \frac{2}{k(k+1)(k+2)} \sum_{i=1}^{k} i(i+1)p_i, \quad p_k = \frac{\lambda^k}{k!}\, e^{-\lambda}.$$

Remark 6

$$x_k \sim \frac{2\lambda(2+\lambda)}{k^3} \quad \textit{as } k \to \infty.$$

Proof: First we will show by induction over k that

$$P\left(\lim_{n\to\infty}\frac{V[n,k]}{n+1}=y_k\right)=1,$$

where the constants y_k satisfy the following recursion:

$$y_0=1,\quad y_k=\frac{k-1}{k+1}\,y_{k-1}+\frac{2q_k}{k+1},\quad q_k=p_k+p_{k+1}+\dots\ .$$

The sequence $U[n,0]$ obeys the strong law of large numbers, because [11], Corollary VII-2-6 implies that

$$U[n,0]=\sum_{i=1}^{n}I(\Delta_i=0)\sim\sum_{i=1}^{n}P\left(\Delta_i=0|\mathcal{F}_{\rangle-\infty}\right)\sim ne^{-\lambda}.$$

Hence we obtain that

$$P\left(\lim_{n\to\infty}\frac{V[n,1]}{n+1}=1-e^{-\lambda}\right)=1.$$

For the induction step suppose our assertion holds true for $k-1$. This time we introduce the normalizing factors

$$d[n,k]=\prod_{i=1}^{n-1}\left(1-I\left(S_i\geq 2(k-1)\right)\frac{(k-1)\lambda}{S_i}\right)^{-1}.$$

Their asymptotic behaviour can be treated similarly to what we have done in 4.1. Thus, with probability 1, we can write

$$d[n,k]=\exp\left((k-1)\lambda\sum_{i=1}^{n-1}I\left(S_i\geq 2(k-1)\right)\frac{1}{S_i}+\frac{(k-1)^2\lambda^2}{2}\sum_{i=1}^{n-1}\frac{1+o(1)}{S_i^2}\right).$$

By Theorem 3 we have $\frac{1}{S_i}=\frac{1}{2\lambda i}(1+o\left(i^{-1/2+\varepsilon}\right))$, thus the exponent differs from $\frac{k-1}{2}\log n$ only by an a.s. convergent term. Therefore $d[n,k]\sim\delta_k\,n^{(k-1)/2}$, with some random variable $\delta_k>0$.

Since

$$V[n+1,k] = V[n,k] + \Delta[n+1,k-1] + I\left(\Delta_{n+1} \geq k\right),$$

it is easy to see that

$$E\left(d[n+1,k]\, V[n+1,k] | \mathcal{F}_n\right) = d[n,k]\, V[n,k] + b[n,k],$$

where

$$b[n,k] = d[n+1,k]\,(k-1)\lambda\, \frac{V[n,k-1]}{S_n} + d[n+1,k]\, P\left(\Delta_{n+1} \geq k | \mathcal{F}_n\right) \leq$$
$$\leq d[n+1,k](\lambda+1).$$

In addition, we have

$$\begin{aligned}
&\mathrm{Var}(d[n+1,k]\, V[n+1,k] | \mathcal{F}_n) = \\
&\quad = d[n+1,k]^2\, \mathrm{Var}(\Delta[n+1,k-1] + I(\Delta_{n+1} \geq k)) \leq \\
&\quad \leq 2d[n+1,k]^2 (\mathrm{Var}(\Delta[n+1,k-1]) + \mathrm{Var}(I(\Delta_{n+1} \geq k))) \leq \\
&\quad \leq 2d[n+1,k] b[n,k] = O(n^{k-1}). \quad (17)
\end{aligned}$$

Let us introduce a square integrable martingale by its differences

$$\xi_n = d[n,k]\, V[n,k] - d[n-1,k]\, V[n-1,k] - b[n-1,k].$$

The increasing process associated with the square of this martingale is

$$A_n = \sum_{i=1}^{n} E\left(\xi_i^2 | \mathcal{F}_{\rangle-\infty}\right) = \sum_{i=1}^{n} \mathrm{Var}\left(d[i,k] V[i,k] | \mathcal{F}_{\rangle-\infty}\right),$$

which is of order $O(n^k)$ by 17. Hence

$$\sum_{i=1}^{n} \xi_i = d[n,k]\, V[n,k] - \sum_{i=1}^{n} b[i-1,k] = o\left(n^{k/2+\varepsilon}\right).$$

From all these we obtain that

$$\frac{V[n,k]}{n+1} = \frac{1}{(n+1)d[n,k]} \sum_{i=1}^{n} b[i-1,k] + o(1).$$

Now, by 4.2 and the induction hypothesis,

$$b[i,k] \sim d[i+1,k]\left(\frac{k-1}{2}\, y_{k-1} + q_k\right).$$

By applying the asymptotics we obtained for $b[i,k]$ we arrive at the formula

$$\frac{V[n,k]}{n+1} = \frac{k-1}{k+1}\, y_{k-1} + \frac{2q_k}{k+1} + o(1),$$

which was to be proved.

