Network Project
A Growing Network Model

Thomas Pye

March 2022

Abstract: This report investigates the Barabasi-Albert network model, specifically the probability distribution of vertex degree and the distribution of largest degree the models. Several different types of attachment/model are investigated: Preferential Attachment, Random Attachment, and the Existing Vertex Model. These properties are numerically investigated by creating a model in `Python`, this was compared to theoretical predictions using Kolmogorov-Smirnov statistical tests. It was found that within 5% confidence the numerical data of the preferential attachment degree distribution only matched the theory for values of $m \leq 8$, however the largest degree did agree with theory for $m = 2$. For random attachment, the degree distribution matched the theoretical distribution for $m \geq 32$ and the largest degree distribution matched the theory. For the existing vertex model, the modelled degree probability distribution matched the theory for all values of m tested. This report also explores finite system size effects on the model

Word Count: `2488` words excluding font page, figure captions, table captions, acknowledgement and bibliography.

0 Introduction

The aim of this project is to investigate the properties of different models of growing networks and compare the numerical results to theoretical predictions using the Barabasi-Albert (BA) network model. The main properties in question are the degree probability distribution and the largest degree for each network. Effects such as finite-system size are also investigated.

Definition

The Barabasi-Albert network model is defined by the following algorithm:

1. Set up an initial network.
2. Increment the time $t \rightarrow t + 1$
3. Add a new vertex to the network
4. Create m new edges, connecting one end to the new vertex and the other end to an existing vertex chosen with probability Π. For preferential attachment, Π is proportional to the degree of each node.
5. Repeat (2) to (4) until N vertices have been added.

There are several rules for the connections made between vertex - each vertex must have a degree $k \geq m$, a vertex cannot be connected to itself, and each edge must be unique.

1 Phase 1: Pure Preferential Attachment Π_{pa}

1.1 Implementation

1.1.1 Numerical Implementation

The algorithm by which the BA model is implemented is as follows:

1. Initialisation: An initial complete simple network is created containing v_0 vertices. The attachment list is initialised as a flattened list of the ends of all edges in the initial graph.
2. Drive: At time t, add a source vertex to the graph, the label of this node is $v_0 + t$.
3. Add edges:
   - For the source vertex, create a 'disallowed list' of vertices to which new edges are not allowed to be connected. Add the new vertex to the disallowed list.
   - Randomly pick a target vertex from the attachment list. If it is in the disallowed list, pick again until it is not. Add the target vertex to the disallowed list to avoid loops.
   - Create an edge between the source vertex and the target vertex.
   - append the ends of this new edge to the attachment list.

- repeat this m times.

4. Repeat 2 and 3 for the required number of added vertices, N.

The degree of each node is stored using the `Networkx` package in python. The method of using a list of the ends of all the edges ("attachment list") to select the ends of the new edges generates preferential attachment because the number of appearances of a given vertex in this list is directly proportional to its degree.

1.1.2 Initial Graph

The initial graph created is a complete simple graph (i.e. a simple graph in which all vertices are connected to all other vertices) with m initial vertices. This is chosen so that all initial vertices have an equal degree and so non are favoured by the preferential attachment. This reduces the effects of the boundary conditions. If there are many more than m initial vertices it was found that the boundary conditions gave greater effects on the data, especially for larger m and smaller N.

1.1.3 Type of Graph

A simple (undirected, unweighted) network is produced by the model as this gave a clear representation of the underlying properties that were being investigated.

1.1.4 Working Code

A visualisation of the network at each time was created in order to see if there were any obvious faults. Whilst this is useful, it does not give a quantitative way of determining if the model is correct. Therefore, several numerical tests were undertaken in which the values in the model were compared to the expected values.

1. A simple test which checks that the number of vertices to be equal to the initial number plus the total number added $(N(t) = N(t=0) + t = v_0 + t)$

2. The number of edges is expected to be equal to the initial number of edges plus the number of edges added per new vertex times the number of new vertices $(E(t) = E(t=0) + mt = \frac{1}{2}v_0(v_0 - 1) + mt)$

3. In the long time limit, the average degree should be twice the number of edges added per vertex $(\lim_{t\to\infty}\langle k\rangle \to 2m)$

Each of these tests was checked for $m = 2, 4, 8, 16, 32, 64, 128$ and $N = 10, 100, 1000, 10000, 100000$ and repeated 100 times. It was found that all the numerical runs of the model agreed with the expectation and no exceptions were found.

