# Networks Project Report

CID: 00923604, Imperial College London

November 18, 2018

Word Count: 2487

**Abstract**

The Barabási-Albert model is solved analytically for preferential, random and mixed preferential/random attachment probabilities to find the degree distributions and variation of expected largest degree with network size. It is found that both the preferential and mixed attachment cases lead to power-law degree distributions and that random attachment leads to a exponential degree distribution. The theoretical results are then compared to numerical simulations, revealing finite size effects in the tail of the distribution. The numerical and theoretical distributions show good agreement through the Kolgomorov-Smirnov test up to a finite size effect cut off point. The numerical largest degree converges to the theoretical value for large system sizes, and it is found that this occurs more quickly for the random and mixed attachment cases.

# 1 Introduction

Complex networks are hard to define precisely, but are often described as networks displaying non-trivial topological characteristics such high clustering coefficients or scale free degree distributions. Such characteristics do not emerge in random networks such as those generated by the Erdös-Réyni model (Erdös & Rényi 1959). Scale free degree distributions are characterised by a power law in the degree $k$: $p(k) \sim k^{\gamma}$. Scale free distributions in stochastic processes were originally studied by Yule and Simon (1955), and a model for the growth of scale free citation networks was first developed by Price (1976). However, significant research interest has been generated in this subject recently due to empirical studies of real-world scale free, complex networks such as the internet and social networks (Albert & Barabási 2002).

The Barabási-Albert (BA) model (Barabási & Albert 1999) is a simple model of growing complex networks displaying fat-tailed degree distributions under certain conditions, and is similar to previous models developed by Yule and Price. In this report, the BA model is solved analytically for a range of attachment probabilities, starting with the original preferential attachment case but also extending to random attachment and mixed preferential-random attachment, in order to determine which attachment probabilities lead to a scale free network. These results are then compared to numerical simulations in order to reveal finite size effects.

# 2 Analytic Results

## 2.1 Master Equation

The BA model algorithm is as follows:

1. Set up an initial network at $t = t_0$, a graph $\mathcal{G}_0$.

2. Increment time $t \to t + 1$.

3. Add a new vertex $\mathcal{V}_t$ to the graph.

4. Add $m$ new edges to the vertex, connecting one end of each edge to $\mathcal{V}_t$ and one end to an existing vertex chosen with probability $\Pi$ which is specified according to the form of attachment used.

5. Repeat step 2 until the network has $N$ vertices.

The BA model is encapsulated in the following 'master equation' (Albert & Barabási 2002) which describes the evolution of the number of vertices with degree $k$ at a given time $t$, $n(k,t)$:

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

where $\delta_{k,m}$ is a delta function, and $\Pi(k,t)n(k,t) = 0$ for $k < m$ (this is because the BA model is defined such that each new vertex has degree $m$ and so all vertices must have $k > m$). Equation (1) can be understood by considering each term. The first term is simply the number of vertices with degree $k$ at time $t$. The next two terms state that a fraction of vertices with degree $k-1$ will be converted to vertices of degree $k$ determined by the attachment probability function $\Pi(k)$, and that similarly a fraction of vertices of degree $k$ will be converted to vertices $k+1$. The last term describes the case where the new vertex is converted to a vertex of degree $k$ if $m = k$.

## 2.2 Preferential Attachment

### 2.2.1 Degree Distribution

The first type of attachment to be considered is that of preferential attachment, where the probability of connecting to an existing vertex is proportional to the degree of the vertex:

$$\Pi_{pa}(k,t) = \frac{k}{2E}, \tag{2}$$

where the probability is normalised by $2E$ as this the total number of 'stubs' (i.e number of edges leaving each vertex) in the network. To solve equation (1) it is assumed that $E(t) = mN(t)$ (which is true in the $t \to \infty$ limit), and the degree probability distribution is defined as $p(k,t) = n(k,t)/N(t)$ where $N(t)$ is the total number of vertices at time $t$. Equation (1) then simplifies to

$$n(k,t+1) = n(k,t) + \frac{m(k-1)n(k-1,t)}{2mN(t)} - \frac{mkn(k,t)}{2mN(t)} + \delta_{k,m} \tag{3a}$$

$$\implies N(t+1)p(k,t+1) = N(t)p(k,t) + \frac{1}{2}[(k-1)p(k-1,t) - kp(k,t)] + \delta_{k,m}. \tag{3b}$$

It is assumed that in the infinite time limit, $p(k,t)$ will tend to a stable form,

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

Taking the limit $t \to \infty$ also allows the substitution $N(t) = t$ into (3b):

$$p_\infty(k) = \frac{1}{2}[(k-1)p_\infty(k-1) - kp_\infty(k)] + \delta_{k,m}. \tag{5}$$

To solve equation (5), consider $k > m$ for which $\delta_{k,m} = 0$:

$$p_\infty(k) = \frac{1}{2}[(k-1)p_\infty(k-1) - kp_\infty(k)] \tag{6a}$$

$$\implies \frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k-1}{k+2}. \tag{6b}$$

