Fifth Colloquium on Mathematics and Computer Science DMTCS proc. **AI**, 2008, 149–162

The Degree Distribution of Thickened Trees[†]

Michael Drmota[1], Bernhard Gittenberger[1], and Alois Panholzer[1]

[1] *Institute of Discrete Mathematics and Geometry, TU Wien, Wiedner Hauptstrasse 8-10, A-1040 Vienna, Austria*

We develop a combinatorial structure to serve as model of random real world networks. Starting with plane oriented recursive trees we substitute the nodes by more complex graphs. In such a way we obtain graphs having a global tree-like structure while locally looking clustered. This fits with observations obtained from real-world networks. In particular we show that the resulting graphs are scale-free, that is, the degree distribution has an asymptotic power law.

Keywords: scale free networks, recursive trees, generating functions

1 Introduction

There has been substantial interest in random graph models where vertices are added to the graph successively and are connected to several already existing nodes according to some given law. The so-called Albert-Barabási model (see Albert and Barabási (2002)) joins a new node to an existing one with probability proportional to the degree. The idea behind is to model various *real-word graphs* like the internet or social networks.

It turns out that the Albert-Barabási model is not unambiguously defined. One rigorous approach is due to Bollobás and Riordan (2004) They introduced a random (multi)graph G_m^n. For example, G_1^n is described in the following way. One starts with an initial node 1 with a loop. (This means that 1 has degree 2). Then at step k we add one node that is connected to exactly one of the k already present nodes, say to $j \le k$, with probability

$$\begin{array}{ll} \dfrac{\deg_{G_1^{k-1}}(j)}{2k-1} & \text{if } j < k, \\[2ex] \dfrac{1}{2k-1} & \text{if } j = k. \end{array}$$

Of course, after n steps we have produced a random (multi)graph with vertex set $\{1, 2, \ldots, n\}$ and n edges.

Now the random graph G_m^n can be constructed from G_1^{mn} by identifying the nodes $\{(\ell-1)m+1, (\ell-1)m+2, \ldots, \ell m\}$ $(1 \le \ell \le n)$ of G_1^{mn} to a new node ℓ (and all edges within the nodes $\{(\ell-1)m+$

[†] This research was supported by EU FP6-NEST-Adventure Programme, Contract n° 028875 (NEMO) and by the Austrian Science Foundation (FWF) NFN-Projects S9604 and S9608

Fig. 1: The multigraph model of Bollobás and Riordan

$1, (\ell-1)m+2, \dots, \ell m\}$ are now loops of the new node ℓ). Of course, this procedure results in a random (multi)graph with vertex set $\{1, 2, \dots, n\}$ and mn edges (see Figure 1).

It turns out that the degree distribution of G_m^n satisfies a *power law*. The probability that a randomly chosen node of G_m^n has degree d is asymptotically $2/d^3$ (see Bollobás et al. (2001)). Graphs with this property are called *scale-free*.

The PORT-model. We first introduce a slightly modified evolution process that always leads to a labelled recursive tree. The process starts with the root that is labeled with 1. Then inductively at step j a new node (with label j) is attached to any previous node of out-degree k with probability proportional to $k+1$. These kinds of trees are also called *plane oriented recursive trees* (PORTs).

This evolution process is quite similar to the process that produces (usual) recursive trees. A (usual) recursive tree is a rooted tree (with n nodes) where the nodes are labeled with $1, 2, \dots, n$ such that all successors of each node have a larger label. In particular, the root has label 1, and every path from the root to a leaf has strictly increasing labels. As above, we can consider a recursive tree as the result of an evolution process. The *process* starts (as above) with the root (that gets label 1). Next, another node is attached to the root (that gets label 2) and in every step a new node is attached to an already existing node (and gets the next label). The labels are the *history* of the tree evolution.

The PORT-model, where we attach a new node according to the degree-distribution of the already existing tree, can be also seen as a planar version of the recursive trees. Namely, if a node of a planar (rooted) tree has out-degree k, that is, it has degree $d = k+1$, then there are precisely d ways to attach there a new node in order to get different planar trees. This explains the name *plane oriented recursive trees*.

We will indicate in the next section that the degree distribution of these random trees is scale-free as in the case of G_m^n. In fact, we have

$$\lim_{n\to\infty} p_n(d) = \frac{4}{d(d+1)(d+2)},$$

where $p_n(d)$ denotes the probability that a random node in random PORT of size n has degree d.