1.1.5 Parameters

Input parameters:

- t, the number of new vertices to be added to the graph. By default this is set to 10^4, as the theoretical predictions about this model use the long-time approximation. This proved to be sufficiently large for any value of m used - such that the approximation could hold.

- m, the number of edges to be attached to each new vertex. By default this is set to 2 (which is the minimum, as $m = 1$ creates a 'tree'). m is not usually set any higher than 256 due to the time it takes for the programme to simulate such a model.

- v_0, the initial number of nodes in the graph. By default this is set to $v_0 = m$. This is the minimum required for the model and best reduces the effect of the initial boundary conditions.

1.2 Preferential Attachment Degree Distribution Theory

Begin with the master equation for a network using this model,

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

where $n(k, t)$ is the number of nodes in the model at time t with degree k, m is the number of edges added at each time-step, $\Pi(k, t)$ is the probability of picking a vertex with degree k at time t, and $\delta_{k,m}$ is the Kronecker Delta [1]. For preferential attachment, the probability of picking a vertex with degree k is

$$\Pi_{\text{pa}}(k, t) = \frac{k}{2mt} \ . \tag{2}$$

The long time ansatz assumes that the number of vertices with degree k is equal to the total number of vertices times the probability of a vertex having degree k:

$$n(k, t) = N(t)p_\infty(k) \tag{3}$$

where $p_\infty(k)$ is the probability of a vertex having degree k as $t \to \infty$. Substituting Equations (2) and (3) into the master equation (Equation (1)) yields

$$N(t+1)p_\infty(k) = N(t)p_\infty(k) + m\frac{k-1}{2mt}N(t)p_\infty(k-1) - m\frac{k}{2mt}N(k, t)p_\infty(k, t) + \delta_{k,m} \ . \tag{4}$$

As the long time limit is being used, it shall be assumed that $N(t = 0) = 0$ for simplicity. This is a valid approximation as the initial number of vertices is negligible as $N \to \infty$. In addition, it is clear that $N(t+1) = N(t)+1$, using this and cancelling in Equation (4) gives

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

First consider the case $k > m$: The fact that $\delta_{m,k} = 0$, and Equation (5) can be rearranged to give

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

which can be solved using the general form

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

where A is an arbitrary constant and Γ is the Gamma function, $\Gamma(n) \equiv (n-1)!$. Using Equation (7), Equation (6) gives

$$p_\infty(k) = \frac{A}{k(k+1)(k+2)} \quad \text{for } k > m \tag{8}$$

Now consider the case $k < m$: the degree of the a vertex cannot be lower than the number of edges added, hence

$$p_\infty(k) = 0 \quad \text{for } k < m \tag{9}$$

Finally the case $k = m$: Using that $\delta_{k,m} = 1$, Equation (5) can be rearranged to

$$p_\infty(m) = \frac{(m-1)p_\infty(m-1)}{m+2} + \frac{2}{m+2} \tag{10}$$

Note that in this case, $k = m - 1 < m$, hence the first term on the RHS is equal to zero according to (9). Hence

$$p_\infty(m) = \frac{2}{m+2} \quad \text{for } k = m \tag{11}$$

Combining (8) and (30) gives $A = 2m(m+1)$. Therefore:

$$p_\infty(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \quad \text{for } k \geq m \tag{12}$$

1.2.1 Theoretical Checks

The degree probability distribution $p_\infty(k)$ given in (12) should be normalised:

$$\sum_{k=m}^{\infty} p_\infty(k) = 1 \tag{13}$$

The form of $p_\infty(k)$ in (30) is written as a sum of partial fractions:

$$\begin{aligned}
\sum_{k=m}^{\infty} p_\infty(k) &= \sum_{k=m}^{\infty} 2m(m+1)\{\frac{1}{k} - \frac{2}{k+1} + \frac{1}{k+2}\} \\
&= 2m(m+1)\{\sum_{k=m}^{\infty}(\frac{1}{k} + \frac{1}{k+1}) - \sum_{k=m}^{\infty}(\frac{1}{k+1} - \frac{1}{k+2})\} \\
&= 2m(m+1)\{(\frac{1}{m} - \cancel{\frac{1}{m+1}} + \cancel{\frac{1}{m+1}} - ...) - (\frac{1}{m+1} - \cancel{\frac{1}{m+2}} + \cancel{\frac{1}{m+2}} - ...) \\
&= m(m+1)\{\frac{1}{m} - \frac{1}{m+1}\} = m(m+1)\{\frac{1}{m(m+1)}\} \\
&= 1
\end{aligned} \tag{14}$$

Hence the distribution is normalised, as expected.

