Network Project
A Growing Network Model

CID: (01413021)

27th March 2020

Abstract: Three models of growing networks were investigated by deriving analytical expressions and comparing it to numerical simulations. First, the BA model was implemented using preferential attachment, giving a power law degree distribution. Second, a similar model but with random attachment was considered. Finally, a model using a random walk to choose the vertex to connect new edges was implemented. For $q = 0$, i.e zero path length, this model gives a random graph. However, as $q \to 1$, i.e the path length tending to ∞, the degree distribution tends towards the behaviour of the preferential attachment model. All results were verified with numerical simulations. The theoretical model for preferential attachment appeared to be a good fit with a confidence of $> 95\%$. The two other model's fit had less accuracy but still managed to represent the overall behaviour.

Word Count: 2492 words excluding font page, figure captions, table captions, acknowledgement and bibliography.

1 Introduction

In this project a simple model of a growing network is implemented to study its degree distribution. First, the Barabási and Albert model (BA model) is implemented using preferential attachment. When a new vertex is added to the system an edge is connected to an existing vertex in the system, which is selected with a probability proportional to its degree. Central to this model is the fat tail behaviour where it forms high degree 'hubs'. In the original paper by Barabási and Albert, the model is used to describe the *World Wide Web*, as a new website is more likely to link to more popular sites.

The overall objective is to create a numerical simulation of the model, using *python* and the *NetworkX* library and compare the simulation against theoretical predictions. With a focus on how the approximations in the numerical model affect the results. In addition, the analysis and results from the BA model will be repeated with a model based on pure random attachment.

Finally, using a similar algorithm as in the BA model a random walk through the vertices is analysed to analyse the degree distribution.

1.1 Definition

2 Phase 1: Pure Preferential Attachment Π_{pa}

2.1 Implementation

2.1.1 Numerical Implementation

An initial graph and list of edges was constructed using the *NetworkX* library. At every time step $t \to t+1$ a new vertex is added, which is connected with m edges to existing vertices. The existing vertex is selected by randomly picking an edge from a list containing all the edges and choosing an end of the edge at random. This way the vertex is selected with a probability proportional to its degree. The list is updated after adding m edges, to avoid self-loops. This procedure is repeated until N nodes are added.

2.1.2 Initial Graph

Since the BA model does not allow self-loops and parallel edges and m edges are added when a new vertex is added, at least m vertices must be present to add a new vertex. Therefore, an initial graph of m vertices was used to minimise the influence of the initial graph. When the first new vertex is added to the system, it can not be biased towards any vertices, which requires all vertices to have the same degree. The BA model is based on an unweighted, undirected graph. To meet all these requirements, a simple, complete graph of m vertices was used as an initial graph.

2.1.3 Type of Graph

The system is initiated by a simple undirected graph, which is a graph that does not have any self-loops or parallel edges and is unweighted. This type of graph is also kept for the rest of the implementation as the BA model only allows one edge between any two pairs of vertices.

2.1.4 Working Code

A series of checks was conducted to ensure that the implementation was working as intended. At time $t = 0$, the number of edges are equal to m, as N vertices are added, $m \times N$ edges are added. Such that the number of edges in the system $E(t)$ at time t is equal to $E(t) = E(0)+m(N(t)-N(0))$, where $N(t)$ is the number of vertices in the system at time t. In the initial graph all vertices have a degree of m and as more edges are added, the degree increases. Therefore, the degree of all vertices must always be greater than m. Care were taken to update the list of edges after adding m nodes to avoid self-loops. The degree distribution was also compared to NetworkX's inbuilt BA graph generator to ensure the implementation was consistent. Since all these checks gave consistent results, it can be assumed that the code is working as intended.

2.1.5 Parameters

The BA model is initialised using the 'BAmodel' class. This takes the input parameters N vertices to add to the network sand m number of edges to add to each new vertex. In addition, the random seed can be specified to reproduce the same network. To find the theoretical predictions for the network, we are required to use large N limit. To demonstrate how this approximation affects the numerical results, networks with values of N between 10^3 and 10^5 were used. m needs to be greater than 1 for preferential attachment. m values between 2 and 64 were used to study the degree distribution for different values of m.

2.2 Preferential Attachment Degree Distribution Theory

2.2.1 Theoretical Derivation

In order to derive a theoretical equation of the degree distribution, we can consider the *master equation* for the BA model

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

where $n(k,t)$ is the number of vertices of degree k at time t and $\Pi(k,t)$ is the probability that one of the new edges at time t is connected to an existing vertex with degree k [2]. This is a difference equation that gives an expression for the number of vertices of degree k at the next time step $(t+1)$, given the distribution at previous times. To solve the master equation we can note that the number of edges at time t is given by

$$E(t) = E(0) + m(N(t) - N(0)) \tag{2}$$

rearranging and taking the long time limit

$$\frac{E(t)}{N(t)} = \frac{E(0)}{N(t)} + m - \frac{N(0)}{N(t)} \quad \therefore \quad E(t) \approx mN(t). \tag{3}$$

In the BA model the probability Π is given by

$$\Pi(k,t) = \frac{k}{2E(t)} = \frac{k}{2mN(t)}, \tag{4}$$

where the approximation in Eq. (3) has been used. Substituting this into the master equation gives

$$n(k,t+1) = n(k,t) + \frac{m(k-1)n(k-1,t)}{2mN(t)} - \frac{mkn(k,t)}{2mN(t)} + \delta_{k,m}. \tag{5}$$

