Ann Inst Stat Math (2008) 60:747–761
DOI 10.1007/s10463-008-0181-5

A preferential attachment model with Poisson growth for scale-free networks

Paul Sheridan · Yuichi Yagahara · Hidetoshi Shimodaira

Received: 17 January 2008 / Revised: 18 April 2008 / Published online: 10 July 2008

Abstract We propose a scale-free network model with a tunable power-law exponent. The Poisson growth model, as we call it, is an offshoot of the celebrated model of Barabási and Albert where a network is generated iteratively from a small seed network; at each step a node is added together with a number of incident edges preferentially attached to nodes already in the network. A key feature of our model is that the number of edges added at each step is a random variable with Poisson distribution, and, unlike the Barabási–Albert model where this quantity is fixed, it can generate any network. Our model is motivated by an application in Bayesian inference implemented as Markov chain Monte Carlo to estimate a network; for this purpose, we also give a formula for the probability of a network under our model.

Keywords Bayesian inference · Complex networks · Network models · Power-law · Scale-free

1 Introduction

Until recent times, modeling of large-scale, real-world networks was primarily limited in scope to the theory of *random networks* made popular by Erdös and Rényi (1959). In the Erdös–Rényi model, for instance, a network of N nodes is generated by connecting each pair of nodes with a specified probability. The *degree distribution* $p(k)$ of a large-scale random network is described by a binomial distribution, where the *degree* k of a node denotes the number of undirected edges incident upon it. Thus, degree in a

P. Sheridan (✉) · Y. Yagahara · H. Shimodaira
Department of Mathematical and Computing Sciences, Tokyo Institute of Technology,
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan
e-mail: sherida6@is.titech.ac.jp

random network has a strong central tendency and is subject to exponential decay so that the average degree of a network is representative of the degree of a typical node.

Over the past decade, though, numerous empirical studies of *complex networks*, as they are known, have established that in many such systems—networks arising from real-world phenomena as diverse in origin as man-made networks like the World Wide Web, to naturally occurring ones like protein-protein interaction networks, to citation networks in the scientific literature; see Albert et al. (1999), Jeong et al. (2001), and Redner (1998), respectively—the majority of nodes have only a few edges, while some nodes, often called *hubs*, are highly connected. This characteristic cannot be explained by the theory of random networks. Instead, many complex networks exhibit a degree distribution that closely follows a *power-law* $p(k) \propto k^{-\gamma}$ over a large range of k, with an exponent γ typically between 2 and 3. A network that is described by a power-law is called *scale-free*, and this property is thought to be fundamental to the organization of many complex systems; Strogatz (2001).

As preliminary experimental evidence mounted (see Watts and Strogatz (1998), for instance), a simple, theoretical explanation accounting for the universality of power-laws soon followed; the network model of Barabási and Albert (1999) (BA) provided a fundamental understanding of the development of a wide variety of complex networks on an abstract level. Beginning with a connected seed network of t_0 nodes, the BA algorithm generates a network over a sequence of iterations $t = t_0 + 1, t_0 + 2, \ldots$ by using two basic mechanisms: *growth*, where at each step t the network is augmented by a single node together with $m \leq t_0$ undirected incident edges; and *preferential attachment* where the m edges emanating from the newly added node are connected to exactly m nodes already in the system such that the probability a node of degree k gets an edge is proportional to $r(k) = k$, the degree of the node. The function $r(k)$ is a specific example of a *preferential attachment function* and we defer a more detailed treatment to the next section. Moving along, when m is fixed throughout, Barabási and Albert (1999) gave a heuristic argument showing that the BA model follows a power-law with exponent $\gamma = 3$ in the limit of large t. The sudden appearance of the this model in the literature, nearly a decade ago, sparked a flurry of research in the field, and, consequently, numerous variations and generalizations upon this prototypal model have been proposed; Boccaletti (2006) provides a thorough survey of the subject.

In this paper, we propose a new *growing model* based on preferential attachment: the Poisson growth (PG) model. Our model, as described in Sect. 2, is an extension of the BA model in two regards. Firstly, we consider the number of edges added at a step to be a random quantity; at each step, we assign a value to m according to a Poisson distribution with expectation $\lambda > 0$. Secondly, we avail ourselves of a more general class of preferential attachment functions $r(k)$ studied by several authors including Krapivsky and Redner (2001) and Dorogovtsev et al. (2000). In Sect. 3 we argue that the degree distribution of the PG model follows a power-law with exponent γ that can be tuned to any value greater than 2; the technical details of our argument are left for the Appendix. In addition, we conducted a simulation study to support our theoretical claims. Our results, provided in Sect. 4, show that the values of γ we estimated from networks generate under the PG model are in agreement with those predicted by our formulae for the power-law exponent.

Our motivation for proposing the PG model, as explained in Sect. 5, arises from a need for a simple, yet realistic scale-free model that is serviceable in applications. In fact, with our model every network has a nonzero probability of being generated, in addition to possessing a tunable power-law exponent. In contrast, the BA model has a fixed γ, and is subject to numerous structural constraints which severely limit the variety of generable networks. We give a simple formula for the probability of a network under the PG model, which can be applied quite naturally in Bayesian inference using Markov chain Monte Carlo (MCMC) methods. Firstly, given a network G, we may estimate the PG model parameters, or engage in model selection in the case when we have more models; or, going against the grain, we may estimate an unknown G from data using our PG model formula as a scale-free prior distribution.

Finally, a cursory note on some important classes of scale-free models amassed in the literature is in order. For brevity we merely outline some generic methodologies and direct the reader to Boccaletti (2006) for model specific details.