The thickened PORT-model. Now, we introduce a substitution process that creates random graphs that have a global tree structure that is governed by plane oriented recursive trees.

For every $k \ge 0$ let $\mathcal{T}_k$ denote a non-empty set of labelled graphs with *half edges* are attached to their nodes in such a way that each graph receives in total $k+1$ half edges. In addition the half edges are ordered. Now consider the following random process. Take a tree T according to the PORT-model. Then we substitute every node v (of out-degree k) in the following way: Cut v and one half of each edge incident with v. Then take a randomly chosen graph G of $\mathcal{T}_k$ and *glue* the $k+1$ half edges of G to those left in T by the cutting of v respecting the given order, i.e., the half edge coming from the predecessor of v is glued to the 0th half edge of G and the 1st, 2nd,…, kth successor of v is attached to the 1st, 2nd, …, kth half edge of G, respectively. Further we relabel all nodes in the new graph $G = G(T)$ in a way that is consistent with the original labelling. We denote the graphs that are obtained by this process *thickened trees* or more precisely *thickened PORTs*.

The idea behind this model is to simulate real networks that are produced by an evolution process

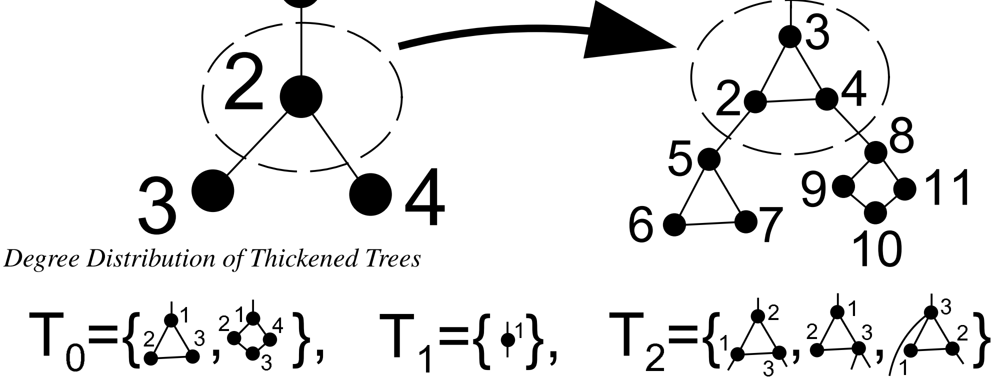

Fig. 2: A simple example of a thickened PORT. The original tree has only nodes of out-degree 0,1, or 2. So the choice of all sets $\mathcal{T}_k$ with $k > 2$ is not relevant for the thickening process. Node 2, e.g., is cut along the circular dashed line. Since it has out-degree 2, we choose one of the three graphs in $\mathcal{T}_2$ (here the first one was chosen) and put it into the place of node 2. Applying the same procedure to all nodes and relabelling afterwards yields the right graph, a thickened PORT.

(following, for example, the Albert-Barabási principle) and have a *global tree structure* with local clusters. Of course, we make a strong simplification. Our model relies on a two-step-procedure. We first let a tree network evolve (following the Albert-Barabási principle) and then replace the tree nodes by random clusters. Hence, the clusters are not produced by an evolution process. Nevertheless we think that our model has several advantages and can be used to explain several properties that are observed in practice:

- There is large flexibility in choosing the structure of local clusters and thus the choice of the sets $\mathcal{T}_k$ can be adapted to the situation.
- The model is feasible for an analytic treatment.
- It can be used to study (analytically) the influence of local changes of the network to the global behaviour.

The main focus of our paper is the degree distribution. We show that under natural conditions the resulting network is scale-free and that the number of nodes of given degree satisfy a central limit theorem.

2 The degree distribution of PORTs

In this section we shortly present a proof that PORTs are scale-free. In particular we show the following property that is due to Mahmoud et al. (1993), see also Bergeron et al. (1992) for similar results and Kuba and Panholzer (2007) for generalizations. The reason for presenting a proof is that the proof of a corresponding property for thickened PORTs will work along similar lines.