Equation (6b) can be rewritten in a more general form

$$\frac{f(z)}{f(z-1)} = \frac{z+a}{z+b}, \tag{7}$$

which has solution

$$f(z) = A\frac{\Gamma(z+1+a)}{\Gamma(z+1+b} \tag{8}$$

where $\Gamma(z)$ is the Gamma function with the following properties

$$\Gamma(z+1) = z\Gamma(z) \tag{9a}$$

$$\Gamma(1) = 1 \tag{9b}$$

$$\Gamma(n) = (n-1)! \qquad n \in \mathbb{Z} \tag{9c}$$

To show this, equation (8) is substituted into equation (7) and equation (9a) is used to simplify the resulting expression:

$$f(z) = A\frac{\Gamma(z+1+a)}{\Gamma(z+1+b)} \tag{10a}$$

$$f(z-1) = A\frac{\Gamma(z+a)}{\Gamma(z+b)} \tag{10b}$$

$$\implies \frac{f(z)}{f(z-1)} = \frac{\Gamma(z+1+a)\Gamma(z+b)}{\Gamma(z+1+b)\Gamma(z+a)} \tag{10c}$$

$$\implies \frac{f(z)}{f(z-1)} = \frac{(z+a)\Gamma(z+a)\Gamma(z+b)}{(z+b)\Gamma(z+b)\Gamma(z+a)} \tag{10d}$$

$$\implies \frac{f(z)}{f(z-1)} = \frac{z+a}{z+b}. \tag{10e}$$

Hence, the solution to equation (6b) can be written as

$$p_\infty(k) = A\frac{\Gamma(k)}{\Gamma(k+3)}. \tag{11}$$

The degree $k$ is a positive integer, and so equation (9c) can be used to simplify (6b):

$$p_\infty(k) = \frac{A}{k(k+1)(k+2)}. \tag{12}$$

To determine the normalisation constant $A$, the case $k = m$ is considered, remembering that in the BA model $p_\infty(k < m) = 0$. Equation (5) simplifies to

$$
\begin{aligned}
p_\infty(m) &= -\frac{m}{2} p_\infty(m) + 1 && \text{(13a)}\\
\implies p_\infty(m) &= \frac{2}{2+m} && \text{(13b)}\\
\implies \frac{A}{m(m+1)(m+2)} &= \frac{2}{2+m} && \text{(13c)}\\
\implies A &= 2m(m+1) && \text{(13d)}
\end{aligned}
$$

Hence, for $t \to \infty$ the exact degree distribution for preferential attachment case is

$$p_\infty(k) = \frac{2m(m+1)}{k(k+1)(k+2)}, \tag{14}$$

which is a power law distribution. This can be compared to the degree distribution derived in (Albert & Barabási 2002). For $k \gg m$, this tends to $p_\infty(k) \sim k^{-3}$. The normalisation of (14) can be verified using partial fractions:

$$
\begin{aligned}
\sum_{k=m}^{\infty} p_\infty(k) &= \sum_{k=m}^{\infty} \frac{2m(m+1)}{k(k+1)(k+2)} && \text{(15a)}\\
\implies \sum_{k=m}^{\infty} p_\infty(k) &= m(m+1) \sum_{k=m}^{\infty} \left( \frac{1}{2k} - \frac{2}{k+1} + \frac{1}{k+2} \right) && \text{(15b)}\\
\implies \sum_{k=m}^{\infty} p_\infty(k) &= m(m+1) \left( \sum_{k=m}^{\infty} \frac{1}{k} - \sum_{k'=m+1}^{\infty} \frac{2}{k} + \sum_{k''=m+2}^{\infty} \frac{1}{k} \right) && \text{(15c)}\\
\implies \sum_{k=m}^{\infty} p_\infty(k) &= m(m+1) \left( \frac{1}{m} + \frac{1}{m+1} - \frac{2}{m+1} \right) && \text{(15d)}\\
\implies \sum_{k=m}^{\infty} p_\infty(k) &= 1, && \text{(15e)}
\end{aligned}
$$

where the indices have been relabelled in equation (15c) in order to cancel the majority of terms in the sums in equation (15d).

#### 2.2.2 Largest expected degree

The largest expected degree, $k_1$, of the BA model network created using preferential attachment can be defined as the degree above which only one vertex would be expected to be found according the theoretical distribution:

$$N \sum_{k=k_1}^{\infty} p_\infty(k) = 1. \tag{16}$$

Decomposing this using partial fractions and relabelling indices to cancel the majority of terms (as in section 2.2.1) results in

$$N \sum_{k=k_1}^{\infty} p_\infty(k) = Nm(m+1)\left(\frac{1}{k_1} + \frac{1}{k_1+1}\right) \tag{17a}$$

$$\implies m(m+1)\left(\frac{1}{k_1} - \frac{1}{k_1+1}\right) = \frac{1}{N} \tag{17b}$$

$$\implies k_1^2 + k_1 - m(m+1)N = 0, \tag{17c}$$

which has solution

$$k_1 = \frac{-1 + \sqrt{(1 + 4m(m+1)N)}}{2} \tag{18}$$

where the positive root is taken to ensure a physical (positive) solution. Hence, $k_1 \sim \sqrt{N}$.

