*Elect. Comm. in Probab.* **12** (2007), 276–282

**ELECTRONIC COMMUNICATIONS in PROBABILITY**

# DEGREE DISTRIBUTION NEARBY THE ORIGIN OF A PREFERENTIAL ATTACHMENT GRAPH

TAMÁS F. MÓRI[1]
*Department of Probability Theory and Statistics, Eötvös Loránd University, Pázmány Péter s. 1/C, Budapest, Hungary H-1117*
email: moritamas@ludens.elte.hu

*Submitted* June 11, 2007, *accepted in final form* September 6, 2007

AMS 2000 Subject classification: 60G42; 05C80
Keywords: Scale free graphs; degree distribution; martingale

*Abstract*

In a 2-parameter scale free model of random graphs it is shown that the asymptotic degree distribution is the same in the neighbourhood of every vertex. This degree distribution is still a power law with characteristic exponent 2, but this exponent is different from the one observed in the whole graph.

## 1 The model

In their paper [1] Barabási and Albert proposed a certain random process of evolving graphs as a model of real-world networks, like the Internet. In their model vertices are added to the graph one by one, and edges connecting the new vertex to the old ones are drawn randomly, with probabilities proportional to the degree of the endpoint. In the particular case where only a single edge is allowed at every step, a recursive tree process, also known as plane oriented recursive trees, is obtained. In fact, that model was introduced more than a decade earlier by Szymański [9], and then a couple of papers have been devoted to it. The interested reader is referred to [2] for a very general model of web graphs.

In [5] the asymptotic degree distribution was obtained for a one-parameter generalization of the Barabási–Albert random tree. In [3] the same degree distribution was proved to exist on each of the largest levels of the tree. Surprisingly, in the neighbourhood of the root, on the lower levels a completely different degree distribution was found to emerge [6].

Consider the following modification of the Barabási–Albert random graph. Starting from the very simple graph consisting of two points and the edge between them, at every step we add a new vertex and some (possibly 0) new edges to the graph. For the new edges each old vertex is selected at random, with probability depending linearly on its degree, and independently of the others; then the selected vertices are connected to the new one.

[1]RESEARCH SUPPORTED BY THE HUNGARIAN SCIENTIFIC RESEARCH FUND (OTKA) GRANT K67961

Let us number the vertices in the order they are added to the graph; thus the vertex set of the graph after $n$ steps is $\{0, 1, \ldots, n\}$. Let $X[n, k]$ denote the number of vertices of degree $k$ after $n$ steps. Then $X[n, 0] + X[n, 1] + \cdots = n + 1$. Let $S_n = \sum_{k \geq 1} kX[n, k]$, the sum of degrees, or equivalently, twice the number of edges. At the $n$th step an old vertex of degree $k$ is connected to the new one with probability $(\lambda_1 k + \lambda_0)/S_{n-1}$, where $\lambda_0$, $\lambda_1$ are nonnegative parameters. This quantity remains below 1, provided $\lambda_0 + \lambda_1 < 2$, which will therefore be assumed in the sequel. In order to preserve the scale free property we also assume that $\lambda_1 > 0$.

This model was investigated in [7] (and a particular case in [4]). It was proved ([7], Theorem 2.1) that

$$S_n = 2sn + o\big(n^{1-\varepsilon}\big) \text{ a.s.,} \tag{1}$$

if $\varepsilon > 0$ is sufficiently small, where

$$s = \frac{1}{2}\Big(\lambda_1 + \sqrt{\lambda_1^2 + 2\lambda_0}\,\Big). \tag{2}$$

Moreover, the following asymptotic degree distribution was found ([7], Theorem 3.1). For every $k = 0, 1, \ldots$, the proportion of vertices of degree $k$ converges a.s. as $n \to \infty$:

$$P\Big(\lim_{n\to\infty} \frac{X[n, k]}{n + 1} = x_k\Big) = 1, \tag{3}$$

where

$$x_k = \frac{1}{t_k + 1} \sum_{i=0}^{k} p_i \prod_{j=i}^{k-1} \frac{t_j}{t_j + 1}, \tag{4}$$

with

$$p_k = \frac{s^k}{k!}\, e^{-s},\ k \geq 0, \quad t_k = \frac{\lambda_1 k + \lambda_0}{2s},\ k \geq 1. \tag{5}$$

This is a power law, that is, $x_k \sim const \cdot k^{-\beta}$ as $k \to \infty$, and the exponent is $\beta = 2 + \sqrt{1 + \lambda_0/\lambda_1^2}$.

