Network Project
A Growing Network Model

CID: 01705049

March 26, 2023

Abstract: Models of network growth can provide insight as to how large-scale characteristics of networks emerge from microscopic attachment rules. This report discusses the numerical simulation of several models and the theoretical expectations for the properties of the generated networks - in particular the degre distribution and maximum-degree scaling with system size. The models investigated were found to be in agreement with theoretical predictions at the 5% significance level apart from one instance of an existing vertices model.

Word Count: 2482 words excluding front page, figure captions, table captions, acknowledgement and bibliography.

0 Introduction

Network science has attracted a lot of interest since the turn of the millennium, due in large part to the rise of the internet. More generally though, networks simply represent bilateral relations between a set of objects, and have therefore seen applications to a wide range of fields such as citation networks and archaeology. Since these are all examples of systems which evolve from the actions of individuals, the question of how large scale properties of these networks emerge from microscopic principles naturally arises.

Models of network growth can provide insight to this matter, and this report analyses the results of the numerical implementations of 3 such models. This report discusses the properties of the generated graphs, as well as any deviations found between theoretical expectations and collected data.

Definition

The definition of the BA model as used in this project was:

1. Generate an initial graph.
2. On an iteration of the algorithm, choose m vertices without replacement with a selection probability proportional to the degree of the vertex. Add a new vertex to the graph and connect edges between it and the chosen vertices.
3. Keep iterating until there are N vertices in the network.

All theoretical results were derived from the master equation (equation (1)). If multiple edges were allowed, then it would be possible for a vertex to have its degree increase than more than unity, which is not accounted for in the master equation. This choice has a disadvantage though; an assumption implicit in the mathematical results derived is that the probability distribution remains constant when choosing nodes. Since this method involves sampling without replacement, this cannot be exactly true. Nonetheless, it is a very good approximation when the number of nodes is large so, on balance, it was decided to disallow multiple edges between nodes.

1 Phase 1: Pure Preferential Attachment Π_{pa}

1.1 Implementation

1.1.1 Numerical Implementation

A standard library was used to generate network objects along with their associated attributes and methods. The algorithm was implemented as follows:

1. **Initialisation**

   - Generate an initial graph G_0. For each edge in the graph, add the vertices attached to a vector, V.

2. **Increment time**

   - Select m random elements from V uniformly. If the elements chosen are not unique, discard them and keep selecting random elements until a distinct set is generated.

- Add a new vertex to G. Connect the vertex to the m vertices chosen. For each edge added, append the vertices attached to them to V.

3. **Iteration**

   - Return to 2 until there are N vertices in the graph.

The number of times a vertex in V is the number of edges it has, and therefore selecting from this vector uniformly leads to a vertex being selected with a probability proportional to its degree, hence this algorithm is a correct implementation of preferential attachment.

1.1.2 Initial Graph

Since the behaviour that results from the algorithm itself rather than the initial conditions is of interest, a small initial graph is ideal. A graph with $m + 1$ vertices was generated, with each vertex having m edges. This guarantees that $p(k, t) = 0$ for $k < m$, where $p(k, t)$ denotes the probability that a node is of degree k at time t. Since this is true in the large time limit, regardless of the initial conditions, the graphs created will always possess this characteristic even with a finite number of iterations, meaning the results should be slightly better aligned to theoretical predictions.

1.1.3 Type of Graph

A simple graph is produced since there is at most one vertex between nodes and there is no direction or weighting to edges. The graph is also sparse; for large N, we have approximately mN edges, so the number of edges is of the order of the number of nodes.

1.1.4 Working Code

A number of sanity tests were performed to ensure the programme was behaving as expected. These included verifying the number of vertices added to the V was equal to $2m$ on each time step, and the number of nodes in the graph at the end of the algorithm was equal to N using print statements. A system was also generated and it was ensured the probability that a node was chosen was actually proportional to its degree, the result of which is demonstrated in Figure 1.

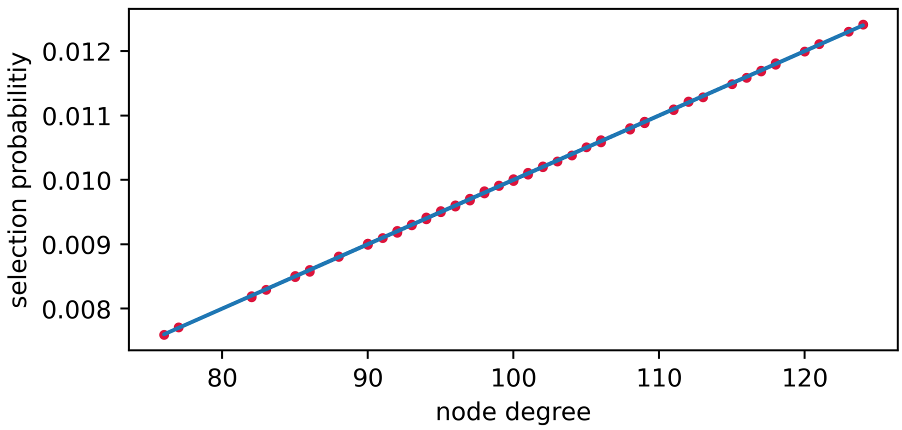

Figure 1: Measured selection probability after sampling 10^8 times for a system of 100 nodes. The line of best fit has an intercept close to 0 (-3.27×10^{-8}) so the relationship is one of directly proportionality as desired for the BA model.

1.1.5 Parameters

The programme only needs 2 parameters, the number of vertices in the final graph desired, N, and the number of vertices added on each time step m. $N = 10^6$ was chosen for this investigation. A large value for N is desirable as it will be closer to the infinite time limit, and the probability distribution for values chosen for m ranged from $m = 2$ to $m = 32$. In a sense, these specific numbers are arbitrary, but very large values for m should be ignored. Recall from the discussion in the definition of the algorithm, it was noted that if multiple edges between vertices are forbidden then the probability for choosing a vertex changes slightly as a consequence of selecting vertices without replacement on a given time step. This behaviour is not accounted for in the theoretical work. Therefore, the higher the value of m, the greater the deviation from our simulation and idealised behaviour imposed by the master equation will be.

