Preferential and random attachment models of a complex network

Paul Secular*
Imperial College London
(Dated: 8th July 2015)

Analytic and numerical results are presented for three models of a complex network exhibiting growth via attachment. The first uses pure preferential attachment (the Barabási-Albert model), the second uses pure random attachment, and the third uses a mixture of both preferential and random attachment. System sizes of 10^n for $n = 1$ to 6 were simulated and results found to be consistent with the analytic asymptotic degree distributions, with cut-offs occurring due to finite system size. The shape of the cut-offs was found to depend on the networks' initial conditions. Numerically, the largest degree in a finite Barabási-Albert network was found to scale with system size like a power law over 5 orders of magnitude. Linear regression on a log-log plot gave an exponent of 0.51 ± 0.01, which is in good agreement with the theoretical prediction of asymptotic power law behaviour with exponent $\frac{1}{2}$.

* paul@secular.me.uk

I. INTRODUCTION

The well-known Barabási-Albert (BA) model describes a complex graph using growth and linear preferential attachment. However, its original description in [1] is actually quite vague, leaving details of implementation open to interpretation. Two variants are considered here: the first being a simple graph and the second an undirected multigraph in which nodes may be connected by more than one edge, but self-loops are not permitted. A third variation, not covered by the present work, is described in [2]. Two other models exhibiting growth are also investigated: one with pure random attachment, and one with a mixture of both preferential and random attachment. The asymptotic degree distribution of each model is derived analytically and compared to results of numerical simulations.

II. ANALYTIC EXPRESSIONS

For each model, we begin by deriving the degree distribution in the limit of infinitely large system size. In all cases we start from the so-called "master equation" given in [3], which describes the growth of the network due to one new node with m edges being added at each time step:

$$n(k, t + 1) = n(k, t) + m\Pi(k - 1, t)n(k - 1, t) - m\Pi(k, t)n(k, t) + \delta_{k,m}. \tag{1}$$

This is a difference equation for $n(k, t)$, the ensemble average over all networks of the number of nodes at time t with a total degree k. Π is the probability that a single edge from the new node is chosen to be attached to a particular node of degree k. The first thing to note about this equation is that it only handles the case where one edge is attached to a particular node at each time step, since modelling multiple edges would require terms in $n(k-2, t)$, $n(k-3, t)$, etc. This master equation thus appears incompatible with a multigraph interpretation of the BA model for $m > 1$. However, if we conjecture (as Dorogovtsev et al. do in [2]) that the number of multiple edges added per timestep becomes vanishingly small as $N \to \infty$, then we can neglect such terms and approximate the growth of the network by (1). Numerical results suggest that this is indeed an excellent approximation: in a BA network with $m = 3$ and $N = 10^6$, only $\sim 0.01\%$ of the edges created were duplicates.

By defining the time-dependent degree probability distribution as

$$p(k, t) = \frac{n(k, t)}{N(t)}, \tag{2}$$

we can write (1) in the form:

$$p(k, t + 1) = m\Pi(k - 1, t)p(k - 1, t)N(t) - m\Pi(k, t)p(k, t)N(t) + \delta_{k,m}, \tag{3}$$

where we have also made use of the fact that $N(t + 1) = N(t) + 1$. We then assume an asymptotic stationary solution for $p(k, t)$,

$$p_\infty(k) = \lim_{t\to\infty} p(k, t), \tag{4}$$

and thus write:

$$p_\infty(k) = m\Pi(k - 1, t)p_\infty(k - 1)N(t) - m\Pi(k, t)p_\infty(k)N(t) + \delta_{k,m}. \tag{5}$$

The solution of this difference equation depends on the form of Π. The solution in the three cases of interest are derived below, making use of the following two lemmas:

Lemma 1. *The solution of the difference equation*

$$\frac{f(k)}{f(k-1)} = \frac{k+a}{k+b}, \qquad k \in \mathbb{N} \mid k \neq -b, \tag{6}$$

is

$$f(k) = A\frac{\Gamma(k+1+a)}{\Gamma(k+1+b)}. \tag{7}$$

Proof. Using the definition of the gamma function, we can write

$$k + c = \frac{\Gamma(k+1+c)}{\Gamma(k+c)}, \tag{8}$$

hence,

$$\frac{f(k)}{f(k-1)} = \frac{\Gamma(k+b)\Gamma(k+1+a)}{\Gamma(k+a)\Gamma(k+1+b)}. \tag{9}$$

We then define

$$g(k) = \frac{\Gamma(k+1+a)}{\Gamma(k+1+b)}, \tag{10}$$

so that

$$\frac{f(k)}{g(k)} = \frac{f(k-1)}{g(k-1)}. \tag{11}$$

If we choose an initial condition $f(1)$, we have by induction that

$$\frac{f(k)}{g(k)} = \frac{f(1)}{g(1)}, \qquad k \in \mathbb{N}. \tag{12}$$

The general solution is thus

$$f(k) = Ag(k), \tag{13}$$

i.e.,

$$f(k) = A\frac{\Gamma(k+1+a)}{\Gamma(k+1+b)},$$

where A is an arbitrary constant. □

Lemma 2. *Irrespective of initial network configuration, the ratio* $\frac{E(t)}{N(t)} \to m$ *as* $t \to \infty$.

Proof. Our network models require an initial configuration with a minimum number of nodes $N(0)$ and edges $E(0)$ in order for them to evolve. At any future time, we therefore have

$$E(t) = E(0) + mt, \tag{14}$$
$$N(t) = N(0) + t. \tag{15}$$

In the limit of large N:

$$\begin{aligned} \lim_{t\to\infty} \left\{ \frac{E(t)}{N(t)} \right\} &= \lim_{t\to\infty} \left\{ \frac{E(0) + mt}{N(0) + t} \right\} \\ &= \lim_{t\to\infty} \left\{ \frac{E(0)}{N(0) + t} + \frac{mt}{N(0) + t} \right\} \\ &= \lim_{t\to\infty} \left\{ \frac{E(0)}{N(0) + t} \right\} + \lim_{t\to\infty} \left\{ \frac{mt}{N(0) + t} \right\} \\ &= \lim_{t\to\infty} \left\{ \frac{m}{N(0)/t + 1} \right\} \\ &= m. \end{aligned} \tag{16}$$