The aim of the present note is to investigate whether this degree distribution is preserved when a certain part of the graph is considered only, namely, the neighbours of a fixed vertex. The answer is negative: in the neighbourhood of each vertex the same asymptotic degree distribution is found, but it differs significantly from (3), having exponent 2.

# 2 Neighbourhood sizes

In this section we approximate the number of neighbours, that is, the degree of a fixed vertex as the size of the graph tends to infinity.

Let $\mathcal{F}_n$ denote the $\sigma$-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 $n$th step, and let $\Delta_n = \sum_{k \geq 0} \Delta[n, k]$ be the total number of new edges. Obviously, the conditional distribution of $\Delta[n, k]$ with respect to $\mathcal{F}_{n-1}$ is binomial with parameters $X[n-1, k]$ and $\dfrac{\lambda_1 k + \lambda_0}{S_{n-1}}$, hence $E\big(\Delta_n \mid \mathcal{F}_{n-1}\big) = \lambda_1 + \dfrac{n}{S_{n-1}}\, \lambda_0$.

In (3.1) of [7] it is proved that the increments $\Delta_n$ are asymptotically independent and asymptotically Poisson distributed. More precisely,

$$\lim_{n\to\infty} \sum_{k=0}^{\infty} \left| P\big(\Delta_{n+1} = k \mid \mathcal{F}_n\big) - \frac{s^k}{k!}\, e^{s} \right| = 0, \quad \text{a.s.} \tag{6}$$

Introduce

$$w = \frac{\lambda_0}{\lambda_1}, \quad \phi = \frac{\lambda_1}{2s}. \tag{7}$$

For $j = 0, 1, \dots$ let $W[n, j]$ denote the weight of vertex $j$ after $n$ steps, defined as $W[n, j] = \text{degree} + w$ with the initial values $W[n, j] = 0$ for $n < j$, $W[1, 0] = W[1, 1] = 1 + w$, $W[j, j] = \Delta_j + w$.

It can happen that $\Delta_j = 0$, i.e., when vertex $j$ is added to the graph, it does not get any edges. If $\lambda_0 = 0$, $j$ will always remain isolated: $W[n, j] = 0$ for all $n$. By (6) the probability of $\Delta_j = 0$ tends to $e^{-\lambda_1}$ as $j \to \infty$.

If $\lambda_0 > 0$, then not even $\Delta_j = 0$ can prevent vertex $j$ from getting edges at later steps.

**Theorem 2.1.** *With probability* 1,

$$W[n, j] \sim \zeta_j \, n^{\phi}, \tag{8}$$

*as $n \to \infty$, where $\zeta_j$ is a positive random variable if $\lambda_0 > 0$, and $\{\zeta_j = 0\} = \{\Delta_j = 0\}$ if $\lambda_0 = 0$.*

*Proof.* (8) is proved in [7], p.41, with a nonnegative $\zeta_j$. Now we are going to show the positivity of $\zeta_j$.

Let $j, k$ be fixed, and $Z[n, j] = \dfrac{I(W[k, j] > 1)}{W[n, j] - 1}$, $n \ge k \ge j$. Define

$$c_n = \prod_{i=1}^{n-1} \Bigl(1 - \frac{\lambda_1}{S_i}\Bigr)^{-1}.$$

Since clearly

$$\begin{aligned} E\bigl(Z[n+1, j] \bigm| \mathcal{F}_n\bigr) &= \Bigl(1 - \frac{\lambda_1 W[n, j]}{S_n}\Bigr) Z[n, j] + \frac{\lambda_1 W[n, j]}{S_n} \cdot \frac{I(W[k, j] > 1)}{W[n, j]} \\ &= \Bigl(1 - \frac{\lambda_1}{S_n}\Bigr) Z[n, j], \end{aligned}$$

we obtain that $\bigl(c_n Z[n, j],\ \mathcal{F}_n\bigr)$, $n \ge k$, is a nonnegative, thus convergent, martingale.

