Networks Project: Growing Networks Model

Kei Tsun Yeung

26-3-2018

Abstract: Three growing network models were studied, namely the (a) Barabási-Albert (BA) Pure Preferential, (b) Pure Random and (c) Mixed Prefrential and Random models. The degree distribution $p(k)$ was derived to be (a) a power-law decay $p(k) \propto k^{-3}$, (b) an exponential decay $p(k) \propto (m/(m+1))^{k-m}$ where m is the edge to vertex ratio, and (c) a power decay of $p(k) \propto k^{-5}$, assuming that the system is infinitely large. Simulation was limited by the finite network size and a fat-tail and cutoff effect can be seen in the BA and mixed models. A scale-free behaviour was observed with the BA model where a data collapse had been successfully implemented.

Word count: 2439 in report

1 Introduction

Three growing network models have been studied, namely the Barabási-Albert (BA) Pure Preferential Attachment, Pure Random Attachment, and Mixed Preferential and Random Attachment models. The BA model [1] [2] employs a *cumulative advantage* principle [3], where vertices with a lot of existing attachments will gain more like a snowballing effect. This model is used to describe citation and website networks, where popular papers or websites gain more citations and links to them respectively [4].

The Pure Random model is one where probability of connecting to any vertex is constant; and the Mixed model introduces an additional probability in choosing the above two models when an edge is added.

2 Theory and Algorithm

2.1 Definition of Model

The model is described as follows.

1. *Initialisation*: Set up the initial network, a **complete** graph with $m+1$ vertices, at time t_0.
2. *Addition*: At time $t \to t+1$, add a new vertex and m edges by
   - Connecting one end of the edge to the new vertex.
   - Connecting the other end of the edge of each new edge to an existing vertex chosen with probability Π which is defined as:
     - **BA**: $\Pi_{\mathrm{pa}}(k) \propto k$, where k is the degree of a vertex.
     - **Pure Random**: $\Pi_{\mathrm{rnd}} \propto 1$, i.e. probabilities of all vertices are equal.
     - **Mixed**: $\Pi_{\mathrm{mix}}(k) = q\Pi_{\mathrm{pa}}(k) + (1-q)\Pi_{\mathrm{rnd}}$, so choosing the BA model with probability q.
3. *Iteration*: Return to step **2** and repeat until the final number of vertices N is reached.

2.2 Master Equation

The master equation describes the number of vertices with degree k at a later time $(t+1)$, i.e. $n(k, t+1)$, given the degree distribution at current time t and probability of choosing an existing vertex with degree k $\Pi(k, t)$, with the form of

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

where $m\Pi(k, t)n(k, t)$ is the number of vertices with degree k and an added edge after adding m edges, meaning they no longer have a degree k. Similarly, $m\Pi(k-1, t)n(k-1, t)$ is the number of vertices which have degree k after adding m edges. $\delta_{k,m}$ accounts for the fact that the new vertex has degree m if $k = m$.

In the long time $n(k,t) \to N(t)p_\infty(k)$ as $t \to \infty$, so

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

Since the number of vertices added at $t \to t+1$ is 1, i.e. $N(t+1) = N(t) + 1$, so

$$p_\infty(k) = mN(t)\left[\Pi(k-1,t)p_\infty(k-1) - \Pi(k,t)p_\infty(k)\right] + \delta_{k,m} \quad (3)$$

is the theoretical degree distribution in the long time limit.

3 Implementation and Test

The initial network was chosen as a complete graph of $m+1$ vertices so that each vertex has exactly m edges and m edges can be added without repeat for the first added vertex. A complete graph was used so that the probability was not biased at the beginning. Repeated edges and self-loops between two vertices were not allowed, as this can bias the distribution of the BA model and is not realistic. As such, a simple graph was generated.

By starting with $m+1$ vertices at the beginning, the theoretical number of edges $E(t) \neq mN(t)$ in the beginning. Moreover, a complete graph means that the initial probability distribution did not follow $p_\infty(k)$. However, for a long enough time where $N \to \infty$, the number of vertices in the initial graph should not matter, as the results presented show.

The master equation (eq. (3)) assumes a negligible probability of choosing the same vertex twice, and this is only true when $N \to \infty$, so there is a finite-size effect which was minimised by investigating $N \gg 1$.

The numerical limitation to this model is due to the fact that N is finite, so probability of any k is limited to a multiple of $1/N$. This can be observed as the fat-tail distribution in Section 4.3, where probabilities of the largest k do not decay as a power law. To minimise this issue of poor statistics, a large number of vertices $N = 10^6$ was added, and this was repeated 100 times to produce 10^8 values of k. The repeating was performed as opposed to adding $N = 10^8$ because for the BA model (and hence mixed), vertices with the largest k at the end of the tail have a larger probability of attracting attachment, so as time goes on, their increase in degree will be greater than the ones with smaller k, hence the fat-tail will be more significant. This was also done for the pure random attachment model so that all the conditions in generating each data set are the same, for comparison purposes. Log-binning can also be performed to extract data from areas with poor statistics by increasing the bin size exponentially.

The implementation of $\Pi_{\text{pa}}(k,t)$ was achieved by saving the two vertices connected to each edge into an edge list. Selecting a vertex in the edge list when a new edge was added would give a probability proportional to its degree, because it is equal to the number of times a vertex appears in the edge list. For Π_{rnd}, since the vertices are added in ascending order from 0 to $N-1$, the existing vertex to be attached was simply one picked from the list of numbers from 0 up to (current vertex -1). For Π_{mix}, a probability q was assigned to determine which model to be used for each edge added.

Visualisation of the networks was used to verify that the implementation was correct. For the BA model, the initial graph is as shown in Figure 1 (a). New vertices are added one by one, with m edges connected to existing vertices, as in Figures 1 (b) and (c). The process continues until N vertices have been added. A notable point is that most of the new edges added in Figure 1 have been added to the same vertices connected to the blue edges, showing a sign of preferential attachment.

