Network Project
A Growing Network Model

CID: 01068099

25th March 2019

Abstract: Models of Pure Preferential Attachment, Pure random Attachment and Existing Vertices networks were implemented computationally and compared to the derived theoretical prediction of degree probability distributions. Agreement was achieved for the first two models but not for the last one. For the first two models, theoretical and computational investigations of the largest degree node were carried out and where found to be in agreement. In general it was shown that the theoretical solution for infinite systems models finite system well excluding the large k region where finite size effects become apparent.

Word Count: 2497 words excluding font page, figure captions, table captions, and bibliography.

1 Introduction

It is common in network science to look for the analytical expression that gives rise to the degree distribution of a particular type of network. This is instrumental in shining light to the mechanism behind the network's formation. This was done by Barabási and Albert[1] in 1999 when they introduced the preferential attachment mechanism to explain power law behaviour. In this project, we computationally implement the Barabási and Albert (BA) model and two others: Pure Random Attachment and Existing Vertices. A theoretical prediction for all three models was obtained analytically and was compared to the computational data graphically and statistically. Additionally, there were investigations of the largest node in the finite systems both numerically and analytically and agreement was generally found. Lastly, a data collapse was applied to the BA networks, highlighting scale-free behavior but also finite size effects.

2 Phase 1: Pure Preferential Attachment Π_{pa}

2.1 Implementation

2.1.1 Numerical Implementation

The models where generally built using the following algorithm:

1. Set up initial network at time t_0, a graph G_0.

2. Increment time $t \to t + 1$.

3. Add one new vertex.

4. Add m edges as follows:

- Connect one end of the new edge to the new vertex.
- Connect the other end of each new edge to an existing vertex chosen with probability Π. This will was specified in different ways for different models.

5. Repeat from 2 until reach final number of N vertices in the network.

The above can be implemented in several ways. Here, the network was implemented as a pair of lists, one holding all existing edges (as pairs of nodes) and the other holding the degree of each node. An example is shown in Fig.1. In truth, all the information about the network is contained in the first list but the second is kept for convenience. Lastly, in this pure preferential attachment model the probability Π of connecting to a certain node is proportional to the existing degree of this node and hence:

$$\Pi_{\text{pa}} = \frac{k}{2E} \tag{1}$$

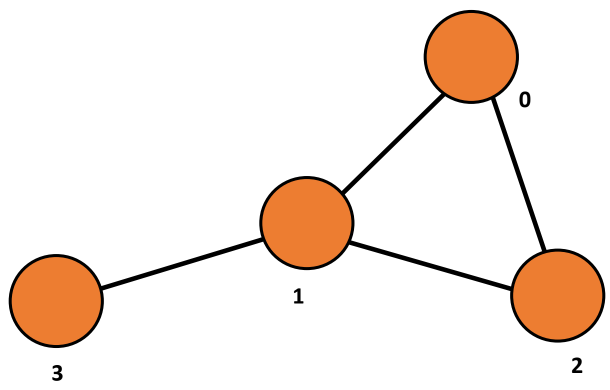


Figure 1: Example of network implementation. The above network would be represented as a list of edges = $[[0, 2], [0, 1][1, 2][1, 3]]$ and the degree distribution $[2, 3, 2, 1]$.

2.1.2 Initial Graph

We use the smallest possible network, i.e. two nodes connected with a single edge. This is done to ensure that the largest possible fraction of the network is built using the desired project. This didn't cause any problems even if m was larger than two as multi-edges where allowed (see below). An alternative approach could have been to produce fully symmetric initial graphs, for example fully connected graphs of size m. However this would bias a large number of existing nodes (think of the case where $m = 50$ and $N = 100$ or 1000).

2.1.3 Type of Graph

The type of graph produced here is undirected with multi-edges allowed. This is the type of graph that most closely resembles the mathematical definition of the network we are building. In the long time limit the probability of connecting to any particular node (and hence to any particular node more than once) tends to zero. Thus, the expected number of multi- edges is not significant enough to cause an observable effect.

