Network Project
A Growing Network Model

CID: 01531221

28th March 2021

Abstract: This report presents an investigation of growing networks that use preferential, random, or mixed-preferential attachment. Theoretical results were derived for the degree distributions and the largest expected degrees of each model, which were then compared to numerical data obtained from a programme designed in Python. Initially, models were compared for the same system size of $N = 10^5$ and different number of edges per new vertex, $m = 2^n$ for $n = 1, \dots, 6$, demonstrating that preferential and mixed-preferential attachment models follow power-laws, thus having degree distributions with a fat tail. On the contrary, random attachment models have an exponentially decaying tail. All models demonstrated a cut-off and a 'bump' – excess of degree probability – due to their finite size, which, however, disappeared as N increased for random-attachment. Three statistical tests were used to determine the goodness of fit of the numerical data, demonstrating that once truncated for the finite-size effects, the models match the theoretical results well, with p-values of approximately 1. Finally, the largest-degree analysis demonstrated that the PA model has the correct N dependence, as a linear regression model of the double-logarithmic plot had a gradient of 0.50 ± 0.04, agreeing with the theoretical expectation of 0.5, while the m dependence of the theoretical result was found incomplete, evident from an offset between numerical data and theory. The discrepancy between the largest expected degrees from theory and numerical data decreased with increasing system size for the random-attachment model, while the discrepancy between theory and data for the degree distributions of the random-attachment model was smaller than that of the preferential-attachment model, as it had smaller finite-size effects.

Word Count: 2428

0 Introduction

Most real large networks demonstrate the 'rich get richer' principle, where for example an actor is more likely to be given a supporting role for an already established actor. This property arises from the continuous growth of networks that use the principle of cumulative advantage – preferential attachment (PA). The main aim of this report is to explore the degree distribution of such growing networks, like the Barabási-Albert (BA) model, compare them to theoretical predictions and demonstrate their scale-free behaviour.

0.1 Definition

The model used in this report is defined as follows.

1. Create initial network G_0 at time t_0.
2. Increment time $t \to t + 1$.
3. Append a new vertex.
4. Add to it m edges and connect each to an older vertex, chosen with probability Π. For the BA model we use linear PA, $\Pi = \frac{k}{2E}$.
5. Iterate processes 2 to 4 until network has N vertices.

1 Phase 1: Pure Preferential Attachment Π_{pa}

1.1 Implementation

1.1.1 Numerical Implementation

All models were implemented using Python, where a network was represented using a dictionary of sets. The dictionary keys represent vertices, while the sets for each key represent neighbours, thus creating an adjacency list. These data structures were used due to their $O(1)$ search time, while they allow only for unique elements, prohibiting duplicates and self-loops. Π was implemented by choosing a vertex using a random number generator for random attachment, while for PA an element was chosen at random from an array containing the vertex indices with multiplicity equivalent to their degree. Finally, the degree distribution was obtained from the distribution of dictionary values once the network was of size N.

1.1.2 Initial Graph

The initial graph for all models was a complete graph – all vertices connected to all other vertices – of size $m + 1$, representing the smallest possible initial graph for a growing network with m edges for new vertices. This ensures that $\Pi(k, t) = 0$ for $k < m$ as assumed in the theoretical derivations, while the initial graph has the smallest possible effect in the growth of the network, by minimising the time it is in a pseudo-random state (where vertices have approximately the same degree).

1.1.3 Type of Graph

The graphs produced were always simple, unweighted, and undirected, as self-loops and multiple edges between vertices are not allowed, to comply with the assumptions of the 'master' equation detailed in section 1.2.1. Moreover, they are sparse, as $\frac{E}{N} = m$ and $O(m) \sim 1$. They exhibited stochastic growth, as each edge is connected using a probability distribution, resulting in a different final graph each time. When using PA, the graphs produced have a fat-tail summarised by a power-law – a distribution with significant probability for large degrees.

1.1.4 Working Code

The programme was checked by: producing graphs and checking them visually for sensible growth at each time t, converting the adjacency list of a graph into an adjacency matrix A, and verifying that $A^T \equiv A$ – thus the network is undirected, has no duplicates or self-loops, as $A_{ii} = 0 \ \forall i$, and is unweighted as entries are 0 or 1, while the implementation of PA was tested using an initial network of fixed seed that was grown 10^6 times with $m = 1$, and the chosen vertices were compared to their respective degrees.

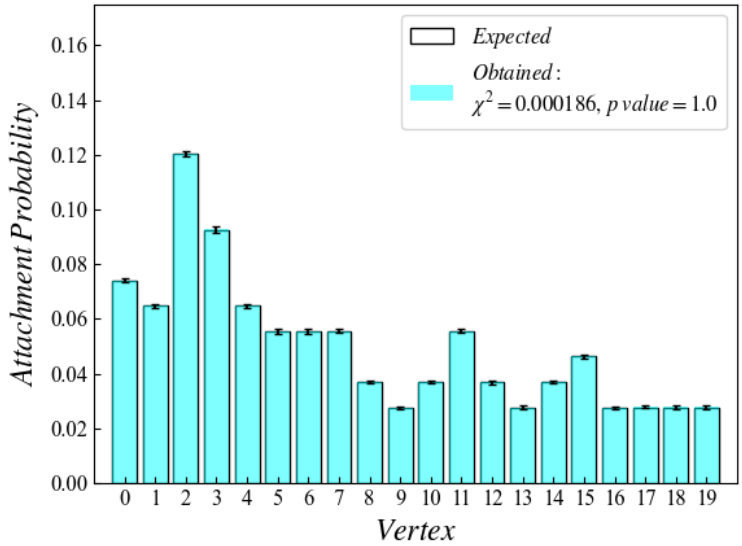

Figure 1.1. Distribution of attachment probabilities for each vertex of a network of fixed seed of size 20, grown 10^6 times with $m = 1$. Expected distribution represents $\Pi(k, t_0)$ for PA. Error bars represent the standard errors of the average probabilities. A p-value of 1 demonstrates that the method for PA works correctly, choosing each vertex the appropriate number of times, reproducing Π for PA.

1.1.5 Parameters

The programme requires 4 parameters: m, N, the method of attachment, and q (the probability of choosing PA in a mixed PA network). Large values of N were used, $N \gg m > 1$ to approximate the degree distribution as $t \to \infty$, while small values of m were used, $1 < m^2 \ll N$, to reduce finite-size effects. Parameters m and N were varied in powers of 2 or 10 to demonstrate properties of the networks in logarithmic plots.

