Bulletin of Mathematical Biology (2006) 68: 1275–1291
DOI 10.1007/s11538-006-9085-9

ORIGINAL ARTICLE

Convergence Properties of the Degree Distribution of Some Growing Network Models

Oskar Hagberg[a], Carsten Wiuf[a,b,*]

[a] *Bioinformatics Research Center, University of Aarhus, Høegh Guldbergsgade 10, Building 90, 8000 Aarhus C, Denmark*
[b] *Molecular Diagnostic Laboratory, Aarhus University Hospital, Skejby, Brendstrupgaardsvej 100, 8200 Aarhus N, Denmark*

Received: 4 July 2005 / Accepted: 27 January 2006 / Published online: 22 April 2006

Abstract In this article we study a class of *randomly grown graphs* that includes some preferential attachment and uniform attachment models, as well as some evolving graph models that have been discussed previously in the literature. The degree distribution is assumed to form a Markov chain; this gives a particularly simple form for a stochastic recursion of the degree distribution. We show that for this class of models the empirical degree distribution tends almost surely and in norm to the expected degree distribution as the size of the graph grows to infinity and we provide a simple asymptotic expression for the expected degree distribution. Convergence of the empirical degree distribution has consequences for statistical analysis of network data in that it allows the full data to be summarized by the degree distribution of the nodes without losing the ability to obtain consistent estimates of parameters describing the network.

Keywords Biological network · Network model · Markov chain · Randomly grown graphs

1. Introduction

Randomly grown graphs (RGGs) have become increasingly popular models to describe real-world networks such as biological networks (e.g. protein and gene regulatory networks), ecological networks (e.g. food-webs) and technological networks (e.g. the World-Wide Web). RGGs are stochastic graphs that grow by the successive addition and/or deletion of nodes and addition, deletion and/or re-wiring of edges at each time step, and an RGG is said to evolve over time by application of rules for its generation. They mimic the real-world in being stochastic and

*Corresponding author.
E-mail address: wiuf@birc.au.dk (Carsten Wiuf).

dynamic, and the generating rules are often interpreted in relation to a real-world network.

Much attention has been devoted to the *degree distribution* of the network. By the degree distribution is meant the sequence that for any $k \geq 0$ gives the proportion of nodes of degree k in the network. Many authors, e.g. Barabási and Albert (1999) and Cooper and Frieze (2003), have concentrated on the degree distribution and shown that the degree distribution behaves in a certain way, e.g. that it converges to a scale free distribution as the network becomes large. While this originally has been motivated by empirical studies of the World-Wide Web (Barabási and Albert (1999); Kumar et al. (1999); Broder et al. (2000)), now also biological systems such as metabolic pathways and protein–protein interaction networks are routinely investigated (Barabási and Oltvai (2004); Stumpf et al. (2005) and references therein).

However, different criteria have been applied for convergence of the degree distribution of an RGG as the size of the network becomes large. Some authors, e.g. Barabási and Albert (1999) and Dorogovtsev and Mendes (2003) consider convergence, for all k, of the *expected proportion* of nodes of degree k. This means that the behaviour of a *single* realization of the RGG is *not* studied, only the average behaviour over many different realizations is considered. Other authors, e.g. Kumar et al. (2000) and Bollobás et al. (2001) use the stronger criterion that, for all k, the proportion of nodes of degree k in a single realization, i.e. the *empirical proportions*, converges in a stochastic sense (e.g., convergence in probability). Only the second criterion makes it *a priori* reasonable to expect that the empirical proportions converge and hence, a statement about convergence of the expected proportions does not imply a similar statement about the empirical proportions. It is possible to construct examples where the degree distribution converges in the first sense but not in the second.

Dorogovtsev and Mendes (2003) and others (see e.g. references in Dorogovtsev and Mendes (2003)) have constructed and used recursion equations (the socalled *master equations*) for the expected degree distribution in an RGG. This gives a simple and unified approach for the study of a wide class of RGGs, but the method suffers from the fact that the results can only be applied to an ensemble of realizations. In this paper, we define a class of RGGs which can be rigorously handled by recursion equations and provide a result that guarantees that the observed degree distribution converges to the expected distribution as time tends to infinity (i.e. the size of the graph). It is seen that the uniform attachment and the preferential attachment models studied by Bollobás et al. (2001), and the linear case of the evolving graph models, studied by Kumar et al. (2000), all belong to this class. Due to the simplicity of the models it is easy to make non-trivial generalisations, including a formalisation of the Barabási–Albert preferential attachment model which is closer to the original description than that of Bollobás et al. (2001). We also find upper limits for the rates of convergence.

A real-world network is one realization out of many possible ones, and convergence of the observed degree sequence to the expected has important consequences for statistical analysis of network data. The likelihood of the full network is difficult to handle due to dependencies between nodes. Our results show that it is sensible to summarize the data in the degree sequence and

optimize the parameters of the model by comparing with the expected degree distribution.

2. The model

We model an RGG as a Markov chain $\{\mathcal{G}_t\}_{t=t_0}^{\infty}$, where $\mathcal{G}_t$ is a random variable taking values in a subspace of the space of all graphs and t denotes discrete time. The subspace usually depends on t, e.g. it can be all graphs of size t. We assume that the graphs have neither multiple edges nor self-loops, but we will consider both directed and undirected graphs.