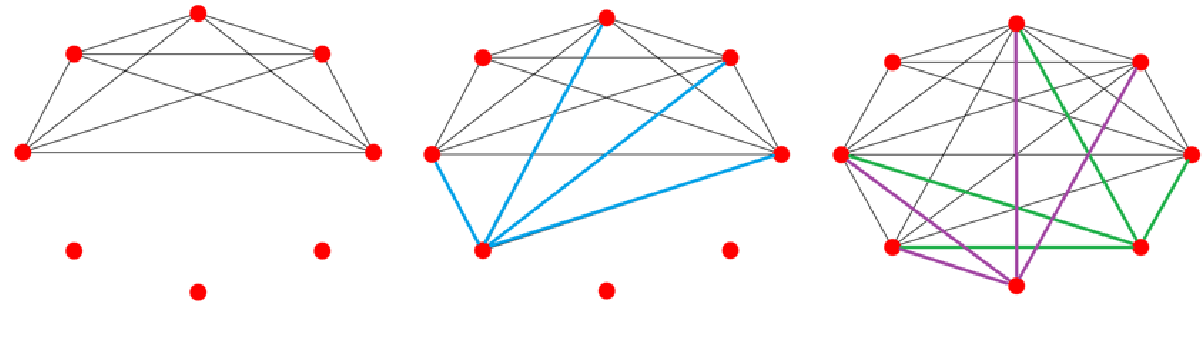

Figure 1: Visualisation of the BA model with $m = 4$ and $N = 8$ vertices. (a) Initial complete graph with $m + 1 = 5$ vertices. (b) Adding one vertex and m edges (blue). (c) Process continues until all vertices have been added. Note that most of the green and purple edges have been added to vertices connected to blue edges, showing a sign of preferential attachment.

More vertices are needed to demonstrate clearly preferential attachment behaviour. Figure 2 shows networks of BA, pure random, and mixed models with $N = 100$. Dense concentration of edges can be seen in (a) and less so in (c) at the top-right corner. There is a more even distribution of edges in (b), although slightly denser to the top. This is because the `networkx draw_circular` function starts adding from the right side then moves anti-clockwise. This means that the top half of vertices are added first. Since they exist in the network for a longer time, they can accumulate more edges, explaining the slightly denser behaviour at the top. Therefore, the models have been implemented correctly.

4 Barabási-Albert (BA) model: Pure Preferential Attachment

4.1 Theoretical degree distribution $p_\infty(k)$

Requiring $\Pi_{\text{pa}} \propto k$ with the normalisation condition that $\sum_{k=0}^{\infty} n(k,t)\Pi_{\text{pa}}(k,t) = 1$ and number of edges $E(t) = \sum_{k=0}^{\infty} n(k,t)\frac{k}{2}$, the probability $\Pi_{\text{pa}} = k/2E(t)$. At $t \to t+1$, 1 new vertex and m new edges are added, so as $t \to \infty$, $E(t) \to mN(t)$ and

$$\Pi_{\text{pa}} = \frac{k}{2mN(t)}. \tag{4}$$

Substituting this form into eq. (3) gives

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

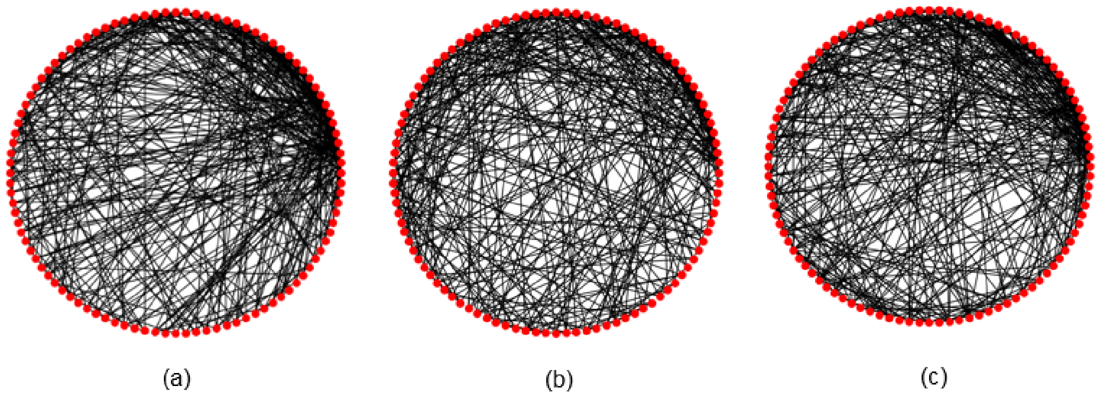

Figure 2: Networks of (a) BA, (b) pure random and (c) mixed attachment models with $N = 100$. (a) There is a very dense concentration of edges to the top-right corner of the graph. (b) There is a more even distribution of edges as the probability $\Pi_{\rm rnd}$ is constant. (c) There is a slightly dense concentration of edges to the top-right corner, although not so much as (a).

For $k > m$, $\delta_{k,m} = 0$, so

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

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

To solve eq. (7), consider

$$\frac{f(z)}{f(z-1)} = \frac{z+a}{z+b}. \tag{8}$$

Using the gamma function $\Gamma(z)$ with the central property $\Gamma(z) = (z-1)\Gamma(z-1)$,

$$f(z) = A\frac{\Gamma(z+1+a)}{\Gamma(z+1+b)} = A\frac{(z+a)\Gamma(z+a)}{(z+b)\Gamma(z+b)} \tag{9}$$

is proposed as a trial solution. Since

$$f(z-1) = A\frac{\Gamma(z+a)}{\Gamma(z+b)}, \tag{10}$$

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

showing that $f(z)$ is a solution to eq. (8).

