Network Project
A Growing Network Model

Benjamin Richards

20th March 2023

Abstract: The model of network growth proposed by Barabási and Albert in 1999, whilst similar to previous models, is a simple model that produces "fat-tailed" distributions, which frequently describe properties of real-world networks. This investigation sought to gain insights into models such as this, by careful comparison between theoretical and numerical analysis. Three similar models of network creation were analysed, where new edges were added preferentially, randomly and also between existing vertices. The theoretically and numerically derived degree distributions were found to agree, with the preferential model agreeing to a p-value of 0.943, the random model to a p-value of 0.318 and the existing vertices model to a p-value of 0.114. Theoretical and numerical results therefore showed reasonable agreement, considering a statistical significance of 5%, inferring the assumptions and limitations of the theoretical derivations and numerical implementations were reasonable.

Word Count: 2498

0 Introduction

Many systems such as social connections, road maps and supply chains can be modelled using networks, also known as graphs. Many properties of real-world networks are not normally distributed, instead having "fat tails", where some small fraction of the graph is disproportionately important. Wealth ownership, paper citations and web links all follow fat-tailed distributions, which can be explained by "rich getting richer" processes.

The aim of this project is to investigate such processes using the model introduced by Barabási and Albert [1]. By theoretical and numerical analysis of various preferential and random attachment models, and careful application of assumptions, approximations and comparisons, we hope to gain insight into this broad class of physical phenomena.

Definition

The BA (Barabási-Albert) model, as used in this investigation, is defined as follows:

1. Increment time by one discrete step ($t \to t+1$)
2. Add one new vertex to the graph
3. Add m new edges to the graph, connecting one end of each to the new vertex
4. Connect the other end of each new edge to an existing vertex with probability, Π, proportional to the vertex's degree, k
5. Repeat, from the first step

The model itself does not specify the initial graph, other than requiring it has at least m vertices, and doesn't require termination. When implemented, it is run until there are N vertices.

1 Phase 1: Pure Preferential Attachment Π_{pa}

1.1 Implementation

1.1.1 Numerical Implementation

All models were implemented in Python, using the NetworkX package for handling graphs [3]. After checking m and N are permissible, a complete graph of size $m+1$ is created, ensuring $k \geq m$. To efficiently pick a vertex in proportion to its degree, each vertex is added k times to a list, e.g. twice for $k=2$, and an edge is randomly selected.

A loop then runs until the desired size is reached. Each cycle a new vertex is added, m unique vertices are selected from the list according to preferential attachment (by repeated sampling until there are no duplicates), m new edges are added between the new vertex and target vertices, and the list of potential vertices is updated.

1.1.2 Initial Graph

An initial graph is required as the algorithm is not defined for $N < m$. The specific initial graph does not impact our investigation, as we are concerned with the long-time limit, which is unaffected if $m \ll N$. The results of section 1.3.2 were found to be independent of the choice of initial graph. A fully connected graph of size $m + 1$ was used, ensuring $k \geq m$.

1.1.3 Type of Graph

Multiple types of graph are valid for the model. The type used, which gives the simplest theoretical analysis, is a simple graph, where one edge is allowed between any pair of vertices, and edges have no direction or weight.

Use of a directed graph, motivated by the distinction between new and existing vertices complicates the theoretical derivation by differentiating between in and out degree, eg. Price [2].

A simple graph, however, means that when adding multiple edges, the available vertices for each subsequent edge decrease by 1, changing the probabilities, Π, significantly for $m \sim N$. Therefore, it is important that $m \ll N$. Allowing multiple edges between pairs of vertices circumvents this, however, this modifies the master equation since degree can change by more than one per time step. This effect is amplified by preferential attachment, with many edges attaching to the largest vertex, making theoretical analysis more complicated.

1.1.4 Working Code

Tests were performed on the preferential, random and existing vertices models, to verify the implementations worked as expected.

Total edge and vertex counts were checked, and for $m \geq N, m < 0, N < 1$ the models raised errors. There were no vertices with $k < m$, meaning edges were added correctly, as both the initial graph and model prevent this. The mean degree, $\langle k \rangle$, was $2m$, as $2m$ stubs are added per vertex, since each edge has two stubs. Finally, the graph was drawn at various stages, allowing identification of unforeseen errors.

1.1.5 Parameters

The model takes two parameters, m and N. Larger N reduces random noise, limited by the available computational resources. The initial graph depends on m, meaning m should be minimised to reduce its effects. Increasing m also requires more computational resources.

Most importantly, finite-size effects (from stopping the model at finite time), depend on m compared to N, due to the finite number of vertices and edges. Minimising these effects, to compare against the long-time limit, requires $m << N$.

Typical values for N were 10^3 to 10^6, and m from 10^0 to 10^2.

1.2 Preferential Attachment Degree Distribution Theory

1.2.1 Theoretical Derivation

To find the long-time limit of the degree distribution, $p_\infty(k)$, we start from the master equation. By considering the number of vertices of degree k, $n(k)$, over one iteration, we find

$$n(k,t+1) = n(k,t) + m\Pi_{PA}(k-1,t)n(k-1,t) - m\Pi_{PA}(k,t)n(k,t) + \delta_{k,m}, \quad (1)$$

where $\Pi_{PA}(k,t)$ is the preferential attachment probability of joining to a vertex of degree k at time t, and $\delta_{k,m}$ is the Kronecker delta. This is since $n(k,t)$ increases by one if $m=k$, gains vertices of degree $k-1$ which gain an edge, and loses vertices of degree k which gain an edge.