For our purposes, it suffices to study the *degree distribution,* $\{N_t\}_{t=t_0}^{\infty}$, of an RGG. This is a random process taking its values in the space of all sequences $\{x_k\}_{k=0}^{\infty}$ of non-negative integers x_k such that only a finite number of them are non-zero. We denote by $N_t(k)$ the kth element in the tth sequence:

$$N_t = \{N_t(k)\}_{k=0}^{\infty}.$$

Depending on the RGG, $N_t(k)$ will either be the number of nodes of degree k at time t or the number of nodes with in-degree k. In the following assume $t \geq t_0 > 0$. Assume the following.

1. $N_{t_0} = \{n_{t_0}(k)\}_{k=0}^{\infty}$ is non-random.
2. $\{N_t\}_{t=t_0}^{\infty}$ is a Markov chain. This condition implies that N_t contains all information avaliable about N_{t+1}— knowledge about $\mathcal{G}_t$ does not add any further information. Note that this does not follow generally from properties of $\mathcal{G}_t$.
3. In every time step one node is added. This implies that there is an integer $u \geq 0$, such that

$$\sum_{k=0}^{\infty} N_t(k) = t + u.$$

   At time $t = t_0$ there are $t_0 + u$ nodes and at each time step the size of the graph increases by one. It follows that $N_t(k) = 0$ for all $k \geq t + u$.
4. There are non-negative numbers a_k and c_k, such that $a_{k+1} \geq a_k$ and

$$\mathsf{E}\,(N_{t+1}(k)|N_t) = \left(1 - \frac{a_k}{t}\right) N_t(k) + \frac{a_{k-1}}{t} N_t(k-1) + c_k, \tag{1}$$

   where it is understood that $a_{-1} = 0$. In all our examples, a_k/t is a probability. Therefore we require that $N_t(k) = 0$ whenever $a_k/t > 1$.
5. The change in the number of nodes of degree k per time unit is uniformly bounded in t and in k; i.e. there is a constant M_1 such that $|N_t(k) - N_{t-1}(k)| \leq M_1$ for all t and k.

The condition that N_{t_0} is non-random can be relaxed somewhat. For example, the results hold unaffected if N_{t_0} is drawn among a finite number of (finite) initial graphs.

Example 1. (*Uniformly grown random graph*) In step $t+1$, a new node is created, and with probability p an edge is sent out from the new node and connected to any of the old nodes with uniform probability. If $N_t(k)$ is the number of nodes with degree k at time t, (1) is fulfilled with

$$a_k = p,$$

and

$$c_k = (1-p)\delta_{k0} + p\delta_{k1}.$$

This model has been studied in detail by Bollobás et al. (2004) and a slightly different version of the special case $p = 1$ by Bollobás et al. (2001).

Example 2. (*Preferential attachment*) Every node in the graph at t_0 has degree $k \geq 1$. Then at time $t+1$ a new node is added and attached to an old node chosen with probability proportional to its degree. Assume that there are t edges at time t; then the probability of choosing a node of degree k is $k/2t$ (the sum of all degrees is $2t$). We have

$$a_k = \frac{k}{2},$$

and

$$c_k = \delta_{k1}.$$

Further, $N_t(0) = 0$ for all t. Barabási and Albert (1999) defined informally a more general model, a fixed number m of nodes are sent out every time. Our model is a formalisation of the case $m = 1$. As Bollobás et al. (2001) pointed out, it was not obvious how to interpret $m > 1$. For their formalisation, they chose to allow for self-loops and multiple edges, even this by construction was forbidden in the Barabási–Albert model. We will return to this in Example 6.

Example 3. (*Evolving copying model, the linear case*) Kumar et al. (2000) start with a directed graph in which all the nodes have out-degree one. In time step $t+1$, a node v_{new} is created and an old node v_{old} is chosen uniformly amongst all old nodes. With probablity p an edge is drawn from v_{new} to v_{old}. With probability $1-p$ a link is drawn to the node that v_{old} links to. (Remember v_{old} has out-degree 1.) The latter form of attachment is called *copying,* from which this model derives its name. Here, $N_t(k)$ is the number of nodes with in-degree k at time t. We have

$$a_k = p + (1-p)k,$$

and

$$c_k = \delta_{0k}.$$

Remark 1. It should be noted that also Kumar et al. (2000) allow a fixed number of edges to be sent out every time. They do this by saying that each node has m out-links, which they carefully keep track of. In step $t+1$ the ith out-link is uniformly attached by probability p, and with probability $1-p$, the ith out-link of the chosen node is "copied" as above. Such a model can be created by generating m independent parallel random graphs as in Example 3 and then lumping all the nodes created a time t together. We refrain from studying this extra complexity, since it is easy to accomplish but rather obscures than sheds light on the structure of the model.