2.1.4 Working Code

The first and most important check was done by manual observation. The network was visualised and the saved lists where printed, checked to agree with each other and meet expectations by hand. Secondly, the limiting case of $m = 0$ yielded the expected behaviour (adding nodes without edges) and lastly, the produced probability distribution was checked to tend to zero for large k.

2.1.5 Parameters

The parameters needed by the programme are the initial network, the choice of m and the time t at which it should stop growing the network. The initial network was not changed through the project but m and t where changed to meet the needs of each section. Lastly, since the size of the initial network was always two we have the relationship $N = t + 2$.

2.2 Preferential Attachment Degree Distribution Theory

2.2.1 Theoretical Derivation

To derive a theoretical form for the degree distribution $p_\infty(k)$ we need to begin by describing the evolution of the number of nodes with degree k at each time step. This is the master equation:

$$n(k,t+1) = n(k,t) + m\Pi(k-1,t)n(k-1,t) - m\Pi(k,t)n(k,t) + \delta_{k,m}, \tag{2}$$

where $n(k,t)$ is the number of nodes that have degree k at time t. We can now use that the probability of n nodes having degree k is $p(k,t) = \frac{n(k,t)}{N(t)}$ where $N(t)$ is the total number of nodes at time t. Taking the long time limit $(t \to \infty)$ we see that if we denote the initial number of nodes and edges N_0 and E_0 respectively:

$$\lim_{t\to\infty} = \frac{E}{N} = \frac{E_0 + mt}{N_0 + t} = m, \tag{3}$$

and therefore $E = mN$. By also realising that $N(t+1) = N(t) + 1$ we can take this long time limit and the expression for the probability to derive:

$$p_\infty(k) = \frac{1}{2}\left[(k-1)p_\infty(k-1) - kp_\infty(k) + \delta_{k,m}\right]. \tag{4}$$

To solve Eq.(4) we will need to prove the following result. That the equation

$$\frac{f(z)}{f(z-1)} = \frac{z+a}{z+b}, \tag{5}$$

has solution

$$f(z) = A\frac{\Gamma(z+1+a)}{\Gamma(z+1+b)}, \tag{6}$$

were $\Gamma(z)$ is the gamma function with key property$\Gamma(z+1) = z\Gamma(z)$. To prove we start from Eq.(6) so that

$$\frac{f(z)}{f(z-1)} = \frac{\Gamma(z+1+a)}{\Gamma(z+1+b)} / \frac{\Gamma(z+a)}{\Gamma(z+b)} \tag{7}$$

Now, applying the aforementioned property we show that this equals

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

Having shown this we take the case $(k > m)$ and rearrange Eq.(4) as

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

recognising $a = -1$ and $b = 2$ from Eq.(6). We now write:

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

finally leading to

$$p_\infty(k) = \frac{A}{k(k+1)(k+2)}. \tag{11}$$

Lastly, we can evaluate A by checking for boundary conditions. Namely, we set $k = m$. Eq.(4) becomes:

$$p_\infty(m) = \frac{1}{2}\left[(m-1)p_\infty(m-1) - mp_\infty(m)\right] + \delta_{m,m} \tag{12}$$

$$= \frac{1}{2}\left[-mp_\infty(m)\right] + 1, \tag{13}$$

which gives $p_\infty = \frac{2}{2+m}$. This is compared to $p_\infty(k) = \frac{A}{m(m+1)(m+2)}$ to give $A = 2m(m+1)$ and finally:

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

2.2.2 Theoretical Checks

The properties of the theoretical solution must be:

1. $\lim_{t\to\infty} p_\infty(k) = 0$
2. $\sum_{k=m}^{\infty} p_\infty(k) = 1$ (normalisation)
3. $p_\infty(k < m) = 0$

The first property is trivial to show as

$$\lim_{k\to\infty} p_\infty(k) = \frac{2m(m+2)}{(\infty)(\infty+1)(\infty+2)} = 0, \tag{15}$$

but the second is more tedious. First, we need to break $p_\infty(k)$ into partial fractions and separate the summation accordingly:

$$\sum_{k=m}^{\infty} p_\infty(k) = m(m+1)\left[\sum_{k=m}^{\infty}\frac{1}{k} - 2\sum_{k=m}^{\infty}\frac{1}{k+1} + \sum_{k=m}^{\infty}\frac{1}{k+2}\right]. \tag{16}$$