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

$$c_n = \exp\Biggl(\lambda_1 \sum_{i=1}^{n-1} \frac{1}{S_i} + \frac{\lambda_1^2}{2} \sum_{i=1}^{n-1} \frac{1 + o(1)}{S_i^2}\Biggr). \tag{9}$$

Since $\dfrac{1}{S_i} = \dfrac{1}{2si}\bigl(1 + o\bigl(i^{-\varepsilon}\bigr)\bigr)$, by (1), we obtain that the exponent in (9) differs from $\phi \log n$ only by a term converging with probability 1. Thus $c_n \sim \gamma\, n^{\phi}$, where $\gamma > 0$; hence $Z[n, j] = O\bigl(n^{-\phi}\bigr)$. From this we get that $n^{\phi}/W[n, j]$ converges a.e. on the event $\{W[k, j] > 1\}$, hence $\zeta_j > 0$ there.

We can complete the proof by showing that $W[n, j] \to \infty$ a.s. if $\lambda_0 > 0$, and $W[n, j] \to \infty$ a.s. on the event $\{\Delta_j > 0\}$ if $\lambda_0 = 0$.

The conditional probability that the weight of vertex $j$ grows at the $n$th step is $\dfrac{\lambda_1 W[n-1, j]}{S_{n-1}}$.

This can be estimated from below by $\dfrac{\lambda_0}{S_{n-1}} \sim \dfrac{\lambda_0}{2sn}$, if $\lambda_0 > 0$, and by $\dfrac{1}{S_{n-1}} \sim \dfrac{1}{2sn}$, if $\lambda_0 = 0$ and $\Delta_j > 0$. Then the Lévy variant of the Borel–Cantelli lemma ([8], Corollary VII-**2**-6) implies just what we needed. $\square$

# 3 Degree distribution in the neighbourhood of a fixed vertex

In this section we prove that the degree distribution among the neighbours of vertex $j$ stabilizes almost surely, as $n \to \infty$, around a power law with exponent 2.
Let $Y[n, j, k]$ be the number of neighbours of vertex $j$ with degree $k$ after $n$ steps, $n > j$. Clearly, $Y[n, j, 1] + Y[n, j, 2] + \cdots = W[n, j] - w$.

**Theorem 3.1.** *Suppose $\lambda_0 > 0$. Then for every $j \geq 0$ and $k \geq 1$ the proportion of vertices of degree $k$ among the neighbours of vertex $j$ converges a.s., as $n \to \infty$:*

$$P\left(\lim_{n\to\infty} \frac{Y[n,j,k]}{W[n,j]} = \frac{1}{(k+w)(k+1+w)} \sum_{i=0}^{k-1} (i+1+w)\, p_i\right) = 1. \tag{10}$$

*If $\lambda_0 = 0$, (10) holds conditionally given that $j$ is not isolated:*

$$P\left(\lim_{n\to\infty} \frac{Y[n,j,k]}{W[n,j]} = \frac{1}{k(k+1)} \sum_{i=0}^{k-1} (i+1) p_i \;\middle|\; \Delta_j > 0\right) = 1. \tag{11}$$