Theorem 1 *Let $p_n(d)$ denote the probability that a random node in a random PORT of size n has degree d. Then*

$$\lim_{n\to\infty} p_n(d) = \frac{4}{d(d+1)(d+2)}.$$

Furthermore, for every $d \geq 1$ let $X_n^{(d)}$ denote the number of nodes of degree d in random PORTs of size n. Then $X_n^{(d)}$ satisfies a central limit theorem

$$\frac{X_n^{(d)} - \mathbb{E}\, X_n^{(d)}}{\mathbb{V}\, X_n^{(d)}} \xrightarrow{\mathrm{d}} N(0,1),$$

where $\mathbb{E}\, X_n^{(d)} \sim 2B(3,d)n = 4n/(d(d+1)(d+2))$ and $\mathbb{V}\, X_n^{(d)} \sim (2B(3,d) + 4B(3,d)^2 - 4B(4,2d-1))n$ and $B(a,b)$ denotes the Beta-function.

Proof: The proof of Theorem 1 is based on a generating function approach. It is well known that the generating function

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

of PORTs satisfies the differential equation

$$y'(z) = 1 + y(z) + y(z)^2 + \cdots = \frac{1}{1-y(z)},$$

that reflects the recursive structure of PORTs. The solution is

$$y(z) = 1 - \sqrt{1-2z} = \sum_{n\ge 1} \frac{1}{2^{n-1}} \frac{(2(n-1))!}{(n-1)!} \frac{z^n}{n!}$$

and, thus, one has

$$y_n = \frac{1}{2^{n-1}} \frac{(2(n-1))!}{(n-1)!}.$$

We now turn to the degree distribution. By obvious reasoning we have

$$p_n(d) = \frac{\mathbb{E}\, X_n^{(d)}}{n}.$$

Thus, we can get the degree distribution with help of the distribution of $X_n^{(d)}$.

Let $y_{n,k}^{(d)}$ denote the number of PORTs of size n with exactly k nodes with degree d. Then the probability generating function $\mathbb{E}\, u^{X_n^{(d)}}$ is given by

$$\mathbb{E}\, u^{X_n^{(d)}} = \frac{1}{y_n} \sum_{k\ge 0} y_{n,k}^{(d)} u^k$$

and the double generating function

$$y(z,u) = \sum_{n,k} y_{n,k}^{(d)} u^k \frac{z^n}{n!} = \sum_{n\ge 1} y_n \left(\mathbb{E}\, u^{X_n^{(d)}} \right) \frac{z^n}{n!}$$

satisfies the differential equation

$$\frac{\partial y}{\partial z} = \frac{1}{1-y} + (u-1) y^{d-1} = \frac{1 + y^{d-1}(1-y)(u-1)}{1-y} \tag{1}$$

Equivalently we have

$$\int_0^y \frac{1-t}{1+t^{d-1}(1-t)(u-1)}\, dt = z \tag{2}$$

We first determine the expected value $\mathbb{E}\, X_n^{(d)}$. For this purpose we set

$$S(z) = \left. \frac{\partial y(z,u)}{\partial u} \right|_{u=1} \sum_{n\ge 1} y_n \cdot \mathbb{E}\, N_{d,n} \cdot \frac{z^n}{n!}.$$

From (2) we get that this function satisfies the differential equation

$$S'(z) = \frac{S(z)}{(1-y(z))^2} + y(z)^{d-1}$$

and has solution

$$S(z) = \frac{1}{1-y(z)} \int_0^{y(z)} (1-t)^2 t^{d-1}\, dt.$$

Recall that $y(z) = 1 - \sqrt{1-2z}$. Thus, it follows that

$$\begin{aligned} \int_0^{y(z)} (1-t)^2 t^{d-1}\, dt &= \int_0^1 (1-t)^2 t^{d-1}\, dt - \int_{y(z)}^1 (1-t)^2 t^{d-1}\, dt \\ &= B(3, d+1) + O\left((1-2z)^{3/2}\right). \end{aligned}$$

Consequently, we have

$$S(z) = \frac{B(3,d+1)}{\sqrt{1-2z}} + O(1) = \frac{2}{d(d+1)(d+2)} \frac{1}{\sqrt{1-2z}} + O(1)$$

which implies (with help of the transfer-lemma of Flajolet and Odlyzko Flajolet and Odlyzko (1990)) that

$$\mathbb{E}\, X_n^{(d)} = n \frac{4}{d(d+1)(d+2)} + O(\sqrt{n}).$$

Consequently, we get the degree distribution

$$\lim_{n\to\infty} p_n(d) = \lim_{n\to\infty} \frac{\mathbb{E}\, X_n^{(d)}}{n} = \frac{4}{d(d+1)(d+2)}.$$