Applying eq. (9) to eq. (7) with $a = -1$ and $b = 2$,

$$p_\infty(k) = A\frac{\Gamma(k)}{\Gamma(k+3)} \tag{12}$$

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

$$= \frac{A}{k(k+1)(k+2)} \qquad \text{for} \quad k \geq m. \tag{14}$$

As $t \to \infty$, $p_\infty(k < m) = 0$ is expected as every new vertex added has at least m edges. For $k = m$, consider

$$p_\infty(m) = \frac{1}{2}\left[(m-1)p_\infty(m-1) - mp_\infty(m)\right] + 1 \tag{15}$$

$$= -\frac{1}{2}mp_\infty(m) + 1 \qquad \text{as} \quad p_\infty(m-1) = 0 \tag{16}$$

$$p_\infty(m) = \frac{2}{m+2}. \tag{17}$$

Consider $k = m+1$,

$$p_\infty(m+1) = \frac{1}{2}[mp_\infty(m) - (m+1)p_\infty(m+1)] \tag{18}$$

$$(m+3)p_\infty(m+1) = mp_\infty(m). \tag{19}$$

Using eq. (14) for $p_\infty(m+1)$ and eq. (17) for $p_\infty(m)$,

$$(m+3)\frac{A}{(m+1)(m+2)(m+3)} = \frac{2m}{m+2}. \tag{20}$$

So

$$A = 2m(m+1) \qquad \text{and} \tag{21}$$

$$p_\infty(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \qquad \text{for} \quad k \geq m. \tag{22}$$

To show that eq. (22) is normalised, partial fraction decomposition can be used to obtain

$$p_\infty(k) = m(m+1)\left(\frac{1}{k} - \frac{2}{k+1} + \frac{1}{k+2}\right). \tag{23}$$

Summing over all possible $k \geq m$ gives

$$\sum_{k=m}^{\infty} p_\infty(k) = m(m+1)\sum_{k=m}^{\infty}\left(\frac{1}{k} - \frac{2}{k+1} + \frac{1}{k+2}\right) \tag{24}$$

$$= m(m+1)\left(\frac{1}{m} - \frac{1}{m+1}\right) = 1, \tag{25}$$

showing that $p_\infty(k)$ is normalised.

For the fat-tail distribution $k \gg 1$, an approximate form of eq. (22) can be obtained as

$$p_\infty(k) = \frac{2m(m+1)}{k^3(1+\frac{1}{k})(1+\frac{2}{k})} \propto k^{-3} \qquad \text{for} \quad k \gg 1, \tag{26}$$

meaning a power-law decay in k as $k \to \infty$ and $t \to \infty$.

4.2 Theoretical largest degree k_1

It follows from the power-law decay that the largest expected degree k_1 should have the smallest probability. Assuming that only 1 vertex has degree k_1, the probability from degree k_1 to infinity is expected to be of the order of $1/N$, i.e.

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

Using eq. (22) for $p_\infty(k)$ and decomposing into partial fractions again,

$$m(m+1)N \sum_{k=k_1}^{\infty} \left(\frac{1}{k} - \frac{2}{k+1} + \frac{1}{k+2} \right) \sim 1 \tag{28}$$

$$m(m+1)N \left(\frac{1}{k_1} - \frac{2}{k_1+1} \right) \sim 1 \tag{29}$$

$$\frac{m(m+1)N}{k_1(k_1+1)} \sim 1. \tag{30}$$

Completing the square and solving for k_1 gives

$$k_1 \sim -\frac{1}{2} \pm \frac{\sqrt{1+4m(m+1)N}}{2} \tag{31}$$

$$k_1 \sim \sqrt{\frac{1}{4} + Nm(m+1)} - \frac{1}{2} \tag{32}$$

$$k_1 \propto \sqrt{Nm(m+1)} \qquad \text{for} \quad N \gg 1, m \gg 1. \tag{33}$$

4.3 Numerical Results and Discussion

4.3.1 Comparison to theoretical distribution

The probability distribution of degree was simulated with a network of size $N = 10^6$ vertices given computational time and memory constraints. The m values used were $m =$ 1, 2, 4, 8, 16 and 32. For each m, the simulation was repeated 100 times to obtain better statistics without causing the fat-tail to enhance, as explained in Section 3. Log-binning was also performed on the data to extract the cutoff feature due to the finite size N, with bin size growing at a scale factor of $a = 1.25$ found by trial and error. This will allow a more even distribution of data points in each bin. The degree distribution of all m values is as shown in Figure 3.

The degree distribution starts from $k = m$ and since the power-law decay is roughly the same, the linear part are parallel to each other. The probability of obtaining the minimum degree $k = m$ decreases as m increases, as the degree distribution is spread over a larger range of k values, so it is less concentrated at the minimum k.

The theoretical distribution (eq. (22)) was compared with the data visually as shown in Figure 4. The binned values are clearly to be seen with a cutoff after a slight 'bump'. The cutoffs are finite size effect of the network, as the maximum degree is limited by the

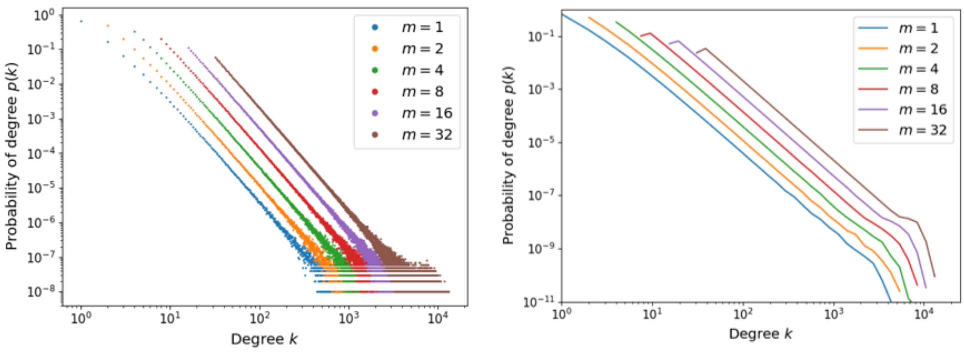

Figure 3: Degree distribution of BA model, unbinned (left) and log-binned (right).