Note that the limit in (10) is $(1 + w + s)k^{-2}$ asymptotically, as $k \to \infty$.
This phenomenon seems to be the same as in the case of scale free trees in [5] and [6]. In both cases we investigated the degree distribution constrained to vertices "close" to the initial configuration. However, level $j$ of a rooted tree can also be characterized as the set of vertices that are of distance $j$ from the root. It would be interesting to know whether Theorem 3.1 remains true in our scale free graph for the set of vertices that are of a certain distance from vertex 0. This looks a little harder, because those sets lack the convenient property that both the neighbourhoods here and the levels in random recursive trees have, namely, that they never decrease as the size of the graph grows.
What is behind this phenomenon? Decrease of the characteristic exponent may be caused by the overlapping of neighbourhoods. Vertices with higher degrees belong to more neighbourhoods at the same time, hence their importance increases, resulting a heavier tail of the asymptotic degree distribution. This sounds plausible, but does not apply to the levels of a tree, for they are disjoint. It is rather due to the observation that nearby the origin there is a relatively large number of old vertices, which are more likely to have higher degrees. But the exponent 2, which appears both here and in [6], independently of the parameters of the two models, is still looking somewhat mysterious. In fact, it seems to be connected with the neighbourhood sizes. The number of neighbours of any fixed node is of the same order of magnitude as the maximal degree of the graph. It has been observed in many scale free graph models that the exponent of the asymptotic degree distribution is in connection with the maxdegree: if the former is equal to $\beta$, then the maxdegree is of order $n^{\phi}$, where $\phi(\beta - 1) = 1$. Suppose we are interested in the asymptotic degree distribution restricted to a subset of nodes, the size of which is a regularly varying function of $n$ with exponent $\alpha$. Under a couple of additional conditions it is true that the restricted asymptotic degree distribution is still a power law, and its exponent is a function of both $\alpha$ and $\beta$. Particularly, when $\alpha(\beta - 1) = 1$, this exponent is equal to 2. The exact theory is still to be worked out (in progress).

*Proof.* The basic idea, that is, the way we apply martingale theory, is reminiscent of the proof of Theorem 3.1 in [7].

For $n > j$ let $A[n, j, k]$ denote the event that when vertex $n$ is added to the graph, it gets exactly $k$ edges, one of them connecting it to vertex $j$. Then clearly

$$P\big(A[n+1, j, k] \bigm| \mathcal{F}_n\big) \leq \frac{\lambda_1 W[n, j]}{S_n}.$$

In addition, let $\Delta[n, j, k]$ be the number of new edges from vertex $n$ into the set of neighbours of vertex $j$ with degree $k$. Then, conditionally on $\mathcal{F}_n$, the distribution of $\Delta[n+1, j, k]$ is binomial with parameters $Y[n, j, k]$ and $\dfrac{\lambda_1 k + \lambda_0}{S_n}$.
We clearly have

$$Y[n, j, k] = Y[n-1, j, k] - \Delta[n, j, k] + \Delta[n, j, k-1] + I(A[n, j, k]), \tag{12}$$

hence

$$E\big(Y[n+1, j, k] \bigm| \mathcal{F}_n\big) = Y[n, j, k] - \frac{\lambda_1 k + \lambda_0}{S_n}\, Y[n, j, k] \\ + \frac{\lambda_1 (k-1) + \lambda_0}{S_n}\, Y[n, j, k-1] + P\big(A[n+1, j, k] \bigm| \mathcal{F}_n\big).$$

We will use the random normalizing factors of [7] defined as

$$d[n, k] = \prod_{i=1}^{n-1} \left(1 - I\big(S_i \geq 2k\big) \frac{\lambda_1 k + \lambda_0}{S_i}\right)^{-1}. \tag{13}$$

If $S_i \geq 2k$, then $\lambda_1 k + \lambda_0 \leq (\lambda_1 + \lambda_0) k < 2k \leq S_i$, thus $d[n, k]$ is well defined and bounded:

$$d[n, k] < \left(1 - \frac{\lambda_0 + \lambda_1}{2}\right)^{-(n-1)}.$$

On the other hand, when $S_i < 2k$, the maxdegree is less than $k$, hence $Y[i, j, k] = 0$. Thus we have

$$E\big(d[n+1, k]\, Y[n+1, j, k] \bigm| \mathcal{F}_n\big) = d[n, k]\, Y[n, j, k] + b[n, j, k], \tag{14}$$

where

$$b[n, j, k] = d[n+1, k] \left( Y[n, j, k-1]\, \frac{\lambda_1 (k-1) + \lambda_0}{S_n} + P\big(A[n+1, j, k] \bigm| \mathcal{F}_n\big) \right).$$