Example 4. Suppose the graph is grown as follows. We let t be the size of the graph at time t. At time $t+1$, a node is added. Consider two "tentative" edges e_1 and e_2 from the new node connected to old nodes, which are chosen uniformly at random *without* replacement. The edge e_1 is added with probability p_1 and e_2 is added independently with probability p_2. This model fits into our framework with

$$a_k = p_1 + p_2,$$
$$c_k = (1 - p_1 - p_2 + p_1 p_2)\,\delta_{0k} + (p_1 + p_2 - 2p_1 p_2)\,\delta_{1k} + p_1 p_2 \delta_{2k}.$$

Example 5. In every time step, a new node is created and m tentative edges are sent out to nodes which are chosen uniformly without replacement from the old nodes. Keep them independently with probability p. We must start with a graph of at least size m, and further

$$a_k = mp,$$
$$c_k = \sum_{j=0}^{m} \binom{m}{j} p^j (1-p)^{m-j} \delta_{jk}.$$

Example 6. (*Preferential attachment of m edges*) In the preferential attachment model by Barabási and Albert (1999), multiple edges and self-loops were not allowed. Bollobás et al. (2001) found themselves forced to allow for both in order to prove convergence of the degree distribution. This is not necessary in our formalism.

What makes this model special is that we choose m edges. Contrary to the other models we have studied, there are constraints on the initial graph. Denote the nodes at some time step by $v_1, \ldots, v_t$. Without replacement, we choose m such that if v_i has degree k_i, its probability, p_i, of being chosen is proportional to k_i. We have

$$m = p_1 + \cdots + p_t,$$

and hence

$$p_i = m\frac{k_i}{k_1 + \cdots + k_t}.$$

Therefore, a necessary condition for preferential attachment without replacement (PAWo) is that

$$\frac{mk_i}{k_1 + \cdots + k_t} \leq 1. \tag{2}$$

It seems to be non-trivial that condition 2 is also sufficient for a rule of preferential attachment without replacement to exist; henceforth called a PAWo rule. We provide a proof in the Appendix that it is possible to choose m nodes with the given marginal probabilities, but stresses that it can be achieved in several ways.

We also need to know that if the graph at time t fulfills the conditions for a PAWo rule then so does the graph at time $t+1$ irrespectively of which nodes have been chosen. Thus suppose (2) is true in step t, and apply the PAWo rule to get a new graph with degrees $k'_1, \ldots, k'_t, k'_{t+1}$, where $k'_{t+1} = m$. Then

$$k'_1 + \cdots + k'_t + k'_{t+1} = k_1 + \cdots + k_t + 2m.$$

The only possible values for k'_i are k_i and $k_i + 1$. Thus if (2) is satisfied, we have

$$mk'_i \leq k'_1 + \cdots + k'_t + k'_{t+1}$$

for $i = 1, 2, \ldots, t$. For this to be true even if $i = t+1$, we must have

$$k'_1 + \cdots + k'_{t+1} \geq m^2,$$

but this is naturally satisfied if we demand that all nodes in the initial graph have degree m.

We assume this is the case and that the total number of edges is a multiple of m. Since the graph grows by the successive addition of m nodes, this is reasonable assumption. If so, the sum of all degrees is $2mt$, and if an edge is selected according to a PAWo rule, we have $a_k = k/2$ and $c_k = \delta_{mk}$.

3. The recursion for the expected degree distribution

Denote $n_t(k) = \mathsf{E}N_t(k)$. Equation (1) yields

$$n_{t+1}(k) = \left(1 - \frac{a_k}{t}\right) n_t(k) + \frac{a_{k-1}}{t} n_t(k-1) + c_k.$$

A solution is found by substitution of $\alpha_k t$ for $n_k(t)$ and solving for α_k:

$$\alpha_k = \frac{a_{k-1}}{1 + a_k}\alpha_{k-1} + \frac{c_k}{1 + a_k}.$$

Then

$$\alpha_k = \sum_{j=0}^{k} \frac{c_j}{1+a_j} \prod_{i=j+1}^{k} \frac{a_{i-1}}{1+a_i}.$$

However, it follows from (1) that there has to be a k_0 such that $c_k = 0$ if $k > k_0$. For $k \geq k_0$, the solution can be simplified

$$\alpha_k = \alpha_{k_0} \prod_{j=k_0+1}^{k} \frac{a_{j-1}}{1+a_j}.$$

The values $n_t(k)$, $k = 0, 1 \ldots$ depend on the initial values $n_{t_0}(k)$, $k = 0, 1, \ldots$, but the next lemma shows the difference between $n_t(k)/t$ and α_k tends to zero for all k as $t \to \infty$.

The following lemma provides a stronger result than required to prove our claim. However, it will be useful in Section 5.

Lemma 1. *Let k and t_1 be given such that $t_1 \geq k$. Then, for all $t \geq t_1$,*

$$\max_{0 \leq j \leq k} \left| n_t(j) - \alpha_j t \right| \leq \max_{0 \leq j \leq k} \left| n_{t_1}(j) - \alpha_j t_1 \right|.$$

Proof: Let $t > t_1$, and suppose $0 \leq i \leq k$. Then

$$|n_t(i) - \alpha_i t| \leq \left(1 - \frac{a_i}{t-1}\right) |n_{t-1}(i) - \alpha_i (t-1)| + \frac{a_{i-1}}{t-1} |n_{t-1}(i-1)$$
$$-\alpha_{i-1}(t-1)| \leq \max_{0 \leq j \leq k} |n_{t-1}(j) - \alpha_j (t-1)|,$$

and the lemma follows by induction.