As $\Pi_{PA}(k) \propto k$, but must be normalised, it must equal $k/2mN(t)$ as $\sum_m^\infty k = 2mN(t)$ given each edge increases the degree of two vertices by 1, and there are $mN(t)$ vertices in the long-time limit. Using this, with $n(k,t) = N(t)\,p_\infty(k)$ in the long-time limit, and $N(t+1) = N(t)+1$, gives

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

Here we are implicitly considering the ensemble-averaged $p_\infty(k)$, since for finite $N(t)$ there are random fluctuations in $n(k,t)$, which is a mean field approximation.

Considering large k, such that $\delta_{k,m} = 0$, and rearranging allows for comparison to a relation with known solutions

$$\frac{p_\infty(k)}{p_\infty(k+1)} = \frac{k-1}{k+2}, \quad (3) \qquad\qquad \frac{f(z)}{f(z+1)} = \frac{z+a}{z+b}, \quad (4)$$

where a and b are constants, and the solution to $f(z)$ can be shown to be

$$f(z) = A\frac{\Gamma(z+1+a)}{\Gamma(z+1+b)}, \quad (5)$$

where A is some constant, by substitution of Stirling's approximation to the Gamma function, itself defined as $\Gamma(z+1) = z\Gamma(z)$.

Using equation (5) as the solution to equation (3), we can simplify by applying the Gamma function's definition, giving $\Gamma(k+3) = k(k+1)(k+2)\Gamma(k)$. To find A, we consider equation (2) for $k=m$. Since no vertex can have $k<m$, $p_\infty(m-1) = 0$. Substituting for $p_\infty(m)$ gives the final solution to be

$$p_\infty(k) = \frac{2m(m+1)}{k(k+1)(k+2)}. \quad (6)$$

1.2.2 Theoretical Checks

As a probability distribution, $p_\infty(k)$ must be normalised such that $\sum_{k=m}^\infty p_\infty(k) = 1$. Expanding equation (6) into partial fractions and cancelling terms that sum to zero shows this to hold. Similarly, all values must be between 0 and 1. From $d\,p(k)/dk$, $p(k)$ decreases for all $k \geq m$. Therefore, the largest value, $p_\infty(k=m) = 2/(m+2)$, is < 1 for all $m > 0$, and the smallest, $\lim_{k\to\infty} p_\infty(k) = 0$, also satisfies this requirement.

The mean degree, $\langle k \rangle$, must equal $2m$, as in the long-time limit each new vertex brings m edges, thus $2m$ stubs. This can be verified by expanding $k\,p_\infty(k)$ into partial fractions, calculating $\sum_{k=m}^\infty k\,p_\infty(k)$ and cancelling terms that sum to 0, leaving only $2m$.

1.3 Preferential Attachment Degree Distribution Numerics

1.3.1 Fat-Tail

Since the "fat tail" is caused by very few vertices having very large degrees, there is a poor signal-to-noise ratio for large k. Logarithmic binning, where data is binned by increasing widths with increasing k, averages out much of the random noise. Running the model for at least 100 repetitions improves the data further, and plotting the binned data on log-log axes makes the tail easier to analyse.

1.3.2 Numerical Results

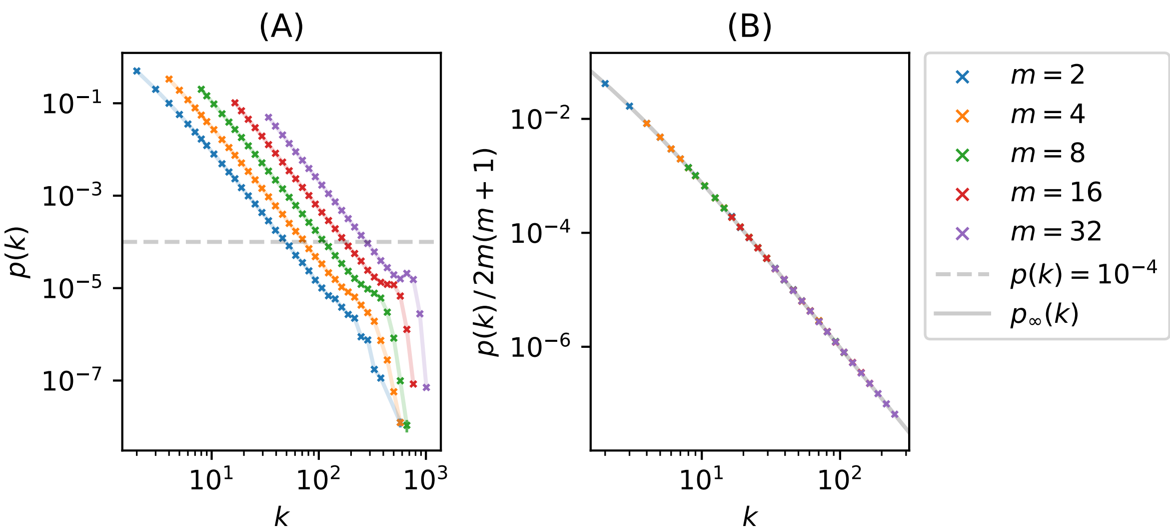

Figure 1: The preferential attachment degree distribution, $p(k)$ against k for $N = 10,000$ and multiple values of m, (A), and $p(k)\,/\,2m(m+1)$ against k, for all $p(k) > 10^{-4}$, to collapse the scaling regions onto one another, (B). Each combination of parameters was repeated 100 times, and the resulting data were logarithmically binned. Error bars on both plots display the standard error on the mean of each bin, although they are only large enough to be visible for the largest values of k in (A). Good visual agreement can be seen between the expected and observed data in (B).