1.2 Preferential Attachment Degree Distribution Theory

1.2.1 Theoretical Derivation

The number of nodes of degree k at time $t+1$, on average, is given by the master equation

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

Assume, for now, that $k \neq m$, and let $N(t)$ be the total number of vertices in the network at time t. Using the relations $n(k,t) = N(t)p_\infty(k)$ and $N(t+1) = N(t)+1$, equation (1) becomes

$$p_\infty(k) = m\Pi(k-1,t)N(t)p_\infty(k) - m\Pi(k,t)N(k,t)p_\infty(k). \tag{2}$$

In preferential attachment, $\Pi(k,t) \propto k$. Requiring this probability distribution to be normalised - and noting that two vertices are required to form an edge - yields

$$\Pi(k,t) = \frac{k}{2E(t)}, \tag{3}$$

where $E(t)$ is the total number of edges at time t. If the number of edges on the initial graph is E_0, then

$$E(t) = mt + E_0. \tag{4}$$

Furthermore, since we add one vertex and m vertex at each time-step

$$N(t) = \frac{E(t) - E_0}{m} + N_0, \tag{5}$$

if N_0 is the number of vertices in our initial graph. In the large time limit, E_0 and N_0 can be neglected when substituting this expression for $N(t)$ into equation (2); doing so yields,

$$\begin{aligned} p_\infty(k) &= m\frac{k-1}{2E(t)}\frac{E(t)}{m}p_\infty(k-1) - m\frac{k}{2E(t)}\frac{E(t)}{m}p_\infty(k) \\ &= \frac{1}{2}\left[(k-1)p_\infty(k-1) - kp_\infty(k)\right]. \end{aligned} \tag{6}$$

Simple rearrangement of this expression leads to

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k-1}{k+2}. \tag{7}$$

Consider the trial solution

$$p_\infty(k) = A\frac{\Gamma(k+1+a)}{\Gamma(k+1+b)}, \tag{8}$$

where $\Gamma(z)$ denotes the gamma function and a and b are constants to be determined. Substituting equation (8) into equation (7) produces

$$\begin{aligned}\frac{k-1}{k+2} &= A\frac{\Gamma(k+1+a)}{\Gamma(k+1+b)}\left(A\frac{\Gamma(k+a)}{\Gamma(k+b)}\right)^{-1} \\ &= \frac{(k+a)(k+a-1)...k\Gamma(k)}{(k+b)(k+b-1)...k\Gamma(k)} \times \frac{(k+b-1)(k+b-2)...k\Gamma(k)}{(k+a-1)(k+a-2)...k\Gamma(k)} \\ &= \frac{k+a}{k+b},\end{aligned} \tag{9}$$

where the property of the gamma function $z\Gamma(z) = \Gamma(z+1)$ has been used. This implies the solution to equation (7) is

$$p_\infty(k) = A\frac{\Gamma(k)}{\Gamma(k+3)} = \frac{A}{k(k+1)(k+2)}. \tag{10}$$

To find the normalisation constant A, equation (6) must be returned to. Consider the case $k = m$. Note that 1 must be added to the RHS since $\delta_{k,m}$ was dropped from going to equation (2) to equation (1) under the assumption $k \neq m$. If $p_\infty(k) = 0$ is taken for $k < m$ then equation (6) can be rearranged to obtain

$$A = 2m(m+1), \tag{11}$$

and thus

$$p_\infty(k) = \begin{cases} \frac{2m(m+1)}{k(k+1)(k+2)} & \text{if } k \geq m \\ 0 & \text{if } k < m \end{cases} \tag{12}$$

1.2.2 Theoretical Checks

It is required that $p_\infty(k)$ is normalised, i.e.

$$\sum_{k=m}^{\infty} p_\infty(k) = \sum_{k=m}^{\infty} \frac{2m(m+1)}{k(k+1)(k+2)} = 1. \tag{13}$$

This can be verified by invoking the method of differences to evaluate the sum

$$\begin{aligned}\sum_{k=m}^{\infty} \frac{2m(m+1)}{k(k+1)(k+2)} &= m(m+1) \lim_{n\to\infty} \sum_{k=m}^{n} \left(\frac{1}{k} - \frac{2}{k+1} + \frac{1}{k+2}\right) \\ &= m(m+1) \lim_{n\to\infty} \sum_{k=m}^{n} \left(\frac{1}{k} - \frac{1}{k+1} + \frac{1}{k+2} - \frac{1}{k+1}\right) \\ &= m(m+1) \lim_{n\to\infty} \left(\frac{1}{m} - \frac{1}{n+1} - \frac{1}{m+1} + \frac{1}{n+2}\right) \\ &= m(m+1)\left(\frac{1}{m} - \frac{1}{m+1}\right) = m(m+1)\left(\frac{1}{m(m+1)}\right) = 1.\end{aligned} \tag{14}$$

In the limit of large k, our expression is also consistent with the expectation from the literature that $p_\infty(k) \propto k^{-3}$ for the BA model. The average degree size should also be equal to $2m$:

$$\begin{aligned}
\sum_{k=m}^{\infty} \frac{2mk(m+1)}{k(k+1)(k+2)} &= \sum_{k=m}^{\infty} \frac{2m(m+1)}{(k+1)(k+2)} \\
&= 2m(m+1) \lim_{n\to\infty} \sum_{k=m}^{n} \left(\frac{1}{k+1} - \frac{1}{k+2} \right) \\
&= 2m(m+1) \lim_{n\to\infty} \left(\frac{1}{m+1} - \frac{1}{n+2} \right) \\
&= 2m(m+1) \left(\frac{1}{m+1} \right) = 2m.
\end{aligned} \tag{15}$$

1.3 Preferential Attachment Degree Distribution Numerics

1.3.1 Fat-Tail

The distribution that has just been derived is an example of a fat-tailed distribution, meaning there is a significant probability of observing vertices with a very large degree. There will be a very small number of observing these vertices, and, therefore, a large associated uncertainty in the probability for $p(k)$ in this range. Moreover, given the large variation in scale in the data collected, there will be many degrees for which there are no observed vertices, which makes conducting certain statistical tests on the data collected problematic. The former was solved by running each system over multiple realisations and then averaging out the data so the probability of observing values for large k is more accurate. The latter was solved by using log-binning, which uses bins increasing exponentially in size, which effectively averages out the degree distribution over , making it more amenable to analysis.