In order to prove the central limit theorem we go back to (2) and similarly expand the integral

$$\begin{aligned} \int_0^y \frac{1-t}{1+t^{d-1}(1-t)(u-1)}\, dt &= \int_0^1 \frac{1-t}{1+t^{d-1}(1-t)(u-1)}\, dt - \int_y^1 \frac{1-t}{1+t^{d-1}(1-t)(u-1)}\, dt \\ &=: C(u) - D(y,u) \end{aligned}$$

Observe that $D(y,u)$ is asymptotically given by

$$D(y,u) = \frac{(1-y)^2}{2} \left(1 + O(|y-1| \cdot |u-1|)\right).$$

Hence (2) translates to

$$\sqrt{2C(u)} \sqrt{1 - \frac{z}{C(u)}} = (1-y)\left(1 + \mathcal{O}\left(|y-1| \cdot |u-1|\right)\right)$$

which can be inverted to

$$y(z,u) = 1 - \sqrt{2C(u)}\sqrt{1-\frac{z}{C(u)}} + \mathcal{O}\left(\left|1-\frac{z}{C(u)}\right|\right).$$

Consequently it follows (compare with Flajolet and Soria (1993)) that $X_n^{(d)}$ satisfies a central limit theorem with mean and variance

$$\mathbb{E}\,X_n^{(d)} = -\frac{C'(1)}{C(1)}n + O(\sqrt{n}) \quad \text{and} \quad \mathbb{V}\,X_n^{(d)} = \left(-\frac{C''(1)}{C(1)} - \frac{C'(1)}{C(1)} + \frac{C'(1)^2}{C(1)^2}\right)n + O(\sqrt{n}).$$

It is an easy exercise to compute $C(1) = 1/2$, $C'(1) = B(3,d)$, and $C''(1) = 2B(4,2d-1)$. □

3 The degree distribution of thickened PORTs

We now deal with a substitution process that creates random graphs that have a global tree structure that is governed by plane oriented recursive trees.

Let us consider the formal solution $y = y(z, x_0, x_1, x_2, \ldots)$ of the differential equation

$$y' = \sum_{k\ge 0} x_k y^k,$$

where $'$ denotes differentiation with respect to z. It is clear that $y = y(z, x_0, x_1, x_2, \ldots)$ can be considered as a power series in $z, x_0, x_1, \ldots$ By construction the coefficient

$$[z^n x_0^{k_0} x_1^{k_1} \ldots]\, y(z, x_0, x_1, x_2, \ldots)$$

is exactly the number of PORTs T of size n and k_j nodes of out-degree j $(j \ge 0)$.

For every $k \ge 0$ let $\mathcal{T}_k$ denote a non-empty set of labelled graphs with $k+1$ additional *half edges* $\tilde{e}_0, \tilde{e}_1, \ldots, \tilde{e}_k$. Further, let

$$t_k(z) = \sum_{G\in\mathcal{T}_k} \frac{z^{|G|}}{|G|!}$$

denote the exponential generating function of these graphs.

We recall that we consider the following random process. For every PORT T we substitute every node v (of out-degree k) by a randomly chosen graph of $\mathcal{T}_k$ where the half edges $\tilde{e}_0, \tilde{e}_1, \ldots, \tilde{e}_k$ are *glued* to the edge coming from the predecessor of v resp. the k successors of v corresponding to the left-to-right order. Further we relabel all nodes in the new graph $G = G(T)$ in a way that is consistent with the original labelling. Thus, the generating function of the number $g(z) = \sum_{n\ge 1} g_n \frac{z^n}{n!}$ of the numbers g_n of graphs that are produced in this way is given by

$$g(z) = y(z, t_0(z)/z, t_1(z)/z, \ldots).$$

The graphs that are obtained by this process are denoted here by *thickened trees* or *thickened PORTs*.