This can be rewritten in terms of of the probability distribution $p(k,t) = n(k,t)/N(t)$ giving

$$N(t+1)p(k,t+1) - N(t)p(k,t) = +\frac{1}{2}(k-1)p(k-1,t) - \frac{1}{2}kp(k,t) + \delta_{k,m}. \tag{6}$$

In the limit that $t \to \infty$ and $N \gg$ one can assume that a stable form of the probability distribution exist, such that

$$\lim_{t\to\infty} p(k,t) = p_\infty(k). \tag{7}$$

Using this and the fact that $N(t+1) = N(t) + 1$, Eq. (6) turns into

$$p_\infty(k) = \frac{1}{2}\left[(k-1)p_\infty(k-1) - kp_\infty(k)\right] + \delta_{k,m} \tag{8}$$

Consider first the case $k > m$, where $\delta_{k,m} = 0$, such that Eq. (8) can be written as

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k-1}{k+2} \tag{9}$$

To solve this equation analytically, we must use the Gamma function $\Gamma(n)$ [3]. The central property of the Gamma function is that

$$\Gamma(z+1) = z\Gamma(z) \tag{10}$$

which means that any function that can be written in the form

$$\frac{f(z)}{f(z-1)} = \frac{z+a}{z+b} \tag{11}$$

can be solved with the Gamma function through

$$f(z) = A\frac{\Gamma(z+1+a)}{\Gamma(z+1+b)} \tag{12}$$

where A is a constant. From Eq. (9), we can identify $a = -1$ and $b = 2$, such that

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

where the last expression comes from the definition of the Gamma function $\Gamma(n) = (n-1)!$. For the case $k = m$, the probability of a vertex having a degree less than m is zero, such that Eq.(8) reduced to

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

By considering the case $k = m+1$ in Eq. (13) we obtain

$$p_\infty(k=m+1) = \frac{A}{(m+1)(m+2)(m+3)}. \tag{15}$$

This can be equated to Eq. (11), using $z = m+1$ such that, such that

$$\frac{p_\infty(m+1)}{p_\infty(m)} = \frac{A}{(m+1)(m+2)(m+3)}\frac{(2+m)}{2} = \frac{m}{(m+3)}. \tag{16}$$

where the expression obtained in Eq. (14) was used for $p_\infty(m)$. Solving for A we obtain

$$A = 2m(m+1). \tag{17}$$

Substituting into Eq. (13) gives the final form of the probability distribution

$$p_\infty(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \qquad \text{for } k \geq m. \tag{18}$$

2.2.2 Theoretical Checks

To ensure that the theoretical solution has the correct properties, the probability distribution is checked using Kolmogorov's axioms. These are

1. The probability of an event occurring is a real number between 0 and 1.
2. The probability of any event occurring in the entire sample space is 1.
3. The probability that an event A or B occurs is given by the sum of their probabilities, that is $P(A \cup B) = P(A) + P(B)$

From Eq. (18), axiom 1 is clearly true and since there are no conditional probabilities in the mode, axiom 3 must also be true. To check axiom 2, we can use the normalisation condition for $p_\infty(k)$

$$\sum_{k=m}^{\infty} p_\infty(k) = 1 \tag{19}$$

This is expressed as

$$\sum_{k=m}^{\infty} p_\infty(k) = 2m(m+1) \sum_{k=m}^{\infty} \frac{1}{k(k+1)(k+2)}. \tag{20}$$

To find the sum, consider the standard result

$$\sum_{n=1}^{X} \frac{1}{n(n+1)(n+2)} = \frac{1}{4} - \frac{1}{2(X+1)(X+2)}. \tag{21}$$

The sum between X and ∞, can be found by taking the difference of two sums and using the standard result to get the expression

$$\sum_{n=X}^{\infty} \frac{1}{n(n+1)(n+2)} = \sum_{n=1}^{\infty} \frac{1}{n(n+1)(n+2)} - \sum_{n=1}^{X-1} \frac{1}{n(n+1)(n+2)} = \frac{1}{2X(X+1)}. \tag{22}$$

Setting $X = m$ and substituting into Eq. (20) gives the required result

$$\sum_{k=1}^{\infty} p_\infty(k) = 2m(m+1) \frac{1}{2m(m+1)} = 1. \tag{23}$$

Additionally, we must ensure that $p_\infty(k) \to 0$ as $k \to \infty$. From Eq. (18), we can see that

$$\lim_{k\to\infty} p_\infty(k) \approx \lim_{k\to\infty} k^{-3} \to 0 \tag{24}$$

as required.