1.2 Preferential Attachment Degree Distribution Theory

1.2.1 Theoretical Derivation

To obtain the degree distribution we begin by writing the 'master' equation outlined in [1], giving the number of vertices with degree k after adding a vertex with m edges at time $t+1$,

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

where Π is the probability that any edge of the new vertex will attach to an existing vertex of degree k. Defining the degree probability distribution as:

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

where $N(t)$ is the number of vertices at time t, we can re-write Eq. (1) as:

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

for which $N(t+1) \coloneqq N(t)+1$. In the long time limit we assume that $p(k,t)$ becomes scale-free and thus has an asymptotic solution

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

allowing to re-write Eq. (3) as:

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

As a result, $p_\infty(k)$ depends on the form of Π, defined specifically for each model.

In the case of PA, we define Π to have linear preferential attachment,

$$\Pi(\mathrm{k,t}) = \frac{k}{2E(t)}, \quad (6)$$

where $E(t)$ is the total number of edges in the network at time t. Substituting Eq. (6) into Eq. (5) gives:

$$p_\infty(k) = m\frac{k-1}{2E(t)}p_\infty(k-1)N(t) - m\frac{k}{2E(t)}p_\infty(k)N(t) + \delta_{k,m}. \quad (7)$$

To simplify Eq. (7) we show that $\frac{E(t)}{N(t)} \to m$ as $t \to \infty$. Since for each vertex we add m edges

$$E(t) = E(0) + mt, \quad (8)$$

where $E(0)$ is the number of edges in the initial graph, and since

$$N(t) = N(0) + t, \quad (9)$$

where $N(0)$ is the number of vertices in the initial graph, then

$$\lim_{t\to\infty}\frac{E(t)}{N(t)} = \lim_{t\to\infty}\frac{E(0)+mt}{N(0)+t} = \lim_{t\to\infty}\frac{E(0)/t+m}{N(0)/t+1} = m. \quad (10)$$

For Eq. (10) to hold in a finite network, as $t = N(t) - N(0)$, we require that $E(0) \ll N$ and that $mN(0) \ll N$. Substituting Eq. (10) into Eq. (7) we get the difference equation:

$$p_\infty(k) = \frac{1}{2}[(k-1)p_\infty(k-1) - kp_\infty(k)] + \delta_{k,m}. \quad (11)$$

If we consider the case when $k \neq m$, we get:

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k-1}{k+2}, \quad (12)$$

for which exists a solution of the form

$$p_\infty(k) = A\frac{\Gamma(k)}{\Gamma(k+3)} = \frac{A}{k(k+1)(k+2)}, \quad (13)$$

where $\Gamma(k+1) = k\Gamma(k), \Gamma(1) = 1$, and A is a constant. To find the complete solution we now also consider the case when $k = m$, for which if we equate Eq. (11) to Eq. (13) we get:

$$p_\infty(m) = \frac{2}{2+m} = \frac{A}{m(m+1)(m+2)}. \quad (14)$$

Thus $A = 2m(m+1)$, and the degree distribution becomes,

$$p_\infty(k) = \frac{2m(m+1)}{k(k+1)(k+2)}, \qquad k \geq m. \quad (15)$$

1.2.2 Theoretical Checks

To ensure that $p_\infty(k)$ has the correct properties we check that $p_\infty(k) = 0$ for $k < m$, which is one of the reasons why models were coded to start with an initial complete graph of size $m + 1$. As every vertex is added with m edges, the largest number of vertices with degree $k < m$ exists at t_0, and the number of vertices with degree k is a monotonically decreasing function:

$$n(k,t) \leq n(k,t_0), \qquad k < m. \quad (16)$$

Substituting Eq. (16) into Eq. (2) we get

$$p(k,t) \leq \frac{n(k,t_0)}{N(t)}, \qquad k < m, \quad (17)$$

thus, as $N \to \infty$

$$p_\infty(k) = 0, \qquad k < m. \quad (18)$$

Moreover, we can check that $p_\infty(k)$ is correctly normalised,

$$\sum_{k=0}^{\infty} p_\infty(k) = \sum_{k=m}^{\infty} p_\infty(k) = 1, \quad (19)$$

where k starts from m, as $p_\infty(k) = 0$ for $k < m$. Substituting Eq. (15) into Eq. (19) we get:

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

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

$$= m(m+1)\left[\left(\frac{1}{m+2} - \frac{1}{m+1} + \frac{1}{m+3} - \frac{1}{m+2} + \ldots\right) - \left(\frac{1}{m+1} - \frac{1}{m} + \frac{1}{m+2} - \frac{1}{m+1} + \ldots\right)\right]$$

$$= m(m+1)\left[-\frac{1}{m+1} + \frac{1}{m}\right] = 1, \tag{20}$$

as expected.

Finally, we ensure that $p_\infty(k)$ is correct as $p_\infty(k) \sim k^{-3}$ for very large k, demonstrating that it follows a power-law, and is thus expected to have a fat-tail distribution.

1.3 Preferential Attachment Degree Distribution Numerics

1.3.1 Fat-Tail

Since the degree distribution is expected to have a fat tail, there will be many very large values of k that will be degrees to no vertex, adding many zeros that create statistical noise and hinder statistical tests. However, these zeros carry crucial information as there is some low probability that vertices with such degrees exist, which is why binning with exponentially increasing bin-sizes has been used to analyse the degree distributions.

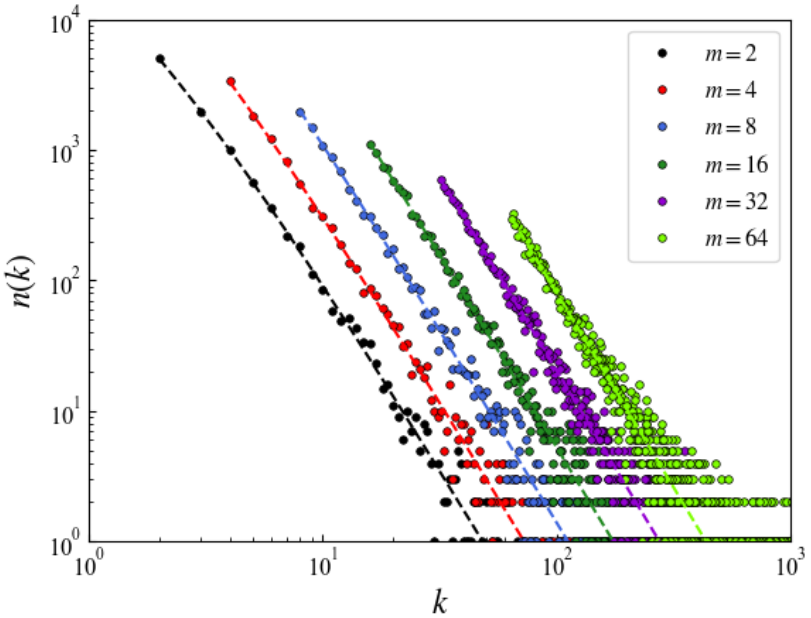

Figure 1.2. Count of vertices $n(k)$ plotted against degree k for six BA models using PA, of size $N = 10{,}000$ and of $m = 2^n$ for $n = 1, \ldots, 6$. Dashed lines, colour-coded in the same way as the numerical data, represent the theoretical degree distributions in the long-time limit $p_\infty(k)$. The distributions demonstrate they have a fat tail, where statistical noise exists for large degrees, making it hard to analyse the numerical data for very large k.