Lemma 1 *Set*

$$F(z,y)=\int_0^y \frac{dt}{\sum_{k\ge 0}\frac{t_k(z)}{z}\,t^k}.$$

Then $g(z)$ *satisfies the functional equation*

$$F(z,g)=z.$$

Proof: We recall that $y=y(z,x_0,x_1,\ldots)$ satisfies the differential equation $y'=\sum_{k\ge 0}x_k y^k$. Hence,

$$\int_0^y \frac{dt}{\sum_{k\ge 0}x_k y^k}=z+c$$

for some constant c. Of course, if we substitute x_k by $t_k(z)/z$ then we immediately get the result. Note further that $g(0)=0$ and $F(0,0)=0$. Thus, we can fix $c=0$. ☐

In a similar way we can deal with parameters. For example, fix a degree d and let

$$t_k^{(d)}(z,u)=\sum_{G\in\mathcal{T}_k}\frac{z^{|G|}}{|G|!}u^{N_d(G)},$$

where $N_d(G)$ denotes the number of nodes in G of degree d including the half-edges $\tilde{e}_0,\ldots,\tilde{e}_k$. Then the generating function

$$g(z,u)=y(z,t_0^{(d)}(z,u)/z,t_1^{(d)}(z,u)/z,\ldots)$$

encodes the distribution of nodes of degree d of thickened trees. Of course, the above lemma extends to this case. It is now convenient to introduce the notation

$$T_d(z,y,u)=\frac{1}{z}\sum_{k\ge 0}t_k^{(d)}(z,u)y^k.$$

Of course, for all d we have $T_d(z,y,1)=\sum_{k\ge 0}t_k(z)/z\,y^k$.

Lemma 2 *Set*

$$G_d(z,y,u)=\int_0^y \frac{dt}{T_d(z,t,u)}.$$

Then $g(z,u)$ *satisfies the functional equation*

$$G_d(z,g,u)=z.$$

Example We just give a simple example. Suppose that $\mathcal{T}_k$ consists for every $k\ge 0$ of exactly two graphs, the first one with one node and the second with two nodes, where all half-edges $\tilde{e}_1,\ldots,\tilde{e}_k$ that will be linked to the k subgraphs are on the second node. Figure 3 depicts the set $\mathcal{T}_3$.

In this example we have for $d\ge 3$

$$t_k^{(d)}(z,u)=\begin{cases} z+\frac{z^2}{2} & \text{if } k\ne d-1\\ uz+u\frac{z^2}{2} & \text{else,}\end{cases}$$

Fig. 3: The substitution set $\mathcal{T}_3$

and consequently

$$T_d(z,y,u) = \left(1+\frac{z}{2}\right)\left(\frac{1}{1-y} + (u-1)y^{d-1}\right).$$

The cases $d=1$ and $d=2$ are similar.

Our main result is the following theorem:

Theorem 2 *Let $\mathcal{T}_k$ be substitution sets (as described above) so that the equation*

$$\rho = \int_0^1 \frac{dt}{T_d(\rho,t,1)}$$

has a unique positive solution in the region of convergence of $T_d(z,y,u)$ and that the $T_d(z,y,u)$ can be represented as

$$T_d(z,y,u) = \frac{C_0(z,y) + C_1(z,y)(1-y)^{r'}y^{d+\alpha}(u-1) + \mathcal{O}\left((1-y)^{r'}(u-1)^2\right)}{(1-y)^r},$$

where r' and r are real numbers with $0 < r' \le r$, α in an integer, $C_0(z,y)$ and $C_1(z,y)$ are power series that contain $z=\rho$ and $y=1$ in their regions of convergence and that satisfy $C_i(\rho,1) \ne 0$ for $i=0,1$, and the $\mathcal{O}(\cdot)$-term is uniform in a neighbourhood of $z=\rho$ and $y=1$.

Let $p_n(d)$ denote the probability that a random node in a thickened PORT of size n has degree d. Then the limits

$$\lim_{n\to\infty} p_n(d) =: p(d)$$

exist and we have, as $d \to \infty$,

$$p(d) \sim \frac{C}{d^{r+r'+1}}.$$

Furthermore, for every $d \ge 0$ let $X_n^{(d)}$ denote the number of nodes of degree d in random thickened PORTs of size n. Then $X_n^{(d)}$ satisfies a central limit theorem

$$\frac{X_n^{(d)} - \mathbb{E}\,X_n^{(d)}}{\mathbb{V}\,X_n^{(d)}} \xrightarrow{\mathrm{d}} N(0,1),$$

where $\mathbb{E}\,X_n^{(d)}$ and $\mathbb{V}\,X_n^{(d)}$ are both asymptotically proportional to n.