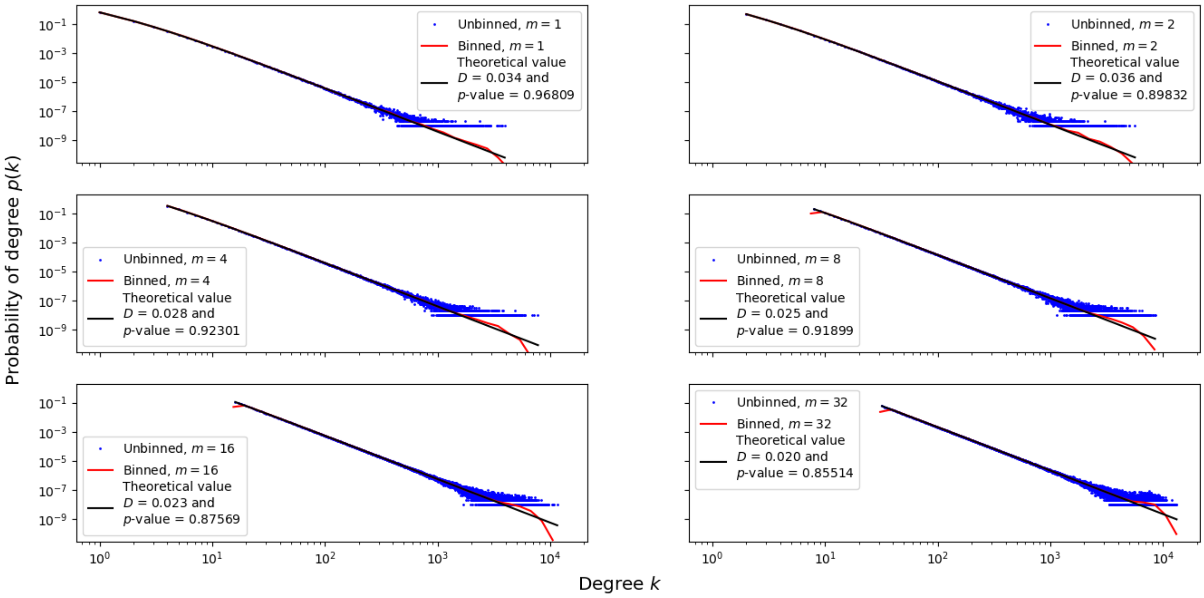

Figure 4: Degree distribution of BA model with $N = 10^6$ vertices for each m studied. Blue dots represent the unbinned values, red lines represent the binned distribution and black lines are the theoretical relationship as given in eq. (22).

number of vertices added N. Before that, the three distributions closely resemble each other.

In order to evaluate the goodness of fit of the binned distribution with the theoretical one, a Kolmogorov–Smirnov (KS) test was performed as it does not limit the type of probability distribution like the chi-squared, which only works for normally distributed data. The cumulative distribution function of the theoretical and binned distributions were compared, and a D-statistic is used to quantify the maximum distance between the two at any given k, so this should be minimised. The null hypothesis is that the two distributions are drawn from the same sample, and a p-value (ranging from 0 to 1) is used to determine whether it is accepted [5]. The p-value for each degree is plotted as the p-value from the start of the binned line up to that degree against degree k as shown in Figure 5.

The p-value is close to 1 when k is small, and up to the 'bump' and cutoff of the

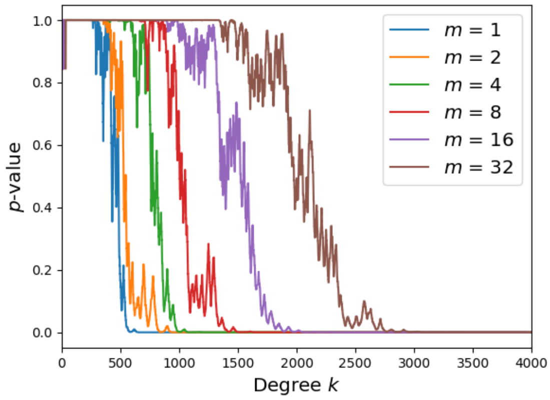

Figure 5: p-values of KS test performed over all sets of data from the BA model between the theoretical distribution and the binned distributions.

binned distribution, where it rapidly decreases to zero. This shows that the linear part of the binned distribution is of a same distribution as the theoretical one. Table 1 shows the maximum degree $k_{\max}$ where the two distributions are in good agreement. $k_{\max}$ is much smaller than the degree k of the 'bump', showing that the effects of fat-tail is significant even before the actual deviation starts. The D-statistics and the p-values of all m values at the maximum degree are shown in Figure 4. From these values, it can be determined that the null hypothesis cannot be rejected at the 15% significance level for all m.

m	1	2	4	8	16	32
$k_{\max}$	411	504	743	958	1316	1906

Table 1: Maximum degree $k_{\max}$ before which the p-value drops rapidly.

4.3.2 Investigation of largest expected degree k_1

The largest expected degree k_1 was taken as the average over the largest degrees of the 100 runs for $m = 16$. Since k_1 was expected to have a relationship of $k_1 \propto \sqrt{N}$ as $N \to \infty$ from eq. (33), a log-log plot of k_1 against N is plotted as in Figure 6, for $N = 10^2, 10^3, 10^4, 10^5, 10^6, 10^7$. A linear fit was then performed on the points except $N = 10^2$ since a small network size means large finite-size effects. The slope was found to be 0.502 ± 0.004, consistent with the theoretical prediction of $k_1 \propto \sqrt{N}$. However, there is a clear parallel offset between the theoretical (eq. (32)) and simulated values. This can be attributed to the finite-size effects due to N not close enough to infinity yet. When deriving the theoretical form from the master equation (eq. (3)), $N \to \infty$ is assumed as a condition which cannot be assumed here.

4.3.3 Data collapse of the degree distribution

To understand how the degree distribution is affected by finite N, log-binned distribution $p_{\text{data}}(k)$ is plotted for $N = 10^2, 10^3, 10^4, 10^5, 10^6$ in Figure 7, with scale factor $a = 1.25$

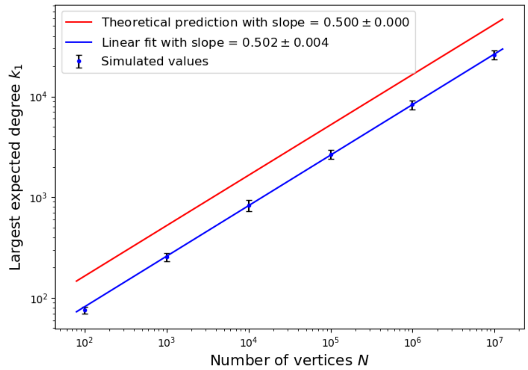

Figure 6: Largest expected degree k_1 against number of vertices N, ranging from 10^2 to 10^7. The theoretical prediction (red) had a slope of 0.5 as expected, and the linear fit (blue) was performed on the points except $N = 10^2$.

and $m = 16$.