## 2.3 Random Attachment

In the random attachment case, the new vertex created at each time step connects to all existing vertices with equal probability, that is

$$\Pi_{rnd}(k,t) = \frac{1}{N(t)}. \tag{19}$$

The master equation (1) now takes the form

$$p_\infty(k) = m(p_\infty(k-1) - p_\infty(k)) + \delta_{k,m}, \tag{20}$$

where $p_\infty(k) = \lim_{t\to\infty} p(k,t)$ as in section 2.2. For $k = m$, equation (20) becomes

$$p_\infty(m) = m(p_\infty(m-1) - p_\infty(m)) + 1 \tag{21a}$$

$$\implies p_\infty(m) = -mp_\infty(m) + 1 \tag{21b}$$

$$\implies p_\infty(m) = \frac{1}{1+m} \tag{21c}$$

Now considering $k = m + 1$, (20) becomes

$$p_\infty(m+1) = \frac{m}{1+m} p_\infty(m) \tag{22a}$$
$$\implies p_\infty(m+1) = \frac{m}{(1+m)^2} \tag{22b}$$

This result can be generalised to any $k$:

$$p_\infty(k) = \frac{1}{1+m}\left(\frac{m}{1+m}\right)^{k-m} \tag{23}$$

which can be proved rigourously by induction, taking the $k = m$ as the base case. The normalisation of (23) can be verified using the standard infinite geometric sum formula which confirms that

$$\sum_{k=m}^{\infty} p_\infty(k) = \frac{1}{1+m}\left(\frac{1+m}{m}\right)^m \sum_{k=m}^{\infty}\left(\frac{m}{1+m}\right)^k = 1. \tag{24}$$

## 2.4 Largest expected degree

The largest expected degree for the random attachment case is obtained by solving equation (16) for $p_\infty(k)$ given by equation (23), which takes the following form in this case:

$$\sum_{k=k_1}^{\infty} p_\infty(k) = \frac{1}{1+m}\left(\frac{1+m}{m}\right)^m \sum_{k=k_1}^{\infty}\left(\frac{m}{1+m}\right)^k = \frac{1}{N}. \tag{25}$$

Using the standard infinite geometric sum formula results in

$$\left(\frac{1+m}{m}\right)^m \left(\frac{m}{1+m}\right)^{k_1} = \frac{1}{N}. \tag{26}$$

Taking logs of both sides to solve for $k_1$,

$$k_1 = m - \frac{\ln N}{\ln m - \ln(1+m)}. \tag{27}$$

For $N \gg m$ this tends to $k_1 \sim \ln N$.

## 2.5 Mixed Preferential and Random Attachment

In the mixed preferential and random attachment case, the new vertex is connected to an existing vertex via preferential attachment with probability $q$ and via random attachment with probability $1 - q$. In other words,

$$\Pi(k) = q\Pi_{pa}(k) + (1-q)\Pi_{rnd}. \tag{28}$$

The master equation in this case takes the form

$$n(k,t+1) = n(k,t)+m\left(\frac{q(k-1)}{2mN(t)}+\frac{1-q}{N(t)}\right)n(k-1,t)-m\left(\frac{qk}{2mN(t)}+\frac{1-q}{N(t)}\right)n(k,t)+\delta_{k,m}, \tag{29}$$

which simplifies to

$$p_\infty(k) = p_\infty(k-1)\left(\frac{1}{2}q(k-1)+m(1-q)\right) - p_\infty(k)\left(\frac{1}{2}qk+m(1-q)\right)+\delta_{k,m} \tag{30}$$

in the limit $t \to \infty$. For $k > m$ this becomes

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k+\frac{2m(1-q)}{q}-1}{k+\frac{2m(1-q)}{q}+\frac{2}{q}}, \tag{31}$$

and for $k = m$:

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

These equations can be combined through normalisation of the solution to give a general solution for any $q$ and $k \geq m$:

$$p_\infty(k) = A\frac{\Gamma(k+\frac{2m(1-q)}{q})}{\Gamma(k+1+\frac{2m(1-q)}{q}+\frac{2}{q})} \tag{33a}$$

$$A = \frac{\Gamma(m+1+\frac{2m(1-q)}{q}+\frac{2}{q})}{\Gamma(m+\frac{2m(1-q)}{q})(1+\frac{qm}{2}+m(1-q))}. \tag{33b}$$

In order to simplify this expression, only the case where $q = 0.5$ is explicitly considered, resulting in the following expression:

$$p_\infty(k) = \frac{12m(3m+3)(3m+2)(3m+1)}{(k+2m+4)(k+2m+3)(k+2m+2)(k+2m+1)(k+2m)}. \tag{34}$$

For $k \gg m$ this tends to $p_\infty(k) \sim k^{-5}$.

### 2.5.1 Largest expected degree

The largest expected degree can be derived for the $q = 0.5$ case for mixed preferential and random attachment in the same was as the previous cases, by solving equation 16 with the theoretical mixed degree distribution given by equation 34. Using Mathematica to carry out the summation and solve for $k_1$:

$$k_1 = \frac{-3+4m+\sqrt{5+4\sqrt{1+9m(1+m)(1+3m)(2+3m)N}}}{2}, \tag{35}$$

where the physical (positive) solution to the resulting quartic equation has been selected. Hence, $k_1 \sim N^{1/4}$ (i.e. the fat tail extends to smaller values of $k$ than for pure preferential attachment, for large $N$).

# 3 Numerical Results

A numerical model of the BA model is now implemented to verify the analytic results. Python is used in conjunction with the NetworkX library to implement the model. An important aspect of the BA model to specify for all attachment schemes is the initial network configuration. In order to avoid biasing a particular vertex in the initial network, a complete simple graph is used. However, it is unlikely that the configuration of the initial graph would make a significant difference in the $t \to \infty$ limit if the size of the initial graph is small enough. Furthermore, the existence of an initial network of any configuration is enough to produce a finite size effect in the degree distribution. The initial graph size used is $m+1$ as this is the smallest vertex number possible that ensures each vertex has degree $m$, so not to violate the BA model boundary conditions used in all analytic derivations. A network size of $N = 10^4$ is mostly used as this is found to be sufficient to exhibit the expected properties of the model, with $N = 10^5$ or $N = 10^6$ used sparingly due to computational limitations.