Now, we need to carefully change the exponents so that each summation has the same argument

$$m(m+1)\left[\sum_{k=m}^{\infty}\frac{1}{k} - 2\sum_{k=m+1}^{\infty}\frac{1}{k} + \sum_{k=m+2}^{\infty}\frac{1}{k}\right], \tag{17}$$

and equalise the indices by taking the extra terms outside the summations:

$$m(m+1)\left[\frac{1}{m} + \frac{1}{m+1}\sum_{k=m+2}^{\infty}\frac{1}{k} - \frac{2}{m+1} - 2\sum_{k=m+2}^{\infty}\frac{1}{k} + \sum_{k=m+2}^{\infty}\frac{1}{k}\right]. \tag{18}$$

Finally, the summations cancel out leaving only

$$\sum_{k=m}^{\infty} p_\infty(k) = m(m+1)\left[\frac{1}{m} - \frac{1}{m+1}\right] = m(m+1)\frac{1}{m(m+1)} = 1. \tag{19}$$

The last property is true by construction.

2.3 Preferential Attachment Degree Distribution Numerics

2.3.1 Fat-Tail

The fat-tailed distribution causes a lot of noise towards large k. This is because the probability of such observations is much smaller, meaning there are not enough observations to represent the probability distribution well. This was addressed through log-binning.Other methods include producing a cumulative distribution function and Zipf plots.

2.3.2 Numerical Results

Networks of size $N = 10^5$ were initialised several times in order to obtain good statistics. The choices of m where 2, 5, 20 and 50. Because of the computational cost of raising m the first two cases were run 100 times but the later two where run only 10. N was chosen so that there can be enough network realisations within acceptable computation time. The initial plot is shown in Fig.2. This is the most straight forward plot that can

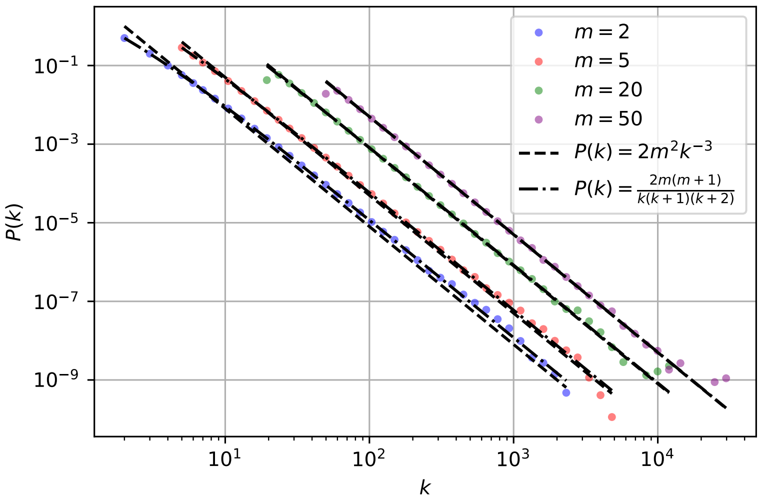
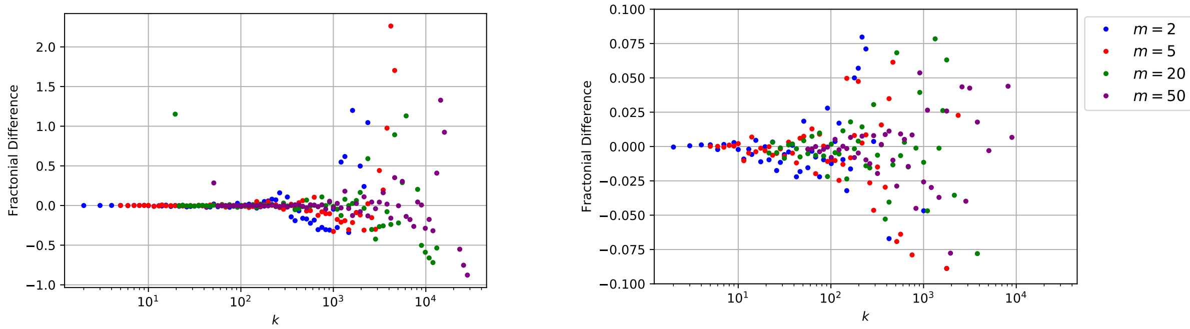