2.3 Preferential Attachment Degree Distribution Numerics

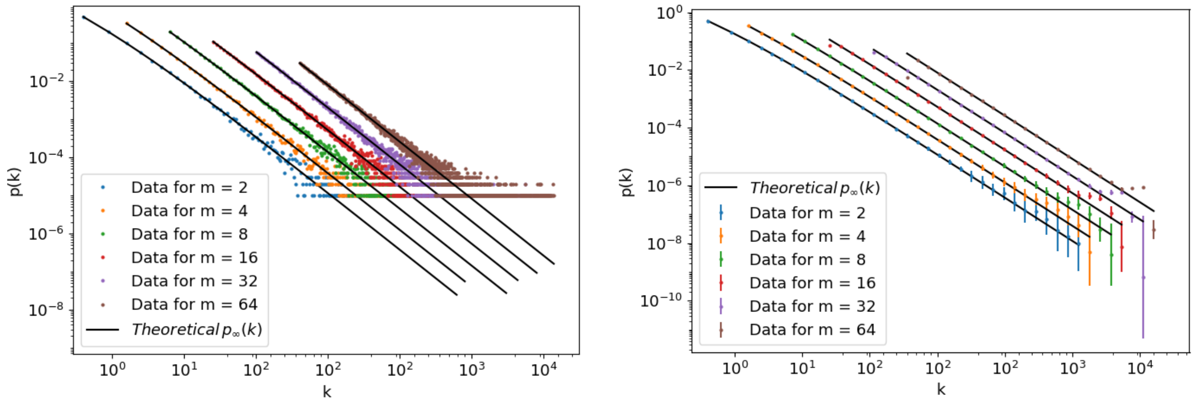
2.3.1 Fat-Tail

A fat-tailed distribution is one which decays slower than an exponential, giving its characteristic 'fat-tail' [2]. The fat-tail distribution has a few large vertices with large values of k. This can cause problems as there will be many large values of k that are missing in the data. These can be crucial for understanding the degree distribution. This issue can be resolved with log-binning, which uses exponentially spaced bins for the degree distribution, where the bins are defined as

$$\frac{b_{i+1}}{b_i} = \exp\{\Delta\} \text{ where } \Delta > 0. \tag{25}$$

As a result, the degree distribution will be equally spaced at intervals of Δ on a log-log plot. Figure 1a shows a the raw data before being log-binned. Figure 1b illustrates the result of using log-binning, resolving the issues caused by a fat-tail. In the project, $exp(\Delta)$ is set to 1.2.

2.3.2 Numerical Results



(a)

(b)

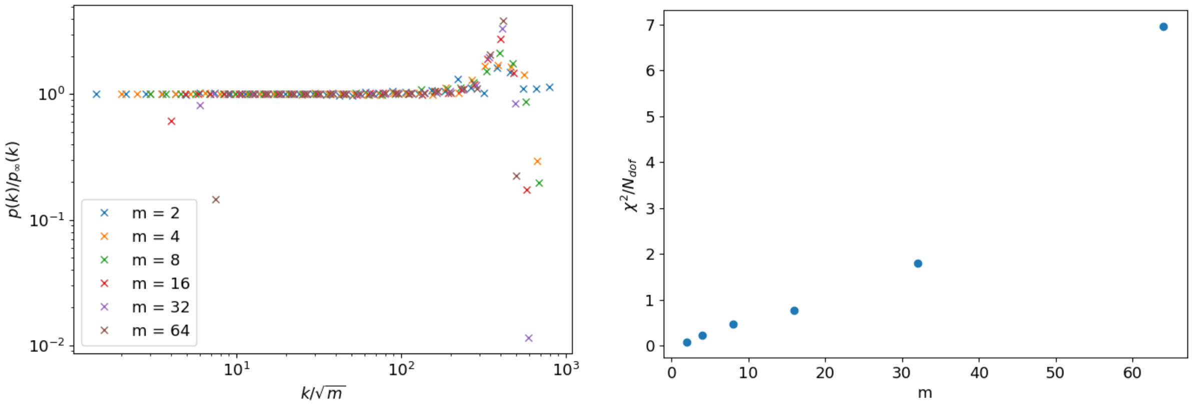
Figure 1: a) is a plot of the raw data from a network of $N = 10^5$ and $m = 2, 4, 8, 16, 32, 64$. b) Is the same network but averaged over 50 realisations and log binned with scale 1.2. The theoretical estimate is shown by black line and the error bars are calculated from the standard deviation of the data.

Figure 1a shows the numerical results for the BA model on a log-log plot of the degree distribution $p(k)$ plotted against k for various values of m. The gaps in the probability distribution arises from the discrete values of k giving many values of k with the same value of $p(k)$, typical for fat-tailed distributions.

In Figure 1b, the same model has been log binned and averaged over 50 realisations. Associated error bars on the probability density has been found from the standard deviation on the different realisations. The numerical result is compared to the theoretical estimate for preferential attachment, given by Eq. (18). The result appears to follow the theoretical estimate, except for the first degree bin, which systematically will have fewer occurrences than it should and for large values of k where finite-size scaling effects become more prevalent. The linear behaviour on the log-log plot, illustrates the power-law relationship of the degree distribution and thus demonstrates that the BA model gives a scale-free network.

To further investigate whether $p_\infty(k)$ from Eq. (18) is a good estimate of the numerical result Figure 2a, shows the ratio of the measured $p(k)$ to the theoretical estimate, as a function of $k/\sqrt{m}$. It shows that $p_\infty(k)$ is a good estimate as the ratio is around 1, except for larger values of k. The discrepancy at large values of k was shown as a bump in the degree distribution. This is a result of the finite-size scaling effect, that limits the growth of the network.

2.3.3 Statistics



(a)

(b)

Figure 2: a) Shows the ratio of the data and theory, which is in agreement for small k. Plotted for $N = 10^5$ vertices and average of 50 realisations. b) χ^2 per degree of freedom excluding the first and last k value, where a value around 1 indicates a good fit.