Let us estimate the conditional variance. By using (12) and the trivial inequality $\operatorname{Var}(u_1 + \cdots + u_n) \leq n(\operatorname{Var} u_1 + \cdots + \operatorname{Var} u_n)$, we obtain that

$$\operatorname{Var}\big(d[n+1, k]\, Y[n+1, j, k] \bigm| \mathcal{F}_n\big) \leq 3\, d[n+1, k]^2 \Big( \operatorname{Var}\big(\Delta[n+1, j, k] \bigm| \mathcal{F}_n\big) + \\ + \operatorname{Var}\big(\Delta[n+1, j, k-1] \bigm| \mathcal{F}_n\big) + \operatorname{Var}\big(I\big(A[n+1, j, k]\big) | \mathcal{F}_n\big) \Big).$$

On the right-hand side each random variable has binomial (conditional) distribution, therefore its (conditional) variance is less than the corresponding expectation. Thus,

$$\operatorname{Var}\big(d[n+1, k]\, Y[n+1, j, k] \bigm| \mathcal{F}_n\big) \leq 3\, d[n+1, k]^2 \bigg( Y[n, j, k]\, \frac{\lambda_1 k + \lambda_0}{S_n} + \\ + Y[n, j, k-1]\, \frac{\lambda_1 (k-1) + \lambda_0}{S_n} + \frac{\lambda_1 W[n, j]}{S_n} \bigg).$$

Similarly to (9), in [7] it is proved that

$$d[n,k] \sim \delta_k\, n^{t_k}, \tag{15}$$

as $n \to \infty$, with some positive random variable $\delta_k$. Hence,

$$\begin{aligned}\operatorname{Var}\bigl(d[n+1,k]\,Y[n+1,j,k] \bigm| \mathcal{F}_n\bigr)&\\ \le 3\,d[n+1,k]^2\biggl(W[n,j]\,\frac{\lambda_1 k+\lambda_0}{S_n}+\frac{\lambda_1 W[n,j]}{S_n}\biggr) &= O\bigl(n^{2t_k+\phi-1}\bigr).\end{aligned} \tag{16}$$

Let us introduce a martingale $\bigl(M_n, \mathcal{F}_n\bigr)$, $n \ge j$, by its differences $\xi_n = M_n - M_{n-1}$ as follows.

$$\begin{aligned} M_j &= d[j,k]\,Y[j,j,k],\\ \xi_n &= d[n,k]\,Y[n,j,k] - E\bigl(d[n,k]\,Y[n,j,k] \bigm| \mathcal{F}_{n-1}\bigr)\\ &= d[n,k]\,Y[n,j,k] - d[n-1,k]\,Y[n-1,j,k] - b[n-1,j,k],\ \ n > j.\end{aligned}$$

Since both $d[n,k]$ and $Y[n,j,k]$ are bounded random variables, $M_n$ is square integrable. The increasing process associated with $M_n^2$ in its Doob decomposition is

$$\sum_{i=j}^{n-1} E\bigl(\xi_{i+1}^2 \bigm| \mathcal{F}_i\bigr) = \sum_{i=j}^{n-1} \operatorname{Var}\bigl(d[i+1,k]\,Y[i+1,k] \bigm| \mathcal{F}_i\bigr),$$

which is of order $O\bigl(n^{2t_k+\phi}\bigr)$ by (16). By Proposition VII-**2**-4 of [8] we have that

$$M_n = o\biggl(\sum_{i=j}^{n-1} E\bigl(\xi_{i+1}^2 \bigm| \mathcal{F}_i\bigr)\biggr)^{1/2+\varepsilon}$$

for all $\varepsilon > 0$, hence

$$M_n = d[n,k]\,Y[n,j,k] - \sum_{i=j}^{n-1} b[i,j,k] = o\bigl(n^{t_k+\phi}\bigr). \tag{17}$$