Since the models cannot be run an infinite number of times to obtain better statistics, the data was binned with bins that increase with a multiple of 1.1; a scale that was found visually to provide a good reach of the underlying data while minimising loss of information from binning.

1.3.2 Numerical Results

The numerical results from the PA model were compared to the theoretical degree distribution $p_\infty(k)$ by fixing N and varying m.

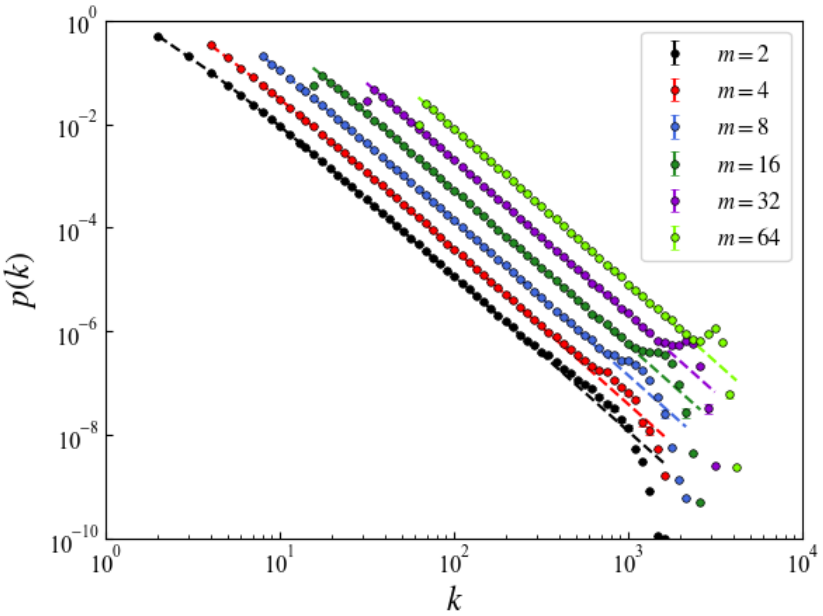

Figure 1.3. Log-binned degree probability distribution $p(k)$ plotted against degree k for six BA models using PA of size $N = 100{,}000$ and of $m = 2^n$, averaged $10 \times 2^{7-n}$ times respectively for $n = 1, \ldots, 6$. Dashed lines represent the theoretical degree distributions $p_\infty(k)$. Error bars represent the standard error on the average values of $p(k)$ – too small for most data points on this scale.

The numerical data follow the theoretical result $p_\infty(k)$ very closely until large values of k, for which there is an exponential cut-off, due to the finite size of the network, starting with a bump that gets taller and narrower with increasing m. A small 'kick' is observed for $m > 8$, arising from the way the data is binned, as the first bin starts at 0 while there are no vertices with $k < m$. The 'kick' increases with m as there are more 'empty' values of k.

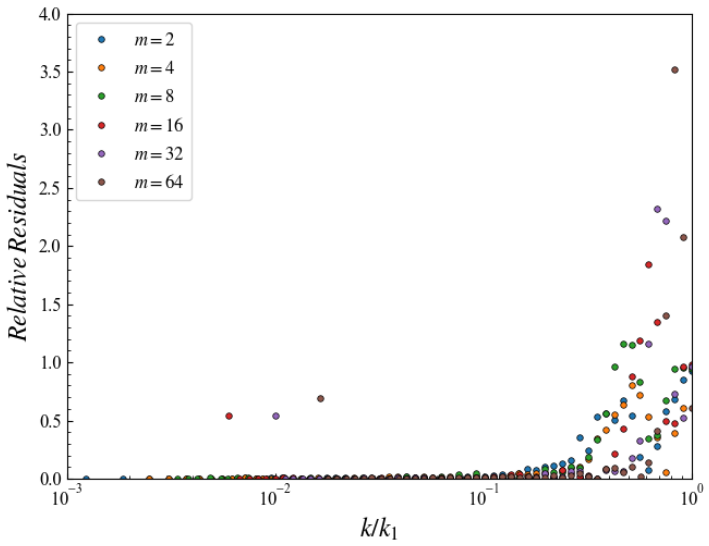

Figure 1.4. Relative residuals of log-binned degree distribution $p(k)$ from theoretical degree distribution $p_\infty(k)$ plotted against degree divided by maximum degree k/k_1, from data of Figure 1.3. The relative residuals are ~ 0 for all except 3 distinct points at low values of k representing the 'kicks', as well as the cut-offs and bumps. However, even those are small; ~ 0.1 to 4 % of the theoretical.

1.3.3 Statistics

Three statistical tests of goodness of fit were used to assess how well the numerical data fits the theoretical predictions, outlined in Table 1.1.

m	R^2	χ^2 (p-value)	KS (p-value)	R^2 (TRUNCATED)	χ^2 (p-value) (TRUNCATED)	KS (p-value) (TRUNCATED)
2	0.99999	1.00000	0.99956	0.99999	1.00000	0.99999
4	0.99999	1.00000	0.99999	0.99999	1.00000	0.99999
8	0.99999	0.99999	0.99940	0.99999	1.00000	0.99999
16	0.84137	0.99999	0.99874	0.99999	1.00000	1.00000
32	0.84690	0.95366	0.99735	0.99999	1.00000	0.99999
64	0.75149	0.01534	0.99444	0.99999	0.99999	1.00000

Table 1.1. R^2 coefficient, Pearson's χ^2 p-value and Kolmogorov-Smirnov (KS) p-value for data in Figure 1.3, given to 5 decimal points. The first 3 columns represent the entire data, while the last 3 columns represent truncated data – the 'kick' and the last ten points (cut-off and bump region) have been removed. A custom-made function was coded to include errors for χ^2 as standard libraries do not provide for this. Statistical tests improve significantly for truncated data for high values of m.

R^2 coefficients showed a good fit, especially once the data was truncated, however linear regression is not valid for data in a double-logarithmic plot. The KS test obtained from Scipy's statistics library does not consider errors and thus does not yield an accurate result, while the χ^2 test does, producing p-values ~ 1 except when $m = 64$ due to the large 'kick' and bump. However, the numerical data fits the predictions well for the region before the cut-off.