1.3.2 Numerical Results

Systems of size $N = 10^6$ were simulated for values of m ranging from $m = 2$ to $m = 32$ over 10 realisations (Figure 2). Note for large k, a power law is observed since the graph is linear on a log-log plot but for small k for $m = 2$ there is a slight divergence as predicted by equation (12). For the largest k, the degree distribution diverges with a characteristic bump; vertices which would have gone on to have a larger degree had N been larger are over-represented and thus vertices with a very high degree are more likely than might be naïvely expected.

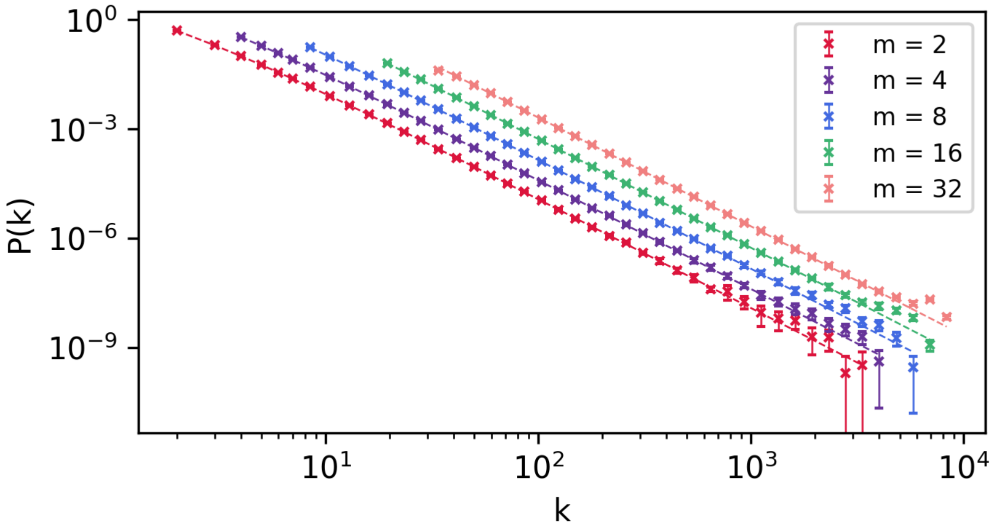

Figure 2: Degree distribution, $P(k)$, for the BA model for various values of m, over 10 realisations for $N = 10^6$. Error-bars represent 95% confidence intervals. Dashed lines represent theoretical fit. Note that for the larger values of m, the error bars do not pass through theoretical distribution, suggesting that the deviation is a genuine finite size effect and not due to insufficient statistics at the tail.

1.3.3 Statistics

In order to compare the data to the theoretical distributions derived, a chi-squared test was performed. This was chosen over Kolmogorov-Smirnov since it assumes a continuous distribution. Although this is perhaps a very good approximation given the large range of scale in the data, the chi-squared test does not require any other assumptions that Kolmogorov-Smirnov does not, such as statistical independence. Furthermore, since the tails are affected by finite-size effects, they were excluded from the hypothesis test, since what is being attempted is to determine how accurate the expression for the infinite limit is. The results are shown in Table 1. All results have a p-value close to 1 and pass at the 5% significance level, suggesting a good fit to the data.

m	reduced χ^2 (3 s.f.)	p-value
2	0.178	>0.999
4	0.200	>0.999
8	0.394	>0.999
16	0.315	>0.999
32	0.410	>0.999

Table 1: Results from chi-squared tests on collected data excluding data from the tail affected by finite-size effects for the BA model.

1.4 Preferential Attachment Largest Degree and Data Collapse

1.4.1 Largest Degree Theory

Assume there is only one vertex with the largest degree, k_1. Since there are N nodes, this implies

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

Adapting the expression obtained from equation (14) transforms equation (15) to give

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

Rearrangement yields the quadratic equation

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

which can readily be solved to obtain

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

The theoretical expectation, therefore, for the scaling of k_1 in the limit of large time is $k_1 \propto N^{\frac{1}{2}}$.

1.4.2 Numerical Results for Largest Degree

To investigate finite size effects, a value of $m = 5$ was chosen and graphs of sizes ranging from $N = 10^2$ to $N = 10^6$ were generated (Figure 3). The value for m was chosen since it was small (much less than the smallest system of $N = 100$) and so the results are reasonably independent from the initial conditions. Each system was realised 60 times and the max degree k_1 was measured and the standard error on the mean was used to estimate the uncertainty in the measured value (Figure 4).

To verify the scaling relation found, a hypothesis test using Pearson's correlation coefficient was used of k_1 against $\sqrt{N}$ was conducted. An r-value of over 0.999 was found, with an associated p-value of $1 - 5.3 \times 10^{-9}$, and thus the simulation is consistent with this theoretical expectation. However, as evident in Figure 3, the proportionality constant between the $\sqrt{N}$ and k_1 derived is incorrect since there is a clear offset. This is likely caused by the assumption used in equation (16). What is implied by it is that there exists one node in the region from k_1 to infinity, and it is not necessarily at the start of the region being summed over, so this approach is not expected to give an exact result.