□

In a finite network, however we have:

$$\lim_{t\to N} \left\{ \frac{E(t)}{N(t)} \right\} = \frac{E(0) + m(N - N(0))}{N} = m + \frac{E(0)}{N} + \frac{mN(0)}{N}. \tag{17}$$

This means we would expect the initial conditions to become unimportant when

$$E(0) \ll N \tag{18}$$

and

$$mN(0) \ll N. \tag{19}$$

The minimum number of initial nodes required to grow a multigraph is 2 (1 if using pure random attachment), whereas for a simple graph it is m if $m > 1$ or 2 if $m = 1$. In general, the initial number of edges required is also non-zero (unless pure random attachment is employed): for a multigraph, this number is 1, whereas for a simple graph it is $\lceil \frac{m}{2} \rceil$. For (16) to hold as a good approximation, we therefore require that

$$N \gg m. \tag{20}$$

Furthermore, for a simple graph using preferential or mixed attachment, we need

$$N \gg m^2 > 1. \tag{21}$$

Proposition 1. *For any model exhibiting growth, defined by the addition of one new node with degree* m *per timestep,* $p_\infty(k) = 0$ *for* $k < m$.

Proof. Since new nodes are only added with degree m, the maximum number of nodes with degree less than m must be a maximum at $t = 0$, i.e.,

$$n(k,t) \leq n(k,0), \qquad k < m. \tag{22}$$

Thus, from (2), we have

$$p(k,t) \le \frac{n(k,0)}{N(t)}. \tag{23}$$

Letting $N \to \infty$ gives

$$p_\infty(k) = 0, \qquad k < m. \tag{24}$$

□

A. Pure preferential attachment

Since we are considering linear attachment, it is obvious (because each edge is attached to two nodes) that for $m = 1$,

$$\Pi(k,t) = \frac{k}{2E(t)}. \tag{25}$$

However, when considering simple graphs with $m > 1$, the total probability of a particular node being chosen in a single timestep is not equal to $m\Pi(k,t)$. Using probability theory and a little algebra, one can show that the appropriate expression when $m = 2$ is in fact:

$$\frac{k}{2E(t)}\left(1 + \sum_{j=1}^{E(t)} \frac{n(j,t)j}{2E(t)-j} - \frac{k}{2E(t)-k}\right). \tag{26}$$

This can be generalised to all m, but rather than try to work with such cumbersome expressions, we conjecture that the probability approaches $m\Pi(k,t)$ in the large N limit for all m. Fortunately, our numerical results confirm that this is a good approximation.

Proposition 2. *The asymptotic degree probability distribution for the BA model is given by*

$$p_\infty(k) = \frac{2m(m+1)}{(k+2)(k+1)k}, \qquad k \ge m. \tag{27}$$

Proof. To derive $p_\infty(k)$, we substitute (25) and (16) into (5) giving:

$$\begin{aligned} p_\infty(k) &= m\frac{k-1}{2E(t)}p_\infty(k-1)\frac{E(t)}{m} - m\frac{k}{2E(t)}p_\infty(k)\frac{E(t)}{m} + \delta_{k,m} \\ &= \frac{k-1}{2}p_\infty(k-1) - \frac{k}{2}p_\infty(k) + \delta_{k,m}. \end{aligned} \tag{28}$$

We first consider the case $m \ne k$, such that $\delta_{k,m} = 0$. Rearranging gives

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k-1}{k+2}, \tag{29}$$

which by lemma 1 implies

$$\begin{aligned} p_\infty(k) &= A\frac{\Gamma(k)}{\Gamma(k+3)} \\ &= \frac{A}{(k+2)(k+1)k}. \end{aligned} \tag{30}$$

We next consider the case $m = k$, such that $\delta_{k,m} = 1$. Substituting into (28) and rearranging gives

$$(m + 2)p_\infty(m) = (m - 1)p_\infty(m - 1) + 2. \tag{31}$$

But by (24), $p_\infty(m - 1) = 0$, so

$$p_\infty(m) = \frac{2}{(m + 2)}. \tag{32}$$

Finally, we impose the probability normalisation condition on $p_\infty(k)$ in order to derive the constant A:

$$\begin{aligned}
1 &= \sum_{k=1}^{\infty} p_\infty(k) \\
&= \frac{2}{(m+2)} + \sum_{m+1}^{\infty} \frac{A}{(k+2)(k+1)k} \\
&= \frac{2}{(m+2)} + \frac{A}{2}\sum_{m+1}^{\infty}\left[\frac{1}{(k+2)} - \frac{2}{(k+1)} + \frac{1}{k}\right] \\
&= \frac{2}{(m+2)} + \frac{A}{2}\sum_{m+1}^{\infty}\left[\frac{1}{(k+2)} - \frac{1}{(k+1)}\right] - \frac{A}{2}\sum_{m+1}^{\infty}\left[\frac{1}{(k+1)} - \frac{1}{k}\right] \\
&= \frac{2}{(m+2)} + \frac{A}{2(m+1)} - \frac{A}{2(m+2)}.
\end{aligned} \tag{33}$$

Hence,

$$A = 2m(m + 1), \tag{34}$$

and so

$$p_\infty(k) = \frac{2m(m+1)}{k(k+1)(k+2)}, \qquad k \geq m.$$

In the limit of large k, we therefore expect power law behaviour: $p_\infty(k) \sim k^{-3}$. □

Proposition 3. *The expected cut-off degree for the BA model is given by*

$$k_1 = \frac{-1 + \sqrt{1 + 4Nm(m+1)}}{2}. \tag{35}$$

Proof. The distribution cut-off k_1 can be defined as the value of k after which only one node is expected to be found on average [3]. This means

$$\sum_{k=k_1}^{\infty} N p_\infty(k) = 1. \tag{36}$$

Using the partial fraction decomposition of p_∞ from (33) we have

$$\frac{2m(m+1)}{2}\sum_{k=k1}^{\infty}\left[\frac{1}{(k+2)} - \frac{1}{(k+1)}\right] - \frac{2m(m+1)}{2}\sum_{k=k1}^{\infty}\left[\frac{1}{(k+1)} - \frac{1}{k}\right] = \frac{1}{N}, \tag{37}$$

which after a little algebra gives

$$\frac{m(m+1)}{k_1(k_1+1)} = \frac{1}{N}. \tag{38}$$

Rearranging shows this is a quadratic equation for k_1:

$$k_1^2 + k_1 - Nm(m+1) = 0. \tag{39}$$

There are two solutions, but only the positive one is physical:

$$k_1 = \frac{-1 + \sqrt{1 + 4Nm(m+1)}}{2}$$

In the limit of large N we therefore expect $k_1 \sim N^{1/2}$. □

B. Pure random attachment

Proposition 4. *The asymptotic degree probability distribution function for the random attachment model is:*

$$p_\infty(k) = \frac{m^{k-m}}{(1+m)^{1+k-m}}, \qquad k \geq m. \tag{40}$$

Proof. For $m = 1$, the probability is defined as

$$\Pi(k,t) = \frac{1}{N(t)}. \tag{41}$$

It is easy to see that this becomes an increasingly good approximation for any m as $N \to \infty$. Substituting into (5) and rearranging gives

$$p_\infty(k)(1+m) = mp_\infty(k-1) + \delta_{k,m}. \tag{42}$$

When $k > m$,

$$p_\infty(k) = \frac{m}{(m+1)} p_\infty(k-1), \tag{43}$$

so we have by induction that

$$p_\infty(k) = \left[\frac{m}{(m+1)}\right]^{k-m} p_\infty(m). \tag{44}$$

Since $p_\infty(k) = 0$ for $k < m$, it is easy to see that for $k = m$,

$$p_\infty(m) = \frac{1}{(1+m)}, \tag{45}$$

and therefore

$$p_\infty(k) = \frac{m^{k-m}}{(1+m)^{1+k-m}}, \qquad k \geq m.$$

□

Proposition 5. *The expected cut-off degree for the random attachment model is given by:*

$$k_1 = m - \frac{\ln N}{\ln(m) - \ln(m+1)}. \tag{46}$$

Proof. Using the same definition of distribution cut-off as in equation (36), we have

$$\sum_{k=k_1}^{\infty} \frac{m^{k-m}}{(1+m)^{1+k-m}} = \frac{1}{N}. \tag{47}$$

Changing the summation index gives

$$\frac{m^m(1+m)^{1-m}}{N} = \sum_{i=0}^{\infty} \left(\frac{m}{1+m}\right)^{i+k_1} \tag{48}$$

$$= \left(\frac{m}{1+m}\right)^{k_1} \sum_{i=0}^{\infty} \left(\frac{m}{1+m}\right)^{i}. \tag{49}$$

Using the standard result for the sum of an infinite geometric series:

$$\left(\frac{m}{1+m}\right)^{k_1} = \frac{m^m(1+m)^{1-m}}{N(1+m)}. \tag{50}$$

Finally, taking logarithms of both sides and rearranging:

$$k_1 \ln\left(\frac{m}{1+m}\right) = \ln\left[\frac{m^m}{N(1+m)^m}\right]. \tag{51}$$

Thus,

$$\begin{aligned} k_1 &= m - \frac{\ln N}{\ln\left(\frac{m}{1+m}\right)} \\ &= m - \frac{\ln N}{\ln(m) - \ln(m+1)}. \end{aligned}$$

□

This result shows that pure random attachment does not give asymptotic power law behaviour, and hence growth alone is not enough to ensure a scale-free network.

C. Mixed preferential and random attachment

For this model,

$$\Pi(k,t) = q\Pi_{BA} + (1-q)\Pi_{rnd}, \tag{52}$$

where q is the probability that an edge is attached using preferential attachment (as in the BA model). Substituting from (25) and (41) we have

$$\Pi(k,t) = q\left(\frac{k}{2E(t)}\right) + (1-q)\left(\frac{1}{N(t)}\right). \tag{53}$$

The master equation then becomes:

$$p_\infty(k) = m\Pi(k-1,t)p_\infty(k-1)N(t) - m\Pi(k,t)p_\infty(k)N(t) + \delta_{k,m}. \tag{54}$$

By exactly analogous arguments to those given for the previous models, we find that for $q \neq 0$,

$$p_\infty(k) = \frac{2}{2 - mq + 2m}, \qquad k = m, \tag{55}$$

and

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k + (\frac{2m}{q} - 2m - 1)}{k + (\frac{2m}{q} - 2m + \frac{2}{q})}, \qquad k > m, \tag{56}$$

which by lemma 1 gives:

$$p_\infty(k) = A\frac{\Gamma(k + \frac{2m}{q} - 2m)}{\Gamma(k + \frac{2m}{q} - 2m + \frac{2}{q} + 1)}, \qquad k > m. \tag{57}$$

This is complicated to normalise in general, but if we choose a particular value for q then we can use the same technique as employed for the BA model. Choosing $q = 0.5$ gives

$$p_\infty(k) = \frac{12m(3m+1)(3m+2)(3m+3)}{(k+2m+4)(k+2m+3)(k+2m+2)(k+2m+1)(k+2m)}, \qquad k \geq m, \tag{58}$$

which in the limit of large k gives power law behaviour: $p_\infty(k) \sim k^{-5}$.

III. SIMULATION

Numerical simulations of the three network models were coded in Python for both simple graphs and undirected multigraphs using the graph object model from the NetworkX library. The basic algorithm is described in figure 1. As noted in the previous section, growth via random or preferential attachment requires an initial network configuration to be in place. Three types were implemented and their effects compared: a "sparse" graph with each node having a degree of 1, a complete graph, and a graph consisting of nodes connected solely to a single hub. A completely disconnected graph (i.e. a network with no edges) was also considered for the pure random attachment model. The default value of $N(0)$ was taken as m (the minimum required for a simple graph) for even values of m, and $m+1$ for odd values. This ensured that the initial graph always had an even number of nodes: a requirement for sparse graphs with all degrees equal to 1.

Random attachment was trivial to implement as it involves simply picking a node using a uniform random number generator. Implementing linear preferential attachment was more involved. The most efficient way identified was to maintain a list of node indices with node multiplicity equal to degree. Choosing an item with uniform probability from this list clearly gives the desired probability of equation (25). An alternative approach was attempted in which a random end of a random edge is chosen, but this was found to be an order of magnitude slower.