1.3 Preferential Attachment Degree Distribution Numerics

1.3.1 Fat-Tail

A fat tailed distribution of $p(k)$ is one that decays slower than an exponential and has a significant probability of finding large values of k. A large range of k (several orders of magnitude) can have the same probability, creating a *'flat-tail'*.

Logarithmic binning (log binning) is used to deal with the problems that a fat-tailed distribution can cause by binning the data into bins with lower bound b_i and upper bound $b_{i+1} - 1$, where,

$$\frac{b_{i+1}}{b_i} = \exp(\Delta) \tag{15}$$

for a scale $\Delta > 0$. The density of data within each bin is returned as the y-data and the geometric mid-point of each bin is returned as the x-data. On a *log-log* plot the data will be equally spaced out on the x-axis at intervals of Δ.

In this study, a scale of $\Delta = 1.1$ was chosen (unless otherwise stated) to produce 'smooth' plots which are easy to interpret whilst still exhibiting important features.

1.3.2 Numerical Results

The model was run for a range of m, and the probability distribution of the degrees, $p_\infty(k)$, was investigated for each. This was done 100 times to give an average, $\tilde{p}_\infty(k)$. The data was log-binned with a scale of 1.1.

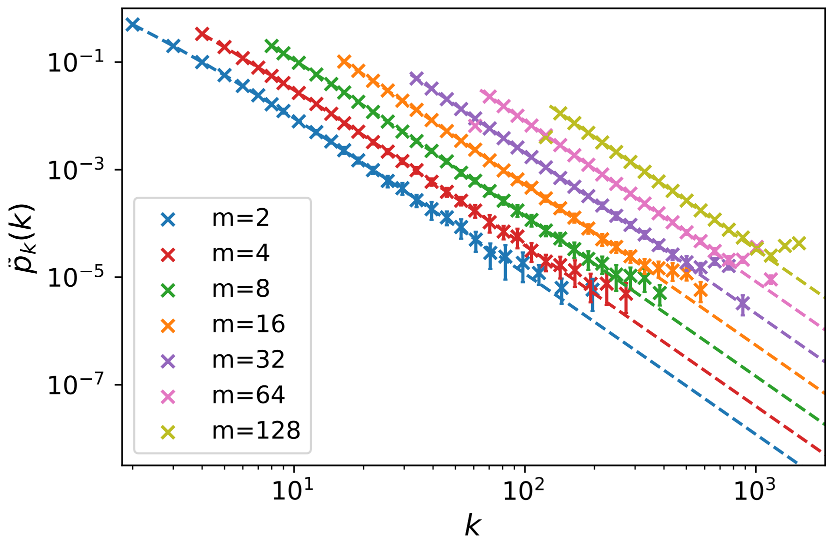

Figure 1: The probability distribution of the degrees of the nodes within the network for $N = 10^4$, plotted on a log-log scale. The dashed lines represent the theoretical distribution given in Equation (12). 100 repeats were undertaken. The error bars represent the standard error on the mean. It appears that for each m the data follows the prediction well until it starts to deviate upwards for smaller $\tilde{p}_\infty(k)$

.

In Figure 1, it can be seen that for larger values of $\tilde{p}_\infty(k)$, the numerical data appears to fit the theoretical prediction. However, for smaller values of $\tilde{p}_\infty(k)$, the data deviates from the theory. This may be attributed to finite-sized effects, where the vertices could have reached larger k if N was infinite; instead they have been cut off and contribute to flat-tail of the distribution at smaller $p_\infty(k)$.

1.3.3 Statistics

A statistical test is used for each value of m to test whether the numerical data fits the theoretical prediction. A two-sampled Kolmogorov-Smirnoff (KS) test is used to find the maximum difference between the observed and theoretical cumulative distributions. The null hypothesis, H_0, is that the observed and theoretical data are derived from the same distribution and the alternate hypothesis, H_1, is that they are not derived from the same distribution. A confidence level of 5% is chosen. Table 1 shows the p-value from the KS test for each value of m shown in Figure 1.

m	KS p-value	Result
2	0.456	Accept H_0
4	0.340	Accept H_0
8	0.148	Accept H_0
16	0.042	Reject H_0
32	0.012	Reject H_0
64	0.006	Reject H_0
128	0.001	Reject H_0

Table 1: The p-value result from the KS test for the data shown in Figure 1.