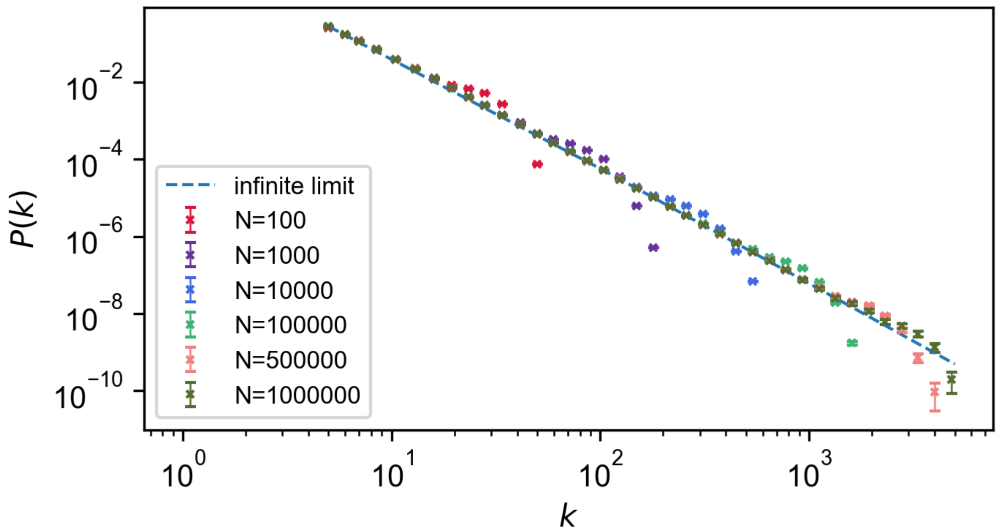

Figure 3: Degree distribution, $P(k)$, for the BA model for various values of system sizes, N, for $m = 5$, averaged over 60 realisations. The distribution for each value of N are coincident until the finite-size effects of each system causes them to diverge.

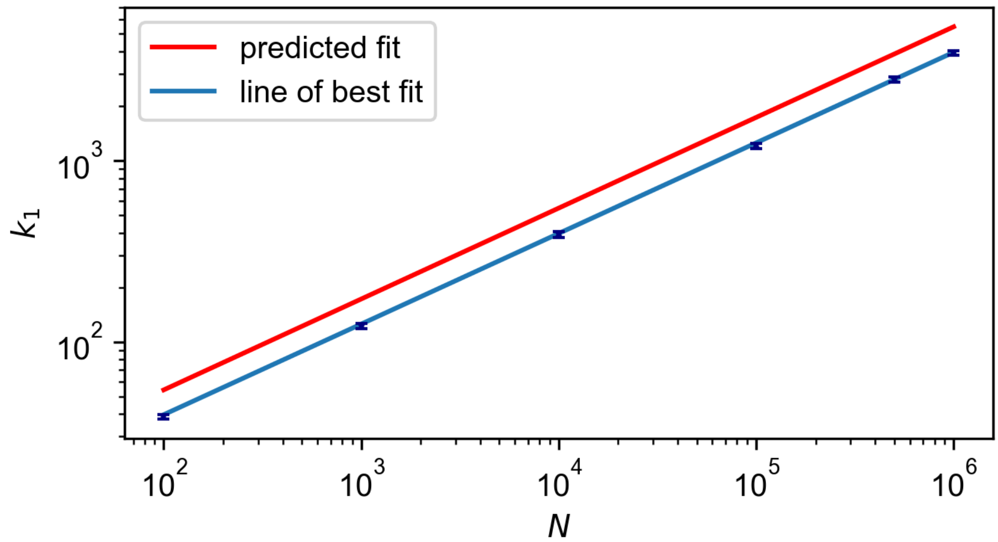

Figure 4: Average largest degree observed k_1, against system size N for $m = 5$. Error bars represent 95% confidence intervals. Despite this, the correlation coefficient of k_1 against $\sqrt{N}$ had an associated p-value very close to one, suggesting the scaling relation was still correct.

1.4.3 Data Collapse

A value of $m = 5$ was used since it was small, so the initial graph is small and thus the finite size effects are not influenced greatly by the initial conditions - a smaller value for m was not used since the data produced tended to be noisy. System sizes ranging from $N = 10^2$ to $N = 10^6$ were used. From equation (12), a data collapse my be expected from multiplying $P(k)$ with $k(k+1)(k+2)$ and plotting against k. However, normalisation of the x-axis to account for the finite size effects. Since k_1 scales with $\sqrt{N}$, $P(k)k(k+1)(k+2)$ was plotted against $k/\sqrt{N}$, the results of which are shown in Figure 5.

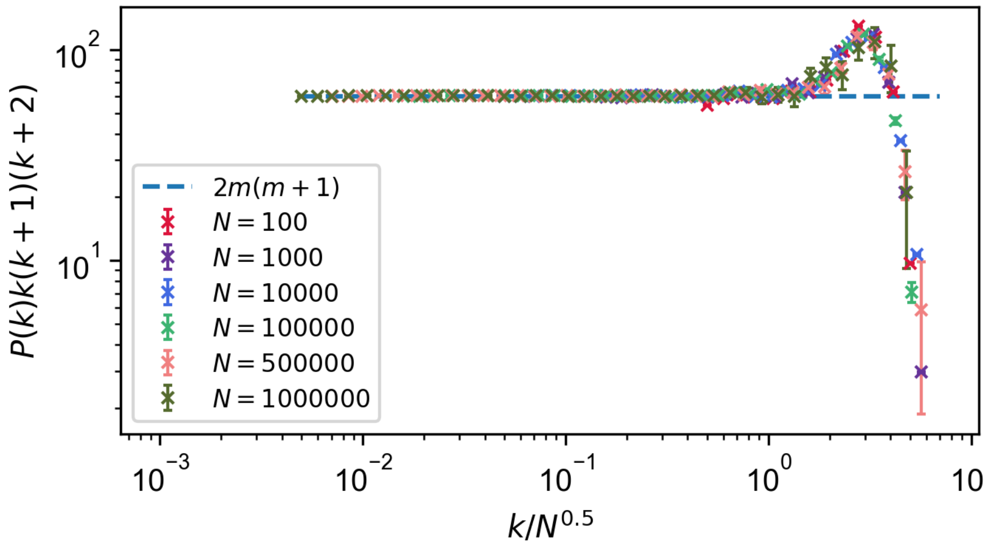

Figure 5: Data collapse for the degree distribution for the BA model.