Remark. The conditions on the generating function $T(z,y,u)$ can be (more of less) interpreted in the following way. If we set $u=1$ then the singularity $1/(1-y)^r$ in $T(z,y,1)$ essentially says that the set $\mathcal{T}_k$ consists $\approx k^{r-1}$ graphs. Furthermore for $k \ge d+\alpha$ there are $\approx k^{r-r'-1}$ graphs with a vertex of degree d, compare also with the examples given in Section 4.

Proof: The proof runs along similar lines as that of Theorem 1. We start by inspecting the generating function $g(z)$ of all thickened PORTs. For simplicity we assume that the substitution sets $\mathcal{T}_k$ are of a form that $g_n > 0$ for sufficiently large $n \ge n_0$, that is, we exclude, for example, the case that the number of nodes of graphs in $\mathcal{T}_k$ are all congruent to 1 modulo some integer $m > 1$.[(i)] Then it follows that $|g(z)| < g(|z|)$ if z is not contained in the positive real line.

We first observe that $\rho > 0$ is the only singularity on the circle of convergence $|z| \le \rho$ and that $g(\rho) = 1$, that is, $g(z)$ is convergent at $z = \rho$. First it is clear that $g(z)$ can be analytically continued starting with $g(0) = 0$ and the functional equation $F(z, g) = z$. However, if $g(z_0) \ne 1$ for some z_0 contained in the region of convergence of $g(z)$ then we have

$$F_g(z_0, g(z_0)) = \frac{1}{T_d(z_0, g(z_0), 1)} = \frac{(1 - g(z_0))^r}{C_0(z_0, g(z_0))} \ne 0.$$

Thus, we can continue analytically with help of the implicit function theorem. Thus, if $g(z)$ has a singularity ρ and if $g(\rho)$ is convergent then $g(\rho) = 1$. Since $g(z)$ is monotone and analytic it certainly reaches a value with $g(\rho) = 1$ where it has to be singular. Further, ρ is characterized by the equation $F(\rho, 1) = \rho$.

Next we characterize the kind of singularity of $g(z)$ at $z = \rho$. By Lemma 1 we have

$$z = \int_0^g \frac{(1-t)^r}{C_0(z,t)}\,dt = \int_0^1 \frac{(1-t)^r}{C_0(z,t)}\,dt - \int_g^1 \frac{(1-t)^r}{C_0(z,t)}\,dt =: G(z) - H(z, g)$$

Hence, by expanding $1/C_0(z, t)$ locally around $t = 1$ we thus get

$$G(z) - z = c_0(z)(1-g)^{r+1}\left(1 + \mathcal{O}\left(|1-g|\right)\right)$$

or

$$\left(\frac{G(z) - z}{c_0(z)}\right)^{1/(r+1)} = (1-g)\left(1 + \mathcal{O}\left(|1-g|\right)\right). \tag{3}$$

Since $G(\rho) = \rho$ and $C_0(z, y)$ is increasing in z we can represent $(G(z) - z)/c_0(z) = K(z)(1 - z/\rho)$. Furthermore, we can invert the relation (3) and obtain

$$g(z) = 1 - K(z)^{1/(r+1)}\left(1 - \frac{z}{\rho}\right)^{1/(r+1)} + \mathcal{O}\left(\left|1 - \frac{z}{\rho}\right|^{2/(r+1)}\right).$$

Since there are no other singularities on the circle $|z| \le \rho$ and $g(z)$ can be analytically continued to a larger range (despite at the point $z = \rho$) it follows from Flajolet and Odlyzko (1990) that

$$g_n \sim K(\rho)^{1/(r+1)} \frac{\rho^{-n} n^{-\frac{r+2}{r+1}}}{-\Gamma\left(-\frac{1}{r+1}\right)}.$$

Next we determine asymptotics on the average value $\mathbb{E}\, X_n^{(d)}$. Set $S(z) = \frac{\partial}{\partial u} g(z, 1)$. Then it follows from Lemma 2 that

$$S(z) = \frac{1}{(1 - g(z))^r} \int_0^{g(z)} C_1(z, t)(1-t)^{r+r'} t^{d+\alpha}\,dt.$$

[(i)] We call this the aperiodic case. In the periodic case we have to deal with m singularities on the boundary of the circle of convergence of $g(z)$ which are all of the same kind.