Table 1 shows that for $m = 2, 4$, the theoretical predictions and the numerical data are likely to be derived from the same distribution (within 5% of confidence), but for $m > 8$ the null hypothesis is rejected. This may be due to finite-sized effects.

1.4 Preferential Attachment Largest Degree and Data Collapse

1.4.1 Largest Degree Theory

The largest degree of all vertices is denoted k_1. The probability of randomly picking the node with the largest degree is $1/N$, where N is the total number of nodes. Hence $p_\infty(k_1) = \frac{1}{N}$. As k_1 is the largest degree, $p_\infty(k_1)$ can be written as the sum from k_1 to ∞:

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

Using the same reasoning as Equation (14)

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

rearranging leads to

$$k_1^2 + k_1 - Nm(m+1) = 0 \tag{18}$$

which is solved using the quadratic formula. Taking only the positive solution (as $k_1 > 0 \; \forall_{N,m}$) gives the theoretical prediction for the largest degree:

$$k_1 = \frac{-1 + \sqrt{1 + 4Nm(m+1)}}{2} \tag{19}$$

1.4.2 Numerical Results for Largest Degree

A numerical study of the largest degree, k_1 is undertaken for $m = 2$. A small value of m is chosen as large values are significantly more affected by the boundary effects, requiring the model to be run for a longer time. 100 runs of the model are carried out to get an average $\tilde{k_1}$. Figure 2 shows the numerical data from the model compared to the theoretical prediction.

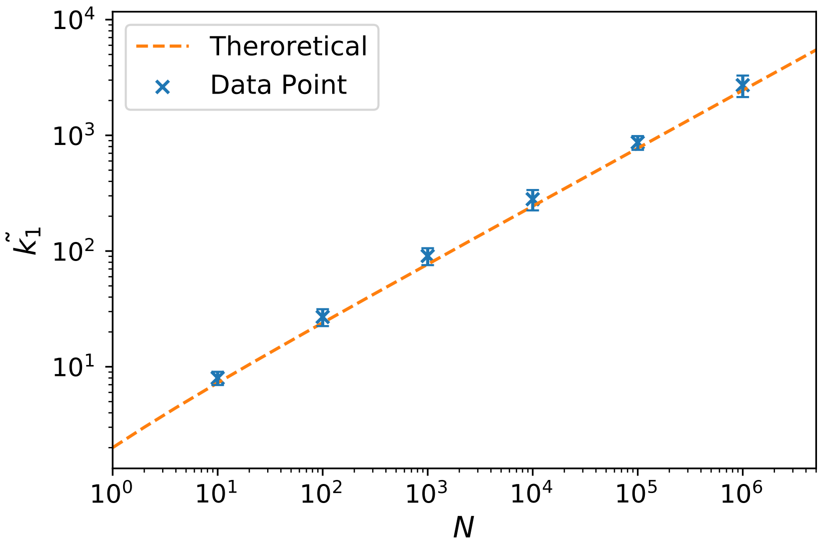

Figure 2: The distribution of k_1 for $m = 2$ over a range of N, shown on a log-log plot. A theoretical distribution is plotted in orange. The data points match the theory within their error ranges. The error bars represent the standard error on the mean.

A power law is fitted to the data in Figure 2 (not shown). This is summarised in Table 2.

Fit: $\tilde{k_1} = aN^b$

Constant	Value	Error
a	2.621	$\pm$ 0.103
b	0.504	$\pm$ 0.018

Table 2: Summary of the power law fit show in Figure 2. The fit of the data appears to be consistent with the theory, which predicts that k_1 scales as $N^{0.5}$ in the long time limit.

1.4.3 Data Collapse

Finite size effects are investigated by plotting a data collapse of the degree probability distribution. $p_\infty(k)$ can be written in terms of a scaling function, $\mathcal{F}$,

$$\tilde{p_\infty}(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \mathcal{F}\left(\frac{k}{k_1}\right). \tag{20}$$