## 3.1 Preferential Attachment

The preferential attachment case is implemented numerically by appending the two vertices associated with each new edge to a list every time an edge is added, leading to a list where the frequency of each vertex is proportional to its degree. Self loops and connecting to the same vertex more than once in the same $m$ connection stage are prevented through checks every iteration. In order to check the correct implementation of the model, model values are printed out for small values of $m$ and $N$, and each attachment stage is checked by hand. The mean degree of the vertices in the network are calculated, and is found to be $2m$ as would be expected (every edge has two stubs). The normalisation of all probability distributions is also verified.

### 3.1.1 Degree Distribution

Figure 1 shows the distribution of vertex degree for the preferential attachment case, for $m = 3^n$, where $n = 0, .., 5$, plotted alongside the theoretically expected distribution given by multiplying equation (14) by network size $N$. Good agreement to the theoretical distribution can be seen visually from this raw data. The noise present in the tail of the distribution is an issue with fat tailed distributions which can be overcome through binning of the data and averaging over many repeats in order to extract the underlying distribution. Unless otherwise stated, all numerical data is produced by averaging over 1000 repeats.

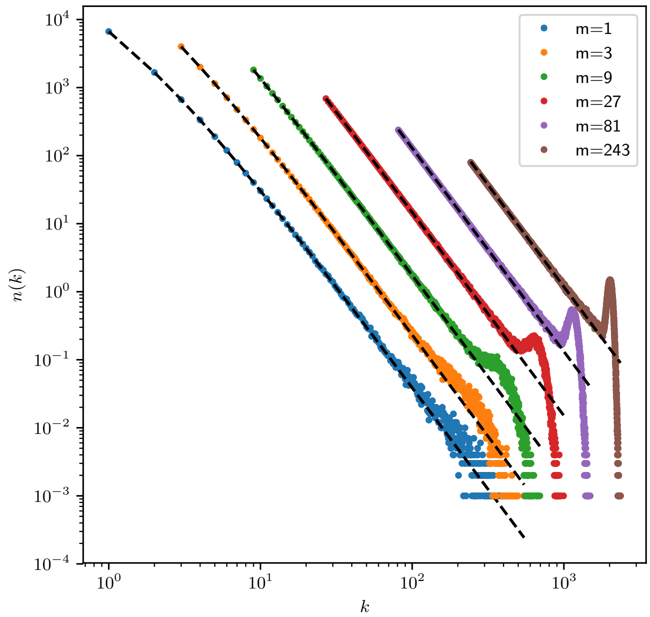
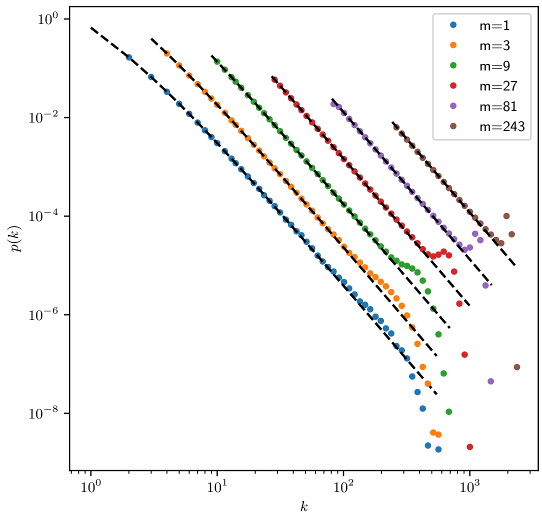


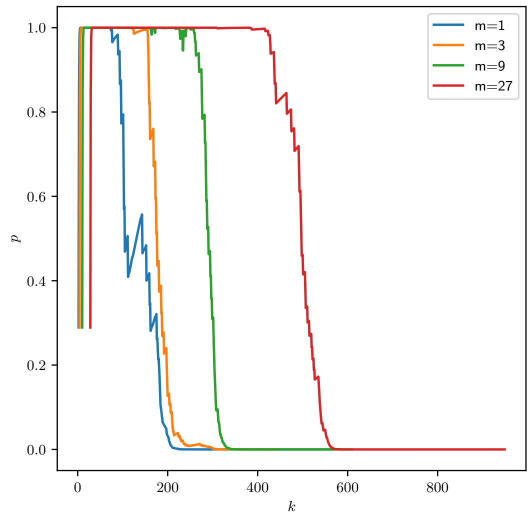

Figure 1: Vertex degree distribution for preferential attachment, plotted alongside the theoretically expected distribution, for $m = 3^n$, with $n = 0, .., 5$, averaged over 1000 repeats. A network size of $N = 10^4$ is used.