To test the code, very small system sizes were employed. Initial and final networks were plotted and checked visually. The algorithm was also checked manually by outputting the

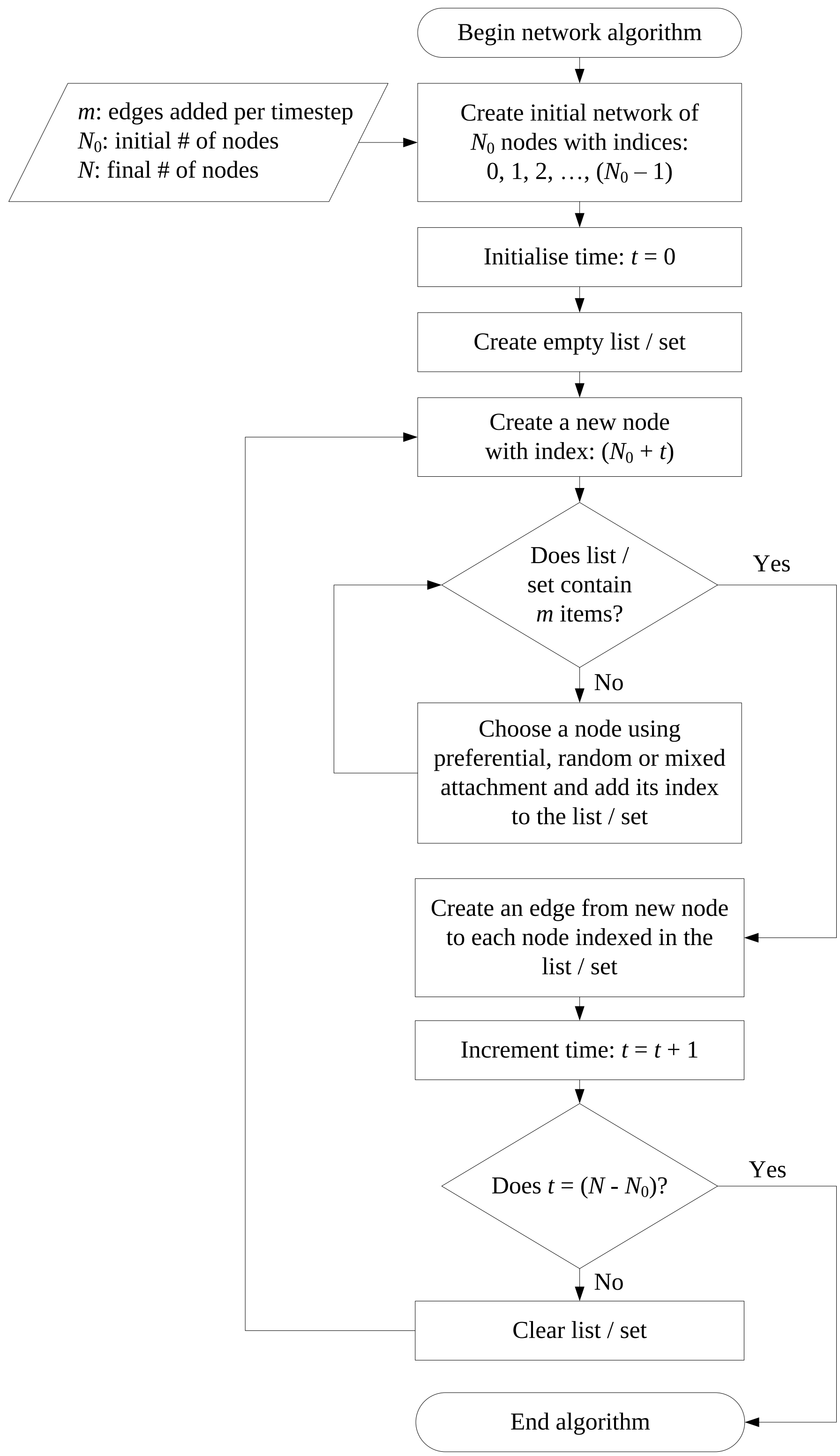

Figure 1. Algorithm to create a complex network of N nodes via growth and attachment. A list data type is used when choosing nearest neighbours for a node in a multigraph, as lists allow duplicate entries. A set data type is instead used for simple graphs, as sets ignore duplicate elements. In both cases, self-loops are not permitted. For notational clarity, N_0 is used here in place of $N(0)$ to represent the initial number of nodes.

following data to the console at each timestep: the nodes chosen as nearest neighbours, the total number of nodes, the total number of edges, and the network degree distribution. For preferential attachment, the node list was also checked at each timestep and the number of times a node appeared was compared with its degree to ensure they matched. For testing purposes, a fixed seed was chosen and fed into Python's random number generator, allowing the same graph to be grown to various different sizes.

IV. NUMERICAL RESULTS

Degree distribution data for various m and $N \gg m$ were were averaged over multiple runs and compared to the analytic results of section II. The largest system sizes achievable were limited by available hardware. $N = 10^6$ was possible for values of $m < 10$, whereas for values of $m > 100$ only $N = 10^4$ was feasible.

Following Milojević [4], logarithmic binning was used to deal with the statistical noise in the tail of the distribution. This was particularly important for the preferential attachment models which were found to have fat tails, as expected from the asymptotic power law form of their degree distributions (see figure 2). The initial bin size was set to 1 and, after experimentation, a bin ratio of 1.1 chosen, meaning each bin is 10% larger than the previous. Visually, this seemed to give the best indication of the underlying distribution, without loss of important information. The node counts were scaled by system size in order to approximate the underlying probability distribution $p(k)$.

To examine finite size effects, data collapse was performed for the BA and random attachment models. k was scaled by the theoretical value for the cut-off, k_1, and $p(k)$ was scaled by the analytic value for $p_\infty(k)$. This maps the theoretical asymptotic distribution to a straight line with $p(k)/p_\infty(k) = 1$ and gives an expected cut-off at $k/k_1 = 1$. The data collapse therefore makes it clear where simulation data deviate from the analytic results. Collapsing data for various m with fixed N shows whether the theoretical expression for k_1 has the correct m dependence for finite size systems. Similarly, collapsing data for various N with fixed m shows whether the N dependence is correct.