Example 1. (*continued, uniformly grown random graph*) We find

$$\alpha_0 = \frac{1-p}{1+p}, \tag{3}$$

and

$$\alpha_k = \frac{2p^k}{(1+p)^{k+1}} \tag{4}$$

for $k = 1, 2, 3, \ldots$. In consequence the distribution is a modified geometric distribution. For $p = 1$, this coincides with the result by Bollobás et al. (2001).

Example 2. (*continued, preferential attachment*) We find $\alpha_0 = 0$ and

$$\alpha_k = \frac{4}{k(k+1)(k+2)}$$

for $k = 1, 2, \ldots$. To get the simple form (1), we assumed that time t is the number of edges – not the number of nodes, which is $n_0 + t$. Here n_0 denotes the number of nodes at time t_0. Therefore $n_t(k)/(n_0 + t)$, not $n_t(k)/t$ is the expected proportion of nodes of degree k at time t. However, we could anyway use Lemma 1 to see that

$$\lim_{t\to\infty} \frac{n_t(k)}{n_0 + t} = \alpha_k.$$

Example 3. (*continued, The copying model*) We find $\alpha_0 = 1/(1 - p)$ and

$$\alpha_k = \frac{1}{1-p} \prod_{j=1}^{k} \frac{2p - 1 + (1-p)j}{1 + p + (1-p)j}$$

for $k = 1, 2, \ldots$. It is not as obvious as in Example 2 that a_k has "scale-free tail," but it can be seen as follows. Define $b_p = (1 + p)/(1 - p)$ and $a_p = (2p - 1)/(1 - p)$. Note that $-1 < a_p < b_p$ and use the "mean value theorem" to see that

$$\begin{aligned}\log \alpha_k &= -\log(1-p) + \sum_{j=1}^{k} \left[\log\left(1 + \frac{a_p}{j}\right) - \log\left(1 + \frac{b_p}{j}\right) \right] \\ &= -\log(1-p) - \frac{2-p}{1-p} \sum_{j=1}^{k} \frac{1}{(1+\xi_j)j}\end{aligned}$$

for $a_p/j < \xi_j < b_p/j$. Hence also

$$\sum_{j=1}^{k} \frac{1}{j + b_p} < \sum_{j=1}^{k} \frac{1}{(1+\xi_j)j} < \sum_{j=1}^{k} \frac{1}{j + a_p}$$

and

$$\alpha_k = \Theta\left(k^{-\gamma_p}\right)$$

as $k \to \infty$, where $\gamma_p = (2 - p)/(1 + p)$. This is exactly the result of Kumar et al. (2000). For nonnegative $g_k = \Theta(f_k)$ we mean that there are non-negative c, C such that $c f_k \le g_k \le C f_k$ for all k large.

Example 4. (*continued*) We get

$$\begin{aligned}\alpha_0 &= \frac{c_0}{1 + p_1 + p_2}, \\ \alpha_1 &= \frac{(p_1 + p_2)c_0}{(1 + p_1 + p_2)^2} + \frac{c_1}{(1 + p_1 + p_2)}, \\ \alpha_2 &= \frac{(p_1 + p_2)^2 c_0}{(1 + p_1 + p_2)^3} + \frac{(p_1 + p_2)c_1}{(1 + p_1 + p_2)^2} + \frac{c_2}{1 + p_1 + p_2},\end{aligned}$$

and

$$\alpha_k = \alpha_2 \left(\frac{p_1 + p_2}{1 + p_1 + p_2} \right)^{k-2}$$

for $k = 3, 4, \ldots$.

Example 5. (*continued*) Also in this case we find that α_k decreases exponentially as $k \to \infty$. However, the expression for α_k is quite involved and is left out.

Example 6. (*continued, Preferential attachment of m edges*) Here $\alpha_{m+j} = 0$ for $j < 0$ and

$$\alpha_{m+j} = \frac{2m(m+1)}{(m+j)(m+1+j)(m+2+j)}$$

for $j = 0, 1, \ldots$.

4. Convergence to the expected values

To show that $N_t(k)/t$ converges to the expected value (Theorem 2), we use the following version of Hoeffding's inequality. (See, e.g. Grimmet and Stirzaker (1992))

Theorem 1. *Let $\{Y_s\}_{s=t_0}^t$ be a martingale (with respect to some filtration) such that for all s with $t_0 < s \leq t$ $|Y_s - Y_{s-1}| \leq M$ for some M (a.s.), then, for $\epsilon > 0$,*

$$\mathsf{P}(|Y_t - Y_{t_0}| > \epsilon) \leq 2 \exp\left(-\frac{\epsilon^2}{2(t - t_0)M^2} \right).$$

Define $Y_{s,t}(k) = \mathsf{E}\left(N_t(k)|N_s\right)$. Since $\{N_t\}_{t=t_0}^{\infty}$ is a Markov chain, for fixed t and k, the sequence $\{Y_{s,t}\}_{s=t_0}^t$ is a Martingale with respect to $\{N_s\}_{s=t_0}^t$. (This is the so-called Doob's martingale; see, again, Grimmet and Stirzaker (1992)).

Lemma 2. *Let $M(k) = M_1 + a_k + u + |c_k|$. Then*

$$|Y_{s,t}(k) - Y_{s-1,t}(k)| \leq M(k) \tag{5}$$

for all s, t with $t_0 < s \leq t$.