1.4 Preferential Attachment Largest Degree and Data Collapse

1.4.1 Largest Degree Theory

To find the theoretical expectation for the largest degree k_1 (the degree of the vertex ranked highest by degree) we can state that:

$$\sum_{k=k_1}^{\infty} N p_\infty(k) = 1, \tag{21}$$

as there are no larger degrees. Writing $p_\infty(k)$ in partial-fraction form as in Eq. (20) we get:

$$Nm(m+1)\sum_{k=k_1}^{\infty}\left(\frac{1}{k+2}-\frac{1}{k+1}\right) - Nm(m+1)\sum_{k=k_1}^{\infty}\left(\frac{1}{k+1}-\frac{1}{k}\right) = 1, \tag{22}$$

which simplifies into:

$$\frac{m(m+1)}{k_1(k_1+1)} = \frac{1}{N}. \tag{23}$$

We can then re-arrange Eq. (23) into a quadratic,

$$k_1^2 + k_1 - Nm(m+1) = 0, \tag{24}$$

which has one positive (physical) solution:

$$k_1 = \frac{-1+\sqrt{1+4Nm(m+1)}}{2}, \tag{25}$$

which for very large N is expected to follow $k_1 \sim \sqrt{N}$.

1.4.2 Numerical Results for Largest Degree

The model was run for a fixed value of m while varying N to see how the largest degree k_1 scales with N. A value of $m = 4$ was chosen to minimise $\frac{m}{N}$, and thus finite-size effects, while ensuring $m > 1$.

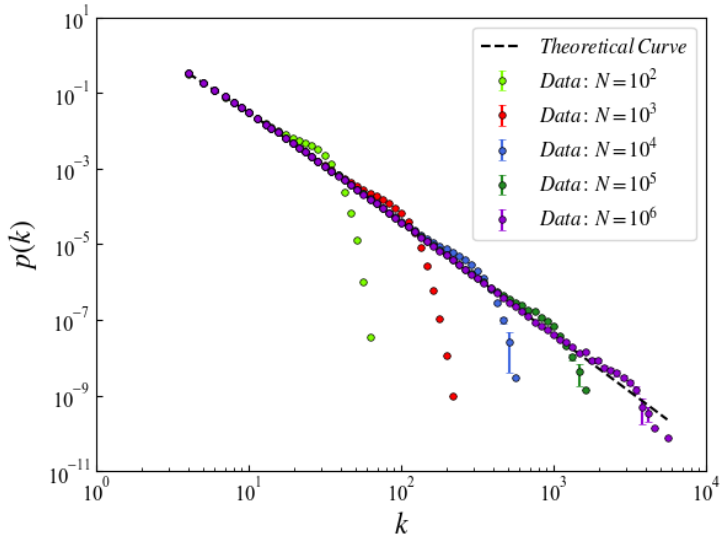

Figure 1.5. Log-binned degree distribution $p(k)$ plotted against degree k for six BA models using PA of sizes $N = 10^n$ and of $m = 4$, averaged $5\times 10^{8-n}$ times respectively for $n = 2,\ldots,6$. Dashed line represents the theoretical degree distribution $p_\infty(k)$. Error bars represent the standard error on the average values of $p(k)$. Data for different values of m overlap until the cut-off region, which depends on N.

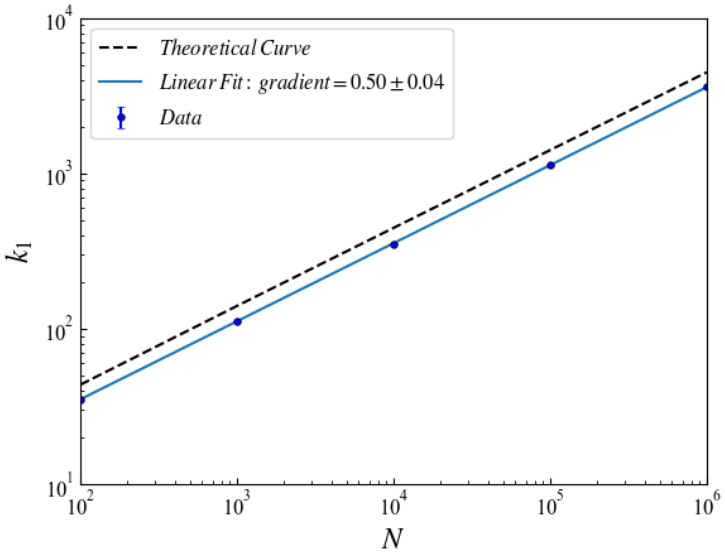

Figure 1.6. Largest degree k_1 plotted against total number of vertices N obtained from data in Figure 1.5. Error bars represent the standard errors in the average values of k_1. The dashed line represents the theoretical curve for k_1 in Eq. (25), while the blue line is a linear fit with gradient of 0.50 ± 0.04.

The uncertainty of the slope is obtained from the range of gradients between all data points, as the curve-fit error is unrealistically small. The gradient obtained agrees with the theoretical prediction in Eq. (25), however, a systematic offset is observed, indicating that the m dependence of Eq. (25) is incorrect for finite-size networks.

1.4.3 Data Collapse

To explore finite-size effects, the data was collapsed by scaling the degree by the maximum degree k_1 and the degree distribution by $p_\infty(k)$ so that the data collapses into a straight line of $p(k)/p_\infty(k) = 1$ and any deviations from that line indicate deviations between numerical data and theory.

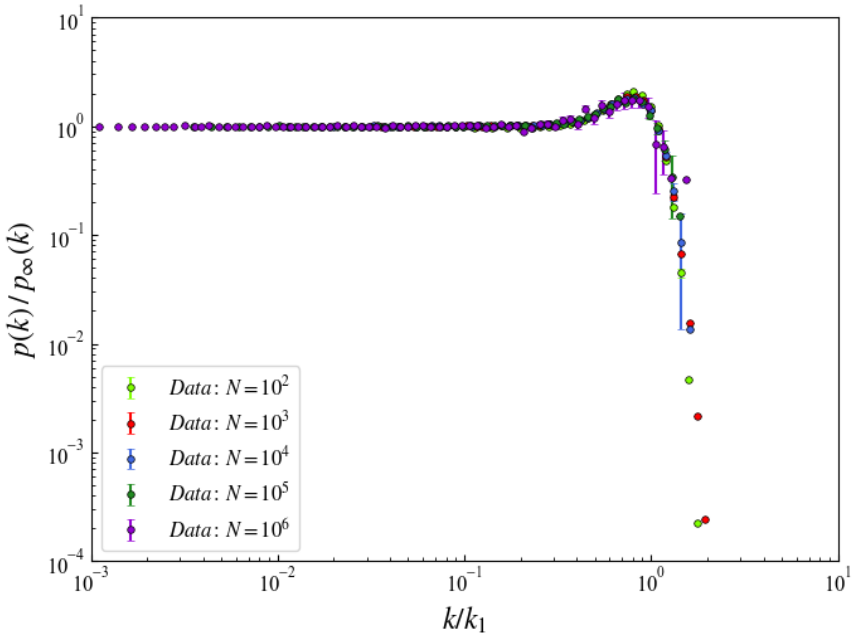

Figure 1.7. Data collapse: log-binned degree distribution over theoretical $p(k)/p_\infty(k)$ plotted against degree over maximum degree k/k_1 obtained from data in Figure 1.5. The error bars represent the standard errors in $p(k)$ scaled by $p_\infty(k)$. The numerical data is a good match when $p(k)/p_\infty(k) = 1$; the exponential decay and the 'bump' that precedes it are regions where the model deviates from theory.