To determine if the numerical data fits the theoretical prediction, a statistical test was performed. The Kolmogorov-Smirnov test is usually applied to continuous distributions and the R^2 squared test is not appropriate for power-law distributions. Therefore, the reduced χ^2 test was used to test the goodness of the fit. The reduced χ^2 squared is defined as the χ^2 per degree of freedom, where the χ^2 is a sum of squared deviations weighted by the standard deviations:

$$\chi^2 = \sum_i \left[\frac{O_i - p_\infty(ki)}{\sigma_i}\right]^2 \tag{26}$$

where O_i is the observed data for bin i, with the corresponding error $\sigma_i$and ki the k value of bin i. The error has been calculated by finding the standard deviation for 50 different realisations. This a valid method since we expect each bin to be normally distributed by the central limit theorem for enough realisations. Some of the uncertainty in the data arises from the use of log-binning, as some information on the value of k is lost when the data is log binned. When averaging over several realisations, this uncertainty will be made less important.

The calculated values of χ^2 per degree of freedom are shown in Figure 2b, plotted as a function of m. For a good fit, $\chi^2/N_{\mathrm{DOF}} \approx 1$, which is the case for $m = 2, 4, 8, 16$. The growing deviation from the theory at higher values of m is likely due to the finite-size scaling effect. In addition, since the graph is initiated with m vertices, small values of m have the least influence from the initial graph.

To obtain a conclusive result on the goodness of the fit, the p-value was calculated from χ^2 and the degrees of freedom. The p-value gives a measure of the probability of observing the measured test statistic. Since the χ^2 value show large fluctuations for large k values, p_{sliced} excludes the first and last point. The p-values are given in the table below.

m	2	4	8	16	32	64
p	0.999	0.999	0.25	0	0	0
p_{sliced}	0.999	0.999	0.999	0.775	0.012	0

The hypothesis that they are independent can be rejected at 95% significance level for $m > 8$ in the whole data set and $m < 16$ for the sliced data set. This illustrates that the theoretical estimate is a good fit to the numerical results, except for larger values of m.

2.4 Preferential Attachment Largest Degree and Data Collapse

2.4.1 Largest Degree Theory

To find the theoretical behaviour of the largest expected degree k_1, we can consider a sum of the number of vertices with degree equal to k_1 and higher. On average you would only expect to find one vertex, such that

$$\sum_{k=k_1}^{\infty} N p_\infty(k) = 1. \tag{27}$$

The system size N can be taken outside the sum as it is independent of k and using the result from Eq. (22) with $X = k_1$ the sum can be evaluated

$$\sum_{k=k_1}^{\infty} p_\infty(k) = 2m(m+1) \times .\frac{1}{2k_1\,(k_1+1)} = \frac{1}{N} \tag{28}$$

Rearranging gives a quadratic equation

$$k_1^2 + k_1 - Nm(m+1) = 0 \tag{29}$$

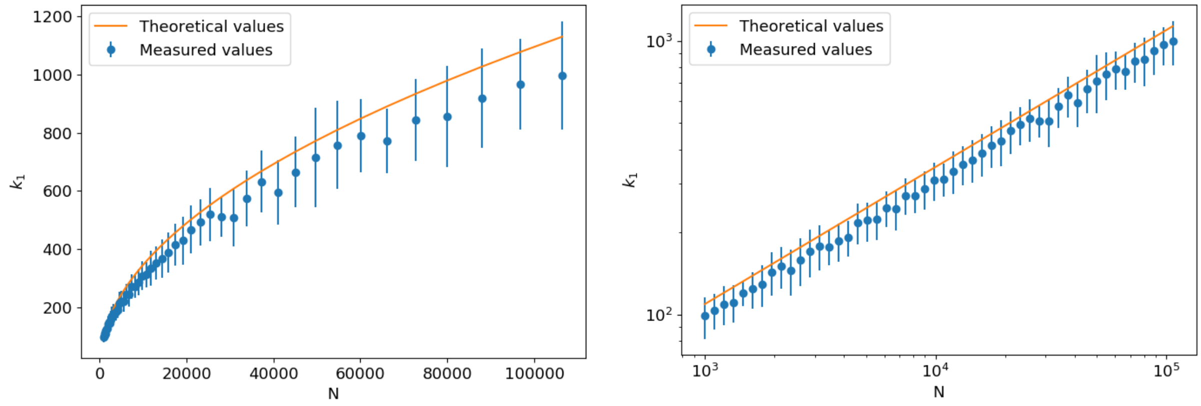
which can be solved for k_1 to give the positive solution

$$k_1 = \frac{-1 + \sqrt{1 + 4Nm(m+1)}}{2}. \tag{30}$$

In the limit $N \to \infty$, the largest degree scales with N as $k_1 \propto N^{1/2}$.

2.4.2 Numerical Results for Largest Degree

Figure 3 shows the largest degree k_1 plotted as a function of N. k_1 has been averaged over 30 realisations and the uncertainty is found from the standard deviation. The values of N were in the range of $10^3 \to 10^5$ in order to demonstrate the the behaviour over several orders of magnitude. The log-log plot shows the linear relationship with a gradient of ≈ 0.5, as predicted earlier. m was set to 3, as this was the value that gave the best fit to the theoretical values. For $m = 3$, all the values lie below the theoretical line. However, the values were found to vary between being over and below the theoretical values for different values of m. This shows that the theoretical estimate is accurate in scaling with N, but not with m. Since the measured values are closely related to the theoretical value, we can conclude that the estimate is accurate and the assumption that there is on average one vertex with degree k_1 is valid.



(a)

(b)

Figure 3: a) shows the mean of the largest degree for different N values. Averaged over 30 realisations. b) Plotted on a log-log scale. The error bars represents the standard deviation.