Log-binning can be applied to the data, where bin size increases for larger degree values, in order to reduce the noise present in the tail of the distribution. The bin size is controlled by a scale parameter $a$, which controls the increase in size of successive bins. Log-binning with a value of $a = 1.1$ is used where this better resolves the underlying distribution. Figure 2a shows the average degree probability distribution of networks with different $m$ values, found by log-binning the degree frequency distribution produced over 1000 repeats and dividing by the network size $N$ and the number of repeats. Visually it can be seen that the log-binned numerical distribution matches the theoretical distribution well up to the tail of the distribution, where the numerical distribution displays finite-size effects in the form of a 'bump', which results from the initial network configuration, as the degree of these initial vertices lead to degrees with a greater probability than expected. It can be seen that this bump decreases in width and increases in height for larger values of $m$, as the effect becomes more significant if many edges are added in the first step, but affects fewer degrees as fewer $m$ connection stages are required to reach the same number of vertices compared to smaller values of $m$.

In order to compare the numerical and analytic data quantitatively, a Kolgomorov-Smirnov (KS) goodness of fit test is carried out (Barlow 1989). This method is chosen as it makes no assumption about the form of the probability distribution being tested. The test is applied to the raw numerical probability distribution data and the corresponding theoretical probability distribution. Unfortunately, this test is usually only valid for continuous distributions. However, the significance levels tend to be conservative when using discrete data (Arnold & Emerson 2011) and so it is hoped that the test still has some validity. A null hypothesis that is made that the numerical and theoretical distributions are not drawn from the same distribution, so a $p$ value close to 1.0 would indicate that the null hypothesis can be rejected and the distributions do match. Figure 2b shows the effect of including larger amounts of data in the KS test. For data taken primarily from the start of the distribution, an excellent ($p \sim 1.0$) goodness of fit is found. However if the tail is included, the $p$ value obtained rapidly deteriorates. Hence, the fit can be said to be good up to a cut-off value that can be determined from the location of the bump.




(a) Log binned degree distribution.




(b) P-value variation.

Figure 2: Figure (a) shows log-binned degree distribution for preferential attachment plotted alongside the theoretically expected distribution for $m = 3^n$, with $n = 0, .., 5$, for $N = 10^4$. Figure (b) shows the resulting variation in KS test p-values for selected values of $m$ when including increasing amounts of data, demonstrating the effect of the finite size effect bump on the fit to the theoretical distribution.

### 3.1.2 Finite size effects

The largest expected degree, $k_1$, is derived in section 2.2.2 for preferential attachment. This can now be compared to the largest degree found through numerical simulations in order to investigate finite network effects. The numerical value of $k_1$ is plotted alongside its theoretical value for various network sizes in figure 3 for $m = 3$.

It can be seen in figure 3 that the theoretical value falls within the error bars of the

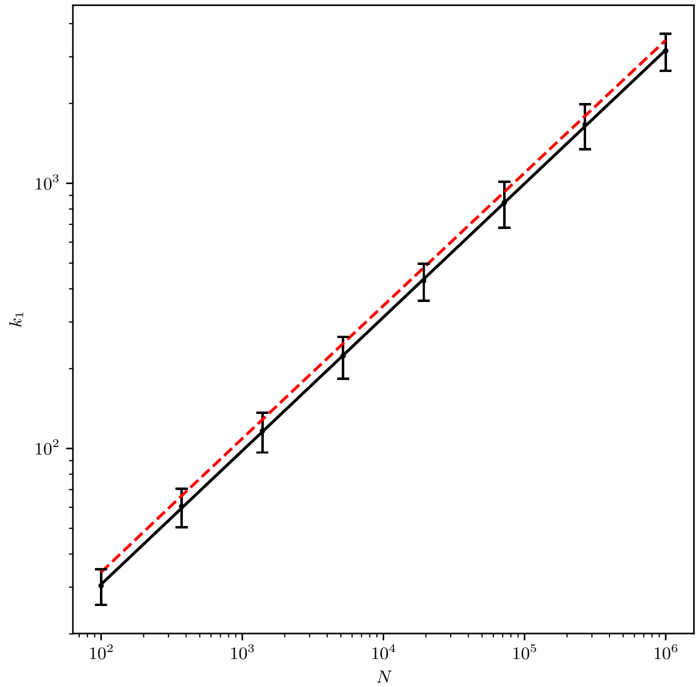



Figure 3: Largest degree for $m = 3$ from numerical results (black) averaged over 200 repeats, with standard deviation represented by the error bars, for $N = 10^n$, where $n$ are 8 equally spaced values between $2-6$ (rounded to integers), alongside theoretical largest expected degree $k_1$ (red dashed line). A linear regression is plotted through the mean numerical values resulting in a gradient of $0.504 \pm 0.002$. Removing small network sizes ($N < 1000$) resulted in a gradient of $0.499 \pm 0.003$.