Nonlinear preferential attachment: Generalizations of the BA model have been proposed in which $r(k) = k^{\nu}, \nu > 0$. It has been established, as we describe in Sect. 2, that when $r(k)$ is asymptotically linear in k, then the algorithm generates scale-free networks with tunable power-law exponent.

Fitness models: An alternate way of tinkering with the attachment function is to assign to each node a *fitness*, or weight, so that $r(k_i) = \eta_i k_i$, where η_i is the fitness and k_i the degree of node i, respectively; see Bianconi (2001).

Dynamic edge rewiring: In the BA model, edges cannot be altered after they have been placed in a network. Therefore, a natural avenue for extending the BA model is to allow for rewiring of existing edges in the network. Albert and Barabási (2000) proposed a BA-like model based on this idea and showed it follows a power-law, again, with tunable exponent.

Other mechanisms: Scale-free models not based on preferential attachment have also been proposed such as the model of Solé et al. (2002), which is a growing model based on node duplication and edge rewiring. Other scale-free models not based on growth have also been put forward; for example, the static model of Lee et al. (2005).

Although extending the BA model via generalizing $r(k)$ can lead to a tunable power-law exponent, such models are still subject to structural constraints similar to that of the BA model itself. On the other hand, dynamic edge rewiring does allow for additional freedom in the network structures that can be generated, but at the expense of using more complex generating algorithms. Hence we contend the PG model is a useful addition to the aforementioned classes of preferential attachment models insomuch as it is serviceable in applications, and its implementation is computationally straight forward in the inferential problems described above.

2 The Poisson growth model

In the PG model, we begin with a small seed network of t_0 nodes. Let $G_t = (V_t, E_t)$ be the network at the onset of time step $t \geq t_0$ where $V_t = \{v_1, v_2, \ldots, v_t\}$ is a set of t nodes and E_t is a multiset of unordered pairs $\{v_i, v_j\}$ taken from V_t such that $i \neq j$. As a multiset the elements of E_t can have more than one membership entailing that

a pair of nodes may be connected by *multiple edges*; however the condition $i \neq j$ prohibits loops, which are edges from a node to itself. The updated network G_{t+1} is generated from G_t as follows:

Poisson growth: A new node v_{t+1} is added to the network together with m_t incident edges; m_t is a random variable assigned according to a Poisson distribution with expectation $\lambda > 0$.

Preferential attachment: Each edge emanating from v_{t+1} is connected with a node already in the network. Node selection can be considered as a series of m_t independent trials, where at each trial the probability of selecting a node from V_t with degree k is

$$q_t(k) = \frac{r(k)}{\sum_{i=1}^{t} r(k_{i,t})}, \tag{1}$$

where $k_{i,t}$ is the degree of node v_i at step t. Define $s_{i,t}$ as the number of times node v_i is chosen at step t. Then the entire selection procedure is equivalent to drawing a vector $(s_{1,t}, s_{2,t}, \ldots, s_{t,t})$ from a multinomial distribution with probabilities $q_t(k_{1,t})$, …, $q_t(k_{t,t})$ and sample size m_t. Equivalently, $s_{i,t}$ has Poisson distribution with expectation $\lambda q_t(k_{i,t})$ independently for $i = 1, \ldots, t$.

The PG model is determined by the choice of $r(k)$; we concentrate on two specifications and discuss their implications in the next section. Firstly, let

$$r(k) = k + a \tag{2}$$

where the offset $a > 0$ is a constant. More generally, we define

$$r(k) = k + a, \quad k \geq 1, \quad \text{and} \quad r(0) = b \tag{3}$$

by taking $a \geq -1$ with extended domain, but in doing so define a threshold parameter $b \geq 0$. Indeed, the latter formulation includes the former as a special case when we take $a = b \geq 0$, so that overall our model is specified by the parameter $\theta = (a, b, \lambda)$.

The BA model can be explained as a reduction of our model by taking $a = b = 0$, and fixing $1 \leq m_t = m \leq t_0$ so that the number of edges added to the system at each step is a constant; the new edges are preferentially attached from the new node to exactly m other nodes. Many structural constrains are implicit in the BA model. Indeed, at step t, a network with t nodes must have $m(t - t_0) + |E_{t_0}|$ edges, none of which are multiple, whereas the number of edges for the PG model can take other values. Technically speaking, in order for the BA model to be well-defined the seed network of t_0 nodes must have at least one edge, otherwise the $q_{t_0}(k)$ values, as calculated following the addition of node $t_0 + 1$, will all be the indeterminate quantity $0/0$. The PG model, too, is subject to constraints on the seed network in the extreme cases when $a = -1$ or $b = 0$. For example, when $a \geq -1$ and $b = 0$ the same restrictions apply as for the BA model. In the case when $a = -1$ and $b > 0$, indeterminacy will arise in the $q_{t_0}(k)$ values if, and only if, all nodes have degree 1. The final case when $a = -1$ and $b = 0$ combines the constraints from the other two cases, and it follows that the same choice of seed network as in the previous case works here, too.

A number of extensions of the BA model based upon generalizing $r(k)$ have been proposed. In particular, Krapivsky et al. (2000) analyzed a version where the preferential attachment function is not linear in the degree k of a node, but instead can be a power of the degree k^ν, $\nu > 0$. They showed that for the scale-free property to hold, $r(k)$ must be asymptotically linear in k. In a subsequent work, Krapivsky and Redner (2001) and Dorogovtsev et al. (2000) independently went on to establish that adding the offset $a > -m$ as in (2) does not violate this requirement, and derived the power-law exponent $\gamma = 3 + a/m$. Their result is analogous with our reported power-law exponent in (6) with $\lambda = m$ as seen in the next section. Furthermore, Krapivsky and Redner (2001) investigated an attachment function similar to (3) defined by $r(k) = k, k \geq 2, r(1) = b, r(0) = 0$. As they took $m \geq 1$ they did not need to be concerned with nodes of degree $k = 0$. The power-law exponent they derived in this case is reminiscent of our result in (5).