The deviation of the numerical data at the tail of the distribution is evidence of the finite size of the models, resulting in an excess of vertices with degrees around k/k_1, before the degree distribution quickly decays towards zero.

2 Phase 2: Pure Random Attachment Π_{rnd}

2.1 Random Attachment Theoretical Derivations

2.1.1 Degree Distribution Theory

When the model uses pure random attachment (RA), vertices are chosen at random with equal probability regardless of k, so

$$\Pi(\mathrm{k}, \mathrm{t}) = \frac{1}{N(t)}. \tag{26}$$

To obtain the degree distribution we substitute Eq. (26) into the 'master' equation, as before, to get:

$$p_\infty(k)(1+m) = mp_\infty(k-1) + \delta_{k,m}. \tag{27}$$

Therefore, when $k > m$,

$$p_\infty(k) = \frac{m}{(m+1)} p_\infty(k-1), \tag{28}$$

removing the recursive dependence from Eq. (28) we get:

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

Finally, $p_\infty(m)$ can be obtained by considering $k = m$,

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

thus,

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

Since every new vertex has m edges, we can use the same argument as used in Eq. (16) to show that for random attachment,

$$p_\infty(k) = 0, \qquad k < m. \tag{32}$$

Moreover, we can check that the solution for $p_\infty(k)$ is normalised:

$$\sum_{k=m}^{\infty} p_\infty(k) = \sum_{k=m}^{\infty} \frac{m^{k-m}}{(m+1)^{k-m+1}}, \tag{33}$$

which once written-out term-by-term becomes clear it is a geometric series of the form,

$$\sum_{k=m}^{\infty} \frac{m^{k-m}}{(m+1)^{k-m+1}} = \sum_{n=0}^{\infty} \frac{1}{m+1}\left(\frac{m}{m+1}\right)^n = \frac{1/(m+1)}{1 - m/(m+1)} = 1, \tag{34}$$

as expected. Finally, we can see that $p_\infty(k)$ does not follow a power-law in k and is thus not expected to have a fat tail but rather an exponential decay tail, as vertices are connected randomly with uniform probability, not producing a scale-free network.

2.1.2 Largest Degree Theory

Using the same argument for the largest degree k_1 as was used for Eq. (21), we get:

$$\sum_{k=k_1}^{\infty} N \frac{m^{k-m}}{(m+1)^{k-m+1}} = 1. \tag{35}$$

We can simplify Eq. (35) by substituting the index $n = k - k_1$, resulting in

$$\sum_{n=0}^{\infty}\left(\frac{m}{m+1}\right)^{n+k_1}=\left(\frac{m}{m+1}\right)^{k_1}\sum_{n=0}^{\infty}\left(\frac{m}{m+1}\right)^{n}=\frac{m^m(m+1)^{1-m}}{N}, \tag{35}$$

which is an infinite geometric series with standard result:

$$\left(\frac{m}{m+1}\right)^{k_1}=\frac{m^m(m+1)^{1-m}}{N(m+1)}. \tag{36}$$

Taking the logarithm of both sides in Eq. (36) with some algebra results in:

$$k_1\,ln\left(\frac{m}{m+1}\right)=ln\left(\frac{m^m}{N(m+1)^m}\right), \tag{37}$$

which can re-arrange to get:

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

2.2 Random Attachment Numerical Results

2.2.1 Degree Distribution Numerical Results

The numerical results from the RA model were compared to the theoretical degree distribution $p_\infty(k)$ by fixing N and varying m, as before.

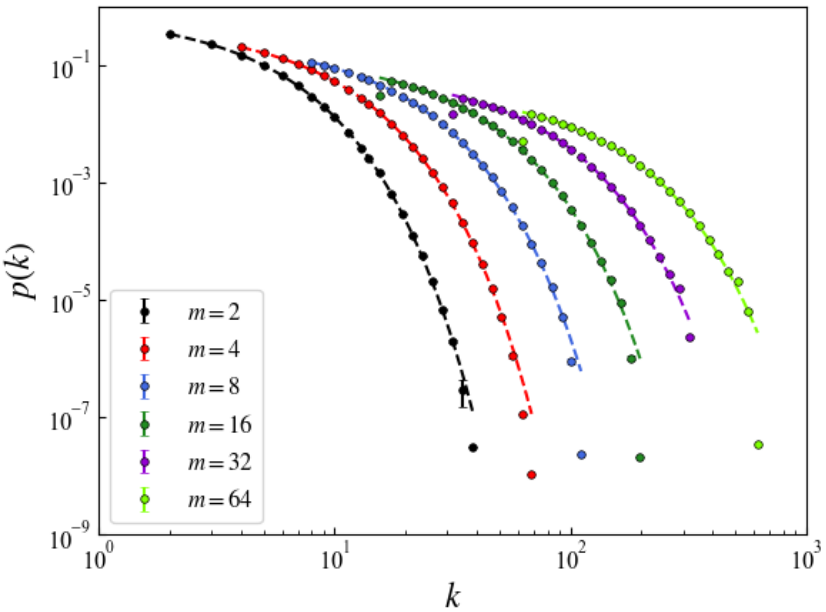

Figure 2.1. Log-binned degree probability distribution $p(k)$ plotted against degree k for six models using RA of size $N = 100{,}000$ and of $m = 2^n$, averaged $10 \times 2^{7-n}$ times respectively for $n = 1, \ldots, 6$. Dashed lines represent the theoretical degree distributions $p_\infty(k)$. Error bars represent the standard error on the average values of $p(k)$ – too small for most data points on this scale.

The degree distribution does not have a fat tail, as expected from the theoretical derivations, but rather, decays exponentially. The numerical data follow the theoretical result $p_\infty(k)$ very closely until the largest values of k, where there is a sharper exponential cut-off preceded by a

very small 'bump', due to the finite size of the network. A small 'kick' arising from binning is also observed for $m > 8$.