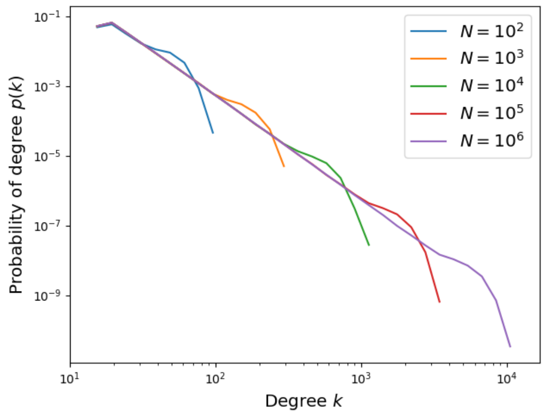

Figure 7: Log-binned degree distribution for finite-size networks with $N = 10^2, 10^3, 10^4, 10^5, 10^6$ and $m = 16$. The power-law decays are the same for small k but have different cutoff positions determined by k_1.

Since $p_{\text{data}}(k)$ is expected to be consistent with $p_{\text{theory}}(k)$, $p_{\text{data}}(k)/p_{\text{theory}}(k)$ should yield a value of 1. To align the values on the x-axis, the cutoff 'bump' at the tail for different N can be used, whose position is determined by k_1. To align the cutoff values, k_{data}/k_1 can be used. This is done as shown in Figure 8, although for smaller system size $N = 10^2$, it can be seen that there are deviations from the trend from the other N. This means that the behaviour of the degree distribution of BA networks is scale free [6].

For such a system, a finite-size scaling ansatz can be proposed as

$$p(k) = k^{-\tau} \mathcal{F}\left(\frac{k}{k_1}\right), \tag{34}$$

such that scale-free variables, such as degree distribution, are identical after a transforma-

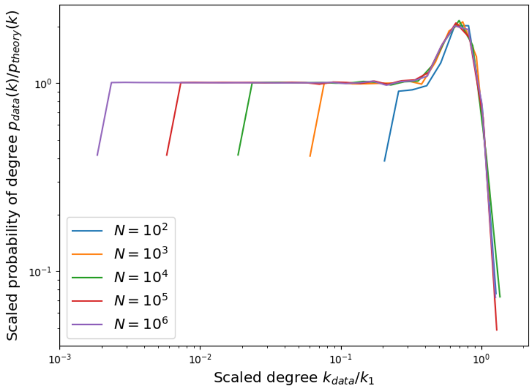

Figure 8: Data collapse of scaled probability of degree $p_{\text{data}}(k)/p_{\text{theory}}(k)$ against scaled degree k_{data}/k_1.

tion by the scaling function. The parameter τ was found to be $\tau \approx 3$, which is consistent with $p(k) \propto k^{-3}$ relationship for large k.

The deviation from the infinite time large degree limit can be seen as the 'bump' and the cutoff at the tail. This is due to the limit of degree as a result of finite size, and so vertices at the tail cannot experience an increase in degree, and more vertices accumulate at the tail as time goes on, explaining the 'bump'.

5 Pure Random Attachment

5.1 Theoretical degree distribution $p_\infty(k)$

Requiring $\Pi_{\text{rnd}} \propto 1$ with the normalisation condition that $\sum_{n=0}^{N} \Pi_{\text{rnd}} = 1$, the probability

$$\Pi_{\text{rnd}} = \frac{1}{N}. \tag{35}$$

Using the master equation (eq. (3)),

$$p_\infty(k) = m[p_\infty(k-1) - p_\infty(k)] + \delta_{k,m} \tag{36}$$

$$p_\infty(k) = \frac{m}{m+1} p_\infty(k-1) + \delta_{k,m}. \tag{37}$$

For $k > m$, $\delta_{k,m} = 0$ so

$$p_\infty(k) = \frac{m}{m+1} p_\infty(k-1). \tag{38}$$

By induction,

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

Again applying the fact that every vertex has degree $\geq m$ as $N \to \infty$, $p_\infty(k < m) = 0$.

Considering $k = m$, as $p_\infty(m-1) = 0$,

$$p_\infty(m) = m[0 - p_\infty(k)] + 1 \tag{40}$$
$$p_\infty(m) = \frac{1}{m+1} \tag{41}$$

Hence the general form is

$$p_\infty(k) = \frac{1}{m+1}\left(\frac{m}{m+1}\right)^{k-m} \qquad \text{for} \quad k \geq m. \tag{42}$$

Checking for normalisation,

$$\sum_{k=m}^{\infty} p_\infty(k) = \frac{1}{m+1}\sum_{k=m}^{\infty}\left(\frac{m}{m+1}\right)^{k-m} \tag{43}$$
$$= \frac{1}{m+1}\sum_{k'=0}^{\infty}\left(\frac{m}{m+1}\right)^{k'} \qquad \text{rewriting} \quad k' = k - m \tag{44}$$
$$= \frac{1}{(m+1)(1-\frac{m}{m+1})} = 1 \qquad \text{using} \quad \sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \tag{45}$$

as required.