For $N = 10,000$ and $m = 2, 4, 8, 16, 32$, Fig. 1 (A), the observed distributions follow power laws until increasing to a "bump" before a cutoff. Increasing m shifts distributions up and to the right, as expected from equation (6).

Plotting $p(k)\,/\,2m(m+1)$ removes m dependence, collapsing the data, Fig. 1 (B). As $p_\infty(k)$ assumes the long-time limit, finite-size effects must be removed before comparison. This is done by truncating the distributions at $p(k) = 10^{-4}$, chosen from visual examination and marked with a grey dashed line, Fig. 1 (A).

1.3.3 Statistics

Different statistical tests can be used, each with its own limitations. The R^2 test would give a measure of agreement, but this is best suited to linear regression. The Kolmogorov–Smirnov test is well suited to comparing two arbitrary distributions, however, it cannot be used on logarithmically binned data without significant alteration, making it difficult to combine the different values of m. Truncated data adds additional complexity, as cumulative distributions are required.

The χ^2 test is capable of working with both truncated distributions and binned data. A modified version of Pearson's χ^2 test allows use of bin averages and errors, rather than frequencies

$$\chi^2(y_i) = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2, \tag{7}$$

where y_i are the mean observed data points at x_i, and $f(x_i)$ is the expected function, assuming no fit parameters [4]. This removes the minimum bin count requirement, and assuming Gaussian errors is a reasonable approximation given the central limit theorem.

Performing a χ^2 test on the combined data of Fig. 1 (B) gives $\chi^2 = 11.6$ with 21 degrees of freedom, corresponding to a p-value of 0.943. This is strong evidence that the data follow the expected distribution, as 94.3 times out of 100 we would expect to see similar or greater deviation. However, visually the expected and observed data are indistinguishable. Small errors from many repetitions mean small deviations from finite-size effects or random noise correspond to large increases in χ^2.

1.4 Preferential Attachment Largest Degree and Data Collapse

1.4.1 Largest Degree Theory

The largest measured degree for one instance of the model will vary stochastically. However, an ensemble average will converge to some mean value, k_1, which has a lower bound according to $n(k \geq k_1) = 1$.

In the long-term limit, this becomes $N \sum_{k=k_1}^{\infty} p_\infty(k) = 1$, into which we substitute $p_\infty(k)$ from equation (6), and rearrange into partial fractions. Combinations of terms sum to zero, leaving a quadratic relation in k_1. The negative solution is negative for all $m, N > 0$, leaving only the positive solution

$$k_1 = -\frac{1}{2} + \sqrt{\frac{1}{4} + Nm(m+1)}, \tag{8}$$

which is positive for $m, N > 0$. For large N, $k_1 \sim \sqrt{Nm(m+1)}$, meaning k_1 scales with $N^{1/2}$ for large N at constant m.

1.4.2 Numerical Results for Largest Degree

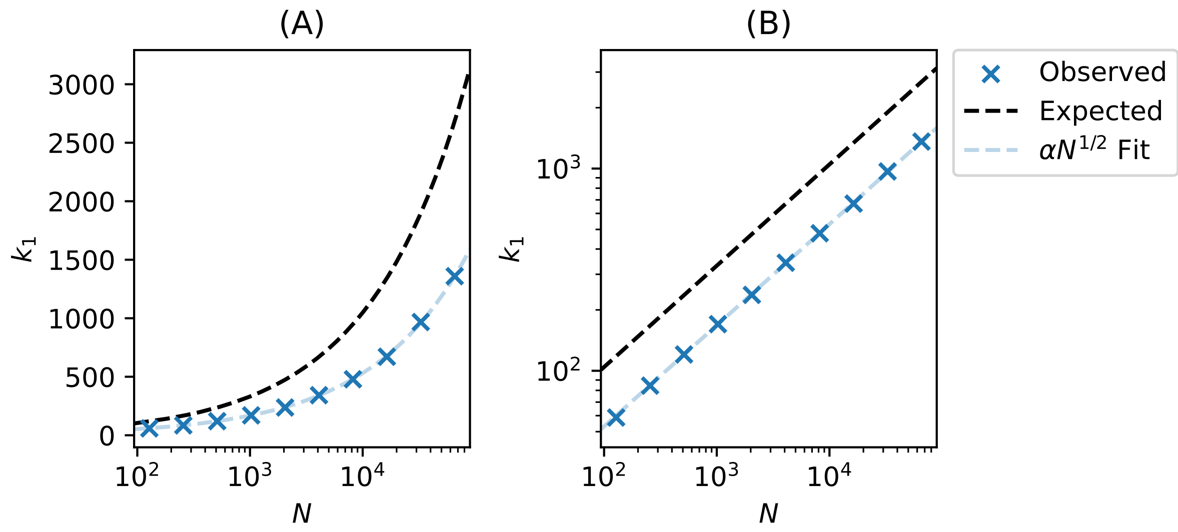

Figure 2: Expected and observed k_1 as a function of N for $m = 10$, for preferential attachment on linear-log axes (A) and log-log axes (B) for direct comparison to the random attachment model. The function $\alpha N^{1/2}$ was fit to the observed data, where α is some constant found to be 5.31 ± 0.01, approximately half of 10.5, the expected value. Each data point is the average of 100 calculations. Errors in k_1 are not plotted, as they would be too small to see, with values ranging from 0.3 to 0.05. The observed scaling behaviour of k_1 matches that expected.