Next we show that the solution of the recursion 4.2 is $y_k = \frac{2}{k(k+1)} \sum_{i=1}^{k} iq_i$.

Let $r_k = q_k + q_{k+1} + \dots$ and $z_k = y_k + y_{k+1} + \dots$, then from the recursion for y_k one can derive that $kz_k = (k-1)z_{k-1} + 2r_k$. From that we have

$$z_k = \frac{2}{k} \sum_{i=1}^{k} r_i,$$

which easily yields the above mentioned explicite form of y_k, and finally, the desired expression for x_k.

□

5 A Nash-equilibrium model of the World Wide Web

Although the Barabási and Albert model is consistent with empirical data for several different types of networks, it has important shortcomings in modeling the World Wide Web. First, it treats the network members as identical, which is a rather unrealistic assumption in the context of Web sites. Furthermore, a model with homogenous network members and thus a fixed number of out-links cannot say much about the distribution of out-degrees in the graph. The empirical evidence shows that this distribution is similar to the one for in-degrees but the Barabási and Albert model does not predict this. Second, and most importantly, this model is 'ad hoc' in the sense that it does not consider the network members as strategic agents acting deliberately in their own interest. Why would nodes (or agents in the nodes) select which links to establish one after the other without any interaction? Why would they consider solely the degrees to make their decisions? From an economic perspective the interesting question is: what incentives drive agents' choices of the nodes and how these choices depend on their inherent characteristics (e.g. their content)?

The model proposed in [22] explicitly addresses these issues, presenting an approach based on game theory. The basic assumption is that the nodes represent rational economic agents (e.g. web sites) who make simultaneous and deliberate decisions on the links they establish between themselves. Agents are strategic and heterogeneous with respect to their endowed "content", which may be thought of as their value in the eyes of the public/market. The utility of a node depends on its content and the structure of the network.

The objective is to find the pure strategy Nash-equilibria of the game where players maximize their utilities and their strategies consist of establishing links from one another in a simultaneous decision. An equilibrium of the game represents a graph and our main interest is in describing the structure of this graph. The results show how the in- and out-degrees depend on the node's contents and that in- and out-degree distributions are similar.

6 Conclusion

In the past few years, the Internet and other large networks became the foci of network research. In order to model these complex researcher in different disciplines proposed to use random methods. Since the original Erdős-Rényi random graph model was found to be incosistent with several empirical patterns in these networks, new models were proposed. The Watts-Strogatz small world model is a simple mixture of a deterministic network and the Erdős-Rényi model. It is still inconsistent with empirical findings regarding the degree distribution. Several examples show that real networks are scale-free with a power law degree distribution. The preferential attachment model proposed by Barabási and Albert solves this problem. Although a variant of the model had been previously known in mathematics, its introduction by Barabási and Albert generated a vast amount of research in the field. The review papers [2] and [5] summarize the empirical findings and the mathematical results in the area.

In this dissertation we studied the random tree process based on the preferential attachment model. We determined the width of the tree and showed that the degree distribution is similar to that of the whole tree on the middle levels. In the original preferential attachment model the new vertex is connected to the graph with a given number of edges randomly with probabilities proportional to the degrees of old vertices. In Section 4, we suggested a modified model where the new node is connected to the graph with edges that are established independently with probabilities proportional to the degrees. We showed that the maximum degree and the degree distribution is similar to the original model.

The papers [17],[14],[18] summarize the work of the author in the field of extremal combinatorics.

7 Acknowledgments

I grateful to my dissertation supervisor, Tamás Móri, for his support and guidance throughout the PhD program. I also appreciate the suggestions and comments of the anonymous reviewers of the papers that the dissertation is based on.

References

[1] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[2] A.-L. Barabási and R. Albert. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47–96, 2002.

[3] A.-L. Barabási, R. Albert, and Hawoong Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A*, 281:69–77, 2000.

[4] A. Benczúr, K. Csalogány, D. Fogaras, E. Friedman, T. Sarlós, M. Uher, and E. Windhager. Searching a small national domain – preliminary report. In *Proceedings of WWW*, 2003.

[5] B. Bollobás and O. Riordan. *Handbook of Graphs and Networks*, chapter Mathematical results on scale-free graphs. Wiley-VCH, Berlin, 2002.

[6] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures and Algorithms*, 18:279–290, 2001.

[7] B. Bollobás. *Random Graphs*. Academic Press, London, 1985.

[8] B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24:5–34, 2004.