5.2 Theoretical largest degree k_1

From eq. (42),

$$\frac{1}{m+1}\sum_{k=k_1}^{\infty}\left(\frac{m}{m+1}\right)^{k-m} \sim \frac{1}{N} \tag{46}$$
$$\left(\frac{m}{m+1}\right)^{k_1-m}\left[\frac{1}{m+1}\sum_{k=k_1}^{\infty}\left(\frac{m}{m+1}\right)^{k-k_1}\right] \sim \frac{1}{N}. \tag{47}$$

Using eq. (44), $\frac{1}{m+1}\sum_{k=k_1}^{\infty}\left(\frac{m}{m+1}\right)^{k-k_1} = 1$ and hence

$$\left(\frac{m}{m+1}\right)^{k_1-m} \sim \frac{1}{N} \tag{48}$$
$$(k_1 - m)\ln\left(\frac{m}{m+1}\right) \sim -\ln N \tag{49}$$

Rearranging gives

$$k_1 \sim m - \frac{\ln N}{\ln m - \ln(m+1)} \tag{50}$$
$$k_1 \sim -\frac{\ln N}{\ln m - \ln(m+1)} \qquad \text{for} \quad k_1 \gg m \quad (\text{true for } N \gg 1). \tag{51}$$

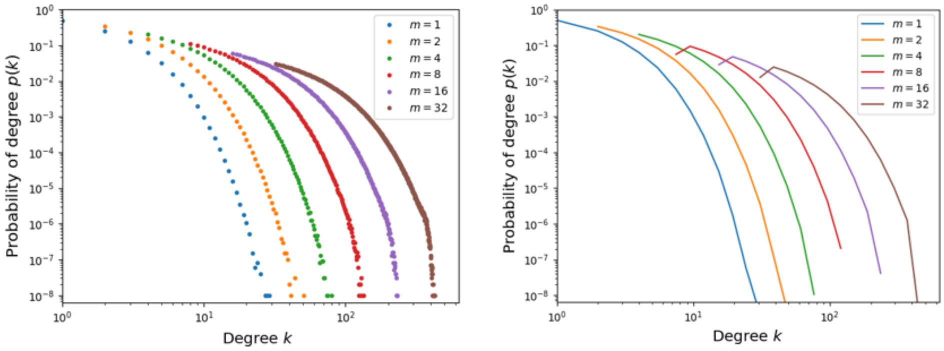

Figure 9: Degree distribution of the pure random attachment model, unbinned (left) and log-binned (right).

5.3 Numerical Results and Discussion

5.3.1 Comparison to theoretical distribution

Similar to Section 4.3, the degree distribution before and after log-binning are shown in Figure 9. The shapes of these curve indicate an exponential decay and the fat-tail is not clearly seen as in the BA model.

Visualisation was used again to verify whether the binned distributions fit with the theoretical ones. Figure 10 shows the comparison between theoretical and experimental distributions. It can be seen that the three distributions are consistent with each other except at the end when k becomes large.

A KS test was again employed to show that the distributions match numerically. The p-value at each degree was plotted against degree as shown in Figure 11. As opposed to the BA model, the p-value do not vary throughout the whole curve, only to rapidly decay at the end. This indicates that the fat-tail is insignificant and also the theoretical and experimental distributions are consistent to a 5% significance level. The maximum degree k_{max} before which p-value drops rapidly for each m is tabulated in Table 2, and the corresponding D-statistic and p-value are shown in Figure 11.

m	1	2	4	8	16	32
k_{max}	25	47	77	129	222	404

Table 2: Maximum degree k_{max} before which the p-value drops rapidly.

5.3.2 Investigation of the largest expected degree k_1

Repeating each simulation for 100 times again for $N = 10^2, 10^3, 10^4, 10^5, 10^6$, and averaging the largest degree in each run gave the largest expected degree k_1. This was plotted against number of vertices N in Figure 12. Again, there is an offset from the theoretical distribution due to the finite size effect, and it is more significant with smaller N.

As the probability decays as exponentially, it was not possible to find a way to produce a data collapse on a log-log plot since it requires a power-law to be done.

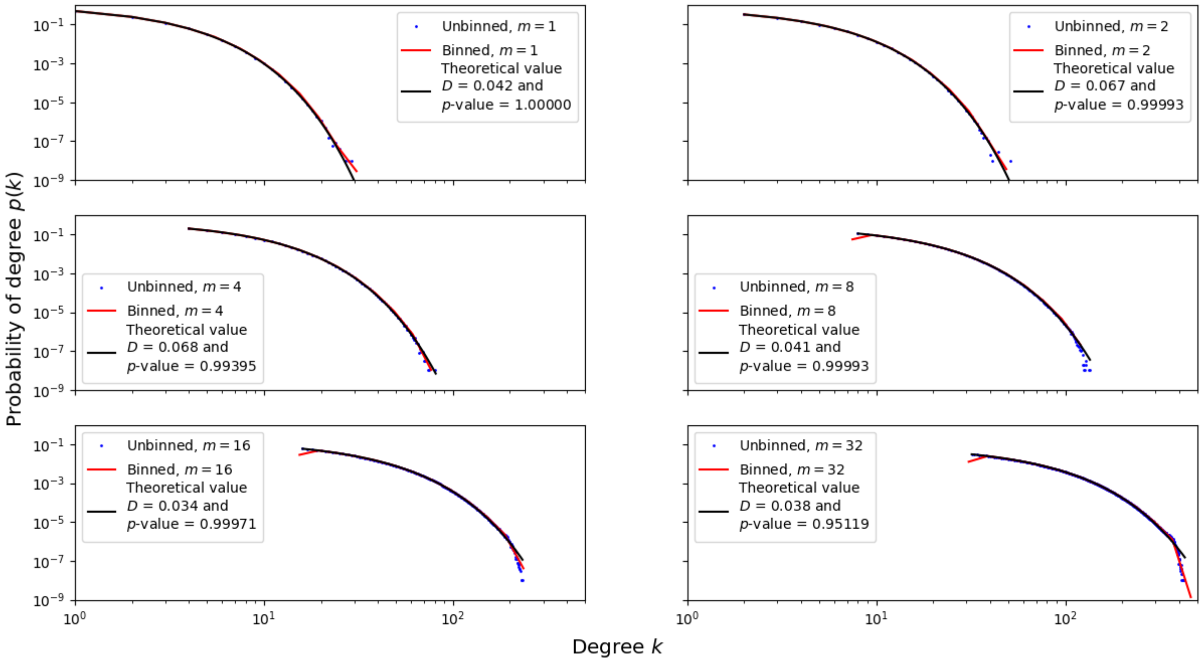

Figure 10: Degree distribution of the pure random attachment model with $N = 10^6$ vertices for each m studied. Blue dots represent the unbinned values, red lines represent the binned distribution and black lines are the theoretical relationship as given in eq. (42).