Comparing the data to the asymptotic distributions using statistical techniques is difficult, not least because of the cut-off phenomenon. Various methods of fitting data to power laws are discussed in the literature [5]. The simplest is to fit log binned data in the scaling region before the cut-off to a straight line using simple linear regression as in [4]. This was attempted for the BA model data shown in figure 3, with the largest network ($N = 10^6$) giving a reasonable value for the exponent of -2.92 ± 0.01 (cf. the expected value of -3). The r^2 coefficient for the fit was found to be 0.9997. However, Clauset et al. have shown in [6] that it is invalid to apply linear regression to a log-log plot and that it cannot be used to evaluate goodness of fit. In other words, the value of r^2 is actually meaningless in telling us how well the data fit the expected asymptotic power law. Clauset et al. outline a superior maximum-likelihood fitting technique with a goodness of fit test based on the Kolmogorov-Smirnov statistic. An application of this method, however, is beyond the scope of the present work.

A. Preferential attachment (BA model)

Some typical results for the BA model are shown in figures 2 to 6. As seen in figure 3, the finite size cut-off is preceded by a bump similar to that observed in the scaling function of self-organized critical systems [7]. However, the exact size and shape of this feature was found to depend heavily on the initial network configuration. For example, when a single hub initial configuration was used, the cut-off was preceded by a double bump. This also occurred for simple graphs with a sparse initial configuration (but not for multigraphs). Networks with $N(0) \sim 10m$ were also simulated and found to give similar results, but with the hump displaced beyond the position of the cut-off decay. Interestingly, these effects remained present in both simple graphs and multigraphs for all values of N. This suggests that for any finite sized BA network, the initial conditions will always have an effect on the shape of the degree distribution.

The average largest node degree was calculated for various N with fixed m in order to compare with the asymptotic cut-off expression given by (35). For this investigation, a value of $m = 3$ was chosen so that the ratio m/N would be as large as possible, whilst avoiding the trivial case of $m = 1$. An example is shown in figure 6. The data appear to scale with N as predicted, but there is a clear offset from the theoretical relationship, suggesting a different m dependence for the cut-off in finite size networks. This offset was present for all network types and initial configurations. Data collapse for systems with fixed N but varying m showed that the m dependence is indeed different from the predicted value for these system sizes.

B. Pure random attachment

Unlike the BA model, the random attachment model does not lead to a fat-tailed degree distribution. Statistical noise was still found at the end of the distribution, however, so log

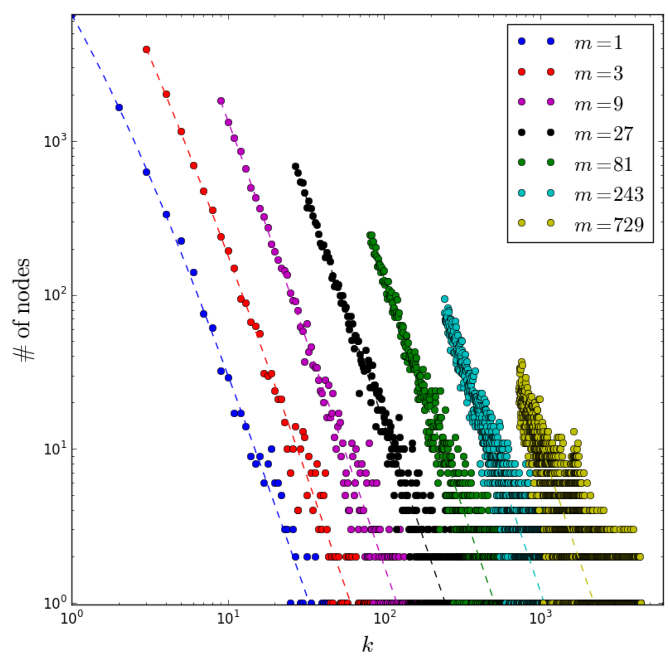

Figure 2. Node count versus degree k for seven BA networks of size $N = 10^4$, each with a different value of m. The initial configuration was a sparse simple graph with $N(0) = m + 1$. The theoretical asymptotic distribution is shown as a dashed line. Notice that the distribution has a fat tail and shows statistical noise for large values of k. For $m = 729$ it is impossible to tell whether the data fit the expected distribution.

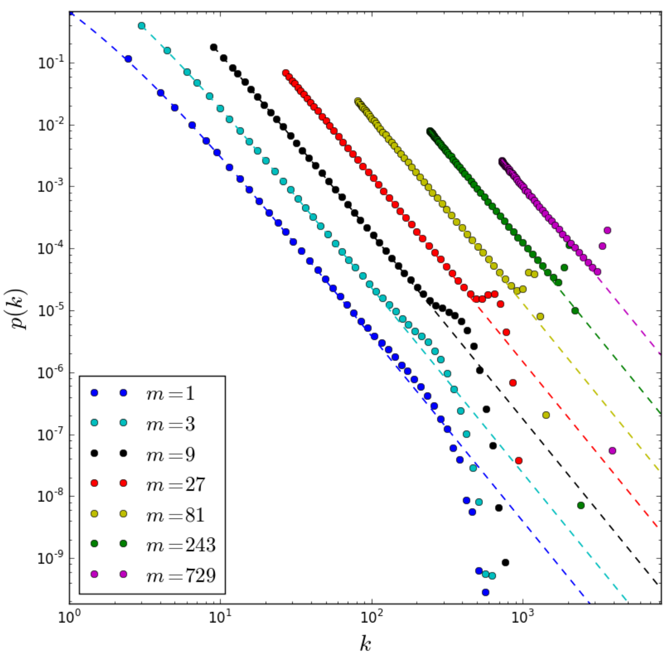

Figure 3. Log binned degree distributions for BA networks with $N = 10^4$ and $m = 3^n$ for $n = 0$ to 6. The data were averaged over $100 \times 2^{6-n}$ runs. The initial configuration was a complete simple graph with $N(0) = m + 1$. The datapoints show the centre of the bins which closely follow the theoretical asymptotic distribution (shown by the dashed line) until a cut-off beginning with a bump which is larger and narrower for increasing m.