3 The degree distribution of the Poisson growth model

In this section, we discuss the degree distribution $p(k)$ for networks generated under the PG model. To make our argument concerning $p(k)$ precise, we have to take into account that G_t is generated randomly and the degree distribution of G_t also varies. Let $n_t(k)$ be the number of nodes in G_t with degree k. Since $\sum_{k=0}^{\infty} n_t(k) = t$, the observed degree distribution of G_t is defined by $p_t(k) = n_t(k)/t$ for $k \geq 0$. Our main result is as follows:

Theorem 1 *Consider the PG model with parameter $\theta = (a, b, \lambda)$ and preferential attachment function as defined in (3). For moderately large k up to $k \sim t^d$ with $d < 1/(2\gamma + 2)$, the limiting distribution $\lim_{t\to\infty} p_t(k) = p(k)$ follows a power-law*

$$p(k) \sim k^{-\gamma}, \tag{4}$$

where $a_k \sim b_k$ indicates these two sequences are proportional to each other so that a_k/b_k converges to a nonzero constant as $k \to \infty$. The power-law exponent is

$$\gamma = 3 + \frac{a + (b - a)p(0)}{\lambda}, \tag{5}$$

where $p(0)$ is the fraction of nodes with degree 0.

The exponent takes the range $\gamma > 2$; the lower limit $\gamma \to 2$ can be attained by letting $a = -1, b = 0$, and $\lambda \to 0$. This lower limit is in fact the limit for any form of $r(k)$ when λ does not depend on t; γ must be larger than 2 to ensure the mean degree $\sum_{k=0}^{\infty} kp(k) = 2\lambda$ converges. For the special case (2), the exponent becomes

$$\gamma = 3 + \frac{a}{\lambda}, \tag{6}$$

and the range is $\gamma \geq 3$.

Proof of Theorem 1 Firstly, we consider a moderately large k for the asymptotic argument as $t \to \infty$. The maximum value of k for consideration is $k \sim t^c$ for a given t with a constant $c = 1/(\gamma + 2 + \epsilon)$ with any $\epsilon > 0$. Then, the expectation of $p_t(k)$ can be expressed as

$$\mathrm{E}(p_t(k)) \sim k^{-\gamma}, \tag{7}$$

which is the power-law we would like to show for $p_t(k)$. The power-law of (7) is an immediate consequence of the recursive formula

$$(k + a - 1 + \gamma)\mathrm{E}\,(p_t(k)) = (k + a - 1)\mathrm{E}\,(p_t(k-1)) \tag{8}$$

for sufficiently large k, and thus $\mathrm{E}\,(p_t(k)) \sim (k+a-1)^{-\gamma} \sim k^{-\gamma}$; see the Appendix for a proof of (7).

The variance of $p_t(k)$ will be shown as

$$\mathrm{V}(p_t(k)) = O(k^{2+\epsilon} t^{-1}), \tag{9}$$

indicating the variance reduces by the factor $1/t$. Note that (9) is not a tight upper bound, and the variance can be much smaller. See the Appendix for the proof of (9). Let $0 < d < 1/(2\gamma + 2 + \epsilon)$, and consider $k = O(t^d)$, which is even smaller than t^c. Then,

$$\frac{\sqrt{\mathrm{V}(p_t(k))}}{\mathrm{E}(p_t(k))} = O(k^{\gamma+1+\epsilon/2} t^{-1/2}) = O(t^{\alpha}) \tag{10}$$

with $\alpha = d(\gamma+1+\epsilon/2) - 1/2 < 0$, indicating the variation of $p_t(k)$ is relatively small compared to the value of $p_t(k)$. Thus the limiting distribution $\lim_{t\to\infty} p_t(k) = p(k)$ follows the power-law of (4). By taking $\epsilon \to 0$, the power-law of $p_t(k)$ is shown up to $k \sim t^d$ with $d < 1/(2\gamma + 2)$. □

It remains to give an expression for $p(0)$ in (5). We will show in the Appendix that $p(0)$ is a solution of the quadratic equation

$$(b-a)x^2 + (2\lambda + a + \lambda b - (b-a)e^{-\lambda})x - (2\lambda + a)e^{-\lambda} = 0. \tag{11}$$

For $a \neq b$, one of the solutions

$$p(0) = \frac{1}{2(b-a)} \left[\left\{ (2\lambda + a + \lambda b - (b-a)e^{-\lambda})^2 + 4(b-a)(2\lambda + a)e^{-\lambda} \right\}^{1/2} - (2\lambda + a + \lambda b - (b-a)e^{-\lambda}) \right] \tag{12}$$

is the unique stable solution with $0 < p(0) < 1$; this can be checked by looking at the sign of $p_{t+1}(0) - p_t(0)$ in the neighborhood of $p(0)$.

4 Simulation study

A small simulation study was conducted to support our theoretical claims of Sect. 3. Specifically, we wish to confirm via simulation that the degree distribution $p(k)$ of (4) as well as its expected value $\mathrm{E}(p_t(k))$ as in (7) follow a power-law with γ as in (5). To that end we generated networks under the PG model for a variety of parameter settings. For each specification of θ we generated $n_{\text{sim}} = 10^4$ networks for three different values of the network of size N, each from a seed network of a pair of connected nodes. We included the BA model, generated under analogous conditions, so as to demonstrate the soundness of our results which are summarized in Table 1.