Numerical values for k_1 were found for N as factors of 2 between 128 and 65,536 with $m = 10$. These parameters were chosen for direct comparison with section 1.4.3, which requires $m > 1$ for greater finite-size effects, while keeping $N \gg m$, constrained by the available computational resources.

The observed and expected values for k_1 do not correspond exactly, but both scale with $N^{1/2}$, Fig. 2 (B). Fitting $\alpha N^{1/2}$ to the observed data, where α is the variable being fitted, we find good visual agreement.

Expected k_1 is a lower bound but observed values are lower still. This discrepancy is caused by finite-size effects, not considered in the derivation. These scale with k_1. From Fig. 1 (A), data follow the expected distribution before increasing at a "bump", meaning k_1 must be smaller than expected to maintain normalisation.

1.4.3 Data Collapse

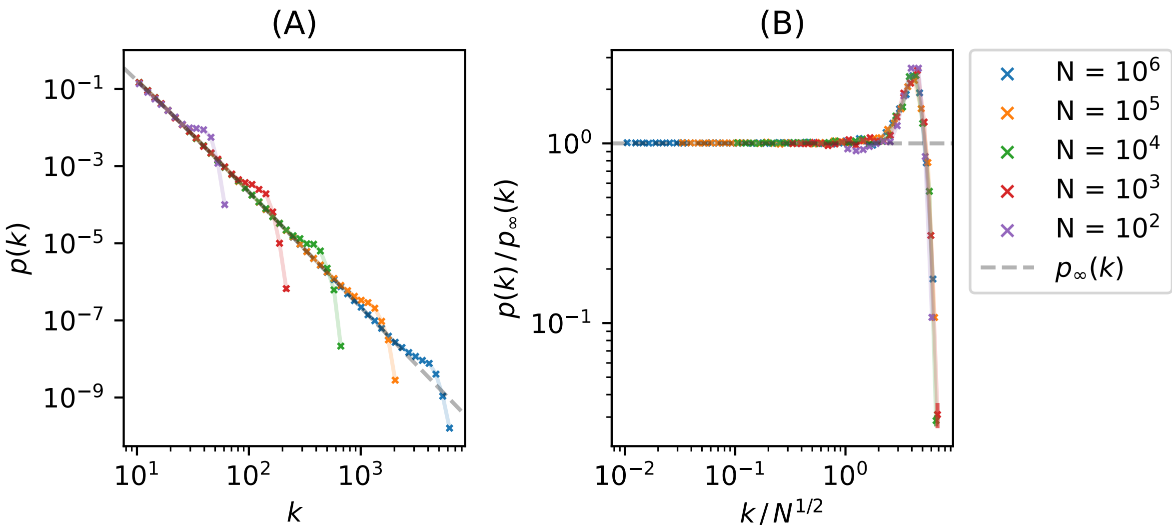

Figure 3: The degree distribution for preferential attachment, $p(k)$, against k for $m = 10$ and various N, (A), and the associated data collapse of the same data, (B), which is visually found to be of good quality. Each combination of parameters was repeated 100 times and the resulting data were logarithmically binned. Error bars indicate the standard error on each logarithmically binned mean, but other than in the tail of the data collapse, they are too small to see.

The parameters $m = 10$ and N varying from 10^2 to 10^6 were chosen as $m > 1$ produces reasonable finite-size effects and $m \ll N$ ensuring our theoretical derivations are still applicable. As predicted from Fig. 2, bump height remains approximately constant on log-log axes for increasing N.

Plotting $p(k)\,/\,p_\infty(k)$ against $k/N^{1/2}$ produces a data collapse of good quality, Fig 3 (B). The "bump" appears above $k \sim N^{1/2}$, before the distribution is cut off.

Vertices with the largest degrees are unable to grow to their expected size, meaning those with expected degrees in the cutoff region are fewer than expected, leading to more just below the cutoff, producing the "bump".

2 Phase 2: Pure Random Attachment Π_{rnd}

2.1 Random Attachment Theoretical Derivations

2.1.1 Degree Distribution Theory

Substituting the preferential attachment probability $\Pi_{PA}(k) \propto k$ for $\Pi_{RA} \propto 1$, the random attachment probability gives the master equation as

$$n(k, t+1) = n(k, t) + m\Pi_{RA}(k-1, t)n(k-1, t) - m\Pi_{RA}(k, t)n(k, t) + \delta_{k,m}. \quad (9)$$

To satisfy normalisation across all vertices, labelled $n_i = 0, 1, \ldots, N$, according to $\sum_{n_i=0}^{N} \Pi_{RA} = 1$, clearly $\Pi_{RA} = 1/N(t)$. This substitution, together with assuming the long-time limit meaning $n(k,t) = N(t)\, p_\infty(k)$, means the master equation becomes

$$p_\infty(k) = m\, p_\infty(k-1) - m\, p_\infty(k) + \delta_{k,m}. \tag{10}$$

Considering large k, such that $\delta_{k,m} = 0$, we find that the normalisation condition of $\sum_{k=m}^{\infty} p_\infty(k) = 1$ becomes a geometric series with proportionality coefficient $1/(m+1)$, noting that the sum starts from m, not 0. This gives the degree distribution as

$$p_\infty(k) = \frac{1}{m+1}\left(\frac{m}{m+1}\right)^{k-m}. \tag{11}$$

This is normalised, given the normalisation condition was used in its derivation. The mean degree, $\langle k \rangle$, is found form $\sum_{k=m}^{\infty} k\, p_\infty(k)$. Collecting common terms of the series, we find the sum converges for positive m to the final value of $2m$, as required.