2 Phase 2: Pure Random Attachment Π_{rnd}

2.1 Random Attachment Theoretical Derivations

2.1.1 Degree Distribution Theory

Return to equation (1). In random attachment, $\Pi(k,t) = 1/N(t)$. Applying this, along with the relations $n(k,t) = p_\infty(k)N(t)$, and $N(t+1) = N(t)$, yields

$$p_\infty(k) = mp_\infty(k-1) - mp_\infty(k) \tag{20}$$

in the case $k \neq m$. Rearrangement obtains,

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{m}{m+1}, \tag{21}$$

and therefore $p_\infty(k)$ is a geometric sequence of the form

$$p_\infty(k) = p_\infty(m)\left(\frac{m}{m+1}\right)^{k-m}, \tag{22}$$

where $p_\infty(k) = 0$ for $k < m$ has been used to fix the normalisation constant as $p_\infty(m)$. By substituting $k = m$ into equation (20) (and also being mindful to insert 1 to the RHS for this special case), $p_\infty(m)$ can be solved for to obtain

$$p_\infty(m) = \frac{1}{m+1}, \tag{23}$$

finally leading to

$$p_\infty(k) = \begin{cases} \frac{1}{m+1}\left(\frac{m}{m+1}\right)^{k-m} & \text{if } k \geq m \\ 0 & \text{if } k < m \end{cases} \tag{24}$$

Evaluating the following sum using the formula for an infinite geometric series gives

$$\begin{aligned}\sum_{k=m}^{\infty} p_\infty(k) &= \sum_{k=m}^{\infty}\left(\frac{1}{m+1}\left(\frac{m}{m+1}\right)^{k-m}\right) \\ &= \frac{1}{m+1}\frac{1}{(1-\frac{m}{m+1})} \\ &= \frac{1}{m+1}\frac{m+1}{1} = 1,\end{aligned} \tag{25}$$

and thus the normalisation of $p_\infty(k)$ has been verified. This expression must also be consistent with the average degree being equal to $2m$. To prove this, let

$$S = \sum_{k=0}^{\infty}\left(\frac{m}{m+1}\right)^{k-m} = m+1. \tag{26}$$

Then

$$\begin{aligned}\sum_{k=m}^{\infty}\frac{k}{m+1}\left(\frac{m}{m+1}\right)^{k-m} &= \frac{1}{m+1}\left(m+(m+1)\left(\frac{m}{m+1}\right)+(m+2)\left(\frac{m}{m+1}\right)^2+...\right) \\ &= \frac{m}{m+1}\left(1+\left(\frac{m}{m+1}\right)+\left(\frac{m}{m+1}\right)^2+...\right) \\ &\quad + \frac{m}{(m+1)^2}\left(1+2\left(\frac{m}{m+1}\right)+3\left(\frac{m}{m+1}\right)^2+...\right) \\ &= \frac{mS}{m+1}+\frac{m}{(m+1)^2}\left(S+S\frac{m}{m+1}+S\left(\frac{m}{m+1}\right)^2+...\right) \\ &= 2m,\end{aligned} \tag{27}$$

as required.

2.1.2 Largest Degree Theory

Similar to the largest degree derivation for the BA model, we assume there is only one vertex with the largest degree, k_1, and so

$$\sum_{k=k_1}^{\infty} p_\infty(k) = \frac{1}{N}. \tag{28}$$

Substituting equation (24) gives

$$\begin{aligned}\frac{1}{m+1}\left(\frac{m}{m+1}\right)^{k_1-m}+\frac{1}{m+1}\left(\frac{m}{m+1}\right)^{k_1-m+1}+... &= \frac{1}{m+1}\left(\frac{m}{m+1}\right)^{k_1-m}\left(1+\frac{m}{m+1}+...\right) \\ &= \frac{1}{m+1}\left(\frac{m}{m+1}\right)^{k_1-m}\left(\frac{1}{1-\frac{m}{m+1}}\right) \\ &= \left(\frac{m+1}{m}\right)^{k_1-m} \\ &= \frac{1}{N}.\end{aligned} \tag{29}$$

Solving for k_1 yields

$$k_1 = m + \frac{\log(N)}{\log(\frac{m+1}{m})}, \tag{30}$$

and therefore $k_1 \propto \log(N)$ in the limit of large N.

2.2 Random Attachment Numerical Results

2.2.1 Degree Distribution Numerical Results

The same parameters and procedure was performed identically to phase 1 aside from using random attachment (Figure 6). Finite-size effects are also observed in the tail of the distributions. The chi-squared test was once again used to test agreement between the measured and theoretical distributions, and the high p-values suggest the theoretical expression found was a good fit (Table 2).

m	reduced χ^2 (3 s.f.)	p-value
2	0.223	>0.999
4	0.119	>0.999
8	0.104	>0.999
16	0.169	>0.999
32	0.149	>0.999

Table 2: Results from chi-squared tests on collected data excluding data from the tail affected by finite-size effects for random attachment model. P-values suggest excellent agreement between numerical data and theory.

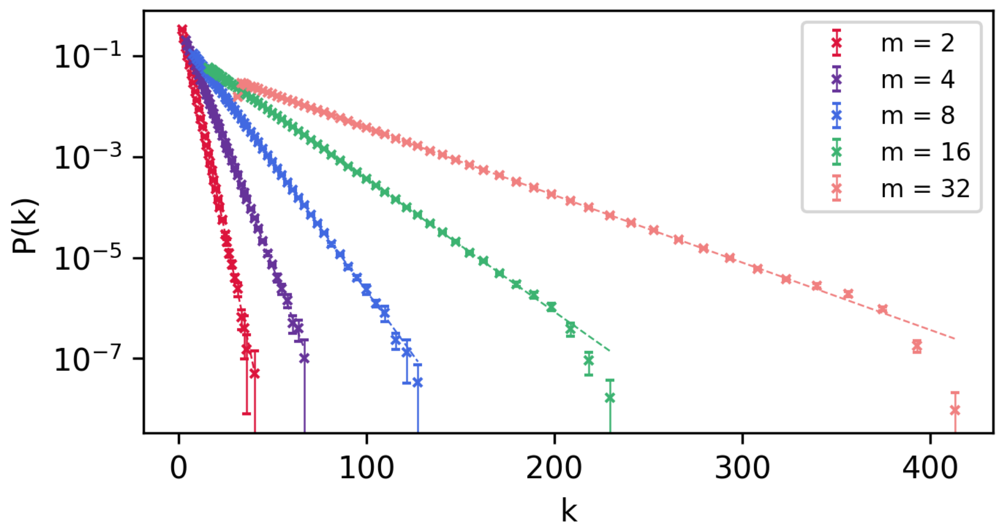

Figure 6: Degree distribution, $P(k)$, for the random attachment model for various values of m, over 10 realisations for $N = 10^6$. Dashed lines represent theoretical fit. Error-bars represent 95% confidence intervals. Note that for larger m, the error bars do not pass through theoretical distribution, suggesting that the deviation is a genuine finite size effect and not due to insufficient statistics at the tail.