In point of fact, estimating γ from a network can be quite tricky and it has been the subject of some attention in the literature; see Goldstein et al. (2004). We sided with using the maximum likelihood (ML) approach described by Newman (2005) for a continuous power-law distribution as an approximation to the discrete case. In this methodology, the ML estimator of γ is given by

$$\hat{\gamma} = 1 + \left(\sum_{k \geq k_{\min}} n(k) \right) \cdot \left(\sum_{k \geq k_{\min}} n(k) \log \frac{k}{k_{\min}} \right)^{-1}$$

where $n(k)$ is the number of nodes with degree k, and $k_{\min}$ is the minimum degree after which the power-law behavior holds. Bauke (2007) studied selecting a value for $k_{\min}$

Table 1 Summary of estimated power-law exponents from simulated networks

Model	Parameters	N	$k_{\min}$	Mean k	Mean $\hat{\gamma} \pm$ SD	$\hat{\gamma}_{\text{avg}}$	γ
BA	$m = 1$	2,500	8	2.0	3.04 ± 0.18	3.04	3
BA	$m = 1$	5,000	10	2.0	3.03 ± 0.15	3.03	3
BA	$m = 1$	10,000	15	2.0	3.01 ± 0.16	3.00	3
PG	$\theta = (0, 0, 1)$	2,500	8	2.0	3.06 ± 0.14	3.04	3
PG	$\theta = (0, 0, 1)$	5,000	10	2.0	3.03 ± 0.12	3.03	3
PG	$\theta = (0, 0, 1)$	10,000	15	2.0	3.01 ± 0.12	3.01	3
PG	$\theta = (-0.95, 0.05, 0.5)$	2,500	8	1.0	2.33 ± 0.17	2.32	2.23
PG	$\theta = (-0.95, 0.05, 0.5)$	5,000	10	1.0	2.28 ± 0.13	2.32	2.23
PG	$\theta = (-0.95, 0.05, 0.5)$	10,000	15	1.0	2.30 ± 0.12	2.30	2.23
PG	$\theta = (-0.9, 0.1, 1)$	2,500	8	2.0	2.70 ± 0.10	2.70	2.44
PG	$\theta = (-0.9, 0.1, 1)$	5,000	10	2.0	2.54 ± 0.10	2.51	2.44
PG	$\theta = (-0.9, 0.1, 1)$	10,000	15	2.0	2.50 ± 0.09	2.50	2.44
PG	$\theta = (-0.9, 0.1, 3)$	2,500	8	6.0	2.89 ± 0.06	2.89	2.72
PG	$\theta = (-0.9, 0.1, 3)$	5,000	10	6.0	2.86 ± 0.05	2.86	2.72
PG	$\theta = (-0.9, 0.1, 3)$	10,000	15	6.0	2.80 ± 0.05	2.80	2.72
PG	$\theta = (0.5, 0.5, 3)$	2,500	8	6.0	3.13 ± 0.06	3.13	3.17
PG	$\theta = (0.5, 0.5, 3)$	5,000	10	6.0	3.15 ± 0.05	3.15	3.17
PG	$\theta = (0.5, 0.5, 3)$	10,000	15	6.0	3.15 ± 0.06	3.15	3.17

The last column is the theoretically predicted γ

by using a χ^2 goodness of fit test over a range of $k_{\min}$; however, we shied away from this level of scrutiny as we found that taking $k_{\min}$ values as shown in Table 1 was reasonable for our examples. We decided on these values based upon visual inspections of the plots of $\hat{\gamma}$ over ranges of $k_{\min}$. Such plots typically exhibit a broad plateau of $\hat{\gamma}$ values for moderately sized $k_{\min}$, and we took our estimates based on this observation. This methodology is illustrated in Fig. 1a and b where we plot the degree distribution with $\hat{\gamma}$ for a typical network generated by the BA and PG model, respectively.

Returning to Table 1, in each case, we confirm (4) has power-law exponent as predicted by (5) and (12). We computed $\hat{\gamma}$ for each network, and calculated the mean and standard deviation of $\hat{\gamma}$ values for $n_\sim$ networks. We observe that the mean $\hat{\gamma}$ agrees well with the predicted γ, and the variation of $\hat{\gamma}$ is relatively small as suggested by (10).

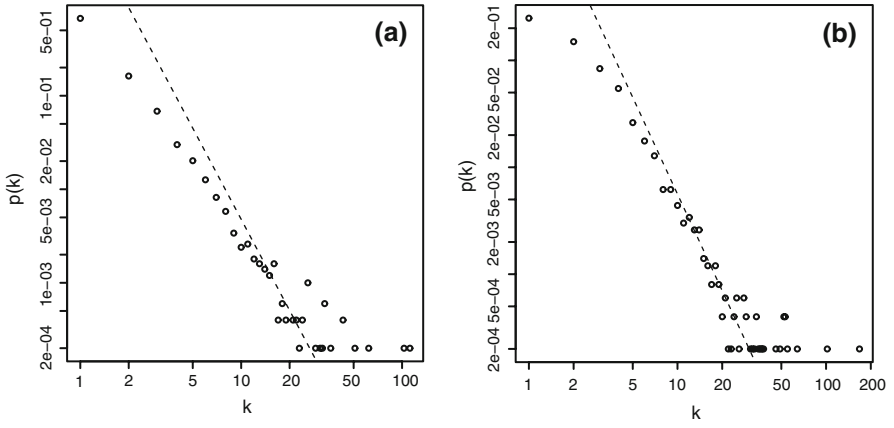

Fig. 1 Degree distribution $p(k)$ of a typical network with $N = 5{,}000$ plotted on a log–log scale with the power-law line using estimated exponent $\hat{\gamma}$. **a** Generated under the BA model; $\hat{\gamma} = 3.03$. **b** Generated under the PG model with $\theta = (0, 0, 1)$; $\hat{\gamma} = 3.01$