To verify the distribution is always between 0 and 1, the derivative shows $p_\infty(k)$ to be decreasing for all $k \geq m$. The largest value, $p_\infty(k=m) = 1/(m+1)$ is therefore < 1 for all $m > 0$, and the smallest value, $\lim_{k\to\infty} p_\infty(k) = 0$, also satisfies the requirement.

2.1.2 Largest Degree Theory

A lower bound on k_1 is again found from $n(k \geq k_1) = 1$. Assuming the long-time limit, $n(k,t) = N(t)\, p_\infty(k)$, and summing over k from k_1 to infinity, we find an infinite geometric series, giving $N = (m/(m+1))^{m-k_1}$. Solving for k_1 by taking logarithms and rearranging for k_1, gives

$$k_1 = m + \frac{\ln(N)}{\ln(m+1) - \ln(m)}, \tag{12}$$

meaning that for large N at constant m, k_1 scales with $\ln(N)$.

2.2 Random Attachment Numerical Results

2.2.1 Degree Distribution Numerical Results

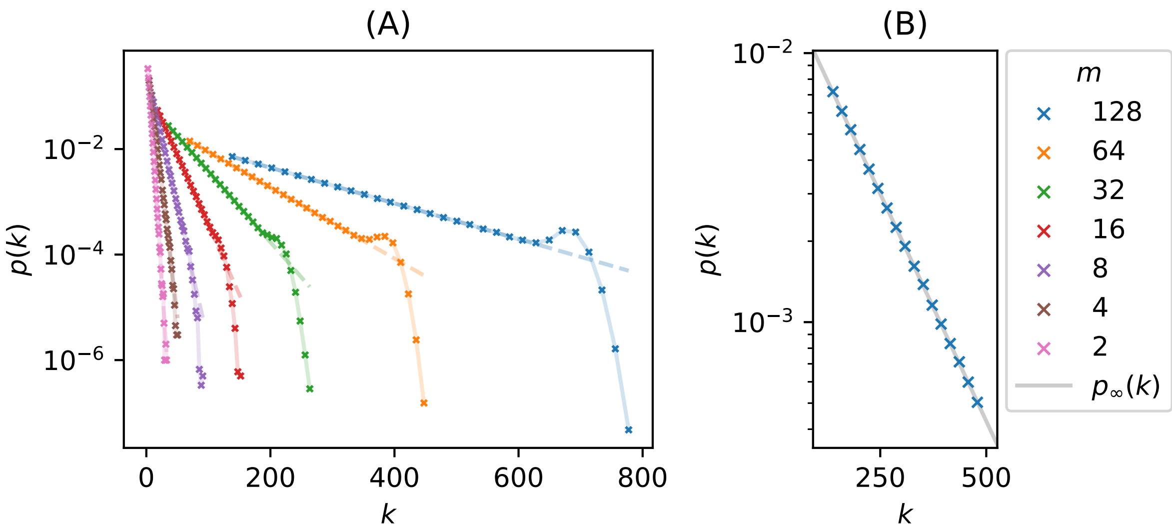

Figure 4: Degree distribution, $p(k)$, against k for random attachment, with $N = 10,000$ and various m, (A), and the section of $m = 128$ data truncated at $k = 500$ against which the χ^2 test was performed, (B), with good visual agreement. Dashed lines show the expected distributions, with line colour corresponding to m value. Each value of m is run for 100 repetitions. Errors are not plotted, since they would be too small to see, with values for $p(k)$ ranging from 6.5×10^{-5} to 4.4×10^{-7}.

For $N = 10,000$ and m ranging from 2 to 128, data follow straight lines as expected, before deviating due to finite-size effects, with increasing gradient for increasing m, Fig. 4 (A).

Choosing $m = 128$ for the large range of k values, for $k < 500$, to remove finite-size effects, we perform a χ^2 test against the expected distribution, Fig. 4 (B). This gives $\chi^2 = 19.3$ with 17 degrees of freedom, with 0.318 the associated p-value. Again low given small errors and theoretical approximations, but above the threshold of $p = 0.05$ and corroborating visual agreement.