2.4.3 Data Collapse

To demonstrate the finite-size scaling effect, a data collapse is performed. Figure 4 shows $p(k)/p_\infty(k)$ plotted as a function of k/k_1, for different values of N. m was set to 3 to limit the effect of the initial graph and minimise the finite-size scaling effect. The data collapse successfully collapses the values onto the same functional form and illustrates better the bump due to the finite-size scaling effect. The values of k near k_1 have the potential to grow much larger for an infinite system size, however for a finite system size they get suppressed down, leaving the bump seen in the data.

3 Phase 2: Pure Random Attachment Π_{rnd}

3.1 Random Attachment Theoretical Derivations

3.1.1 Degree Distribution Theory

In a similar manner as before, the theoretical derivation of the degree distribution is obtained by considering the master equation. For random attachment, the probability of attaching the new edge to a given vertex is defined as

$$\Pi(k,t) = \frac{1}{N(t)} \tag{31}$$

substituting this into the master equation (1) gives

$$n(k,t+1) = n(K,t) + \frac{m}{N}n(k-1,t) - \frac{m}{N}n(k,t) + \delta_{k,m}. \tag{32}$$

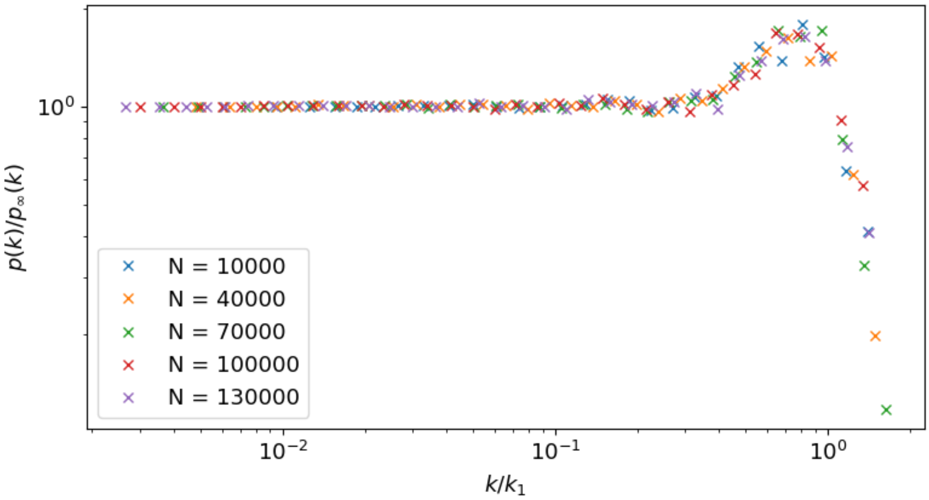


Figure 4: Data collapse produced by plotting $p(k)/p_\infty(k)$ vs $k/k1$. Notice the bump at the tail of the distribution due to finite-size scaling effects.

Using the probability distribution $p(k,t) = n(k,t)/N(k,t)$ this can be rewritten as

$$p(k,t+1)N(k,t+1) = p(k,t)N(k,t) + mp(k-1,t) - mp(k,t) + \delta_{k,m}. \tag{33}$$

In the long time limit

$$p_\infty(k)(N(t)+1) = p_\infty(k)N(t) + mp_\infty(k-1) - mp_\infty(k) + \delta_{k,m} \tag{34}$$

which simplifies to a recurrence relation for the probability distribution

$$p_\infty(k) = \frac{mp_\infty(k-1) + \delta_{k,m}}{(m+1)} \tag{35}$$

To find the functional form of the probability distribution, we first consider the case $k = m$ where $p_\infty(k = m-1) = 0$ such that

$$p_\infty(k = m) = \frac{1}{m+1} \tag{36}$$

. Using this result and noting that $\delta_{k,m} = 0$ for $k > m$ the case where $k = m+1$ gives

$$p_\infty(k = m+1) = \frac{m}{(m+1)^2}. \tag{37}$$

From the recurrence relation we can therefore conclude that the probability distribution can be written as

$$p_\infty(k) = \frac{m^{k-m}}{(m+1)^{k-m+1}} \qquad \text{for } k \geq m. \tag{38}$$

To verify that the probability has the correct properties, Kolmogorov's axioms can be used again. Axiom 1 is clearly true as $k, m \in^+$ and $0 \leq p_\infty(k) \leq 1$. Axiom 3 is also true as the model does not have conditional probabilities. Finally, we must verify axiom 2 by checking the normalisation condition

$$\sum_{k=m}^{\infty} p_\infty(k) = \sum_{k=m}^{\infty} \frac{m^{k-m}}{(m+1)^{k-m+1}} = 1. \tag{39}$$

Taking a factor of 1/m outside of the sum gives

$$\sum_{k=m}^{\infty} p_{\infty}(k) = \frac{1}{m}\sum_{k=m}^{\infty}\left(\frac{m}{m+1}\right)^{k-m+1}. \tag{40}$$

By performing a change of variable $n = k - m + 1$ and finding the sum of the geometric series we find

$$\sum_{k=m}^{\infty} p_{\infty}(k) = \frac{1}{m}\sum_{n=1}^{\infty}\left(\frac{m}{m+1}\right)^{n} = \frac{1}{m}\cdot m = 1. \tag{41}$$

as required.