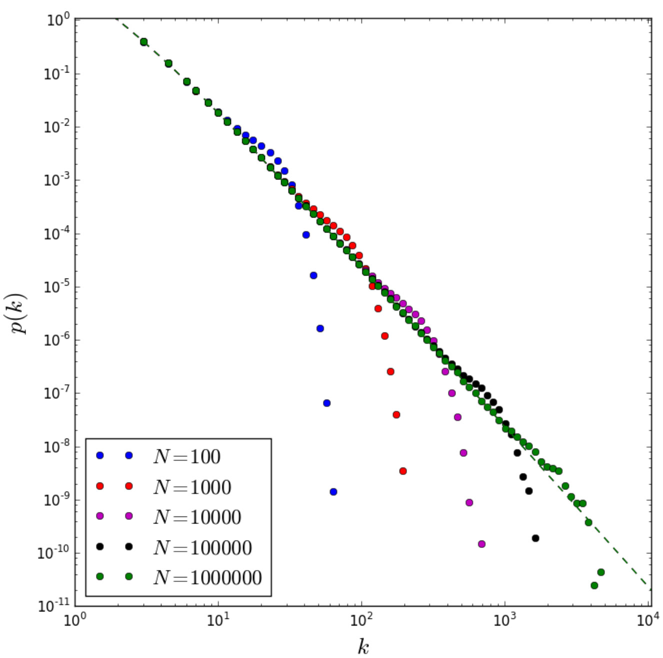

Figure 4. Log binned degree distributions for BA networks with $m = 3$. Networks with $N = 10^n$ for $n = 2$ to 6 were averaged over 10^{8-n} runs. The initial configuration was a complete simple graph with $N(0) = 4$. The dashed line shows the theoretical asymptotic distribution. Notice that the data points lie on top of one another and follow the predicted PDF until the cut-offs.

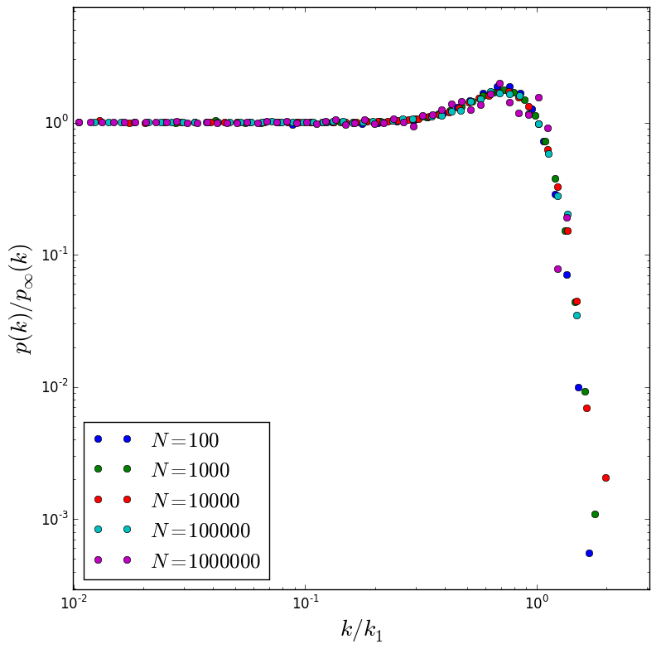

Figure 5. Data collapse of the log binned data from figure 4 clearly showing the distribution cut-off. The expected rapid decay is preceded by a characteristic bump which peaks just before $k/k_1 = 1$. Similar results were found for multigraphs with the same initial network configuration.

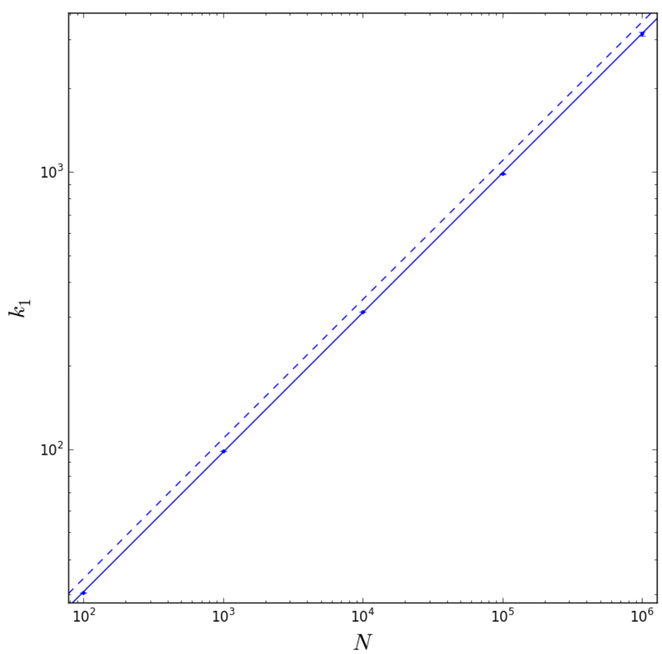

Figure 6. Largest node degree plotted against system size for the data shown in figure 4. Vertical error bars showing the standard error on the mean value of k_1 cannot be seen at this scale. The solid line is a linear fit, which gave a slope of 0.51 ± 0.01. The dashed line shows the theoretical asymptotic scaling relation. Although the data appear to scale as predicted (as seen by the similar slopes of the two lines), there is a clear offset from the theoretical curve.

binning was again employed. Some results are shown in figures 7 through 9. The cut-off was found to be preceded by a bump, but this was only obvious for large m, making it unfeasible to investigate via a data collapse over large ranges of N. Little difference was found between the degree distributions of simple graphs and multigraphs with complete or sparse initial conditions. A completely disconnected initial configuration was also tested, but again the results were similar. Even a multigraph with an initial configuration of just a single node gave rise to a similar bump. The only exception was for both simple graphs and multigraphs grown from the initial single hub configuration. In this case, a hump was found to occur far earlier than the cut-off, with the distribution returning to the theoretical prediction before decaying rapidly.

The scaling of the average largest node degree was investigated using $m = 3$ but found to differ from the asymptotic cut-off expression given by (46) for all network types. In fact, a good fit was only found for $m = 1$. Results for $m = 1$, 3 and 5 are shown in figure 9. It seems plausible that the cut-off scaling may approach the asymptotic relationship in the limit of large N for any m (taking longer the larger the value of m), but it was not possible to test this hypothesis as system sizes larger than 10^6 could not be simulated.

C. Mixed attachment

As expected, the degree distribution for the mixed attachment model with $q = 0.5$ had a fat tail, but decayed faster than the BA model. Log binned simulation data followed the theoretical asymptotic distribution for up to two orders of magnitude (for a system size of $N = 10^6$). The finite size cut-off was again preceded by a characteristic hump, but this was only noticeable for $m \geq 9$ (see figure 10).

The average cut-off degree was plotted against system size and found to scale approximately as a power law. A linear fit gave an exponent of 0.286 ± 0.007 (see figure 12).