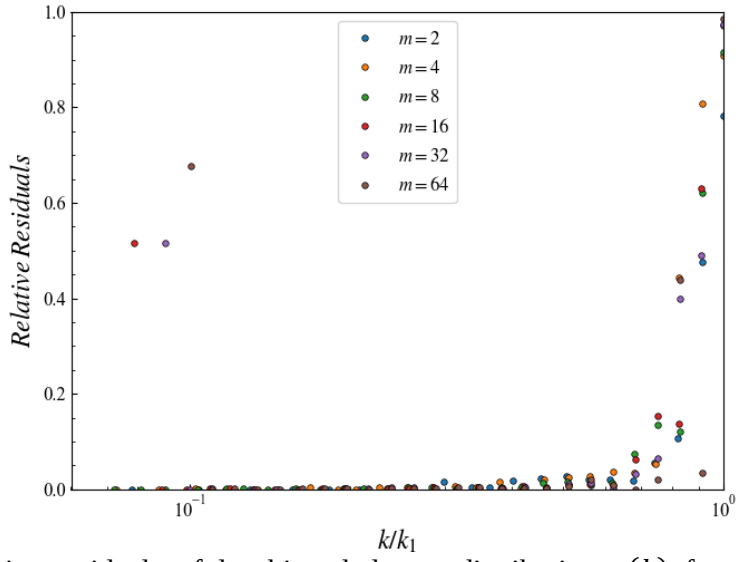

Figure 2.2. Relative residuals of log-binned degree distribution $p(k)$ from theoretical degree distribution $p_\infty(k)$ plotted against degree divided by maximum degree k/k_1, from data of Figure 2.1. The relative residuals are ~ 0 for all except 3 distinct points at low values of k representing the 'kicks', as well as the cut-offs and bumps. However, even those are small; ~ 0.05 *to* 1 % of the theoretical.

m	R^2	χ^2 (p-value)	*KS* (p-value)	R^2 (TRUNCATED)	χ^2 (p-value) (TRUNCATED)	*KS* (p-value) (TRUNCATED)
2	0.99999	0.99986	1.00000	0.99999	0.99874	0.99999
4	0.99999	0.99997	0.99999	0.99999	0.99093	0.99999
8	0.99999	0.99952	1.00000	0.99999	0.99971	1.00000
16	0.89480	0.75834	0.99999	0.99999	0.99985	1.00000
32	0.89503	0.53163	0.99999	0.99999	0.99681	0.99999
64	0.81625	0.28341	0.99999	0.99999	0.97838	1.00000

Table 2.1. R^2 coefficient, Pearson's χ^2 p-value and Kolmogorov-Smirnov (KS) p-value for data in Figure 2.1, given to 5 decimal points. The first 3 columns represent the entire data, while the last 3 columns represent truncated data – the 'kick' and the last ten points (cut-off and bump region) have been removed. A custom-made function was coded to include errors for χ^2 as standard libraries do not provide for this. Statistical tests improve significantly for truncated data for high values of m.

R^2 coefficients showed a good fit, especially once the data was truncated, however linear regression is not valid for data in a double-logarithmic plot. The KS test obtained from Scipy's statistics library does not consider errors and thus does not yield an accurate result, while the χ^2 test does, producing p-values ~ 1 except when $m = 16$ *to* 64 due to the large 'kick'. However, the numerical data fits the predictions well for the region before the cut-off.

2.2.2 Largest Degree Numerical Results

A fixed value of m and different values of N were used to see how the largest degree k_1 scales with N.

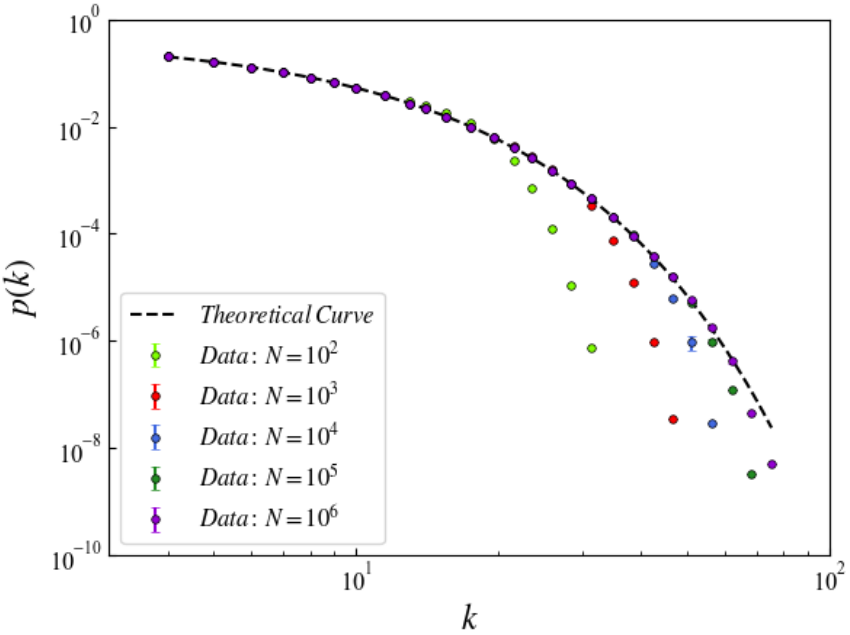

Figure 2.3. Log-binned degree distribution $p(k)$ plotted against degree k for six models using RA of sizes $N = 10^n$ and of $m = 4$, averaged $5\times 10^{8-n}$ times respectively for $n = 2,\ldots,6$. Dashed line represents the theoretical degree distribution $p_\infty(k)$. Error bars represent the standard error on the average values of $p(k)$. Data for different m overlaps until the cut-off region which depends on N.

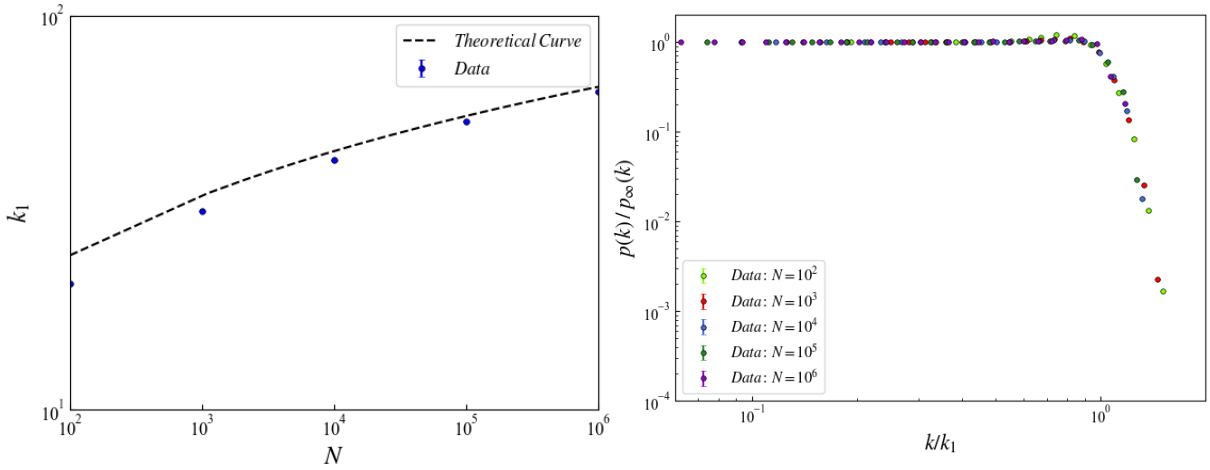

Figure 2.4. a) Largest degree k_1 plotted against number of vertices N obtained from data in Figure 2.3. Error bars represent the standard errors in the average values of k_1. The dashed line represents the theoretical curve for k_1, Eq. (38). b) Data collapse: log-binned degree distribution over theoretical $p(k)/p_\infty(k)$ plotted against degree over maximum degree k/k_1 obtained from data in Figure 2.3. The error bars represent the standard errors in $p(k)$ scaled by $p_\infty(k)$. The numerical data is a good match when $p(k)/p_\infty(k) = 1$. Notice that the bump is much smaller compared to PA and vanishes as N increases, as the number counts for all vertices converge due to the random element of RA.