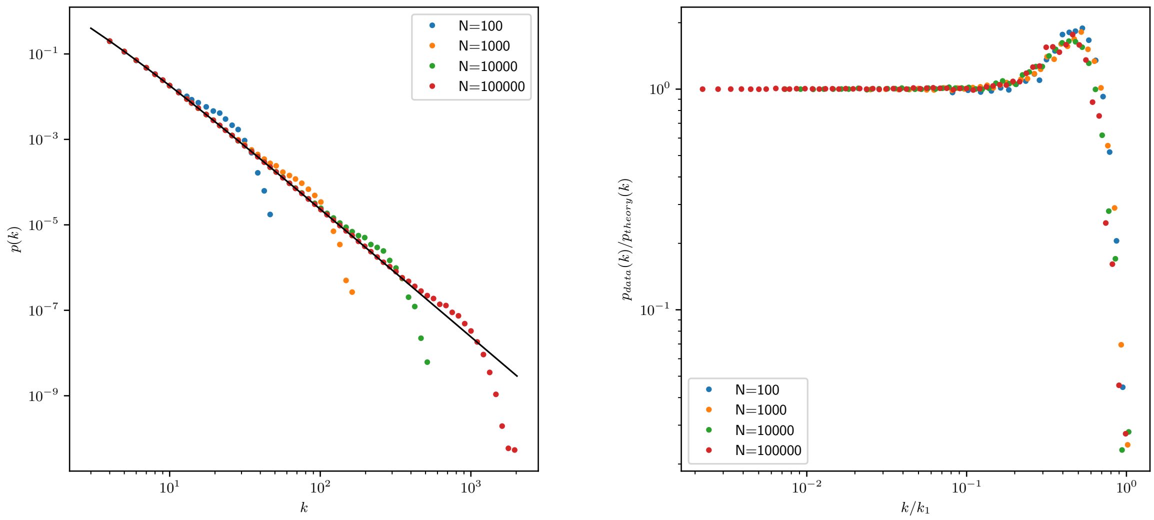
numerical results. Carrying out a linear regression on the numerical data results in a gradient of $0.504 \pm 0.002$ which is close to the theoretically expected $k_1 \sim N^{0.5}$. However there is a consistent trend that the numerical value for $k_1$ is smaller than the theoretical value. The numerical value is averaged over 200 repeats and the standard deviation on this value is used as the error. This result is more extreme for larger values of $m$, resulting in theoretical values that are larger than the upper error bars. This can be related to the assumption made in deriving the theoretical results that the probability distribution must reach a steady state that is invariant in time. However, this is only true in the $t \to \infty$ limit, which is analogous to $N \to \infty$. It would also be expected that larger values of $m$ require a longer time to reach the steady state, and so a value of $m = 3$ is used as this results in a small value of the ratio $m/N$ while also avoiding the most simple cases of $m = 1$ and $m = 2$. Removing small network sizes ($N < 1000$) from the linear regression in figure 3 results in a gradient of $0.499 \pm 0.003$ which confirms that larger networks tend towards the theoretical result.

The degree distribution for networks of different sizes is now investigated. Figure 4a shows that the width of the bump resulting from the finite size of the network in the tail of the distribution increases for larger network sizes, with the height of the bump also decreasing (bearing in mind that this is a log-log plot), showing how the finite size effects diminish for larger networks. In figure 4b the degree distributions from networks of different sizes are scaled by the largest degree found numerically and by the expected theoretical degree distribution. Thus the data is collapsed to $k/k_1 = 1$ and a deviation from $p_{data}(k)/p_{theory}(k) = 1$ implies a finite size effect. The finite size effects in the tail of the distribution can clearly be seen in this form.

## 3.2 Random Attachment

### 3.2.1 Degree distribution

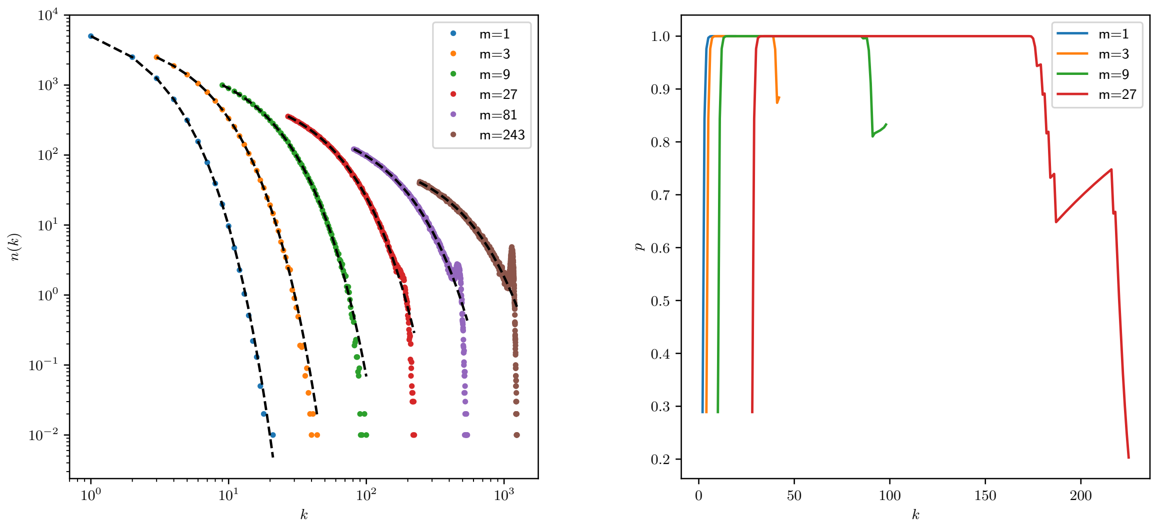
In the random attachment case, picking the existing vertex to connect to the new vertex is straightforward as this is just picked with uniform probability from a list of all existing vertices. In figure 5a the degree distribution is plotted for different values of $m$. It can be seen that the distribution is no longer fat-tailed and decays much faster than a power law, as suggested by the analytic distribution which is an exponential. Finite size effects are still observed in the tail of the distribution. In figure 5b a good fit is seen across a range of $m$ values, with the p-values from the KS test once again staying extremely close to 1.0 until the tail of the distribution.




(a) Log binned degree distribution.

(b) Data collapse.