3.1.2 Largest Degree Theory

Give your best theoretical estimate of how the largest expected degree, k_1 (subscript 1 indicating the degree of the vertex ranked first by degree size) depends on the number of vertices N in a finite size system. 4

Using the same method as for preferential attachment, Eq. (27) for random attachment gives

$$\sum_{k=k_1}^{\infty} N\frac{m^{k-m}}{(m+1)^{k-m+1}} = 1. \tag{42}$$

Taking N and a factor of m outside gives

$$\frac{N}{m}\sum_{k=k_1}^{\infty}\left(\frac{m}{1+m}\right)^{k-m+1} = 1. \tag{43}$$

By performing a change of variable, $n = k - k_1$ we get

$$\frac{N}{m}\left(\frac{m}{1+m}\right)^{k_1-m+1}\sum_{n=0}^{\infty}\left(\frac{m}{1+m}\right)^{n} = 1 \tag{44}$$

The sum can again be evaluated by calculating the sum of the geometric series (now including $n = 0$), giving

$$\frac{N}{m}\left(\frac{m}{1+m}\right)^{k_1-m+1}(m+1) = 1. \tag{45}$$

Taking the logarithm and rearranging, an expression for k_1 is obtained

$$k_1 = m - \frac{\log N}{\log m - \log(m+1)}. \tag{46}$$

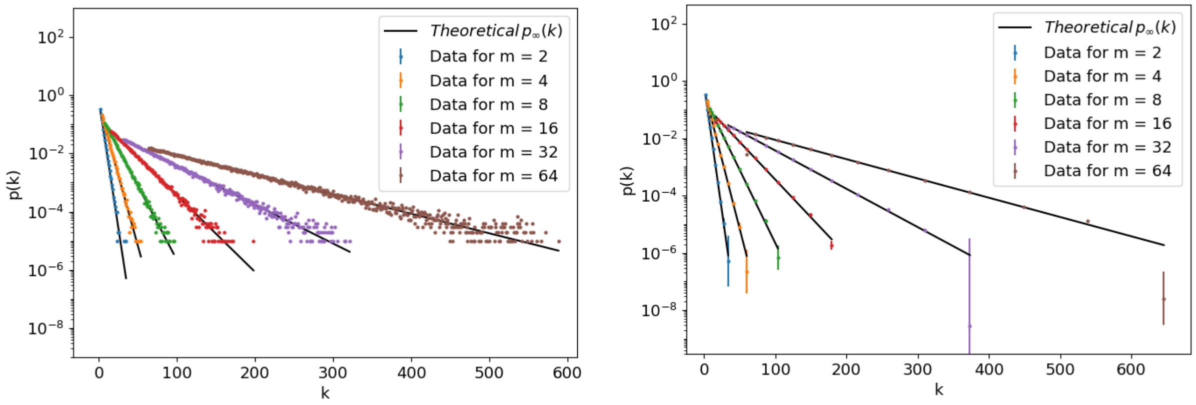
In the limit $N \to \infty$, $k_1 \propto \log N$ which means that k_1 will always be smaller for random attachment, compared to preferential attachment. This is as expected since preferential attachment is proportional to the degree of the vertex, such that hubs of high degrees are formed.

3.2 Random Attachment Numerical Results

3.2.1 Degree Distribution Numerical Results

Figure 5 a and b, shows the raw and log-binned probability distribution for random attachment, respectively. The fat-tail effect is less prominent for random attachment, which is as expected from the derived $p_\infty(k)$ as it decays more slowly than exponential. The effect becomes more prevalent for larger values of m. It is plotted on a semi-log plot to get a linear relationship. The black line represents the theoretical estimate derived in Eq. (38). It shows good agreement with theory, except for the first value of k and for large values of k. The fact that the network is linear on a semi-log plot and not log-log plot, demonstrates that this is a random network rather than a scale-free.

Again the numerical result is compared to theoretical estimate by taking the ratio in Figure 6a. It shows little difference for small values of k (except for the first value of k) but have large fluctuations for large values of k.



(a)

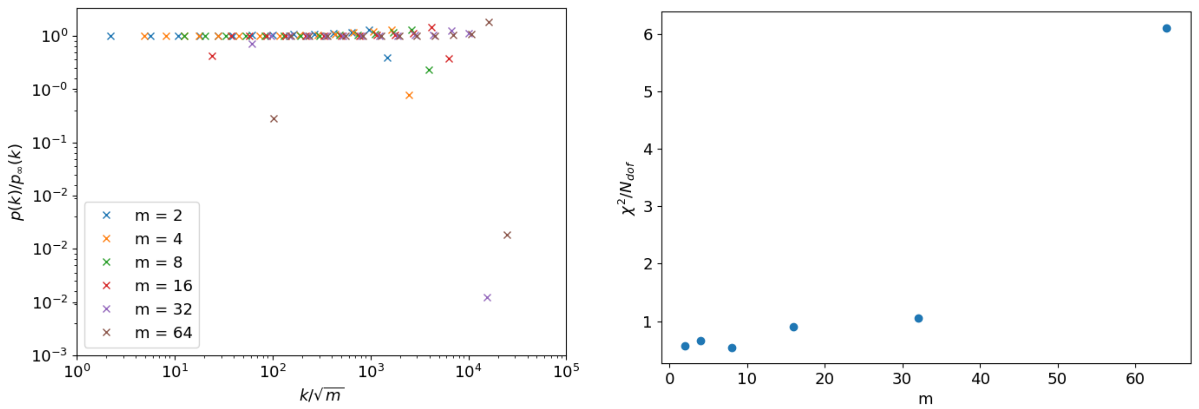
(b)

Figure 5: a) shows unbinned data for random attachment for $N = 10^5$ vertices. a) Log-binned data using scale of 1.2. The data has been averaged over 50 realisations. Notice the linear behaviour on the semi-log plot, indicating not a scale-free network.