Proof: We use induction in the difference $d = t - s$. If $d = 0$, consider $Y_{t,t}(k) - Y_{t-1,t}(k)$. We have $Y_{t,t}(k) = N_t(k)$ and

$$Y_{t-1,t}(k) = \left(1 - \frac{a_k}{t-1}\right) N_{t-1}(k) + \frac{a_{k-1}}{t-1} N_{t-1}(k-1) + c_k. \tag{6}$$

Hence,

$$|Y_{t,t}(k) - Y_{t-1,t}(k)| \le |N_t(k) - N_{t-1}(k)| + \frac{a_k}{t-1}\left\{N_{t-1}(k) + N_{t-1}(k-1)\right\} + |c_k| \le M_1 + a_k + u + |c_k| = M(k).$$

Now fix a $d > 0$ and suppose (5) is true for all s, with $t - s = d - 1$. By conditioning on N_s in relation (6), we get

$$\mathsf{E}(N_t(k)|N_s) = \left(1 - \frac{a_k}{t-1}\right)\mathsf{E}(N_{t-1}(k)|N_s) + \frac{a_{k-1}}{t-1}\mathsf{E}(N_{t-1}(k-1)|N_s) + c_k.$$

This is true also if s is replaced by $s - 1$, and we have

$$\mathsf{E}(N_t(k)|N_s) - \mathsf{E}(N_t(k)|N_{s-1}) = \left(1 - \frac{a_k}{t-1}\right)\left\{\mathsf{E}(N_{t-1}(k)|N_s) - \mathsf{E}(N_{t-1}(k)|N_{s-1})\right\} + \frac{a_{k-1}}{t-1}\left\{\mathsf{E}(N_{t-1}(k-1)|N_s) - \mathsf{E}(N_{t-1}(k-1)|N_{s-1})\right\}.$$

If $a_k/(t-1) > 1$, then $N_{t-1}(k-1) = N_{t-1}(k) = 0$ almost surely. If not,

$$|\mathsf{E}(N_t(k)|N_s) - \mathsf{E}(N_t(k)|N_{s-1})| \le \left(1 - \frac{a_k}{t-1}\right)|\mathsf{E}(N_{t-1}(k)|N_s) - \mathsf{E}(N_{t-1}(k)|N_{s-1})| + \frac{a_{k-1}}{t-1}|\mathsf{E}(N_{t-1}(k-1)|N_s) - \mathsf{E}(N_{t-1}(k-1)|N_{s-1})| \le M(k),$$

since the inductive assumption is valid for $t - 1$ and s, whose difference is $d - 1$. The lemma follows.

The following theorem shows that for fixed k, $N_t(k)/t$ tends to α_k in probability as t increases.

Theorem 2. *Let k and $\epsilon > 0$ be fixed. If $t > t_0$ is so large such that*

$$|n_t(k)/t - \alpha_k| < \epsilon/2, \tag{7}$$

then

$$\mathsf{P}\left(\left|\frac{N_t(k)}{t} - \alpha_t\right| > \epsilon\right) \le 2\exp\left(-\frac{t^2\epsilon^2}{8(t-t_0)M(k)^2}\right).$$

Proof: We know that $Y_{t_0} = n_{t_0}(k)$ and $Y_t = N_t(k)$. Thus Theorem 1 gives for $\eta > 0$

$$\mathsf{P}\left(\left|\frac{N_t(k)}{t} - \frac{n_t(k)}{t}\right| > \eta\right) = \mathsf{P}(|Y_t - Y_{t_0}| > t\eta) \le 2\exp\left(-\frac{t^2\eta^2}{2(t-t_0)M(k)^2}\right).$$

Then

$$\mathsf{P}\left(\left|\frac{N_t(k)}{t}-\alpha_k\right|>\epsilon\right)\leq \mathsf{P}\left(\left|\frac{N_t(k)}{t}-\frac{n_t(k)}{t}\right|>\epsilon/2\right)\leq 2\exp\left(-\frac{t^2\epsilon^2}{8(t-t_0)M(k)^2}\right)$$

and the theorem is proven.

The Borel–Cantelli lemma provides the stronger result that $N_t(k)/t$ converges a.s. It is shown in the following corollary.

Corollary 1. *For fixed k,*

$$\frac{N_t(k)}{t}\to\alpha_k \text{ a.s.}$$

as $t\to\infty$.

Proof: Fix k and t_1 such that $a_k\leq t_1$. If $\max_{0\leq j\leq k}|n_{t_1}(j)-\alpha_j t_1|>0$, let

$$\varepsilon_t=2t^{-1/4}\max_{0\leq j\leq k}|n_{t_1}(j)-\alpha_j t_1|^{1/4}.$$

Choose $t_2\geq t_1$ large such that $\varepsilon_t<2$ when $t\geq t_2$. Then, for all $t\geq t_2$,

$$|n_t(k)-\alpha_k t|\leq\left(\frac{\varepsilon_t}{2}\right)^4<\frac{\varepsilon_t}{2},$$

where the first inequality comes from Lemma 1. Thus (7) is satisfied for $t\geq t_2$ if $\epsilon=\epsilon_t$.

If $\max_{0\leq j\leq k}|n_{t_1}(j)-\alpha_j|=0$, let $\varepsilon_t=t^{-1/4}$. Then Lemma 1 shows that $|n_t(k)-\alpha_k|=0$, and (7) is trivially satisfied with $\epsilon=\epsilon_t$.