As in the proof of Theorem 1 it follows that

$$S(z) = \frac{1}{K(z)^{\frac{r}{r+1}} (1 - z/\rho)^{\frac{r}{r+1}}} \int_0^1 C_1(z,t)(1-t)^{r+r'} t^{d+\alpha}\, dt + \mathcal{O}(1)$$

which proves that

$$\mathbb{E}\, X_n^{(d)} \sim n \cdot \frac{r+1}{K(\rho)} \int_0^1 C_1(\rho,t)(1-t)^{r+r'} t^{d+\alpha}\, dt.$$

Thus, the limit $p(d) = \lim_{n\to\infty} \mathbb{E}\, X_n^{(d)}/n$ exists and is asymptotically given by

$$p(d) = \frac{r+1}{K(\rho)} \int_0^1 C_1(\rho,t)(1-t)^{r+r'} t^{d+\alpha}\, dt \sim \frac{C}{d^{r+r'+1}}$$

for some constant $C > 0$.

Finally the proof that the limiting distribution is normal is very similar to the corresponding proof of Theorem 1. We skip the details. □

4 Examples

4.1 Substituting by one or two nodes

Let us continue the example preceding Theorem 2. Here we have (for $d \geq 3$):

$$T_d(z,y,u) = \left(1 + \frac{z}{2}\right) \frac{1 + (1-y)(u-1)y^{d-1}}{1-y}.$$

Thus, Theorem 2 applies with $r = r' = 1$ and $\alpha = -1$. The degree distribution $p(d)$ is scale-free with tail $p(d) \sim C/d^3$. A detailed computation (including the instances $d < 3$ also) shows that the degree distribution is given as follows:

$$p(d) = \frac{2+\sqrt{2}}{d(d+1)(d+2)}, \quad \text{for } d \geq 3, \qquad p(1) = \frac{2+\sqrt{2}}{6}, \quad p(2) = \frac{14 - 5\sqrt{2}}{24}.$$

4.2 Thickening with triangles 1

We consider two examples where we substitute each node of the original tree by a triangle. More precisely, each node of out-degree k is then substituted by a triangle with $k+1$ half-edges $\tilde{e}_0, \ldots, \tilde{e}_k$ attached to it. The $\tilde{e}_0$ is glued to the predecessor of the original nodes and the other half-edges to the successors of the original nodes according to their naturally given order (e.g. left to right in the case of plane trees). Let the parameter of interest be the number of nodes of degree d. Then we have

$$t_k(z,u) = (a_{0,k} + a_{1,k}u + a_{2,k}u^2 + a_{3,k}u^3)\frac{z^3}{3!}$$

where $a_{i,k}$ is the number of configurations (triangles with k half-edges) containing i nodes of degree d

We first consider the special case where the ingoing edge of each triangle is separated from the outgoing edges. That means that $\tilde{e}_0$ connects the predecessor of the triangle to a vertex of degree 3 (the edges are then $\tilde{e}_0$ and two edges of the triangle) while $\tilde{e}_1, \ldots, \tilde{e}_k$ are connected to the other two vertices of the triangle. Clearly we have then

$$t_k(z) = (k+1)\frac{z^3}{3!}.$$

When focusing on the number of nodes of degree d, then it is easy to see that (for $d \geq 4$) we have

$$a_{0,k} = \begin{cases} k+1 & \text{if } k < d-2 \\ k & \text{if } k = 2d-4 \\ k-1 & \text{else.} \end{cases}$$

This holds because of the following argument. Let us label the nodes of the triangle by $0, 1, 2$ where $\tilde{e}_0$ is attached to 0. Then ℓ_1 of the edges $\tilde{e}_1, \ldots, \tilde{e}_k$ are attached to 1 and $\ell_2 = k - \ell_1$ to 2. The configuration contains at least one node of degree d if and only if $\ell_1 = d-2$ or $\ell_2 = d-2$. Exactly these cases do not contribute to $a_{0,k}$. Moreover we get $a_{3,k} = 0$,

$$a_{1,k} = \begin{cases} 0 & \text{if } k = 2d-4 \text{ or } k < d-2 \\ 2 & \text{else,} \end{cases} \qquad \text{and} \qquad a_{2,k} = \begin{cases} 0 & \text{if } k \neq 2d-4 \\ 1 & \text{if } k = 2d-4. \end{cases}$$