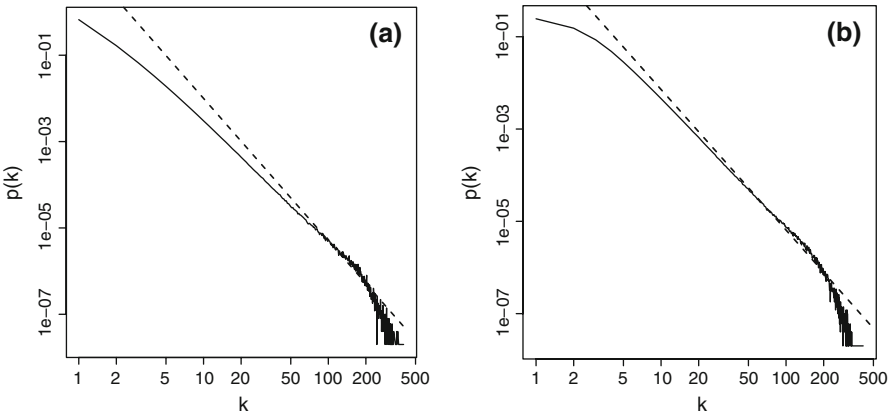

Fig. 2 Average degree distribution $\mathrm{E}(p(k))$ of the simulation with the power-law line using estimated exponent $\hat{\gamma}_{\text{avg}}$ in the case when $N = 5{,}000$. Plotted for **a** the BA model and for **b** the PG model with $\theta = (0, 0, 1)$, where $\hat{\gamma}_{\text{avg}} = 3.03$ for both cases

In addition, to show that the same holds for (7), in each case we computed the average degree distribution of the n_{sim} networks as an estimate of $\mathrm{E}(p(k))$. Then we estimated the degree exponent $\hat{\gamma}_{\text{avg}}$ as seen in the table and Fig. 2. Again, the simulated results match well with theory.

5 Discussion

The PG model has a special place in the class of preferential attachment models. It has a tunable power-law exponent and a simple implementation, yet it can generate any network. In contrast, the BA model and its generalizations described in Sect. 2 have serious restrictions on the types of networks that can be generated because m is held constant. For example, at step t an instantiation of the BA model will consist of a t node network with the number of edges equal to exactly $m(t - t_0)$, plus the number of edges in the seed network. The simple design of our model makes computing the probability of a network straightforward. This in combination with its modeling potential gives rise to several useful applications in Bayesian inference.

In explicit terms, let $G = (V, E)$ be a network with $N = |V|$ nodes where $V = \{v'_1, \ldots, v'_N\}$. Furthermore, let $G_N = (V_N, E_N)$ be a network generated under PG model after step $N - 1$ so that $V_N = \{v_1, \ldots, v_N\}$, where the seed network consists of a single node. The association between V and V_N is defined by a permutation $\sigma = (\sigma_1, \ldots, \sigma_N)$ so that $v_i = v'_{\sigma_i}$. Given G, once we specify σ, then it is straightforward to compute $k_{i,t}, s_{i,t}$ for $i = 1, \ldots, t; t = 1, \ldots, N - 1$. Then the probability of G given $\theta = (a, b, \lambda)$ and σ is

$$\mathrm{P}(G|\theta, \sigma) = \prod_{t=1}^{N-1} \left(\prod_{i=1}^{t} e^{-\lambda q_t(k_{i,t})} \frac{(\lambda q_t(k_{i,t}))^{s_{i,t}}}{s_{i,t}!} \right).$$

One application is when G is known and we wish to estimate θ. This can be done by assigning a prior $\pi(\theta)$ for θ and the uniform prior on σ. The posterior probability of (θ, σ) given G is

$$\pi(\theta, \sigma|G) \propto \mathrm{P}(G|\theta, \sigma)\pi(\theta).$$

Using MCMC to produce a chain of values for (θ, σ), the posterior $\pi(\theta|G)$ is simply obtained from the histogram of θ in the chain. Moreover, this procedure can be used for model comparison, if we have several models for generating the network.

Another application is when we wish to make inference about G from data D with likelihood function $P(D|G)$. The posterior probability of (G, θ, σ) given D is

$$\pi(G, \theta, \sigma|D) \propto P(D|G)\mathrm{P}(G|\theta, \sigma)\pi(\theta).$$

Then the posterior $\pi(G|D)$ is simply obtained from the frequency of G in the chain. Indeed, we used this approach for inferring a gene network from microarray data in Sheridan et al. (2007).

Recall that the PG model can produce networks with multiple edges, but, in practice, we often wish to restrict our attention to networks without this feature. As an approximation, we could apply the formula for $P(G|\theta, \sigma)$ just as well in this case. Though

the number of multiple edges will be quite small in a network generated from a λ of reasonable size so that their presence may have little effect on the overall inference.