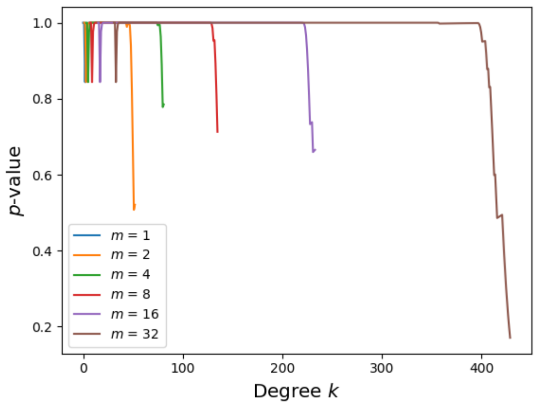

Figure 11: p-values of KS test performed over all sets of data from the random attachment model between the theoretical distribution and the binned distributions.

6 Mixed Preferential and Random Attachment

6.1 Theoretical degree distribution $p_\infty(k)$

The probability of attachment is given by

$$\begin{aligned}\Pi_{\text{mix}}(k,t) &= q\Pi_{\text{pa}}(k,t) + (1-q)\Pi_{\text{rnd}}(t) &(52)\\ &= q\left(\frac{k}{2E(t)}\right) + (1-q)\left(\frac{1}{N(t)}\right), &(53)\end{aligned}$$

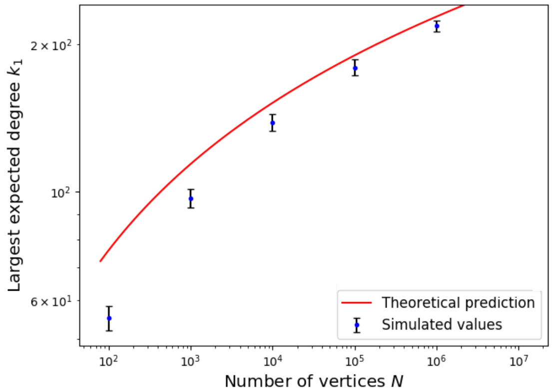

Figure 12: Largest expected degree k_1 against number of vertices N, ranging from 10^2 to 10^6.

where q is the probability of selecting preferential attachment. Since individually $\Pi_{\text{pa}}(k,t)$ and $\Pi_{\text{rnd}}(t)$ are normalised, and $q+(1-q)=1$, so $\Pi_{\text{mix}}(k,t)$ is normalised.

Using the master equation (eq. (3)),

$$p_\infty(k) = mN(t)\left[\left(\frac{q(k-1)}{2mN(t)}+\frac{1-q}{N(t)}\right)p_\infty(k-1)-\left(\frac{qk}{2mN(t)}+\frac{(1-q)}{N(t)}\right)\right]p_\infty(k)+\delta_{k,m} \tag{54}$$

$$= \left[\frac{q(k-1)}{2}+m(1-q)\right]p_\infty(k-1)-\left[\frac{qk}{2}+m(1-q)\right]p_\infty(k)+\delta_{k,m}. \tag{55}$$

For $k=m$, assuming again $p_\infty(k<m)=0$ as $N\to\infty$,

$$p_\infty(m) = -\left[\frac{qk}{2}+m(1-q)\right]p_\infty(k)+1 \tag{56}$$

$$p_\infty(m) = \frac{2}{2m-mq+2}. \tag{57}$$

For $k>m$, rearranging yields

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k+\frac{2m}{q}-2m-1}{k+\frac{2m}{q}-2m+\frac{2}{q}}. \tag{58}$$

Using the gamma function trial solution in eq. (9) with $a=2m/q-2m-1$ and $b=2m/q-2m+2/q$,

$$p_\infty(k) = A\frac{\Gamma(k+\frac{2m}{q}-2m)}{\Gamma(k+\frac{2m}{q}-2m+\frac{2}{q}+1)}. \tag{59}$$

Choosing $q=0.5$ for this analysis and considering

$$p_\infty(m+1) = A\frac{\Gamma(3m+1)}{\Gamma(3m+6)} \tag{60}$$

$$= \frac{A}{(3m+5)(3m+4)(3m+3)(3m+2)(3m+1)}, \tag{61}$$

comparing with eq. (58) with $k = m + 1$,

$$p_\infty(m+1) = \frac{3m}{3m+5} p_\infty(m) = \frac{12m}{(3m+5)(3m+4)}, \tag{62}$$

and equating these two expressions yields

$$A = 12(3m+3)(3m+2)(3m+1)m. \tag{63}$$

So the general form for $q = 0.5$ is given by

$$p_\infty(k) = \frac{12(3m+3)(3m+2)(3m+1)m}{(k+2m+4)(k+2m+3)(k+2m+2)(k+2m+1)(k+2m)} \tag{64}$$

for $k \geq m$, and for $k \gg 1$, $p_\infty(k) \propto k^{-5}$.

6.2 Numerical Results and Discussion

The degree distribution before and after log-binning are shown in Figure 13. As expected, the shape of this curve follows an exponential decay at small k and a power-law decay at large k. This is mainly due to preferential attachment leading to larger degrees than pure random, so dominant over large k.

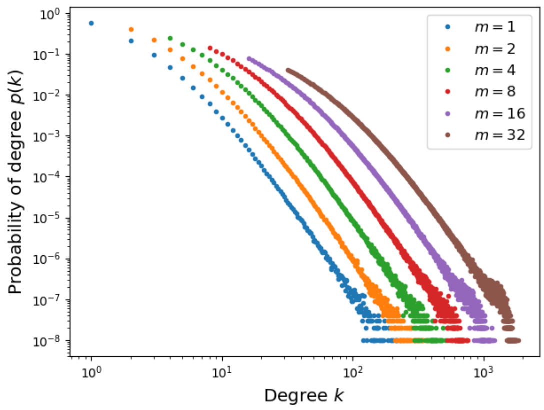

Figure 13: Degree distribution of the mixed attachment model. The behaviour at small k follows the random attachment model, while for large k it follows the preferential attachment.