Figure 4: Data collapse of numerical degree distributions for preferential attachment, for $m = 3$ and $N = 10^n$, where $n = 2, ..., 5$. Figure (a) shows the log-binned degree distribution data alongside the theoretical distribution (black line). Figure (b) shows the data collapse performed by plotting $k/k_1$ against $p_{data}(k)/p_{theory}(k)$.



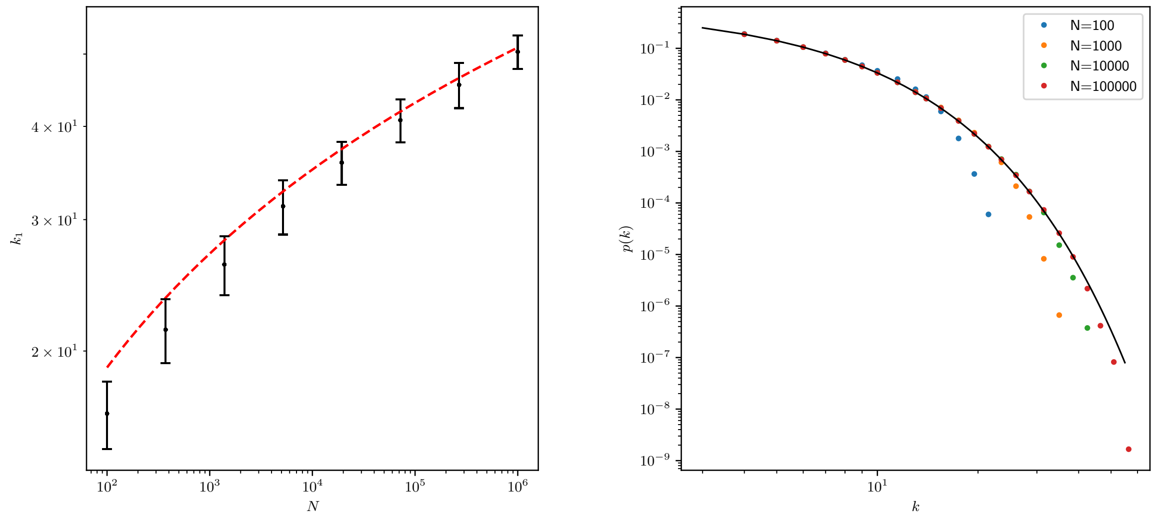

(a) Raw degree distribution.

(b) P-value variation.

Figure 5: Figure (a) shows the raw degree distribution data for random attachment plotted alongside the theoretically expected distribution for for $m = 3^n$, with $n = 0, .., 5$, for a network size of $10^4$. Figure (b) shows the resulting variation in KS test p-values for selected values of $m$ when including increasing amounts of data, demonstrating the effect of the finite size effect bump on the fit to the theoretical distribution.

### 3.2.2 Finite size effects

The largest expected degree from equation (27) is plotted alongside the numerical value in figure 6a for several $N$ values. In the random attachment case the numerical values get visibly closer to the expected values for larger network sizes. Plotting the numerical distribution for different $N$ values in figure 6b shows that this because the numerical distribution rapidly approaches the theoretical distribution for large $N$ values.



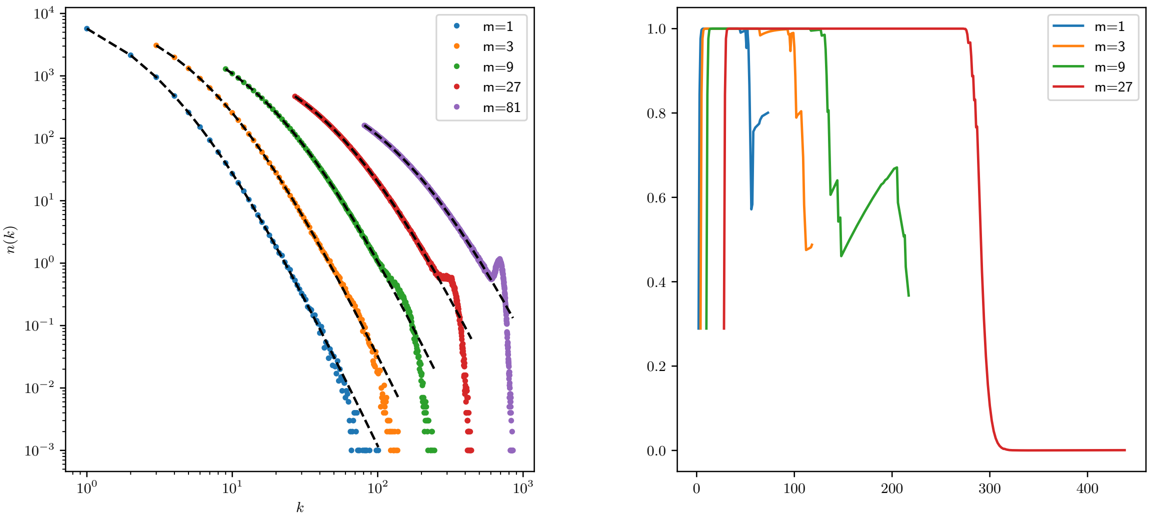

(a) Largest degree variation.

(b) Log-binned degree distribution.