Figure 3 shows this data collapse,

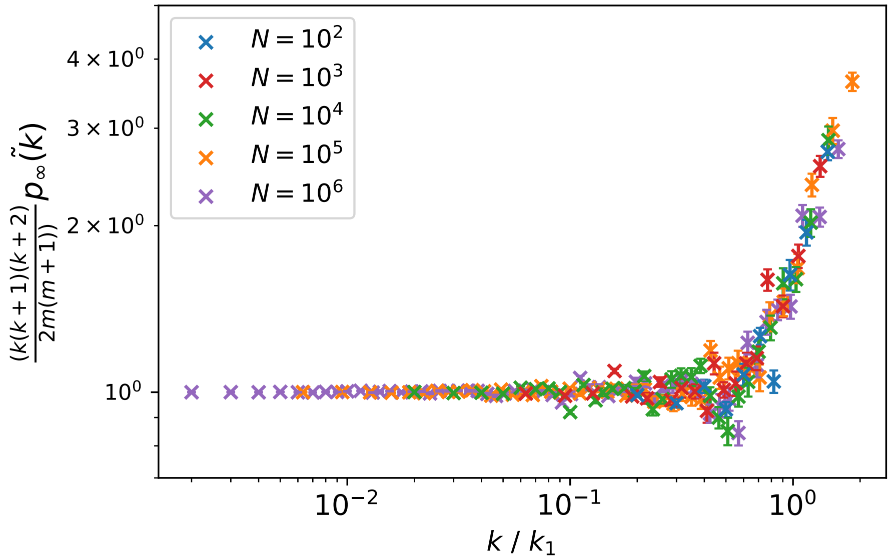

Figure 3: A data collapse of the degree probability distribution for $m = 2$ over a range of N. Notice that for smaller values on the x-axis, the data is collapsed onto $y = 1$. For larger values of x the data deviates away from this flat line.

In Figure 3, the presence of the deviation from $y = 1$ demonstrates the presence of the finite size effects discussed at the end of Section 1.3.2.

2 Phase 2: Pure Random Attachment Π_{rnd}

2.1 Random Attachment Theoretical Derivations

2.1.1 Degree Distribution Theory

For random attachment, the probability of picking a vertex is

$$\Pi_{\text{rnd}}(t) = \frac{1}{N(t)}. \tag{21}$$

Starting with the master equation (Equation 1), $\Pi_{\text{rnd}}(t)$ and the long time ansatz (Equation 3) can be substituted to give

$$N(t+1)p_\infty(k) = N(t)p_\infty(k) + m\frac{1}{N(t)}N(t)p_\infty(k-1) - m\frac{1}{N(t)}N(t)p_\infty(k) + \delta_{k,m}. \tag{22}$$

As the long time limit is being used, it shall be assumed that N (t = 0) = 0 for simplicity. One can also see that $N(t+1) = N(t) + 1$. Using this, Equation (22) becomes

$$p_\infty(k) = m(p_\infty(k-1) - p_\infty(k)) + \delta_{k,m}. \tag{23}$$

First consider the case $k > m$: $\delta_{k,m} = 0$, Equation (23) can be rearranged to

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{m}{m+1}. \tag{24}$$

Relabelling $k \to k-1$, (assuming $k-1 \geq m$), leads to:

$$p_\infty(k-1) = \frac{m}{m+1} p_\infty(k-2). \tag{25}$$

By combining Equations 24 and 25, it can be seen that

$$\begin{aligned} \frac{p_\infty(k)}{p_\infty(k-1)} &= (\frac{m+1}{m}) \frac{p_\infty(k)}{p_\infty(k-2)} \\ \implies \frac{p_\infty(k)}{p_\infty(k-2)} &= (\frac{m}{m+1}) \frac{p_\infty(k)}{p_\infty(k-1)} \end{aligned} \tag{26}$$

making a substitution using Equation (24),

$$\frac{p_\infty(k)}{p_\infty(k-2)} = \left(\frac{m}{m+1}\right)^2. \tag{27}$$

In fact,

$$\frac{p_\infty(k)}{p_\infty(k-n)} = \left(\frac{m}{m+1}\right)^n \text{ for } k-n > m \tag{28}$$

which can be proved by induction [2]. In the case that $k-n = m$, one sees that $n = k-m$, giving

$$\frac{p_\infty(k)}{p_\infty(m)} = \left(\frac{m}{m+1}\right)^{k-m} \text{ for } k > m \tag{29}$$

For the case $k < m$, $p_\infty(k) = 0$ as it is not possible for a vertex to have fewer nodes than the number of edges added.

Finally for the case $k = m$: Equation (23) becomes

$$p_\infty(m) = \frac{1}{m+1} \tag{30}$$

Substituting Equation (30) into Equation (29) results in the random attachment degree probability distribution,

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

This agrees with the literature [2].