Figure 2.4 a) demonstrates that the numerical data match the theoretical predictions for k_1, however, there is a deviation that gets smaller with increasing N, suggesting that the theoretical derivation is accurate for infinite-size networks.

3 Phase 3: Mixed Preferential and Random Attachment

3.1 Mixed Attachment Model Theoretical Derivations

For mixed preferential attachment (MPA), vertices attach using PA with probability q, or with RA with probability $1-q$, so that:

$$\Pi(\mathrm{k},\mathrm{t}) = \mathrm{q}\Pi_{pa} + (1-\mathrm{q})\Pi_{rnd}. \tag{39}$$

Substituting Eq. (39) into the 'master' equation, as before, we get:

$$p_\infty(k) = m\Pi(\mathrm{k}-1,\mathrm{t})p_\infty(k-1)N(t) - m\Pi(\mathrm{k},\mathrm{t})p_\infty(k)\,N(t) + \delta_{k,m}, \tag{40}$$

which we can expand in terms of q:

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

Employing the same arguments as used for previous models, when $k = m$, for $q \neq 0$,

$$p_\infty(m) = \frac{2}{2m+2-mq}, \qquad k = m, \tag{42}$$

and when $k > m$, we get:

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k+\frac{2m}{q}-2m-1}{k+\frac{2m}{q}-2m+\frac{2}{q}}, \qquad k > m, \tag{43}$$

for which exists a solution of the form:

$$p_\infty(k) = A\,\frac{\Gamma\left(k+\frac{2m}{q}-2m\right)}{\Gamma\left(k+\frac{2m}{q}-2m+\frac{2}{q}+1\right)}, \qquad k > m, \tag{44}$$

where $\Gamma(k+1) = k\Gamma(k), \Gamma(1) = 1$, and A is a constant. To normalise Eq. (44) we can choose specific values for q, for example: $q = 0.5\ or\ q = 2/3$, and find A by equating Eq. (44) to Eq. (42) when $k = m$, to get:

$$p_\infty(k) = \frac{6m(2m+1)(2m+2)}{(k+m)(k+m+1)(k+m+2)(k+m+3)}, \qquad q = 2/3,\ k \geq m, \tag{45}$$

and

$$p_\infty(k) = \frac{12m(3m+1)(3m+2)(3m+3)}{(k+2m)(k+2m+1)(k+2m+2)(k+2m+3)(k+2m+4)}, \qquad q = 1/2,\ k \geq m. \tag{46}$$

Wolfram Alpha's Sums function [2] was used, to check that the solutions are correctly normalised, while both models follow power laws in the large k limit, with $p_\infty(k) \sim k^{-4}$ and $p_\infty(k) \sim k^{-5}$ respectively, demonstrating that the closer q is to 0, the faster $p_\infty(k)$ decays as it uses less PA and more RA.

3.2 Mixed Attachment Model Numerical Results

Numerical results from the MPA model were compared to the theoretical degree distributions $p_\infty(k)$ for $q = 2/3$ and $q = 1/2$, by fixing N and varying m.

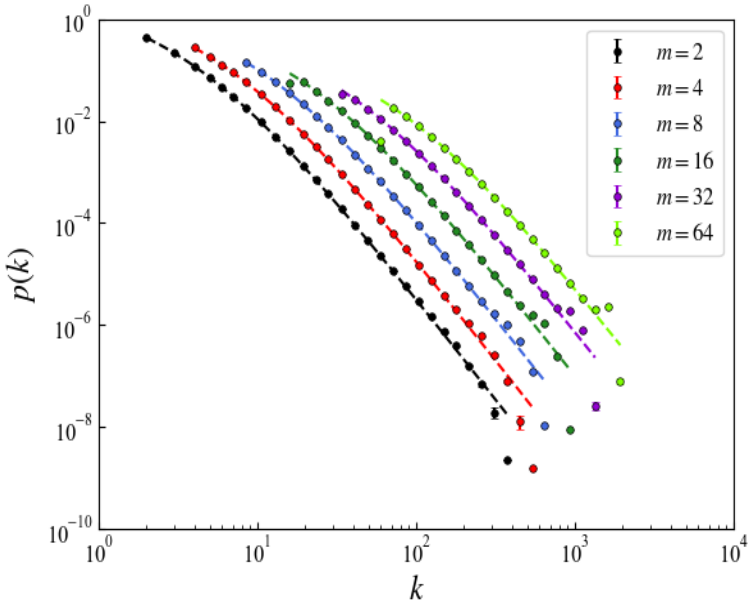

Figure 3.1. Log-binned degree probability distribution $p(k)$ plotted against degree k for six models using MPA and $q = 2/3$, of size $N = 100{,}000$ and of $m = 2^n$, averaged $10 \times 2^{7-n}$ times respectively for $n = 1, \ldots, 6$. Dashed lines represent the theoretical degree distributions $p_\infty(k)$. Error bars represent the standard error on the average values of $p(k)$ – too small for most data points on this scale.

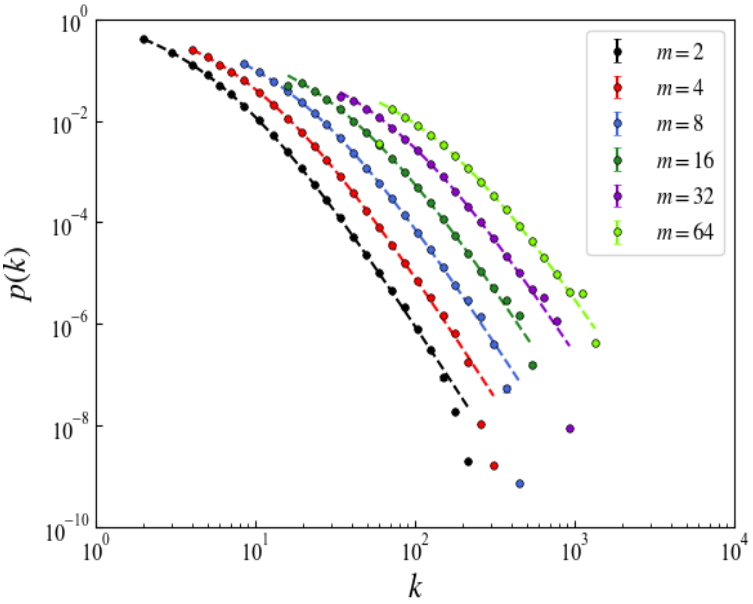

Figure 3.2. Log-binned degree probability distribution $p(k)$ plotted against degree k for six models using MPA and $q = 1/2$, of size $N = 100{,}000$ and of $m = 2^n$, averaged $10 \times 2^{7-n}$ times respectively for $n = 1, \ldots, 6$. Dashed lines represent the theoretical degree distributions $p_\infty(k)$. Error bars represent the standard error on the average values of $p(k)$ – too small for most data points on this scale.