[9] A.Z. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J.L. Wiener. Graph structure in the web. *Computer Networks*, 33:309–320, 2000.

[10] B. Chauvin, M. Drmota, and J. Jabbour-Hattab. The profile of binary search trees. *The Annals of Applied Probability*, 11:1042–1062, 2001.

[11] C. Cooper and A. Frieze. A general model of web graphs. *Random Structures and Algorithms*, 22:311–335, 2003.

[12] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *Comp. Comm. Rev.*, 29:251–262, 1999.

[13] P. Flajolet and A.M. Odlyzko. Singularity analysis of generating functions. *SIAM J. Discrete Math.*, 3:216–240, 1990.

[14] Z. Furedi and Zs.Katona. Multiply intersecting families of sets. *Journal of Combinatorial Theory, Series A*, 106(2):315–326, 2004.

[15] E.N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30:1141–1144, 1959.

[16] A. Joffe, L. Le Cam, and J. Neveu. Sur la loi des grands nombres pour des variables aléatoires attachées à un arbre dyadique. *C. R. Acad. Sci. Paris. Ser. A*, 277:963–964, 1973.

[17] Zs. Katona. Intersecting families of sets, no l containing two common elements. *Discrete Mathemathics*, 226:233–241, 2001.

[18] Zs. Katona. 3-wise exactly 1-intersecting families of sets. *Graphs and Combinatorics*, 21(1):71–76, 2005.

[19] Zs. Katona. The width of a scale-free tree. *Journal of Applied Probability*, 42.3:839–850, 2005.

[20] Zs. Katona. Levels of a scale-free tree. *Random Structures and Algorithms*, 29(2):194 – 207, 2006.

[21] Zs. Katona and T.F. Móri. A new class of scale free random graphs. *Statistics and Probability Letters*, 76(15):1587–1593, 2006.

[22] Zs. Katona and M. Sarvary. Network formation and the structure of the commercial world wide. *Insead working paper*, 2006/04/MKT.

[23] L. LeCam. An approximation theorem for the poisson binomial distribution. *Pacific J. Math.*, 10:1181–1197, 1960.

[24] A.J. Lotka. The frequency distribution of scientific productivity. *J. of the Washington Acad. of Sci.*, 16:317, 1926.

[25] J. Lu and Q. Feng. Strong consistency of the number of the vertices of given degrees in nonuniform random recursive trees. *Yokohama Math*, 45:61–69, 1998.

[26] H. Mahmoud, R. Smythe, and J. Szymański. On the structure of plane-oriented recursive trees and their their branches. *Random Structures & Algorithms*, 4:151–176, 1993.

[27] R. Smythe Mahmoud, H. A survey of recursive trees. *Theory of Probability and Mathematical Statistics*, 51:1–29, 1996.

[28] T.F. Móri. On random trees. *Studia Sci. Math. Hungar.*, 39:143–155, 2000.

[29] T.F. Móri. The maximum degree of the barabási random tree. *Combin. Probab. Comput.*, 14:339–348, 2005.

[30] J. Neveu. *Discrete-Parameter Martingales*. North-Holland, Amsterdam, 1975.

[31] B. Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures Algorithms*, 5(2):337–347, 1994.

[32] A. Rudas, B. Tóth, and B. Valkó. Random trees and general branching processes. *Random Structures and Algorithms*, forthcoming, 2006.

[33] T. Szele. Combinatorial investigations concerning complete directed graphs. *Matematikai és Fizikai Lapok*, 50:287–293, 1943.

[34] J. Szymański. On the nonuniform random recursive tree. *Annals of Discrete Math.*, 33:297–306, 1987.

[35] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-worlds' networks. *Nature*, 393:440–442, 1998.

Random Graph Models

Summary

Zsolt Katona

The dissertation studies random graphs models used in describing complex real-world networks, focusing on random trees. The theory of random graphs was introduced by Erdős and Rényi in the early 60's after Erdős employed random methods to solve extremal graph theory problems. In the 90's, as the Internet became an important way of communication, many researchers (physicists, computer scientists) began to study its structure. In order to overcome the problem of power-law degree distributions and the natural growth of real networks Barabási and Albert suggested the preferential attachment model. 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 a specific case of the Barabási-Albert model, the resulting graph is a tree. 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. 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 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 dissertation, I summarize the results obtained on the Barabási-Albert type tree obtained with the generalization of Móri. First, I study the shape of the tree. Starting from the root (0th level), we divide the tree to

levels. The neighbors of the root will be on level 1, the neighbors 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 $\alpha = \frac{1+\beta}{2+\beta}$. Then we show that as $n \to \infty$, $W_n \sim \frac{n}{\sqrt{2\alpha\pi \log n}}$. In addition our results also yield that the width 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. 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.