This distribution must be normalised, this can be verified using Equation (13):

$$\begin{aligned} \sum_{k=m}^{\infty} p_\infty(k) &= \sum_{k=m}^{\infty} \frac{m^{k-m}}{(m+1)^{k-m+1}} \\ &= \frac{m^{-m}}{(m+1)^{1-m}} \sum_{k=m}^{\infty} \left(\frac{m}{m+1}\right)^k \end{aligned} \tag{32}$$

relabeling $k \to m+n$ allows the sum to take the form of a geometric sum:

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

where in the second line, the geometric sum to infinity, $\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$, is used. Hence the distribution given by Equation (31) is normalised as expected.

2.1.2 Largest Degree Theory

The probability of randomly picking the node with the largest degree is $1/N$, hence $p_\infty(k_1) = \frac{1}{N}$. As k_1 is the largest degree, $p_\infty(k_1)$ can be written as the sum from k_1 to ∞:

$$
p_\infty(k_1) = \sum_{k=k_1}^{\infty} p_\infty(k) = \frac{1}{N} \tag{34}
$$

Using (31) for $p_\infty(k)$,

$$
\begin{aligned}
p_\infty(k_1) &= \sum_{k=k_1}^{\infty} \frac{m^{k-m}}{(m+1)^{k-m+1}} \\
&= \frac{m^{-m}}{(m+1)^{m-1}} \sum_{k=k_1}^{\infty} \frac{m^{k}}{(m+1)^{k}} = \frac{1}{N}
\end{aligned} \tag{35}
$$

Relabeling $k \rightarrow k_1 + n$ allows the sum to take the form of a geometric sum:

$$
\begin{aligned}
p_\infty(k_1) &= \frac{m^{-m}}{(m+1)^{m-1}} \sum_{n=0}^{\infty} \frac{m^{n+k_1}}{(m+1)^{n+k_1}} \\
&= \frac{m^{-m}}{(m+1)^{m-1}} \frac{m^{k_1}}{(m+1)^{k_1}} \sum_{n=0}^{\infty} (\frac{m}{m+1})^{n} \\
&= \frac{m^{-m}}{(m+1)^{m-1}} \frac{m^{k_1}}{(m+1)^{k_1}} \frac{1}{1-\frac{m}{m+1}} \\
&= \frac{m^{k_1-m}}{(m+1)^{k_1+m-1}} (m+1) = \frac{1}{N}
\end{aligned} \tag{36}
$$

where in the second line, the geometric sum to infinity, $\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$, is used. Rearranging gives

$$(\frac{m}{m+1})^{k_1} = \frac{(\frac{m}{m+1})^m}{N}. \tag{37}$$

Now taking logarithms,

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

Hence

$$k_1 = \frac{m \ln\left(\frac{m}{m+1}\right) - \ln\left(N\right)}{\ln\left(\frac{m}{m+1}\right)} \tag{39}$$

which is confirmed by the literature [2].

2.2 Random Attachment Numerical Results

2.2.1 Degree Distribution Numerical Results

The model was run for a range of different m, and the probability distribution of the degrees was investigated for each. This was done 100 times to give an average, $\tilde{p}_\infty(k)$. The results are shown in Figure 4.

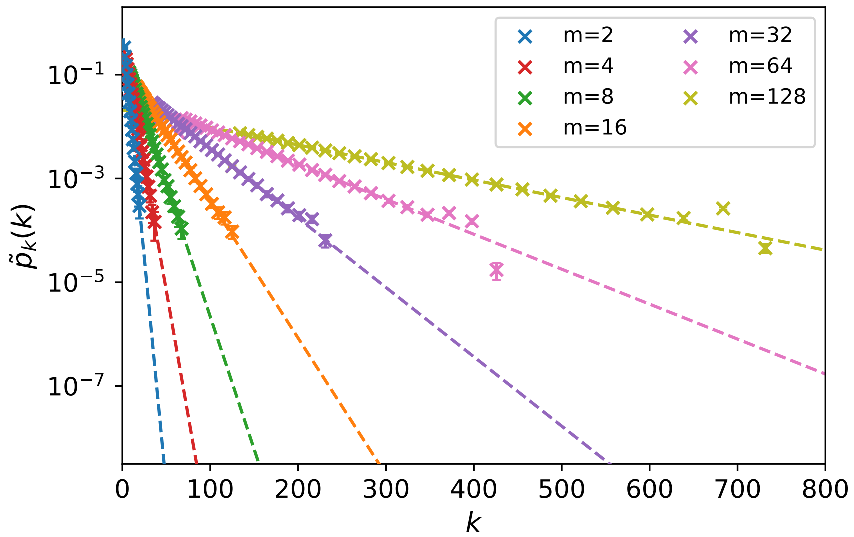

Figure 4: The probability distribution of the degrees of the nodes within the network for $N = 10^4$, plotted on a log-linear scale. The dashed lines represent the theoretical distribution given in Equation (31). 100 repeats were undertaken. The error bars represent the standard error on the mean.