2.2.2 Largest Degree Numerical Results

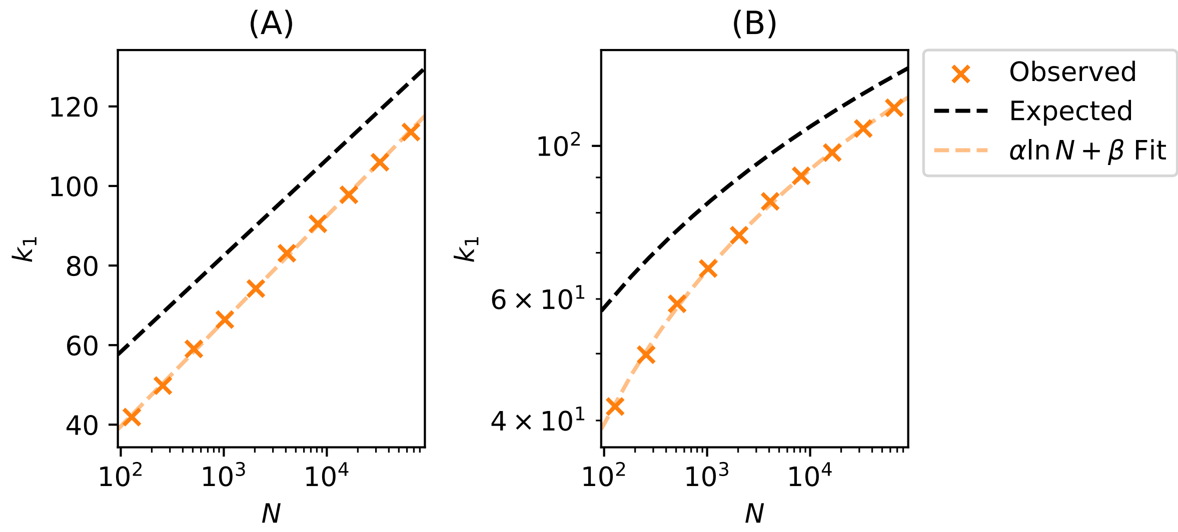

Figure 5: Expected and observed k_1 as a function of N for $m = 10$, for random attachment on linear-log axes (A) and log-log axes (B) for direct comparison to the preferential attachment model. Each value is the mean of 100 calculations. The function $\alpha \ln(N) + \beta$ was fit to the observed values, with parameters found to be $\alpha = 11.5 \pm 0.1$, higher than the expected value of 10.5, and $\beta = -13.3 \pm 0.7$, lower than the expected value of 10. Errors in k_1 are not plotted as they would be too small to see, with values ranging from 5×10^{-2} to 3×10^{-2}.

For $m = 10$ and N ranging from 128 to $65,536$ for the same reasons as previously explained, the expected and observed data follow a linear relation on linear-log axes, as expected from equation (12), Fig. 5 (A).

Observed k_1 values are again below those estimated, due to finite-size effects. These no longer scale with k_1, but are constant and gradually decreasing for increasing N. Fitting $\alpha \ln(N) + \beta$, where α and β are constants, we visually confirm that k_1 follows a logarithmic distribution.

For preferential attachment, even for small m the largest vertices rapidly approach the limits of the graph, whereas for random attachment fewer vertices grow as large, meaning fewer reach the limits of the graph unless $m \sim N$, explaining the diminishing finite-size effects for random attachment.

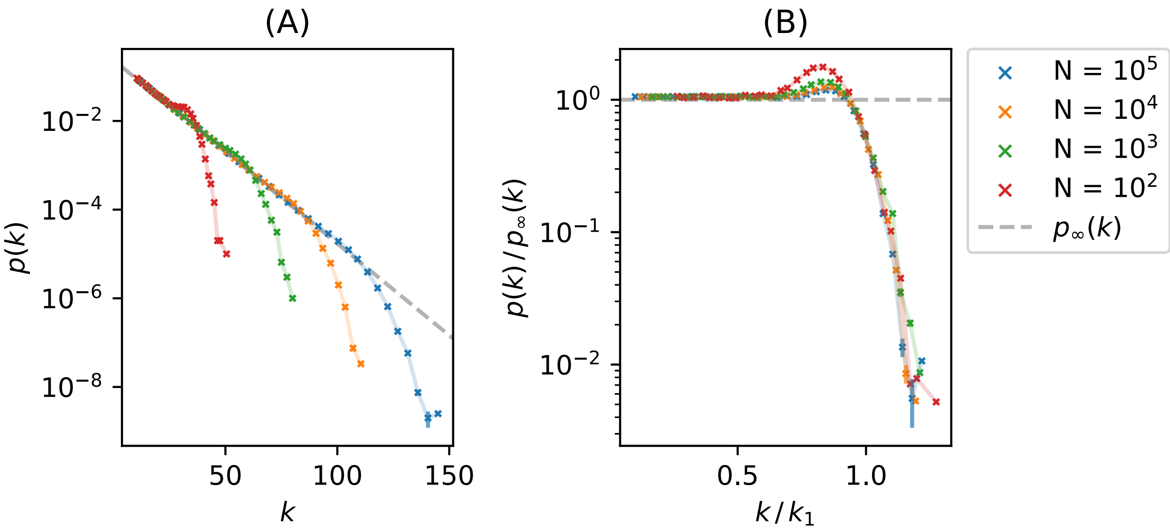

Figure 6: The degree distribution for random attachment, $p(k)$ against k for $m = 10$ and various N, (A), and the associated data collapse, (B), which is visually found to be of reasonable quality. Each combination of parameters was repeated 1,000 times, and the resulting data were logarithmically binned. Error bars indicate the standard error on each logarithmically binned mean, but other than in the tail of the distribution, they are too small to be seen.

Taking $m = 10$ and N between 10^2 and 10^5, the distributions follow those expected before deviating due to finite-size effects. As expected from Fig. 5, bump height, and thus finite-size effects, diminish with increasing N. Plotting $p(k)\,/\,p_\infty(k)$ against $k\,/\,k_1$, where k_1 is the observed value, we find a data collapse of reasonable quality, Fig. 6 (B).