Figure 6: Figure (a) shows the variation of $k_1$ with $N$ for $m = 3$, averaged over 200 repeats, with the error bars representing the standard deviation of $k_1$ values obtained, alongside the theoretical values (red dashed line). $N = 10^n$ where $n$ are 8 equally spaced values between $2 - 6$ (rounded to integers). Figure (b) shows the numerical degree distribution for $m = 3$ from numerical results (coloured points) for $N = 10^n$, where $n = 2, ..., 5$, alongside the theoretical degree distribution (black line).

## 3.3 Mixed Attachment

In the mixed attachment case, figure 7a shows that plotting the numerical and theoretical degree distributions for the $q = 0.5$ case reveals a distribution that falls off more quickly than the preferential attachment case but slower than the pure random case. Finite size effects are again seen in the tail of the distribution in the form of the characteristic bump. The fit is seen to be excellent, with the KS test resulting in a p-value of close to 1.0 up to this cut off in figure 7b. In figure 8 the numerical and theoretical largest expected degree is plotted as a function of network size. As in the preferential and random attachment cases, these values grow closer for larger values of $N$. A linear regression carried out on all numerical data points results in a gradient of $0.291 \pm 0.006$ and omitting the smaller network sizes ($N < 1000$) results in a gradient of $0.272 \pm 0.006$. Hence the value of the gradient approaches the expected 0.25 expected from the theoretical derivation for larger networks.




(a) Unbinned degree distribution.

(b) P-value variation.

Figure 7: Figure (a) shows log-binned degree distribution for mixed attachment plotted alongside the theoretically expected distribution for for $m = 3^n$, with $n = 0, .., 4$, for a network size of $10^4$. Figure (b) shows the resulting variation in KS test p-values for selected values of $m$ when including increasing amounts of data, demonstrating the effect of the finite size effect bump on the fit to the theoretical distribution.

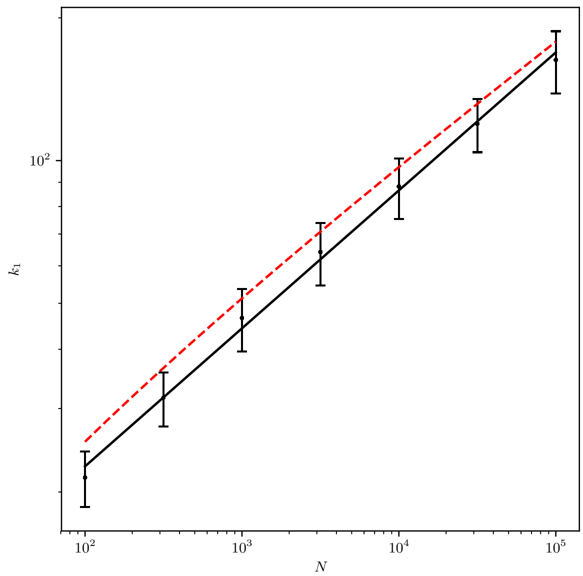



Figure 8: Largest degree for $m = 3$ from numerical results for mixed attachment (black) averaged over 100 repeats, with standard deviation represented by the error bars, for $N = 10^n$, where $n = 2, ..., 5$, alongside theoretical largest expected degree $k_1$ (red dashed line). A linear regression is plotted through the mean numerical values resulting in a gradient of $0.291 \pm 0.006$. Removing smaller network sizes ($N < 1000$) resulted in a gradient of $0.272 \pm 0.002$.

# 4 Conclusion

In this report, the BA model has been solved analytically for three attachment probabilities: preferential, random and mixed preferential/random. Preferential and mixed attachment resulted in fat-tailed, power law degree distributions wereas random attachment resulted in an exponentially decaying tail. These theoretical results were then compared to numerical simulations of the BA model, which matched well up to finite network size effects. It was found that in all cases, the finite size effects exhibited by the tail of the distribution that were quantified by the the largest degree $k_1$ diminished for larger network sizes, as long as a small $m$ value was selected. This convergence occurred more quickly for the random and mixed attachment compared to pure preferential attachment. Future work could analyse the large $N$ behaviour for larger $m$ values and investigate the effect of initial network configurations on finite size effects. An improved statistical analysis through goodness of fit tests better suited to discrete distributions would be beneficial.

# References


Albert, R. & Barabási, A.-L. (2002), 'Statistical mechanics of complex networks', *Rev. Mod. Phys.* **74**, 47–97.
**URL:** *https://link.aps.org/doi/10.1103/RevModPhys.74.47*

Arnold, T. B. & Emerson, J. W. (2011), 'Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions', *The R Journal* **3**(2), 34–39.
**URL:** *https://journal.r-project.org/archive/2011/RJ-2011-016/index.html*

Barabási, A.-L. & Albert, R. (1999), 'Emergence of Scaling in Random Networks', *Science* **286**(5439), 509–512.
**URL:** *http://science.sciencemag.org/content/286/5439/509*

Barlow, R. (1989), *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, Wiley.

Erdös, P. & Rényi, A. (1959), 'On random graphs, I', *Publicationes Mathematicae (Debrecen)* **6**, 290–297.
**URL:** *http://snap.stanford.edu/class/cs224w-readings/erdos59random.pdf*

Price, D. D. S. (1976), 'A general theory of bibliometric and other cumulative advantage processes', *Journal of the American Society for Information Science* **27**(5), 292–306.
**URL:** *https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.4630270505*

Simon, H. (1955), 'On a class of skew symmetric functions', *Biometrika* **42**, 425–440.
**URL:** *https://academic.oup.com/biomet/article/42/3-4/425/296312*