This implies (for $d \geq 4$):

$$\begin{aligned} T(z,y,u) &= \frac{z^2}{6}\left(\sum_{k\geq 0}(k-1+2u)y^k + \sum_{k=0}^{d-3} 2(1-u)y^k + (1-u)^2 y^{2d-4}\right) \\ &= \frac{z^2}{6}\,\frac{1 + 2(1-y)y^{d-2}(u-1) + (1-y)^2 y^{2d-4}(u-1)^2}{(1-y)^2}. \end{aligned}$$

Here Theorem 2 applies with $r' = 1$, $r = 2$ and $\alpha = -2$. Hence the degree distribution $p(d)$ is scale-free with tail $p(d) \sim C/d^4$. A detailed computation (including the instances $d < 4$ also) shows that the degree distribution is given as follows:

$$p(d) = \frac{12}{(d-1)d(d+1)(d+2)}, \quad \text{for } d \geq 4, \qquad p(1) = 0, \quad p(2) = \frac{1}{2}, \quad p(3) = \frac{13}{30}.$$

4.3 Thickening with triangles 2

In the case of general triangles the edges $\tilde{e}_0, \ldots, \tilde{e}_k$ can be attached to all the vertices of the triangles. This gives, after somewhat tedious considerations, the following result for $t_k^{(d)}(z,u)$.

Lemma 3 *Let $d > 3$. Then we have*

$$t_k^{(d)}(z,u) = \frac{z^3}{6} a_k^{(d)}(u),$$

with

$$a_k^{(d)}(u) = \begin{cases} \binom{k+3}{3} & \textit{if } k \le d-4 \\ \binom{k+3}{3} - (k+1)(k-d+4) + (k+1)(k-d+4)u & \textit{if } d-3 \le k \le 2d-6 \\ \binom{k+3}{3} - \frac{(k+1)^2}{2} + \frac{(k-1)(k+1)}{2}u + (k+1)u^2 & \textit{if } k = 2d-5 \\ \binom{k+3}{3} - (k+1)(k-d+3) + (k+1)(k-d+2)u + (k+1)u^2 & \textit{if } k \ge 2d-4 \\ & \textit{and } k \ne 3d-7 \\ \binom{k+3}{3} - \frac{(k+1)(2k+3)}{3} + \frac{2(k+1)^2}{3}u + \frac{k+1}{3}u^3 & \textit{if } k = 3d-7. \end{cases}$$

This implies (for $d \ge 4$):

$$T_d(z,u,v) = \frac{z^2}{6}\frac{1}{(1-y)^4}\Big[1 + (1-y)y^{d-3}((4-d)y + d - 2)(u-1) \\ - (1-y)^2 y^{2d-5}((2d-5)y - 2d + 4)(u-1)^2 + (d-2)(1-y)^4 y^{3d-7}(u-1)^3\Big].$$

Now Theorem 2 applies with $r' = 1$, $r = 4$ and $\alpha = -3$. Hence the degree distribution $p(d)$ is scale-free with tail $p(d) \sim C/d^6$. A detailed computation (including the instances $d < 4$ also) shows that the degree distribution is given as follows:

$$p(d) = \frac{1600}{(d-1)d(d+1)(d+2)(d+3)(d+4)}, \quad \text{for } d \ge 4, \qquad p(1) = 0, \quad p(2) = \frac{5}{9}, \quad p(3) = \frac{20}{63}.$$

References

R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Modern Phys.*, 74(1): 47–97, 2002. ISSN 0034-6861.

F. Bergeron, P. Flajolet, and B. Salvy. Varieties of increasing trees. In *CAAP '92 (Rennes, 1992)*, volume 581 of *Lecture Notes in Comput. Sci.*, pages 24–48. Springer, Berlin, 1992.

B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004. ISSN 0209-9683.

B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, 18(3):279–290, 2001. ISSN 1042-9832.

P. Flajolet and A. M. Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3:216–240, 1990.

P. Flajolet and M. Soria. General combinatorial schemas: Gaussian limit distributions and exponential tails. *Discrete Math.*, 114(1-3):159–180, 1993. ISSN 0012-365X.

M. Kuba and A. Panholzer. On the degree distribution of the nodes in increasing trees. *J. Combin. Theory Ser. A*, 114(4):597–618, 2007. ISSN 0097-3165.

H. M. Mahmoud, R. T. Smythe, and J. Szymański. On the structure of random plane-oriented recursive trees and their branches. *Random Structures Algorithms*, 4(2):151–176, 1993. ISSN 1042-9832.

B. Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures Algorithms*, 5(2):337–347, 1994. ISSN 1042-9832.