Once again, visualisation and KS test were used to compare between theoretical and simulated distribution. Figure 14 shows the theoretical and simulated values for each m, and Figure 15 shows the p-values at each degree k. The binned distribution was found to be consistent with the theoretical distribution up to the fat-tail and cutoff, which is less pronounced than the BA model. Larger m allowed a more clear cutoff to be seen. Also, the decay is not entirely a power-law decay, and exponential behaviour dominates at small k. The fluctuations in p-values are more than the random model but less than the BA, as expected from the nature of the mixed model. As before, since p-value is constant at

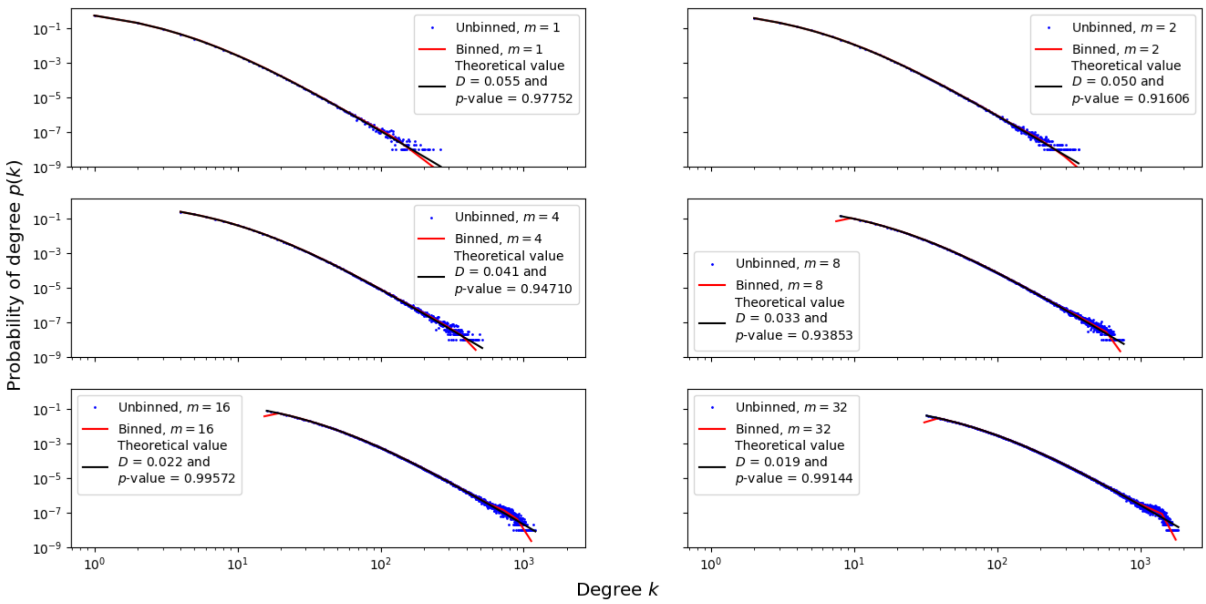

Figure 14: Degree distribution of mixed model with $N = 10^6$ vertices for each m studied. Blue dots represent the unbinned values, red lines represent the binned distribution and black lines are the theoretical relationship as given in eq. (64).

around 1 before it rapidly decays, the binned distributions and the theoretical distributions are consistent. The maximum degree k_{max} before which p-value drops rapidly for each m is as tabulated in Table 3, and the corresponding D-statistic and p-value are shown in Figure 14.

m	1	2	4	8	16	32
k_{max}	147	240	319	524	687	1099

Table 3: Maximum degree k_{max} before which the p-value drops rapidly.

7 Conclusion

Three growing network models were studied, namely the (a) Barabási-Albert (BA) Pure Preferential, (b) Pure Random and (c) Mixed Prefrential and Random models. The degree distribution $p(k)$ was derived to be (a) a power-law decay, (b) an exponential decay, and (c) a power-law decay, assuming an infinitely large system. As this is not the case, the models were limited by the finite size and a fat-tail and cutoff effect can be seen in the BA and mixed models. A scale-free behaviour was observed with the BA model where a data collapse had been successfully implemented.

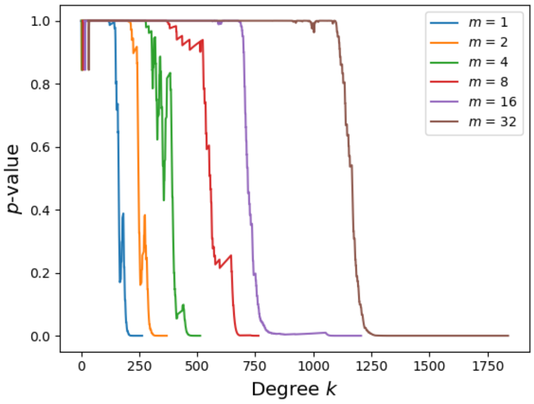

Figure 15: p-values of KS test performed over all sets of data from the mixed attachment model between the theoretical distribution and the binned distributions.

References

[1] Barabási AL, Albert R. *Emergence of Scaling in Random Networks Science.* The Structure and Dynamics of Networks. 1999;286(173).

[2] Price DJDS. *The Scientific Foundations of Science Policy.* Nature. 1965;206(4981):233–238.

[3] Evans T. *Network Project Notes.* Imperial College London; 2018.

[4] Evans T. *Price/BA Networks Model.* Imperial College London; 2018.

[5] *How to interpret p-value of Kolmogorov-Smirnov test (python)?*; 2015. Available from: `https://stats.stackexchange.com/questions/57885/how-to-interpret-p-value-of-kolmogorov-smirnov-test-python`.

[6] Christensen K, Moloney NR. *Complexity and Criticality.* Imperial College Press; 2005.