The next results are about the following modification of the Barabási–Albert random graph. At every step a new vertex is added to the graph. It is connected to the old vertices randomly, with probabilities proportional to the degree of the other vertex, and independently of each other. Since we are interested in asymptotic analysis, the initial configuration can be arbitrary but, for the sake of simplicity, let us start from the very simple graph consisting of two points and the edge between them. We study some asymptotic properties of this random graph as the number of vertices tends to infinity. First we we prove a strong law of large numbers for the maximum degree. Let M_n denote the maximum degree after step n. Then we have $\lim_{n\to\infty} M_n/\sqrt{n} = \mu$ with probability 1, where the limit μ differs from zero with positive probability. It is also shown that the proportion of vertices of degree k converges a.s. to a constant, which, as a function of k, decreases in the order of k^{-3} as $k \to \infty$.

Véletlen Gráf Modellek

Összefoglaló

Katona Zsolt

A dolgozat témája komplex hálózatok modellezésére használt véletlen gráfok tulajdonságainak vizsgálata. A véletlen gráfok elméletét Erdős és Rényi vezette be a korai 60-as években, miután Erdős véletlen módszerekkel oldott meg egy extremális kombinatorikai problémat. A 90-es évek végén, ahogy az Internet elterjedt, sok tudós kezdte tanulmányozni a struktúráját. A mindenhol tapasztalt skála-független fokszámelolszások megmagyarázása és a hálózatok természetes növekedésének modellezése érdekében Barabási és Albert bevezette a következőmodellt.

Az új véletlen fa modellben egy pontból indulunk ki, ez a gyökér. Az első lépésben hozzáveszünk egy új pontot és összekötjük a gyökérrel. Minden további lépésben hozzáveszünk egy új pontot, amelyet egy régi ponttal kötünk össze, minden régi pontot a fokszámával arányos, azaz

$$\Pi(i) = \frac{d_i}{\sum_j d_j}$$

valószínűséggel választva.

Móri Tamás általánosította az eljárást a következő módon. Válassznuk egy tetszőleges monoton növő, pozitív egészeken értelmezett, pozitív értékű φ fügvényt, jelölje S_n a $\sum_{j=1}^{n} \varphi(d_j)$. Az eddigi fokszámarányos valószínűség helyett most

$$\Pi(i) = \frac{\varphi(d_i)}{S_n}$$

valószínűséggekel kössük az új csúcsot az i., már csúcshoz. Különösen érdekes a $\varphi(k) = k + \beta$ eset ($\beta > -1$), hiszen ekkor $S_n = 2n + \beta(n + 1)$, tehát nem véletlen.

Móri Tamás megmutatta, hogy a k fokú csúcsok aránya 1 valószínűséggel konvergál egy c_k határértékhez, amely, mint k függvénye $k^{(-3+\beta)}$ alakban csökken, ami megfelel az empirikus eredményeknek.

A disszertációban meghatározom a fa szélességét. A fát a gyökértől indulva szintekre bontva, a legtöbb csúcsot tartalmazó szint adja a fa szélességét. Ha $n \to \infty$, akkor n lépés után a fa szélessége 1 valószínűséggel

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

ahol $\alpha = \frac{1+\beta}{2+\beta}$. Tehát a $\beta = 0$ esetben

$$W_n \sim \frac{n}{\sqrt{\pi \log n}}.$$

Természetesen vetődik fel a kérdés, hogy a fokszámeloszlás, hogy változik szintenként. Móri megmutatta, hogy az első szinten nem ugyanaz, mint az egész fában, tehát nem lehet ugyanaz szintenként. A disszertációban azonban megmutatom, hogy a középső szinteken ($\alpha \log n$ körül), melyek a csúcsok többségét tartalmazzák, a fokszámeloszlás ugyanaz, mint az egész gráfban.

A további eredmények egy módosított modellre vonatkoznak, amelyben továbbra is egyenként adjuk a csúcsokat a gráfhoz, de az éleket máshogy választjuk. Az eredeti Erdős-Rényi modellhez hasonlóan az új csúcsból az éleket egymástól függetlenül, de továbbra is fokszámarányos valósznűséggel húzzuk be. A gráf aszimptotikus tulajdonságait tanulmányozva meghatározzuk a gráf maximális fokszámát és fokszámeloszlását. Pontosabban, ha M_n jelöli a maximális fokszámot n lépés után, akkor $\lim_{n\to\infty} M_n/\sqrt{n} = \mu$ 1 valószínűséggel, ahol μ pozitív valószínűséggel nem zérus. A fokszámeloszlás pedig hasonló az eredeti modellhez, azaz a k fokú pontok aránya egy konstanshoz konvergál, amint $n \to \infty$, és a konstans k^{-3} nagyságrendben csökken.