Alternatively, we propose a slight modification to our model where we generate m_t edges at step t according to a binomial distribution with parameter $p = \lambda/t$ and sample size t. In this formulation the seed network must be selected such that $\lambda \leq t_0$, otherwise $p > 1$ may occur. Then by sampling nodes without replacement, multiple edges are avoided. In our simulation (results not included) we found that these modifications do not change the power-law, though an airtight derivation of this is more complicated than for the PG model itself. In fact, entertaining multiple edges in a scale-free model as a means of facilitating mathematical arguments about the underlying power-law has precedent in Bollobás et al. (2001). They extended the BA model by allowing for multiple edges—and loops—and provided a mathematically rigorous argument establishing that the power-law exponent γ is 3.

Finally, though we made specific choices for $r(k)$ in our arguments, the PG model can be generalized to a wider class of preferential attachment functions. For instance, Dorogovtsev and Mendes (2001) investigated *accelerated growth* models where m_t increases as the network grows. It should be possible to incorporate accelerated growth into PG model by gradually increasing the value of λ over time. Another line of generalizations of the PG model is via the incorporation of dynamic edge rewiring.

Appendix: Proofs

The expected value of $p_t(k)$

Here we give the proof of (7). We assume that the functional form of $r(k)$ is (2), and a modification to handle (3) is mentioned at the bottom.

Let $I(A)$ denote the indicator function of the event A, so $I(A) = 1$ if A is true and $I(A) = 0$ if A is false. We use the notation $\mathrm{P}(\cdot)$, $\mathrm{E}(\cdot)$ and $\mathrm{V}(\cdot)$ for the probability, expectation and the variance, and also $\mathrm{P}(\cdot|A)$, $\mathrm{E}(\cdot|A)$ and $\mathrm{V}(\cdot|A)$ for those given a condition A. By noting

$$n_{t+1}(k) = \sum_{i=1}^{t} I(k_{i,t} + s_{i,t} = k) + I(m_t = k),$$

the conditional expectation of $n_{t+1}(k)$ given G_t is

$$\begin{aligned}
\mathrm{E}(n_{t+1}(k)|G_t) &= \sum_{i=1}^{t} \mathrm{P}(k_{i,t} + s_{i,t} = k|G_t) + \mathrm{P}(m_t = k|G_t) \\
&= \sum_{i=1}^{t} e^{-\lambda q_t(k_{i,t})} \frac{(\lambda q_t(k_{i,t}))^{k-k_{i,t}}}{(k-k_{i,t})!} + e^{-\lambda}\frac{\lambda^k}{k!} \\
&= \sum_{s=0}^{k} n_t(k-s) e^{-\lambda q_t(k-s)} \frac{(\lambda q_t(k-s))^s}{s!} + e^{-\lambda}\frac{\lambda^k}{k!}. \qquad (13)
\end{aligned}$$

The last term $e^{-\lambda}\lambda^k/k! \sim (e\lambda/k)^k$ can be ignored for a large k, since it is exponentially small as k grows. We examine the terms in the summation over $s = 0, 1, \ldots, k$ for $k = O(t^c)$ as $t \to \infty$. For a fixed s, $q_t(k-s) \sim k/t$ for a linear preferential attachment model. More specifically, for $r(k) = k + a$, $k \geq 0$,

$$q_t(k-s) = \frac{r(k-s)}{\sum_{i=1}^{t} r(k_{i,t})} = \frac{k-s+a}{t(2\lambda+a)}(1 + O(t^{-1/2})),$$

because the mean degree of G_t is

$$\frac{1}{t}\sum_{i=1}^{t} k_{i,t} = \frac{2}{t}\left(|E_{t_0}| + \sum_{t'=t_0}^{t-1} m_{t'}\right) = 2\lambda + O(t^{-1/2}),$$

and the denominator of $q_t(k)$ is

$$\sum_{i=1}^{t} r(k_{i,t}) = \sum_{i=1}^{t}(k_{i,t} + a) = t(2\lambda + a + O(t^{-1/2})). \tag{14}$$

Thus the sum in (13) over $s = 0, 1$ becomes

$$n_t(k)\left(1 - \frac{\lambda(k+a)}{(2\lambda+a)t} + O(kt^{-3/2})\right) + n_t(k-1)\left(\frac{\lambda(k+a-1)}{(2\lambda+a)t} + O(kt^{-3/2})\right).$$

For $s \geq 2$, each term is $\sim n_t(k-s)(k/t)^s$. By noting $\sum_{s=2}^{k} n_t(k-s) \leq t$, the sum over $s = 2, \ldots, k$ becomes $O(k^2 t^{-1})$.

Next, we take the expectation of (13) with respect to G_t to obtain the unconditional expectation $\mathrm{E}(n_{t+1}(k))$, and replace $n_t(k) = tp_t(k)$. Using the results of the previous paragraph, we get

$$\begin{aligned}\mathrm{E}(p_{t+1}(k)) = \mathrm{E}(p_t(k)) - \frac{\lambda}{(2\lambda+a)t}\Big((k' + \gamma + O(kt^{-1/2}))\mathrm{E}(p_t(k)) \\ -(k' + O(kt^{-1/2}))\mathrm{E}(p_t(k-1)) + O(k^2 t^{-1})\Big)\end{aligned} \tag{15}$$

with $k' = k + a - 1$ and the γ of (6). Let us assume $\mathrm{E}(p_t(k-1)) \sim (k-1)^{-\gamma}$, and remember $c < 1/(\gamma+2)$. By taking the limit $t \to \infty$ and equating $\mathrm{E}(p_{t+1}(k)) = \mathrm{E}(p_t(k))$, we get

$$(k' + \gamma + o(1))\mathrm{E}(p_t(k)) = (k' + o(1))\mathrm{E}(p_t(k-1)).$$

So that, for sufficiently large t,

$$\mathrm{E}(p_t(k)) \sim k^{-\gamma}$$

also holds for k. Since $\mathrm{E}(p_t(k)) = O(1)$ for a fixed k, the power-law holds for any k by induction up to $k \sim t^c$.