From (15) and (17) we obtain that

$$Y[n,j,k] = \frac{1}{d[n,k]}\sum_{i=1}^{n-1} b[i,j,k] + o\bigl(n^{\phi}\bigr). \tag{18}$$

We are going to prove by induction over $k$ that $\lim_{n\to\infty} Y[n,j,k]/W[n,j]$ exists (on the event $\{\Delta_j > 0\}$ when $\lambda_0 = 0$), it is a constant, and does not depend on $j$. We will denote it by $y_k$. Since $Y[n,j,0] = 0$, this holds for $k = 0$ with $y_0 = 0$. For the induction step we shall need the asymptotics of $P\bigl(A[n+1,j,k] \bigm| \mathcal{F}_n\bigr)$. The (conditional) probability that vertex $n+1$ gets connected to $j$ is $\dfrac{\lambda_1 W[n,j]}{S_n}$, and, independently of it, we require $k-1$ further edges toward the other vertices. The number of such edges, being a sum of (conditionally) independent indicators, is asymptotically Poisson with parameter $s$. This can be shown similarly to (6, by applying LeCam's theorem on Poisson approximation. Hence

$$P\bigl(A[n+1,j,k] \bigm| \mathcal{F}_n\bigr) \sim \phi\, p_{k-1} n^{-1} W[n,j].$$

Making use of the induction hypothesis we obtain that

$$\begin{aligned} b[n,j,k] &\sim d[n+1,k]\, n^{-1} W[n,j]\big(y_{k-1}t_{k-1}+\phi p_{k-1}\big) \\ &\sim \big(t_{k-1}y_{k-1}+\phi\, p_{k-1}\big)\,\delta_k\,\zeta_j\, n^{t_k+\phi-1}. \end{aligned}$$

Plug it back into (18) to get

$$\begin{aligned} Y[n,j,k] &\sim \frac{1}{\delta_k n^{t_k}} \cdot \frac{\big(t_{k-1}y_{k-1}+\phi\, p_{k-1}\big)\,\delta_k\zeta_j\, n^{t_k+\phi}}{t_k+\phi} \\ &\sim W[n,j]\cdot \frac{t_{k-1}y_{k-1}+\phi\, p_{k-1}}{t_{k+1}}. \end{aligned}$$

This is just what we wanted to prove. It also yields the following recursive formula for the constants $y_k$.

$$y_k = \frac{t_{k-1}y_{k-1}+\phi\, p_{k-1}}{t_{k+1}} = \frac{(k-1+w)\,y_{k-1}+p_{k-1}}{k+1+w}.$$

Finally, it is not hard to see that the solution to this recursion is given by (10). □

# References


[1] A.-L. Barabási, and R. Albert. Emergence of scaling in random networks. *Science* **286** (1999), 509–512. MR2091634 MR2091634

[2] C. Cooper, and A. Frieze. A general model of web graphs. *Random Structures Algorithms* **22** (2003), 311–335. MR1966545 MR1966545

[3] Z. Katona. Levels of a scale-free tree. *Random Structures Algorithms* **28** (2006), 194–207. MR2245500 MR2245500

[4] Z. Katona, and T. F. Móri. A new class of scale free random graphs. *Statist. Probab. Lett.* **76** (2006), 1587–1593. MR2248845 MR2248845

[5] T. F. Móri. On random trees. *Studia Sci. Math. Hungar.* **39** (2002), 143–155. MR1909153 MR1909153

[6] T. F. Móri. A surprising property of the Barabási–Albert random tree. *Studia Sci. Math. Hungar.* **43** (2006), 265–273. MR2229623 MR2229623

[7] T. F. Móri. On a 2-parameter class of scale free random graphs. *Acta Math. Hungar.* **114** (2007), 37–48. MR2294512 MR2294512

[8] J. Neveu. *Discrete-Parameter Martingales.* North-Holland, Amsterdam, 1975. MR0402915 MR0402915

[9] J. Szymański. On a nonuniform random recursive tree. *In: Random graphs '85 (Poznań, 1985)*, 297–306, North-Holland Math. Stud. **144** (1987), North-Holland, Amsterdam. MR0930497 MR0930497