Figure 4 shows that the random attachment model numerical data fits the theory well for larger values of $p_\infty(k)$ but is affected by finite size effects for smaller values.

A KS statistical test is carried out in the same way as Section 1.3.3, the results are summarised in Table 3.

m	KS p-value	Result
2	0.752	Accept H_0
4	0.700	Accept H_0
8	0.706	Accept H_0
16	0.536	Accept H_0
32	0.162	Accept H_0
64	0.025	Reject H_0
128	0.014	Reject H_0

Table 3: The p-value result from the KS test for the data shown in Figure 4

These results suggest that for all the values of $m \leq 32$ modelled, the numerical data is derived from the same distribution as the theoretical predictions. Thus for these m, the theoretical result is a good fit for the numerical data,

2.2.2 Largest Degree Numerical Results

The random attachment B.A. model is run for a range of N and the largest degree, k_1, is obtained from each. 100 runs are carried out for each N in order to obtain an average, $\tilde{k_1}$.

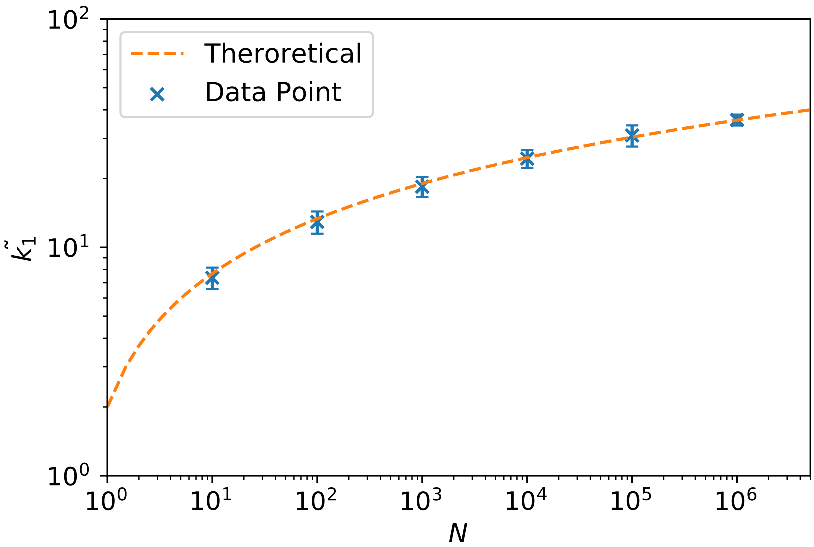

Figure 5: The largest degree of a network with $m = 2$ after various N vertices have been added using random attachment. The theoretical distribution (Equation 39) is plotted in orange. The error-bars represent the standard error on the mean.

A χ^2 statistical test, with a 5% confidence level, is undertaken to determine how well the data in Figure 5 matches the theoretical prediction (Equation 39). A statistic of $\chi^2 = 0.294$ is obtained, giving a p-value of 0.997. Therefore the data is consistent with the theory within a 5% level of confidence.

3 Phase 3: Existing Vertices Model

3.1 Existing Vertices Model Theoretical Derivations

For this model, the master equation becomes

$$\begin{aligned} n(k,t+1) = n(k,t) &+ r\Pi_{\mathrm{rnd}}(k-1,t)n(k-1,t) - m\Pi_{\mathrm{rnd}}(k,t)n(k,t) + \delta_{k,r} \\ &+ 2(m-r)\Pi_{\mathrm{pa}}(k-1,t)n(k-1,t) \\ &- 2(m-r)\Pi_{\mathrm{pa}}(k,t)n(k,t) \end{aligned} \tag{40}$$

where the first 4 terms represent the the random attachment of the new vertices and the last two terms represent the interconnection of existing vertices using preferential attachment.

Using Equations (2) and (21) for Π_{pa} and Π_{rnd} respectively, the long time ansatz (Equation (3)) can be used in this master equation to obtain

$$(k+m+2)p_\infty(k) = (k-m)p_\infty(k+1) + 2\delta_{k,\frac{m}{2}} \tag{41}$$

where the fact that this model uses $r = \frac{m}{2}$ has also been used.