In both cases, we use Theorem 2 to conclude that

$$\sum_{t=t_2}^{\infty}\mathsf{P}\left(\left|\frac{N_t(k)}{t}-\alpha_k\right|>\varepsilon_t\right)<\infty,$$

and the Borel–Cantelli lemma (see. e.g. Shiryaev (1996)) gives a.s. convergence.

For completeness, we add the following two corollaries; the first is proven by noting that $|N_t(k)/t-\alpha_k|\leq u+1+\alpha_k$, and that if a sequence of uniformly limited random variables converges in probability, then it also coverges in p-norm for all p.

Corollary 2. *For all k,*

$$\frac{N_t(k)}{t}\to\alpha_k \text{ in } \mathcal{L}_p$$

for any $0\leq p<\infty$.

Corollary 3. *The process* $(N_t(0)/t, N_t(1)/t, \ldots)$ *converges weakly as* $t \to \infty$ *to the constant process* $(\alpha_0, \alpha_1, \ldots)$.

Proof: It follows from Corollary 2 that $(N_t(0)/t, \ldots, N_t(k)/t)$ converges in probability $(\mathcal{L}_0)$ for any k; hence the process $(N_t(0)/t, N_t(1)/t, \ldots)$ converges weakly.

5. Range of convergence

Since the process starts with a finite number of nodes and grows by one node being added in every time step, it follows for all $t \geq t_0$ that there is k such that $N_t(j) = 0$ for all $j \geq k$. Hence the convergence is non-uniform in the sense that

$$\min_k \mathsf{P}\left((1-\epsilon)\,\alpha_k \leq \frac{N_t(k)}{t} \leq (1+\epsilon)\,\alpha_k\right) = 0$$

for all ϵ with $0 < \epsilon < 1$. However, as we will see, if we fix a number C between 0 and 1, for every $t \geq t_0$, there is a $\kappa_C(t)$ such that

$$\min_{0\leq j\leq \kappa_C(t)} \mathsf{P}\left((1-\epsilon)\,\alpha_j \leq \frac{N_t(j)}{t} \leq (1+\epsilon)\,\alpha_j\right) \geq C. \tag{8}$$

It will be seen that we can choose $\kappa_C(t)$ such that it tends to infinity as t tends to infinity, and the rate at which $\kappa_C(t)$ tends to infinity will be a measure of the *range of convergence* of the process. The inequality (8) can also be written

$$\max_{0\leq j\leq \kappa_C(t)} \mathsf{P}\left(\left|\frac{N_t(j)}{t} - \alpha_j\right| > \alpha_j\epsilon\right) < 1 - C.$$

Looking at Theorem 2, we see that $\kappa_C(t)$ can be chosen as the largest k fulfilling

$$\left|\frac{n_t(j)}{t} - \alpha_j\right| \;<\; \epsilon\alpha_j/2, \tag{9}$$

and

$$2\exp\left(-\frac{t^2\epsilon^2\alpha_j^2}{8(t-t_0)M(j)^2}\right) \;<\; 1 - C \tag{10}$$

for all $j \leq k$. If there is no such k, we let $\kappa_C(t) = -1$.

Example 1. (*continued, Uniformly grown random graph*) For the uniform attachment model, $\alpha_j = p$, and we can use Lemma 1 with $t_1 = t_0$:

$$\left|\frac{n_t(j)}{t} - \alpha_j\right| \leq \frac{1}{t}|n_{t_0}(j) - \alpha_j|.$$

Hence inequality (9) is true if

$$\frac{1}{t}|n_{t_0}(j) - \alpha_j| < \epsilon\alpha_j/2.$$

From (4):

$$\frac{1}{t}|n_{t_0}(j) - \alpha_j|(1-p) < \epsilon\left(\frac{p}{1+p}\right)^j.$$

Solving this inequality, we can let k grow as $O(\log(t))$ and make (9) true for all $j \le k$. For the uniformly grown random graph, $M(k)$ is constant. Hence (10) is true if we let k grow as $O(\log(k))$. In summary we can choose

$$\kappa_C(t) = O(\log t).$$

Example 2. (*continued, Preferential attachment*) Here Lemma 1 can be used with $t_1 = k$, and thus

$$\max_{0\le j\le k}\left|\frac{n_t(j)}{t} - \alpha_j\right| \le \frac{1}{t}\max_{0\le j\le k}\left|\frac{n_{t_1}(j)}{t} - \alpha_j\right| \le \frac{n_{t_0}+k}{t}.$$

Since $\alpha_k = O(1/k^3)$, we can let k grow as $O(t^{1/4})$ to make (9) true. However, we also know that $M(j) = O(j)$. Hence, to make inequality (10) true, we cannot let k grow faster than $t^{1/8}$. In summary, we can define $\kappa_C(t)$ such that $\kappa_C(t) = O(t^{1/8})$.

Example 3. (*continued, The copying model*) Repeating the argument in the previous example, we find that $\kappa_C(t)$ can be chosen such that

$$\kappa_C(t) = O\left(t^{\frac{1}{2\gamma p+2}}\right).$$

6. Discussion

This paper demonstrates a simplified and uniform formalism, which combines some already known results about the degree distribution into a larger framework, from which we easily derive almost sure convergence and convergence in norm, including a general asymptotic expression as $t \to \infty$ for the degree distribution, and a description of the range of convergence.