For $r(k)$ of (3), the preferential attachment is modified to

$$r(k) = k + a + (b - a)I(k = 0), \quad k \geq 0.$$

This changes the the denominator of $q_t(k)$ in (14) to

$$\sum_{i=1}^{t} r(k_{i,t}) = t\left(2\lambda + a + (b - a)p_t(0) + O(t^{-1/2})\right), \tag{16}$$

and thus $2\lambda + a$ in the updating formula (15) is replaced with $2\lambda + a + (b - a)p(0)$, leading to (5). Note that $p_t(0) = p(0) + O(t^{-1/2})$ from (9) shown in the next section.

The variance of $p_t(k)$

Here we give the proof of (9) by working on $\mathrm{V}(n_t(k)) = t^2\mathrm{V}(p_t(k))$. Although $r(k)$ of (2) is again assumed, the argument is basically the same for (3). By noting the identity

$$\mathrm{V}(n_{t+1}(k)) = \mathrm{E}(\mathrm{V}(n_{t+1}(k)|G_t)) + \mathrm{V}(\mathrm{E}(n_{t+1}(k)|G_t)), \tag{17}$$

we evaluate the two terms on the right hand side.

The conditional variance of $n_{t+1}(k)$ given G_t is evaluated rather similarly as the conditional expectation of (13). By noting $\mathrm{V}(I(A)) = \mathrm{P}(A) - \mathrm{P}(A)^2$, $\mathrm{V}(n_{t+1}(k)|G_t)$ is expressed for $k = O(t^c)$ as

$$\sum_{s=0}^{k} n_t(k - s)\left\{e^{-\lambda q_t(k-s)}\frac{(\lambda q_t(k - s))^s}{s!} - \left(e^{-\lambda q_t(k-s)}\frac{(\lambda q_t(k - s))^s}{s!}\right)^2\right\}$$

$$\approx n_t(k)\frac{\lambda(k + a)}{(2\lambda + a)t} + n_t(k - 1)\frac{\lambda(k + a - 1)}{(2\lambda + a)t}, \tag{18}$$

where terms from $I(m_t = k)$ are ignored for a large k. Thus, the first term in (17) is

$$\mathrm{E}(\mathrm{V}(n_{t+1}(k)|G_t)) = O(k^{-\gamma+1}).$$

On the other hand, the second term in (17) is evaluated by considering the variance of (13) as

$$\begin{aligned}\mathrm{V}(\mathrm{E}(n_{t+1}(k)|G_t)) \le \mathrm{V}(n_t(k))\left(1-\frac{2\lambda(k+a)}{(2\lambda+a)t}+O(kt^{-3/2})\right)\\ +2\sqrt{\mathrm{V}(n_t(k))}\sqrt{\mathrm{V}(n_t(k-1))}\left(\frac{\lambda(k+a-1)}{(2\lambda+a)t}+O(kt^{-3/2})\right)\\ +\mathrm{V}(n_t(k-1))O(k^2t^{-2})+\sqrt{\mathrm{V}(n_t(k))}O(k^2t^{-1})+O(k^4t^{-2}).\end{aligned}$$

We substitute these two expressions for those in (17). We will show, by induction, that

$$\mathrm{V}(n_t(k)) < Ak^{2+\epsilon}t \tag{19}$$

holds for all (t,k) with $k = O(t^c)$ using some constant A. Let us assume that (19) holds for (t,k) and $(t,k-1)$. By taking a sufficiently large A, we have

$$\mathrm{V}(n_{t+1}(k)) \le Ak^{2+\epsilon}(t-(2\lambda+a)^{-1})+o(k^{1+\epsilon/2}) < Ak^{2+\epsilon}(t+1), \tag{20}$$

implying that (19) also holds for $(t+1,k)$.

On the other hand, for any random variable $0 \le n \le t$ with its expectation $\mathrm{E}(n)$ fixed, the largest possible variance $O(t)\mathrm{E}(n)$ is attained if the probability concentrates on the extreme values 0 and t. Applying this upper bound to $n_t(k)$ with $k \sim t^c$, we obtain $\mathrm{V}(n_t(k))/t = O(\mathrm{E}(n_t(k))) = O(k^{-\gamma}t) = O(k^{2+\epsilon})$, implying that (19) holds for any (t,k) with $k \sim t^c$.

For induction with respect to k, we only have to show

$$\mathrm{V}(n_t(k)) < v(k)t \tag{21}$$

for a sufficiently large k so that terms from $I(m_t = k)$ in (18) can be ignored. $v(k)$ is an arbitrary constant depending on k. We start from $k = 0$. First note that

$$n_{t+1}(0) = \sum_{i=1}^{t} I(k_{i,t} = 0 \cap s_{i,t} = 0) + I(m_t = 0).$$

Thus $\mathrm{E}(n_{t+1}(0)|G_t) = n_t(0)e^{-\lambda q_t(0)} + e^{-\lambda}$, and so

$$\mathrm{V}(\mathrm{E}(n_{t+1}(0)|G_t)) = \mathrm{V}(n_t(0))\left(1-\frac{2\lambda a}{(2\lambda+a)t}+O(t^{-3/2})\right).$$

On the other hand, $\mathrm{V}(n_{t+1}(0)|G_t)$ is expressed as

$$n_t(0)(e^{-\lambda q_t(0)} - e^{-2\lambda q_t(0)}) + e^{-\lambda} - e^{-2\lambda} + 2n_t(0)(1-e^{-\lambda q_t(0)})e^{-\lambda}.$$

By substituting these two expressions for those in (17), we observe that the increase of the variance, i.e., $\mathrm{V}(n_{t+1}(0)) - \mathrm{V}(n_t(0))$ is bounded by a constant, and we have $\mathrm{V}(n_t(0)) = O(t)$.