First, the case $k > m$: $\delta_{k,\frac{m}{2}} = 0$, so Equation (41) can be written as

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k+m}{k+m+2}. \tag{42}$$

This can be solved using Equation (7) to give

$$p_\infty(k) = \frac{A}{(k+m+2)(k+m+1)(k+m)} \text{ for } k > m \tag{43}$$

for some constant A.

Now the case $k < m$: It does not make sense for a node to have a smaller degree than the minimum edges added, hence

$$p_\infty(k) = 0 \text{ for } k < m \tag{44}$$

Finally the case $k = m/2$: $\delta_{k,\frac{m}{2}} = 1$, so Equation (41) becomes

$$p_\infty(\frac{m}{2}) = \frac{1}{3m+4}. \tag{45}$$

Equating this to Equation (43) reveals that $A = 3m(3m+2)$, hence the degree probability distribution is

$$p_\infty(k) = \frac{3m(3m+2)}{2(k+m)(k+m+1)(k+m+2)} \text{ for } k \geq m. \tag{46}$$

3.2 Existing Vertices Model Numerical Results

A numerical simulation was run for the existing vertices model in order to obtain the degree probability distribution. Various values of m were run for a fixed value of N. For each value of m, the model was run 70 times to give an average $p_\infty(k)$ (fewer runs were used here due to the time taken for each run to complete). Figure 6 shows these results.

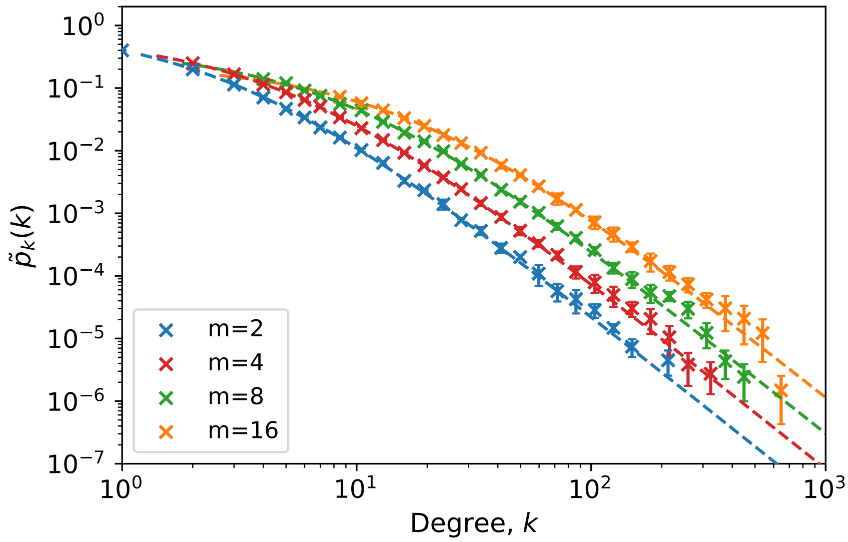

Figure 6: The degree probability distribution of the mixed attachment existing vertices model for $N = 10^4$. The dashed lines represent the theoretical prediction of the distribution given by Equation (46).

A KS statistical test - with 5% significance level - is carried out to determine how well the data fits the theoretical predictions for each m, the results are summarised in Table 4.

m	KS p-value	Result
2	0.620	Accept H_0
4	0.573	Accept H_0
8	0.222	Accept H_0
16	0.102	Accept H_0

Table 4: The p-value result from the KS test between the theoretical and numerical data shown in Figure 6.

Table 4 shows that for all of the values of m tested, the numerical data are a good fit for the theoretical predictions.

4 Conclusions

This project has investigated the distribution of degree probability and the largest degree of several different network models - preferential attachment, random attachment, and mixed attachment in the existing vertices model. The numerical data was compared to theoretical predictions and it was found that for preferential attachment, the data only matched the theory for smaller values of m, whilst for random and mixed attachment the data matched the theory for larger values of m, although quite a limited range was tested. Possible future work on this project could include a further investigation into the different dynamics of the existing vertices model (e.g. different weightings of random and preferential attachment).

References

[1] Evans, T. S. (2022). Complexity and Networks Course, Level 3 course. Imperial College London..

[2] Secular, P. (2015). Preferential and random attachment models of a complex network. Imperial College London.