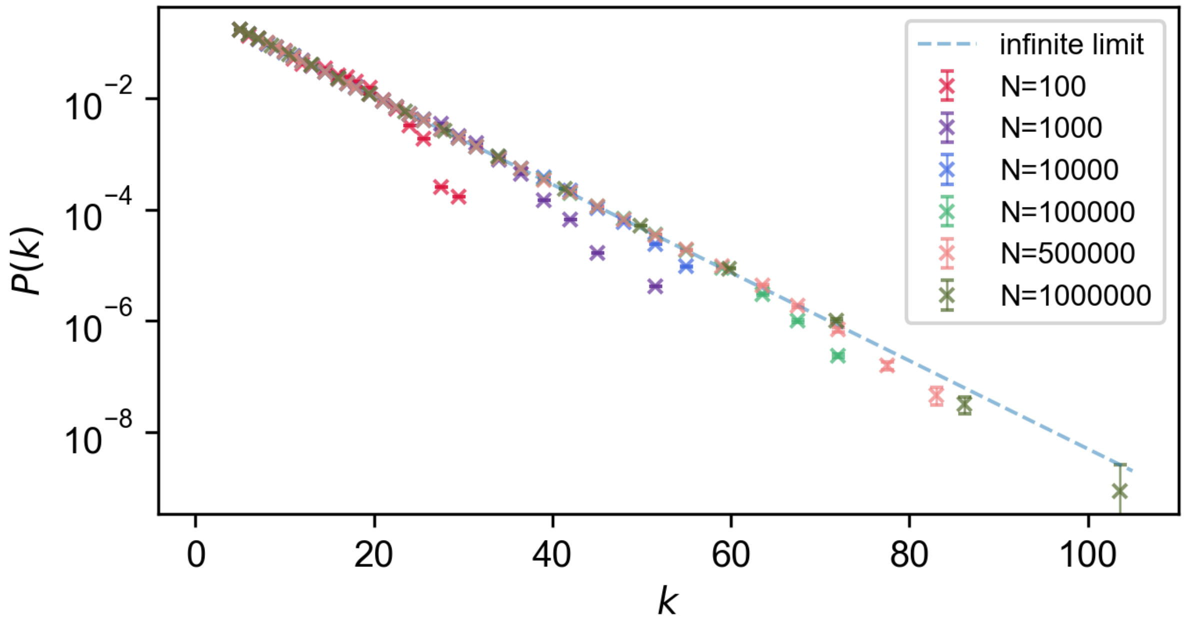

Figure 7: Degree distribution, $P(k)$, for the random attachment model for various values of system sizes, N, for $m = 5$, averaged over 60 realisations. The distribution for each value of N are coincident until the finite-size effects of each system causes them to diverge.

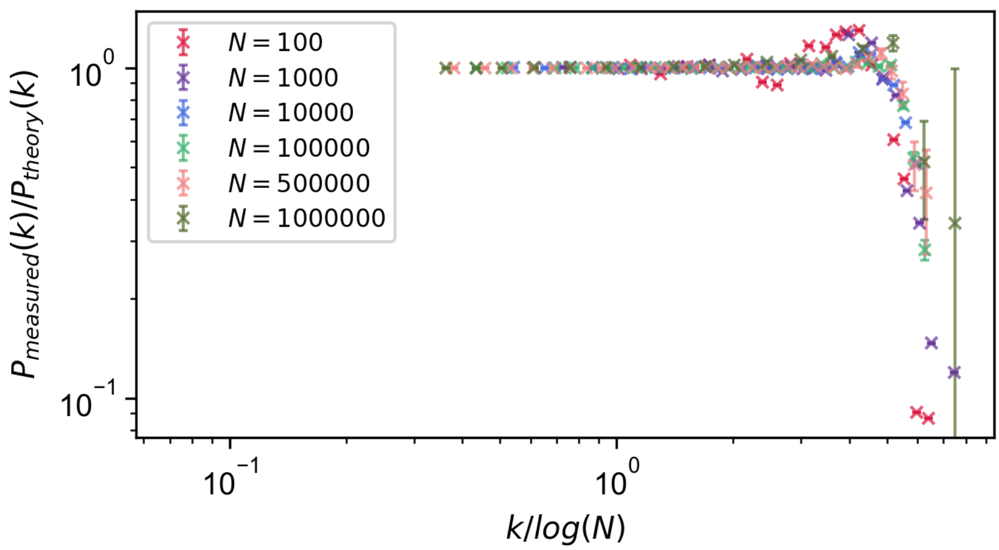

Figure 8: Data collapse for the degree distribution for the random attachment model.

2.2.2 Largest Degree Numerical Results

The same parameters were used as phase 1, for essentially the same reasons as before. To ascertain whether the scaling relation derived was consistent with the numerical results, Pearson's correlation coefficient of was once again used with $\log(N)$ against k_1 with an associated p-value of 0.9997 and thus is a good fit, but once again the proportionality constant is not correct for the same reason given in section 1.4.2. Following similar logic to phase 1, a data collapse was produced by plotting the measured probability distribution divided by the theoretical distribution against $k/\log(N)$ (Figure 8).