Using the same method as in the preceding section, the χ^2 value was found. Figure 6b, shows the χ^2 per degree of freedom, calculated by excluding the first and last value of k. It shows that the theoretical model is accurate for m values up to 32. Again the, best results are obtained for small values of m.

The hypothesis that they are independent can be rejected at 95% significance level for $m > 8$ in the whole data set and $m < 32$ for the sliced data set. This illustrates that the theoretical estimate is a good fit to the numerical results, except for larger values of m. However, the results are not as good as for the preferential attachment.

m	2	4	8	16	32	64
p	0.923	0.74	0.880	0	0	0
p_{sliced}	0.833	0.780	0.858	0.526	0.396	0



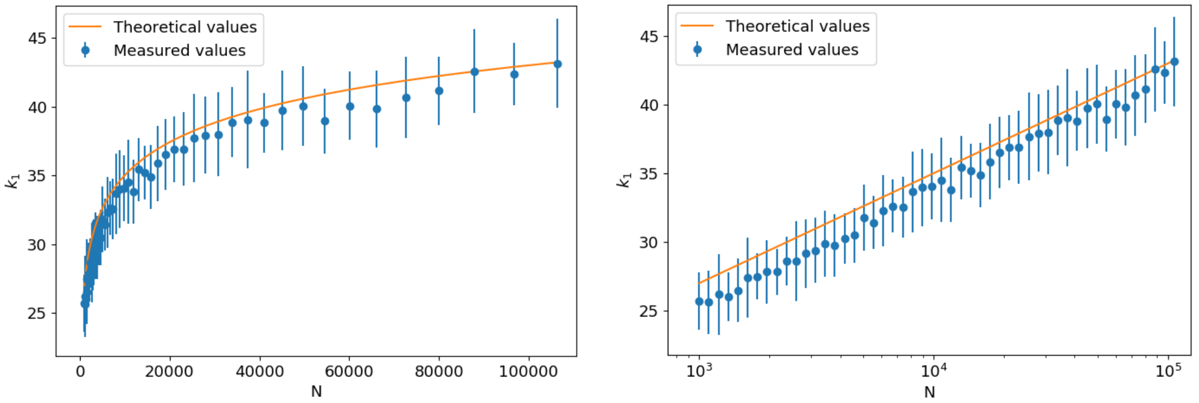
(a)

(b)

Figure 6: a) shows the ratio of $p(k)/p_\infty(k)$ vs $k/\sqrt{m}$ for random attachment. Averaged over 50 realisations on a network of $N = 10^5$ vertices. b) χ^2 per degree of freedom for the sliced data.

3.2.2 Largest Degree Numerical Results

Figure 7 shows the measured values of k_1 and the theoretical estimate from Eq. (46) plotted on semi-log plot for different values of N. Again the $m = 3$ is used as it gives the smallest deviation from the line. The same effect of scale-free with N but not m was observed for random attachment. Since the measured values are closely related to the theoretical value, we can again conclude that the revised estimate for random attachment is valid. However, the deviation from the line and standard deviation appears to be larger than for preferential attachment.



(a)

(b)

Figure 7: a) shows the mean of the largest degree for different N values, averaged over 30 realisations. b) Plotted on a semi-log scale. The error bars represents the standard deviation.

4 Phase 3: Random Walks and Preferential Attachment

4.1 Implementation

The random walk model was implemented using a similar algorithm as in the BA model, but with some differences:

- Rather than keeping track of the edges, it has a list of the existing vertices to choose from
- A vertex is initially chosen at random from the existing vertices
- With a probability $1 - q$, the random walk is ended
- With a probability q a neighbouring vertex is selected for a step in the random walk
- Before ending the random walk, the vertex at which it stopped at is checked against the existing edges, to avoid self-loops
- Using this method, m edges are added for each vertex, up to N vertices

The network was again initialised with a complete, simple graph. For this part the log-binning scale $exp(\Delta)$ was set to 1.1, to allow more points for the statistics test.

4.2 Numerical results

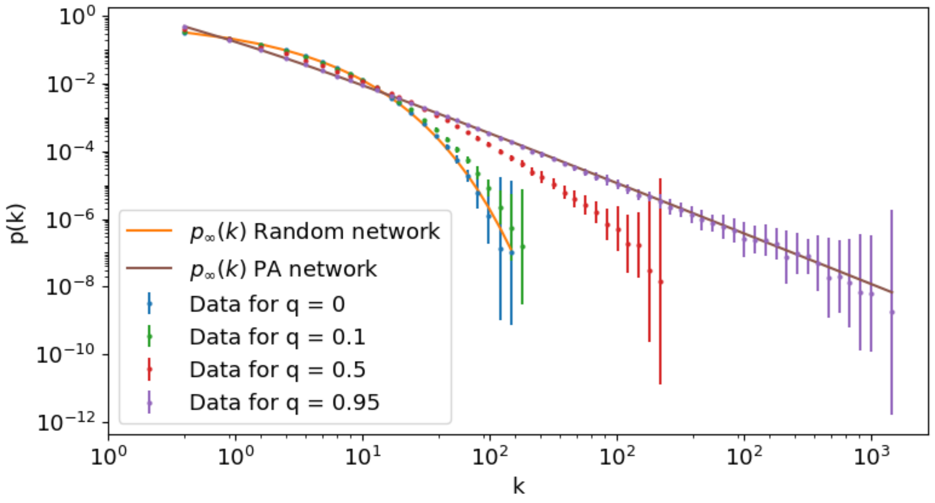


Figure 8: Shows $p(k)$ and $p_\infty(k)$ for a random walk model. With parameters $N = 10^5$ $m = 2$. Averaged over 50 realisations and plotted for q values $0, 0.1, 0.5, 0.95$.