V. CONCLUSION

Preferential and random attachment models have been explored analytically and numerically. Finite size networks for even fairly small N are seen to fit the asymptotic degree distributions up to a finite size cut-off. These cut-offs were preceded by one or more bumps whose shape depended on the initial configuration and type of network. The scaling of the cut-off degree with system size was also investigated. The BA simulation data appear to have the same N dependence as the theoretical cut-off, but a different m dependence. The random attachment data appear to have the correct N dependence only for $m = 1$. With additional computing resource it should be possible to see numerically whether or not the largest degree in random attachment networks approaches the theoretical value given by (46) for large N, as conjectured. Other future work could include:

- Statistical comparison of the simulation data to the analytic asymptotic distributions using the methods of [6].

- Further investigation of the mixed attachment model. Analytically, this should include normalisation of the degree distribution for all values of the preferential attachment probability q, as well as a derivation of the largest expected node degree.

- An attempt to better understand the finite size cut-offs using extreme value theory as discussed by Boguñá et al. in [8].

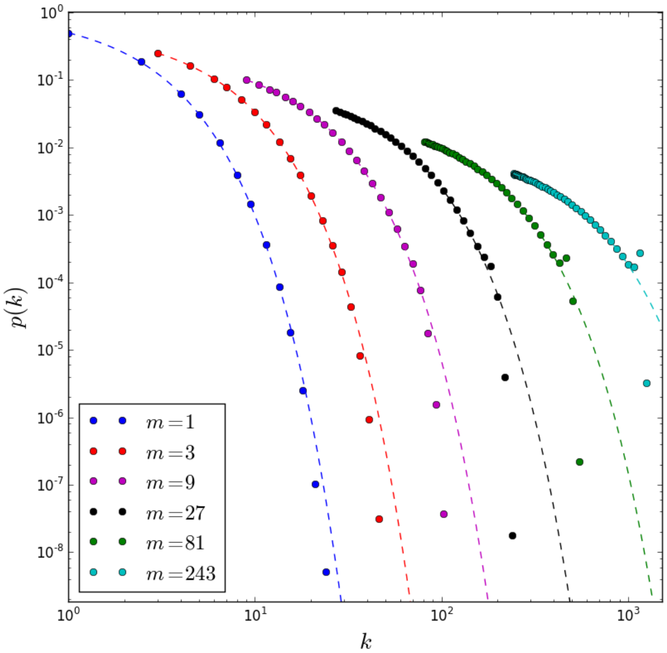

Figure 7. Log binned degree distributions for random attachment networks with $N = 10^4$ and $m = 3^n$ for $n = 0$ to 5. The data were averaged over $100 \times 2^{6-n}$ runs. The initial configuration was a complete simple graph with $N(0) = m + 1$. The dashed lines show the theoretical asymptotic distributions.

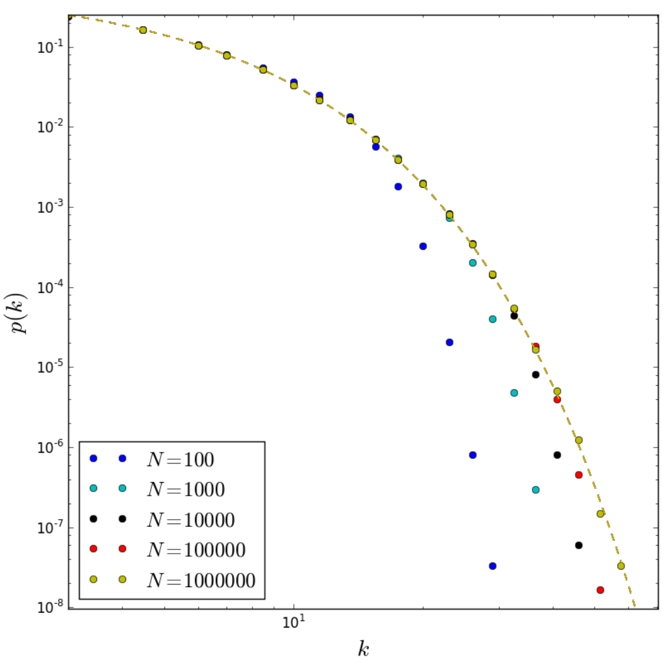

Figure 8. Log binned degree distributions for random attachment networks with $m = 3$ and $N = 10^n$, for $n = 2$ to 6. The data are averages over 10^{7-n} runs. The initial configuration was a complete simple graph with $N(0) = 4$. Even for a network of size $N = 10^6$, the data only follow the asymptotic degree distribution over one and a half orders of magnitude (cf. the fat tail of the BA model).

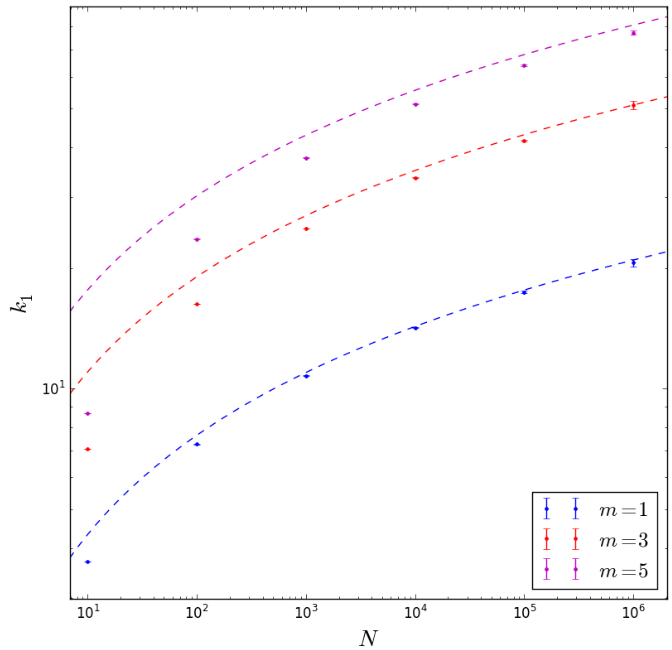

Figure 9. Comparison of average cut-off degree in the random attachment model for $m = 1$, 3 and 5. Networks of size $N = 10^n$, for $n = 1$ to 6 were averaged over 10^{7-n} runs. The initial configuration was a complete simple graph with $N(0) = 4$. Notice how the data appear to approach the theoretical asymptotic relationship (shown by dashed lines) with increasing N.