Let us assume (21) holds up to $k-1$. Then $\mathrm{V}(n_{t+1}(k))$ can be expressed quite similarly as (20), but $\mathrm{E}(\mathrm{V}(n_{t+1}(k)|G_t))$ includes additional terms from $I(m_t = k)$; $\mathrm{V}(I(m_t = k)) = O(1)$ and $\mathrm{E}(\sum_{i=1}^{t} \mathrm{Cov}(I(k_{i,t} + s_{i,t} = k), I(m_t = k)|G_t))$. For $k_{i,t} = k$, the covariance term $\leq \mathrm{P}(m_t = k)(1 - \mathrm{P}(s_{i,t} = 0|G_t)) = O(t^{-1})$, and for $k_{i,t} = k - s$ with $s \geq 1$, the covariance term $\leq \mathrm{P}(s_{i,t} = s)(1 - \mathrm{P}(m_t = k)) = O(t^{-s})$. Thus, by taking the sum over $i = 1, \ldots, t$, it becomes $O(t \cdot t^{-1}) = O(1)$. Therefore, $\mathrm{V}(n_{t+1}(k)) - \mathrm{V}(n_t(k))$ is bounded by a constant, and (21) holds for k. By induction, (21) holds for any k.

The equation of $p(0)$

Here we derive (11) for the $r(x)$ of (3). By taking the expectation of $\mathrm{E}(n_{t+1}(0)|G_t) = n_t(0)e^{-\lambda q_t(0)} + e^{-\lambda}$ with respect to G_t, and using (16), we get

$$\mathrm{E}(n_{t+1}(0)) = \mathrm{E}(n_t(0)) \left(1 - \frac{\lambda b}{(2\lambda + a + (b-a)p(0))t} + O(t^{-3/2})\right) + e^{-\lambda}.$$

By substituting $n_t(0) = t p_t(0)$ and taking the limit $t \to \infty$, we get a formula for $f(x) = (t+1)(\mathrm{E}(p_{t+1}(0)) - \mathrm{E}(p_t(0)))$ as a function of $x = p(0)$

$$f(x) = -x \left(1 + \frac{\lambda b}{2\lambda + a + (b-a)x}\right) + e^{-\lambda}.$$

The quadratic equation (11) is obtained by letting $f(x) = 0$. In addition, the condition $df(x)/dx < 0$ was checked for the stable solution.

References

Albert, R., Jeong, H., Barabási, A.-L. (1999). Diameter of the world-wide web. *Nature*, *401*, 130–131.

Albert, R., Barabási, A.-L. (2000). Topology of evolving networks: local events and universality. *Physical Review Letters*, *85*, 5234–5237.

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

Bauke, H. (2007). Parameter estimation for power-law distributions by maximum likelihood methods. *The European Physical Journal B Condensed Matter and Complex Systems*, *58*(2), 167–173.

Bianconi, G., Barabási, A.-L. (2001). Competition and multiscaling in evolving networks. *Europhysics Letters*, *54*(4), 436–442.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U. (2006). Complex networks: structure and dynamics. *Physics Reports 4-5*, *424*, 175–308.

Bollobás B., Riordan, O., Spencer, J., Tusanády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, *18*, 279–290.

Dorogovtsev, S.N., Mendes, J.F.F., Samukhin, A.N. (2000). Structure of growing networks with preferential linking. *Physical Review Letters*, *85*, 4633–4636.

Dorogovtsev , S. N., Mendes, J. F. F. (2001). Effect of accelerated growth of communications networks on their structure. *Physical Review E*, *63*, 025101.

Erdös, P., Rényi, A. (1959). On random graphs I. *Publicationes Mathematicae*, *6*, 290–297.

Goldstein, M. L., Morris, S. A., Yen, G. G. (2004). Problems with fitting to the power-law distribution. *The European Physics Journal B*, *41*, 255–258.

Jeong, H., Mason, S., Barabási, A.-L., Oltvai, Z.N. (2001). Lethality and centrality in protein networks. *Nature*, *411*, 41–42.

Krapivsky, P. L., Redner, S., Leyvraz, F. (2000). Connectivity of growing random networks. *Physical Review Letters*, *85*, 4629–4632.

Krapivsky, P. L., Redner, S. (2001). Organization of growing random networks. *Physical Review E*, *63*, 066123.

Lee, D. S., Goh, K. I., Kahng, B., Kim, D. (2005). Scale-free random graphs and Potts model. *Pramana Journal of Physics*, *64*, 1149–1159.

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323–351.

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *The European Physics Journal B*, *4*, 131–134.

Sheridan, P., Kamimura, T., Shimodaira, H. (2007). Scale-free networks in Bayesian inference with applications to bioinformatics. *Proceedings of The International Workshop on Data-Mining and Statistical Science* (*DMSS2007*), 1–16, Tokyo.

Solé, R. V., Pastor-Satorras, R., Smith, E., Kepler, T. B. (2002). A model of large-scale proteome evolution. *Advances in Complex Systems*, *5*, 43–54.

Strogatz, S. H. (2001). Exploring complex networks. *Nature*, *410*, 268–276.

Watts, D. J., Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, *393*, 440–442.