3 Phase 3: Existing Vertices Model

3.1 Existing Vertices Model Theoretical Derivations

After adding a new vertex, we now add r edges between the new vertex and existing vertices according to random attachment, and add the remaining $(m - r)$ vertices between pairs of existing vertices, chosen by preferential attachment. The master equation therefore includes, $\Pi_{PA}(k)$, Π_{RA}, and each edge between existing vertices has two opportunities to impact $n(k, t)$, giving

$$\begin{aligned} n(k, t+1) = \delta_{k,r} + r\Pi_{RA}(k-1, t)n(k-1, t) \quad &+2(m-r)\Pi_{PA}(k-1, t)n(k-1, t) \\ - r\Pi_{RA}(k, t)n(k, t) \quad &-2(m-r)\Pi_{PA}(k, t)n(k, t), \end{aligned} \tag{13}$$

where terms corresponding to random and preferential attachment have been aligned. Substituting $\Pi_{RA} = 1/N(t)$ and $\Pi_{PA}(k) = k/2mN(t)$, and taking the long-time limit gives

$$p_\infty(k) = rp_\infty(k-1) - rp_\infty(k) + \frac{m-r}{m}(k-1)p_\infty(k-1) - \frac{m-r}{m}kp_\infty(k) + \delta_{k,r}. \tag{14}$$

Considering large k, and comparing to equations (4) and (5), gives

$$p_\infty(k) = A\frac{\Gamma\left(k+\frac{mr}{m-r}\right)}{\Gamma\left(k+1+\frac{m(r+1)}{m-r}\right)}, \tag{15}$$

where A is a constant. We find A by considering equation (14) for $k = r$, as $k \geq r$. This gives the final solution as

$$p_\infty(k) = \frac{1}{1+\frac{r}{m}(2m-r)} \frac{\Gamma\left(r+1+\frac{m(r+1)}{m-r}\right)}{\Gamma\left(\frac{r(2m-r)}{m-r}\right)} \frac{\Gamma\left(k+\frac{mr}{m-r}\right)}{\Gamma\left(k+1+\frac{m(r+1)}{m-r}\right)}, \tag{16}$$

which simplifies as we only consider $m = 3r$, giving

$$p_\infty(k) = \frac{1}{1+\frac{5}{3}r} \frac{\Gamma\left(\frac{5}{2}(r+1)\right)}{\Gamma\left(\frac{5}{2}r\right)} \frac{\Gamma\left(k+\frac{3}{2}r\right)}{\Gamma\left(k+\frac{3}{2}r+\frac{5}{2}\right)}. \tag{17}$$

The mean degree is $2m$ as required, found by numerically calculating $\sum_{k=r}^{\infty} k\, p_\infty(k)$. This solution is normalised, as expected given use of the normalisation constraint. Numerically calculating the gradient, we find $p(k)$ decreases for all $k \geq r$, meaning the largest value, $p_\infty(k=r) = 1/(1+\frac{5}{3}r)$, is less than 1 and the smallest value, $\lim_{k\to\infty} p_\infty(k)$ can be found numerically to be 0, as required.

3.2 Existing Vertices Model Numerical Results

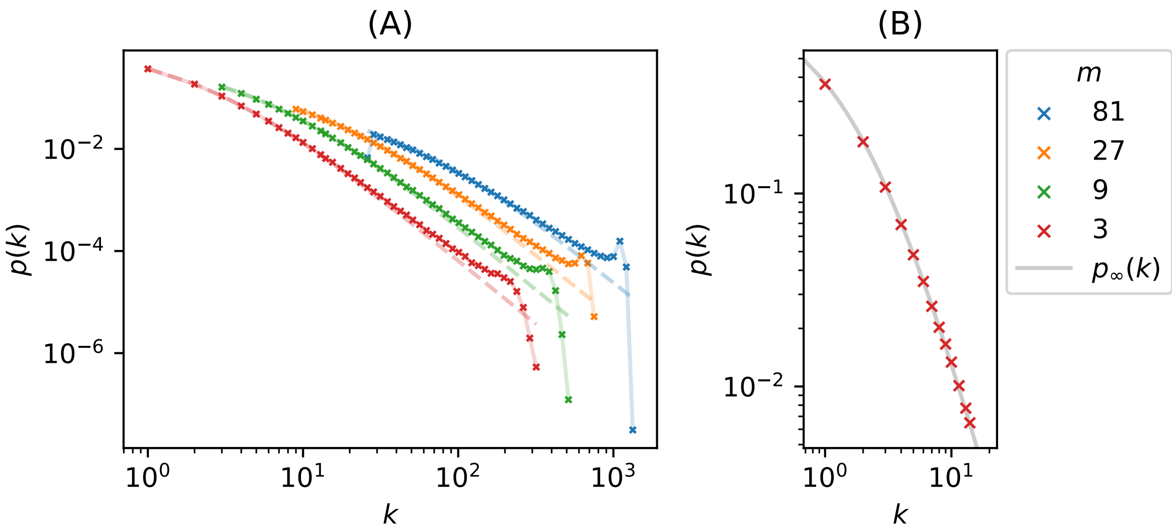

Figure 7: The existing vertices model degree distribution, $p(k)$ for $n = 10,000$ and various m, (A), and the section of $m = 3$ data, truncated at $k = 15$, against which the χ^2 test was performed, (B), with good visual agreement. Dashed lines show the expected distributions, with colour corresponding to m value. Each combination of parameters was run for 100 repetitions and data was logarithmically binned. Errors are not plotted, as they would be too small to see, with values for $p(k)$ ranging from 2.2×10^{-5} to 1.2×10^{-8}.