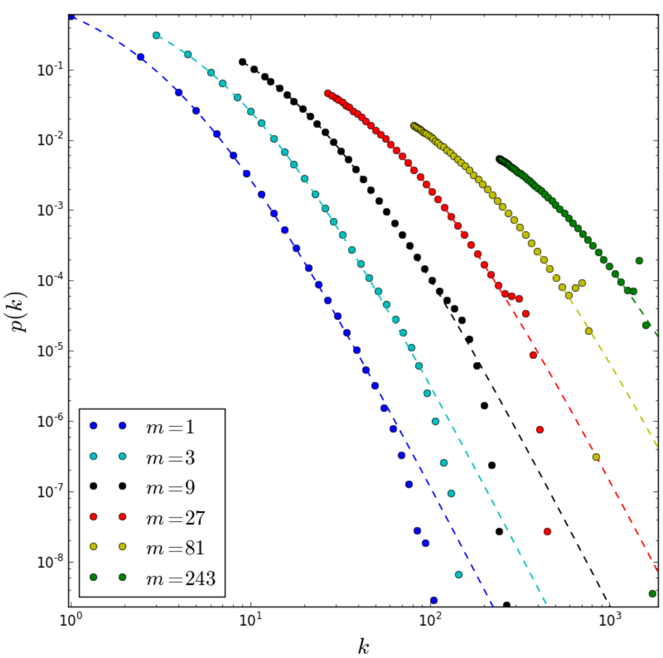

Figure 10. Log binned degree distributions for $q = 0.5$ mixed attachment networks with $N = 10^4$ and $m = 3^n$ for $n = 0$ to 5. The data were averaged over $100 \times 2^{6-n}$ runs. The initial configuration was a complete simple graph with $N(0) = m + 1$. The dashed lines show the analytic asymptotic distributions.

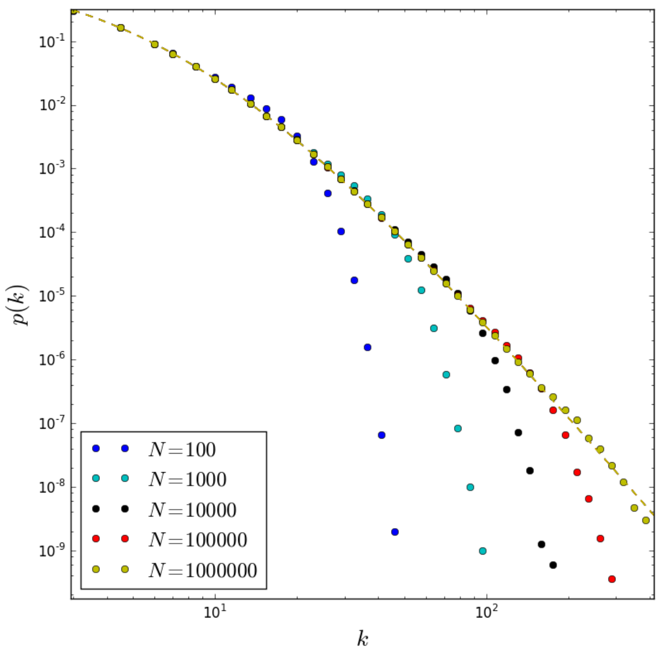

Figure 11. Log binned degree distributions for mixed attachment networks with $m = 3$ and $N = 10^n$, for $n = 2$ to 6. The data are averages over 10^{7-n} runs. The initial configuration was a complete simple graph with $N(0) = 4$. For $N = 10^6$, the data follow the asymptotic degree distribution for approximately two orders of magnitude.

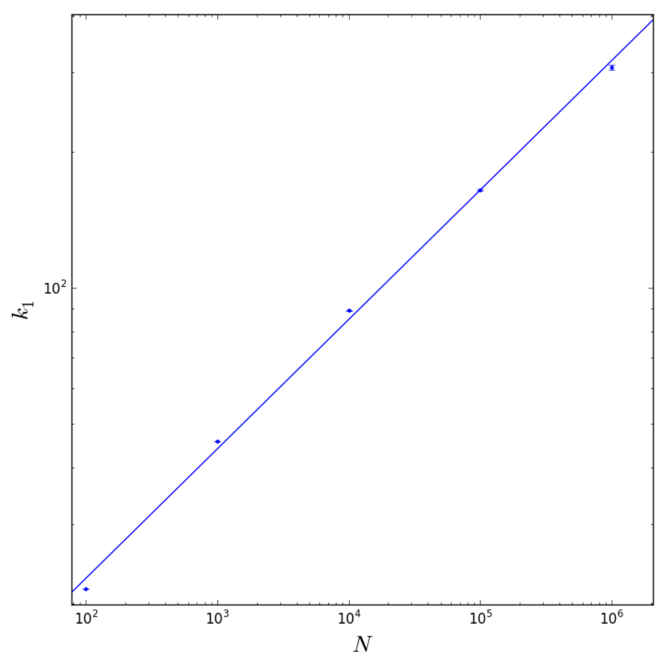

Figure 12. Largest node degree plotted against system size for the data shown in figure 11. The vertical error bars showing the standard error on the mean cannot be seen at this scale. The solid line shows the linear fit. Linear regression gives a slope of 0.286 ± 0.007, which is likely an overestimate of the asymptotic behaviour due to the inclusion of small network sizes.

ACKNOWLEDGMENTS

The author thanks Dr. Tim Evans for his help and support, and Ms. Jana Smutná and Mr Imran Rashid for stimulating discussions.

---

[1] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[2] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Physical Review Letters **85**, 4633 (2000).
[3] S. N. Dorogovtsev and J. F. F. Mendes, Advances in Physics **51**, 1079 (2002).
[4] S. Milojevic, Journal of the American Society for Information Science and Technology **61**, 2417 (2010).
[5] M. Newman, Contemporary Physics **46**, 323 (2005).
[6] A. Clauset, C. Shalizi, and M. Newman, SIAM Review **51**, 661 (2009).
[7] G. Pruessner, *Self-organised criticality : theory, models and characterisation* (Cambridge University Press, Cambridge, 2012).
[8] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, The European Physical Journal B - Condensed Matter and Complex Systems **38**, 205 (2004).