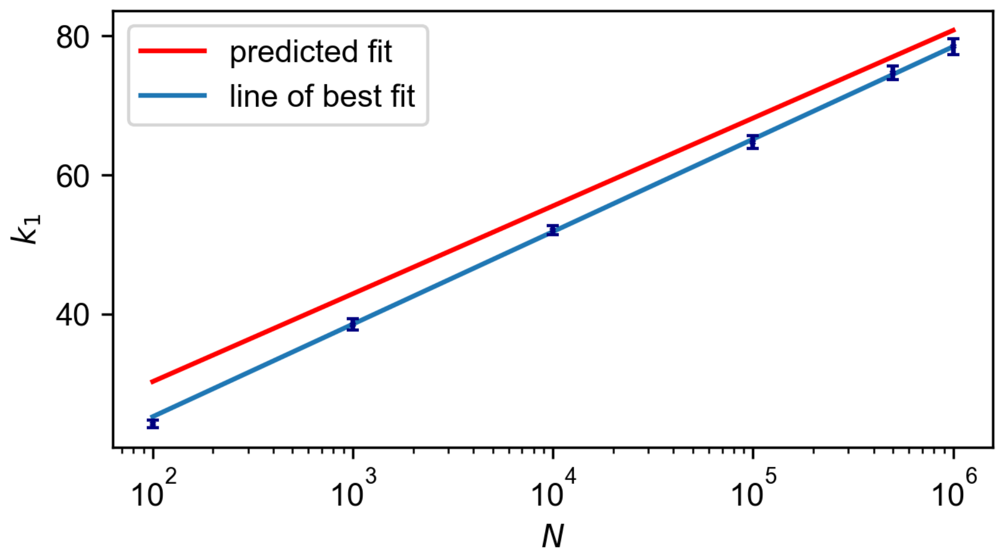

Figure 9: Average largest degree observed k_1, against system size N for $m = 5$. Error bars represent 95% confidence intervals. Despite the theoretical derivation being a poor fit to the data, the correlation coefficient of k_1 against $\log(N)$ had an associated p-value very close to one, suggesting the scaling relation was still correct.

.

3 Phase 3: Existing Vertices Model

3.1 Existing Vertices Model Theoretical Derivations

In the existing vertices model, a new node is added and m edges at each time step as usual but with a caveat; r edges attached between an existing node chosen with probability Π_1 and the newly added node and $m-r$ nodes between existing vertices chosen with probability Π_2. For this project, $\Pi_1 = \Pi_{RA}$ and $\Pi_2 = \Pi_{PA}$. The master equation in this case becomes

$$\begin{aligned} n(k,t+1) &= n(k,t) + r\Pi_{RA}(k-1,t)n(k-1,t) - r\Pi_{RA}n(k,t) \\ &\quad + (m-r)\Pi_{PA}(k-1,t)n(k-1,t) - (m-r)\Pi_{PA}(k,t)n(k,t) + \delta_{k,r}. \end{aligned} \tag{31}$$

In order to find the degree distribution in the long time limit, the results of sections 1 and 2 will be reused. In particular, remembering that $\Pi_{RA}(k,t) = 1/N(t)$ and $\Pi_{PA}(k,t) = k/2mN(t)$, and the ansatz $n(k,t) = p_\infty(k)N(t)$, equation (31) can be transformed to obtain

$$p_\infty(k) = r\left(p_\infty(k-1) - p_\infty(k)\right) + 2(m-r)\left(\frac{k-1}{2m}p_\infty(k-1) - \frac{k}{2m}p_\infty(k)\right) \tag{32}$$

for $k \neq r$. Simple rearrangement yields

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k + \frac{mr}{m-r} - 1}{k + \frac{m(r+1)}{m-r}}. \tag{33}$$

Adapting the trial solution proposed in section 1 (equation (8)), the degree distribution must be of the form

$$p_\infty(k) = A\frac{\Gamma(k + \frac{mr}{m-r})}{\Gamma(k + \frac{m(r+1)}{m-r} + 1)}. \tag{34}$$

Using the boundary condition for $k = r$ to find A in a similar manner to previous derivations using the master equation, it can be shown the full solution including normalisation is given by

$$p_\infty(k) = \begin{cases} \frac{m}{m(1+r)+r(m-r)} \frac{\Gamma(r+\frac{m(r+1)}{m-r}+1)}{\Gamma(r+\frac{mr}{m-r})} \frac{\Gamma(k+\frac{mr}{m-r})}{\Gamma(k+\frac{m(r+1)}{m-r}+1)} & \text{if } k \geq m \\ 0 & \text{if } k < m \end{cases} \tag{35}$$

Verifying this distribution is normalised is highly nontrivial. However, note that for large k, this function behaves as a power-law with an exponent of $-1 - \frac{m}{m-r}$, and therefore the integral of $p_\infty(k)$ decays fast enough to be normalisable in principle.

3.2 Existing Vertices Model Numerical Results

For this investigation, $m = 3r$ was used and system sizes were simulated for $m = 3$ to $m = 81$ over 10 realisations (Figure 10). Once again, results are consistent with theoretical expectations at the 5% significance level aside from $m = 3$ (Table 3), and there are finite-size effects in the tail observed.

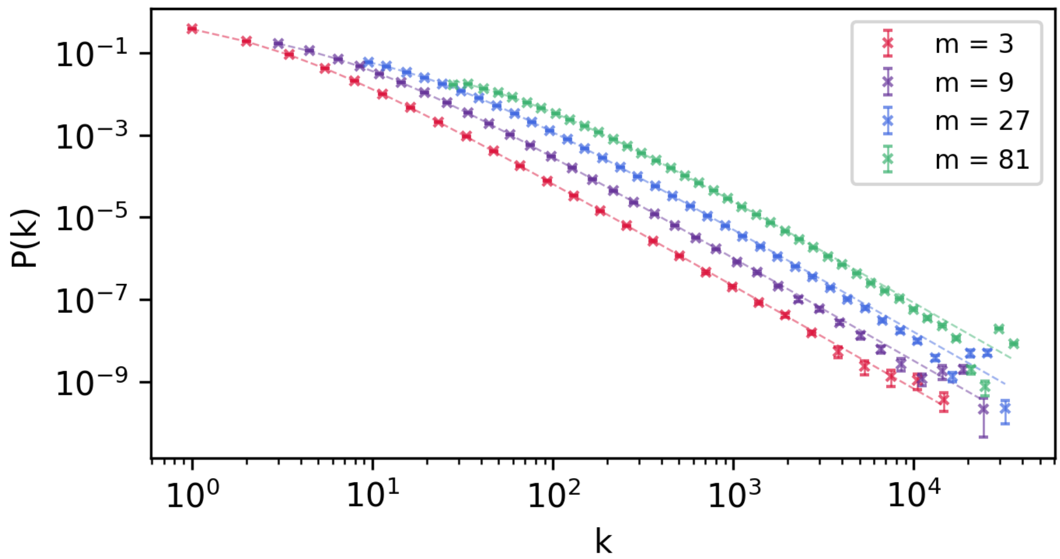

Figure 10: Degree distribution, $P(k)$ for the existing vertices model for various values of m, over 10 realisations for $N = 10^6$. Error-bars represent 95% confidence intervals. Dashed lines represent the theoretical distributions. Note that for larger values of m, the error bars do not pass through theoretical distribution, suggesting that the deviation is a genuine finite size effect and not due to insufficient statistics at the tail.