Figure 8 shows the degree distribution $p(k)$ for the random walk model, log-binned and averaged over 50 realisations. It is plotted for different values of q. For $q = 0$ the length of the random walk is 0 and so the model becomes a random network as a vertex is chosen at random every time a vertex is added. For $q = 0.5$ the degree distribution tends more towards a linear distribution on the log-log plot. Finally, for $q = 1$, the path length

$\rightarrow \infty$. The figure shows this approximated by taking $q = 0.95$, which is close enough to 1 but still does not take too long to run. It has almost a linear relationship, demonstrating that it is a scale-free network.

Since you are more likely to arrive at a high degree vertex than a low degree one, we expect the random walk model to tend towards the preferential attachment degree distribution as $q \rightarrow \infty$. To test this hypothesis, the theoretical estimate earlier derived for the BA model in Eq. (18) is plotted on top of the measured data in Figure 8. It appears to give an adequate fit to the data, but deviates above the line for small k and below for large k. Using the same χ^2 test and ignoring the first and last point the p-value is found to be $p = 0.929$. This shows that the random walk model has a similar behaviour to the preferential attachment in this case. As q gets closer to 1, we would expect the deviation to become smaller.

Similarly, for $q = 0$, the measured data can be compared to the theoretical estimate found in Eq. (38). The figure shows that the line follows the points closely, which is confirmed by the p-value of 0.999. This demonstrates that for $q = 0$, the random walk model is indeed a random network.

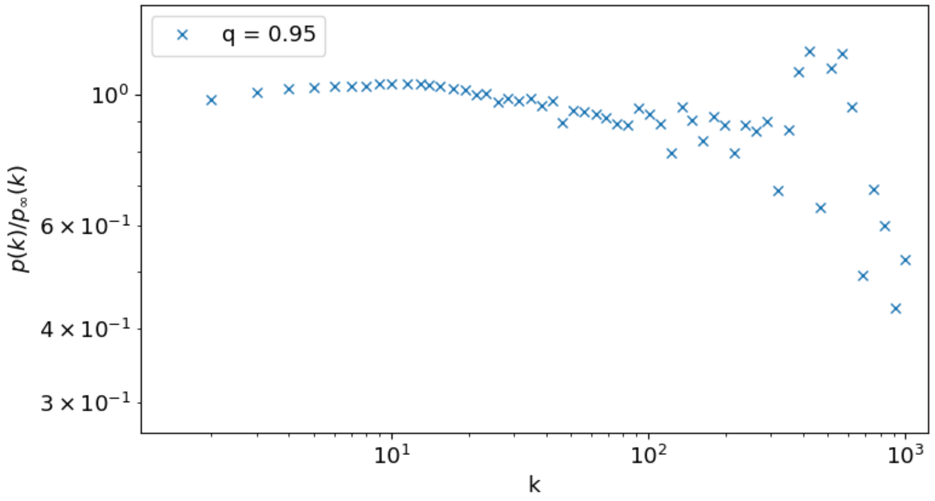


Figure 9: Shows the ratio of the measured data $p(k)$ for $q = 0.95$ to the expected preferential attachment theory $p_\infty(k)$. The bump at the tail of the distribution is due to to finite-size scaling effects.

4.3 Discussion of Results

The probability of arriving at a vertex is proportional to the number of ways of arriving at the vertex, which again is given by its degree [4]. This explains why the degree distribution follows the same power-law as preferential attachment.

To investigate the finite-size scaling effects of this model, Figure 9 shows the ratio of the measured data to the preferential attachment theoretical estimate, plotted as a function of k. It shows that the estimate is accurate for smaller values of k, while it has large fluctuations at larger values of k.

Finally, it is worth noting that the random walk model only uses local information, contrary to the BA model which uses global information for its normalisation. This network of a network structure makes this model a self-organising mechanism.

5 Conclusions

Three models of growing networks were investigated by deriving analytical expressions and comparing it to numerical simulations. First, the BA model was implemented using preferential attachment, giving a power law degree distribution. Second, a similar model but with random attachment was found to follow a geometric series degree distribution. Finally, a model using a random walk to choose the vertex to connect new edges was implemented. For $q = 0$, i.e zero path, this model gives a random graph. However, as $q \to 1$, i.e the path length tending to ∞, the degree distribution tends towards the behaviour of the preferential attachment model. All results were verified with numerical simulations, using the χ^2 test. The theoretical model for preferential attachment appeared to be a good fit with a p-value of $> 95\%$. The two other model's fit had less accuracy but still managed to represent the overall behaviour. The random walk model, uses structure of a network to produce the networks, making it a scale-free network and self-organising mechanism.

References

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

[2] T.S. Evans, *Complexity and Networks Lecture Notes*, Imperial College London Department of Physics, (2020).

[3] T.S. Evans, *Complexity and Networks Problem Sheet 2*, Imperial College London Department of Physics, (2020).

[4] T.S. Ecans, *Random Walks and Networks*, Imperial College London, Presentation from workshop on Complex Systems, USP São Paulo, (2011).