For $N = 10,000$ and m as 3, 9, 27 and 81, so that $r = m/3$ is an integer, we find the distributions follow those expected, before deviating due to finite-size effects, Fig. 7 (A).

As vertices are added between existing vertices, a fully connected initial graph is invalid, so a star graph was used. As in section 1.1.2, choice of initial graph has no impact in the long-time limit. For $k > 10^2$, Stirling's approximation is used to reduce the expected distribution to a Puiseux series (a complex power series with fractional powers) of which the first two terms are used.

Compared to random and preferential attachment, finite-size effects start earlier. Visual agreement is good and a χ^2 test was run on the $m = 3$ data, as it showed the least finite-size effects, for $k < 15$ in order to remove finite-size effects, Fig. 7 (B). From this, $\chi^2 = 19.6$ with 13 degrees of freedom, corresponding to a p-value of 0.114, still above a 5% significance level.

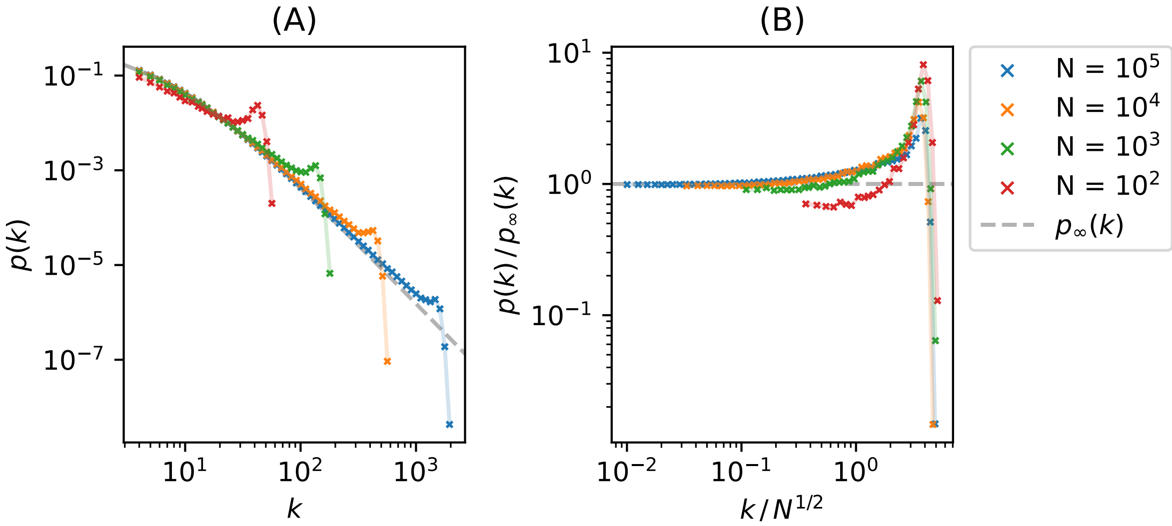

Figure 8: Degree distribution for the existing vertices model, $p(k)$, against k for $m = 12$, $r = 4$ and various N, (A), and the associated data collapse, (B), which breaks down for $N = 100$ as m approaches N. Each combination of parameters was repeated 100 times, and the resulting data were logarithmically binned. Errors are not plotted as they would be too small to see, with values ranging from 1.7×10^{-5} to 1.2×10^{-9}.

Data were collected for $m = 12$, $r = 4$ and N ranging from 10^2 to 10^5, Fig. 8 (A). The data follow the expected distribution until deviation due to finite-size effects.

The observed k_1 were found to scale with $N^{1/2}$ for the data in Fig. 8 (A), meaning plotting $p(k)\,/\,p_\infty(k)$ against $k\,/\,N^{1/2}$ produced a data collapse of good quality, Fig. 8 (B). For $N = 10^2$ this starts to break down as m and N are within an order of magnitude. Finite-size effects are found to start much earlier than for the other models, with a slow increase from $k \sim 0.1 \times N^{1/2}$, due to attachment between existing vertices.

4 Conclusions

This project sought to investigate networks that are governed by the Barabási-Albert model, and models similar to it. Through theoretical and numerical analysis of preferential attachment, random attachment and a combination of the two, by attaching existing vertices together, insights were gained into the approximations, limits and features of such models.

Acknowledgements

The author would like to thank the demonstrators and professor for their help throughout this project.

References

[1] A.-L. Barabási and R. Albert, *Emergence of scaling in random networks*, Science, (1999) **286** 173.

[2] D. J. de S. Price, *A general theory of bibliometric and other cumulative advantage processes*, J. Am. Soc. Inf. Sci., **27** 292–306 (1976).

[3] A. A. Hagberg, D. A. Schult, P. J. Swart, *Exploring network structure, dynamics, and function using NetworkX*, Proceedings of the 7th Python in Science Conference (SciPy2008, Pasadena, CA USA), pp. 11–15, (2008)

[4] M. Richards, *Second Year Statistics of Measurement - Lecture 7 The chi-squared estimation method*, Imperial College London, (2018)