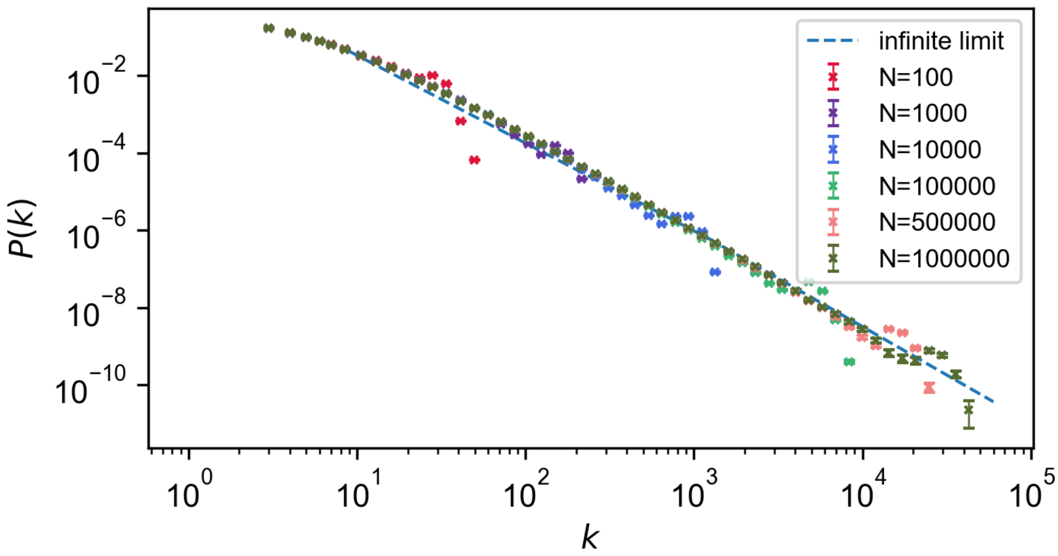

Figure 11: Degree distribution, $P(k)$, for the existing vertices model for various values of system sizes, N, for $m = 9$, averaged over 50 realisations. The distribution for each value of N are coincident until the finite-size effects of each system causes them to diverge

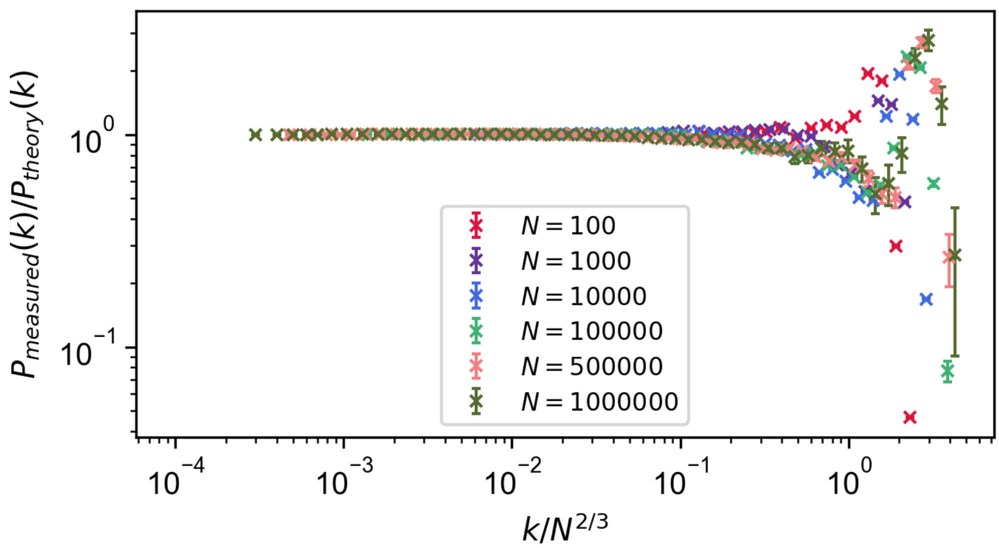

Figure 12: Data collapse for the existing vertices model. $N = 100$ exihibits corrections to scaling since the tail starts earlier than the other systems.

m	Reduced χ^2 (3 s.f.)	p-value
3	4.89	1.28×10^{-7}
9	1.40	0.144
27	0.283	0.998
81	0.457	0.940

Table 3: Results from chi-squared tests on collected data excluding data from the tail affected by finite-size effects for the existing vertices model. P-values suggest excellent agreement between numerical data and theory for $m = 27$ and $m = 81$; the case $m = 9$ has a reduced chi-squared greater than 1 suggesting a poor fit but is not significant enough at the 5% level to reject the proposed ditribution. The distribution for $m = 3$ is a very poor fit.

Finite-size effects were also investigated using the same variety of system sizes in previous sections, but using $m = 9$ (Figure 11). For $m = 3r$, the distribution behaves as a power-law with exponent $-5/2$, and therefore the system-size is proportional to $k_1^{\frac{3}{2}}$ for large N and k. This can be seen if the cumulative probability distribution function from k_1 to infinity is approximated as an integral and is equated to $1/N$ as in the previous phases. To produce a data collapse, the measured probability over the theorerical probability distribution was therefore plotted against $k/N^{2/3}$ (Figure 12).

4 Conclusions

Results were broadly in line with theoretical expectation. Measured degree distributions were generally in line with the derived long-time limit up to finite size effects, and corrections to scaling for the smallest system sizes. Even though the master equation imposed idealised behaviour finite systems can never emulate in practice, the data collected suggests they were very good approximations nonetheless.

Acknowledgements

I would like to thank Dr. Evans and the other demonstrators for their guidance throughout the project, along with Thomas Travis and Manuel Coelho for stimulating discussion.