In earlier papers, like Kumar et al. (2000) and Bollobás et al. (2001), when a fixed number $m > 1$ of new edges have been added in every time step, it has been a problem that sampling nodes without replacement introduces dependence between the choices, a problem that has been solved by accepting multiple edges. We show here a solution without multiple edges.

We have studied a very simple class of models. Extensions include letting c_k and/or a_k depend on t or allowing for terms of order $j < k-1$ in Equation 1.

However, it is clear that the key features of the model are that 1) the degree distribution as a process is Markovian and 2) the change in the number of nodes of a certain degree is bounded in t. Also, more complicated models would fulfil these requirements and consequently similar results to those presented here could be proven. We have chosen to stick to the simple models in order to make the presentation clear.

In our construction, the asymptotic independence of the initial graph stands out clear. For applications this is important, since it is not reasonable to assume that the generating process has been going on all the time since a very small initial graph. Rather, it seems more appropriate to ignore the first period of evolution and start the process with some intermediate graph, which is chosen such that the generating process has already stabilised. Our result shows that how the graph was generated before this time is asymptotically unimportant, at least with respect to the degree distribution.

Growth, copying and preferential attachments are biological mechanisms by which biological systems, such as protein interaction networks, metabolic pathways and transcriptional networks, evolve (see e.g. Barabási and Oltvai (2004) for a review of biological networks). Gene duplication or copying produce genes that share some of the interaction partners with the duplicated gene and highly connected genes are more likely to become partners in new processes and interactions than nodes with few connections (preferential attachment). The models discussed in this paper reflect these biological mechanisms for network growth (Barabási and Oltvai (2004)).

It has been argued Burda et al. (2001) that the mechanistic process underlying network evolution is less important than the ability to qualitatively describe the data. It is for example well known, that protein interaction networks do not grow according to the simple preferential attachment model. For example, millions of years ago, a whole genome duplication occurred in the lineage ancestral to *S. cerevisiae* and other related yeast species. Such exceedingly rare events—even though they are of fundamental biological importance—cannot be easily incorporated into a probability model. But even if the model is oversimplified it still allows us to gain new insights. It is for example possible to estimate the parameters of the models; e.g. a copying probability would be interpreted in relation to duplication activity in the network.

With the abundance of biological network data that are becoming available probabilistic models for its analysis are in demand. The apparatus for statistical analysis of such data is however still lacking behind. In particularly, findings on scale free networks have been debated and are not generally agreed upon (e.g. Stumpf et al. (2005)). It has also become clear that many other aspects and features (such as small-scale structures) of a network are important in the study of the evolution and dynamics of real-world networks (e.g. Milo et al. (2004) and Rice et al. (2005)). For correct interpretation and analysis of network data, a sound mathematical and statistical methodology is important. One such methodology is likelihood theory. However, computation of the likelihood of $\mathcal{G}_t$ can be a huge burden computationally (see Wiuf et al. (2005)) that renders application of likelihood methods difficult. Our results demonstrate that at least for a large class of RGGs it is sound to summarize the network data into the degree sequence and optimize

parameters of the model by comparing the emipircal degree sequence to the expected degree sequence. The empirical degree sequence is very easy to calculate and a reduction of the network data to the emiprical degree sequence is thus likely to reduce the computational burden of the statistical analysis without making it invalid or untrustworthy.

Appendix The existence of a PAWo rule

To be able to construct the PAWo rule, we need the following theorem, which has been known a long time (see e.g. Fellegi, I.P. (1963)). However, we are not aware of a simple proof and provide one here for completeness.

Theorem 3. *Let $p_1, \ldots, p_t$ be probabilities with*

$$p_1 + \cdots + p_t = m$$

for an integer m. Then there is (at least) one procedure to choose m values out of $\{1, 2, \ldots, t\}$ such that the marginal probability of choosing i is p_i.

Proof: We suppose without loss of generality that

$$p_1 \leq p_2 \leq \ldots \leq p_t.$$

We will use induction in t. If $t = 1, 2$ or if $m = 1$ the statement is trivially true. Fix $t_0 \geq 3$ and assume the statement is true for $t < t_0$ and $m \leq t$. Now let $t = t_0$. If $m = t$ we are done. Otherwise it is sufficient to prove it for $m \geq t/2$, since when we know there is such a procedure for this case, if $m < t/2$, we can use it to choose the $t - m$ elements which are *not* chosen with probabilities $1 - p_i$.