The degree distributions have fat tails as expected, which, decay faster than that of the PA model. Moreover, Figure 3.2 demonstrates that the tail for $q = 1/2$ decays faster than that for $q = 2/3$, expected, as that model 'chooses' PA and 'advantages' vertices of large k less often.

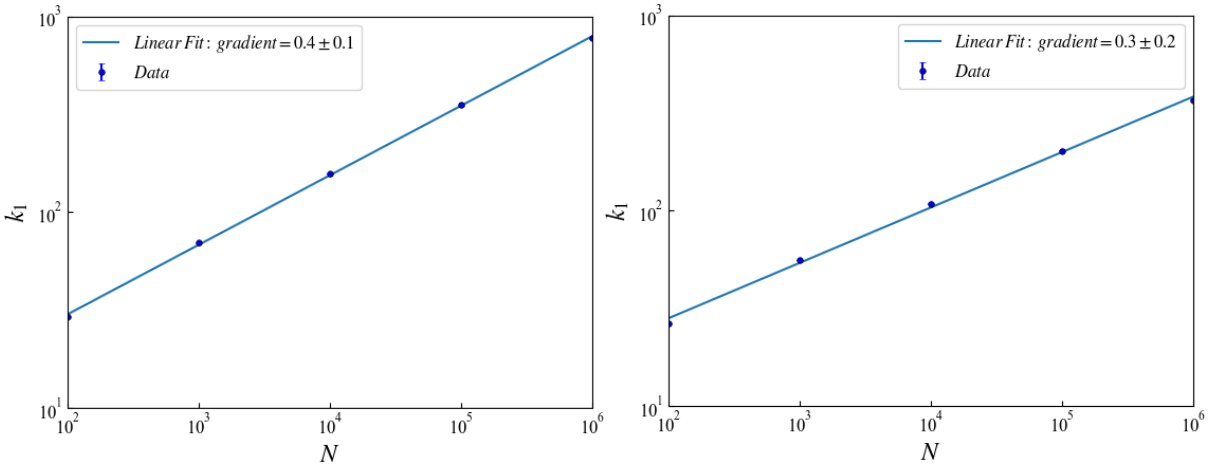

Figure 3.3. a) Largest degree k_1 plotted against total number of vertices N obtained from data in Figure 3.1. Error bars represent the standard errors in the average values of k_1. The blue line is a linear fit with gradient of 0.4 ± 0.1. b) Largest degree k_1 plotted against total number of vertices N obtained from data in Figure 3.2. Error bars represent the standard errors in the average values of k_1. The blue line is a linear fit with gradient of 0.3 ± 0.2. Notice how as $q \to 1$ the gradient tends to 0.5 (expected for PA).

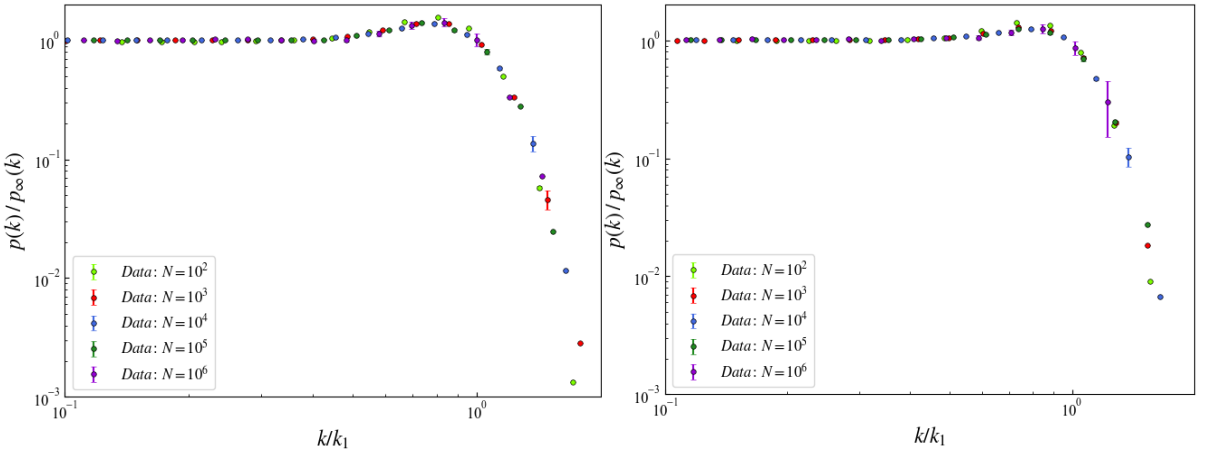

Figure 3.4. a) Data collapse: log-binned degree distribution over theoretical $p(k)/p_\infty(k)$ plotted against degree over maximum degree k/k_1 obtained from data in Figure 3.1. b) Data collapse: log-binned degree distribution over theoretical $p(k)/p_\infty(k)$ plotted against degree over maximum degree k/k_1 obtained from data in Figure 3.2. The error bars represent the standard errors in $p(k)$ scaled by $p_\infty(k)$. The numerical data is a good match when $p(k)/p_\infty(k) = 1$; the exponential decay and the 'bump' that precedes it are regions where the model deviates from theory.

Finally, the data collapse for both values of q show that the numerical data fits the theoretical predictions well until the 'bump' and cut-off region, while $q = 2/3$ has a bigger bump than $q = 1/2$.

4 Conclusions

All models were found to follow the theoretical results even for small system sizes, until the degree distribution decayed quickly in a cut-off region caused by finite size effects. The BA model with PA demonstrated it follows a power-law and has a fat tail, while RA degree distributions decayed exponentially. Many real-life networks are a form of MPA models, which demonstrated to have a fat tail distribution as long as $q > 0$, however, decaying faster than PA.

References

[1] T. S. Evans, *Network Notes: Complexity and Networks Course, Level 3 course*, Imperial College London, London, (2021), pg. 49

[2] Wolfram Alpha, *Calculus and Analysis – Sums*, Accessed: 21/03/2021, available at: https://www.wolframalpha.com/examples/mathematics/calculus-and-analysis/sums/