Thus assume $m \geq t/2$. Choose t with probability p_t. If t is *not* chosen, choose $t - 1, \ldots, t - m$ always. If t is chosen, we define probabilities $\tilde{p}_1, \ldots, \tilde{p}_{t-1}$ with sum $m - 1$ such that

$$\begin{aligned} \tilde{p}_i\, p_t &= p_i, \quad i = 1, 2, \ldots, t - m - 1 \\ \tilde{p}_i\, p_t + (1 - p_t) &= p_i, \quad i = t - m, \ldots, t - 1. \end{aligned}$$

Then we use the inductive hypothesis to conclude that there is a procedure to choose $m - 1$ elements without replacement from $\{1, \ldots, t - 1\}$ with probabilities $\tilde{p}_1, \ldots, \tilde{p}_{t-1}$ and hence also to choose m out of t To convince ourselves that the $\tilde{p}_1, \ldots, \tilde{p}_{t-1}$ fulfill the requirements, solve the equations to obtain:

$$\begin{aligned} \tilde{p}_i &= \frac{p_i}{p_t} \quad i = 1, 2, \ldots, t - m - 1 \\ \tilde{p}_i &= \frac{p_t + p_i - 1}{p_t} \quad i = t - m, \ldots, t - 1. \end{aligned}$$

The $\tilde{p}_i$ sum to $m-1$ and $\tilde{p}_i \leq 1$. We are done if we can show that the nominator in the last equation is non-negative. To show this, consider, for $i = t-m, \ldots, t-1$,

$$\begin{aligned} m(p_t + p_i - 1) &= mp_t + mp_i - (p_1 + \cdots + p_t) \\ &= mp_t - (p_t + \cdots + p_{t-m+1}) + mp_i - (p_{t-m} + \cdots + p_1). \end{aligned}$$

By assumption $m \geq t/2$ and therefore there are at most m terms in the last parenthesis. Since $i \geq t-m$ they are all less than or equal to p_i. Hence the expression is non-negative.

This proof guarantees the existence of a PAWo rule. If node t is chosen the construction requires a non-stochastic choice of the remaining $m-1$ nodes, which introduce strong dependencies between some nodes.

Looking at the proof, it is reasonable to assume that there are other procedures fulfilling the theorem. This is also the case. Just as in the proof, let $m \leq t/2$ and make the same assumption of induction. If t is not chosen, choose m elements out of $1, \ldots, t$ such that i is chosen with probability $\hat{q}_i$. If t is chosen, choose $m-1$ such that i is in the sample with probability $\hat{p}_i$. From

$$\hat{p}_i\, p_t + \hat{q}_i(1 - p_t) = p_i$$

follows

$$\hat{p}_i = \frac{p_i - (1-p_t)\hat{q}_i}{p_t}.$$

For $i = 1, 2, \ldots, t-m-1$ we can choose $\hat{q}_i$ arbitrary except for the condition

$$0 \leq \hat{q}_i \leq \min\left(1, \frac{p_i}{1-p_t}\right),$$

which is interpreted as 1 if $p_t = 1$. Note that the sum of $\hat{q}_i$ with $i = 1, 2, \ldots, t-m-1$ is less than or equal to $m-1$. Then we can choose $\hat{q}_i$ for $i = t-m, \ldots, t-1$ arbitrary except that

$$\sum_{i=t-m}^{t-1} \hat{q}_i = m - 1 - \sum_{i=1}^{t-m-1} \hat{q}_i.$$

References

Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science, 286 509–512.

Barabási, A.L., Oltvai, Z.N., 2004. Network biology: Understanding the cell's functional organization. Nat. Rev. Genet. 5, 101–113.

Bollobás, B., Janson, S., Riordan, O., 2004. The phase transition in the uniformly grown random graph has infinite order. Random Struct. Algorithms 26(1–2), 1–36.

Bollobás, B., Riordan, O., Spencer, J., Tusnady, G., 2001. The degree sequence of a scale-free random graph process. Random Struct. Algorithms 18(3), 279–290.
Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., Wiener, J., 2000. Graph structure in the web. Random Comput. Netw. 33(3), 309–320.
Burda, Z., Diaz-Correia, J., Krzywicki, A., 2001. Statistical ensemble of scale-free random graphs. Phys. Rev. E 64, 046118.
Cooper, C., Frieze, A., 2003. A general model of web Graphs. Random Struct. Algorithms 22(3), 311–335.
Dorogovtsev, S.N., Mendes, J.F.F., 2003. Evolution of Networks: From Biological Nets to the Internet and WWW. Oxford University Press, Oxford.
Fellegi, I.P., 1963. Sampling with varying probabilities without replacement: Rotating and non-rotating samples. J. Am. Stat. Assoc. 58, 183–201.
Grimmet, G.R., Stirzaker, D.R., 1992. Probability and Random Processes. 2nd edn. Oxford University Press, Oxford.
Kumar, S.R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A., Upfal, E., 2000. Stochastic models for the web graph. Proc. 41st Ann. Symp. Found. Comput. Sci. 57–65.
Kumar, S.R., Raghavan, P., Rajagopalan, S., Tomkins, A., 1999. Trawling the web for emerging cyber-communities. Proc. Intl. Conf. Comb. Comput., 1–18.
Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M., Alon, U., 2004. Superfamilies of evolved and designed networks. Science 303, 1538–1542.
Rice, J.J, Kershenbaum, A., Stolovitzky, G., 2005. Lasting impressions: Motifs in protein–protein maps may provide footprints of evolutionary events. Proc. Natl. Acad. Sci., 102(9), 3173–3174.
Shiryaev, A.N., 1996. Probability, 2nd edn. Spinger-Verlag, Berlin.
Stumpf, M.P.H., Ingram, P.J., Nouvel, I., Wiuf, C., 2005. Statistical model selection methods applied to biological network data. Trans. Comput. Sys. Biol. 3, 65–77.
Wiuf, C., Brameier, M., Hagberg, O., Stumpf, M.P.H., 2005. A likelihood approch to analysis of network data, PNAS, to appear.