Figure 2: Numerical data for different values of m along the corresponding theoretical predictions. It is evident that the prediction fits the data but it is difficult to see exactly how. We see that the fit is worst towards large k which is a sign of finite size effects. Note that an alternative suggestion for the theoretical distribution is shown for reference. This was suggested in the original publication of the BA model[1].

be produced for these results but it is not the most instructive. To see what is really happening we plot the fractional difference of theory and numerics against k in Fig.3. Namely, we are plotting

$$\frac{p_\infty(k) - p_{\text{data}}}{p_{\text{data}}}. \tag{20}$$

Examining this figures we see that - ignoring finite size effects - the data points are symmetrically distributed about the $y = 0$ line which suggests that the discrepancies are random and not systematic. Additionally - again, ignoring finite size effects - the fractional discrepancy is capped at approximately 7.5% while more of the points are located within the 2.5% margin. In summary, the figures indicate a good fit. As an additional point, we see that in Fig.3a the points in the large k region are not randomly allocated but rather follow a specific shape which is characterising of the excess of observations that builds



(a) We notice that the fractional difference is minimum at low k but grows rapidly as k grows larger. This is a manifestation of the finite size of the network but also of the fact that probability itself is becoming smaller.

(b) Magnified plot to show observations in the 0.1 range. We see that the points in the range are symmetrically scattered about $y = 0$ which indicates random error but a good fit.

Figure 3: Fractional difference between data and theoretical distribution as defined in Eq.(20) against degree k.

up in a finite system. These are the nodes that would have had larger degree had the network been larger. This is particularly visible in the $m = 2$ case.

2.3.3 Statistics

The Pearson χ^2 test is specifically designed for binned data but it is known to not work well for observations spanning many orders of magnitude. The Kolmogorov-Smirnov (KS) test, on the other hand, is designed for comparison of probability distributions and it is not known to have any trouble across a long observation range. Additionally, the KS test is the test of choice in [2] where many different real networks are checked for scale-free (power law) behaviour. Both tests are carried out for completeness (and as means of verification). It was noticed that the test results are very sensitive to the *scale* variable which controls the growth of the bins when log binning is applied to the data. Hence, the results for a scale of 1 (regular binning), 1.01 and 1.1 are shown in Tables 1, 2, and 3.

m	KS Test p-value	χ^2 Test p-value
2	7.824343556359249e-34	0.9998881166990833
5	6.202083435706684e-180	0.20166699307658584
20	0.0	8.153184526043496e-16
50	0.0	2.4701240078754954e-97

Table 1: Statistical tests between theoretical prediction and computational implementation for the BA model. Scale = 1

Before inferring any results it is important to understand what we are measuring. Both tests employed here test the null hypothesis (H_0) that the two sets we compare (data and theory) come from the same distribution. The $p-$value we measure gives us the probability that we will be making an error if we reject H_0. Hence, the smaller the p-value the more certain we are that theory and data **do not** agree. Yet, interpreting large p-value is more ambivalent. In general, the closer the value is to one the larger the probability we will be making a mistake if we discard H_0 but whether we can take this to

m	KS Test p-value	χ^2 Test p-value
2	0.8627699719741215	0.9999894567138532
5	0.00014135515017276622	0.20220610613986847
20	1.2320763850752001e-21	2.6924297147372662e-15
50	8.815925441021169e-38	5.723333469430018e-85

Table 2: Statistical tests between theoretical prediction and computational implementation for the BA model. Scale = 1.01

m	KS Test p-value	χ^2 Test p-value
2	0.9999999999999993	0.9999988068822425
5	0.716074655886033	0.2384764090476358
20	0.021710454943598634	1.4982297311731315e-05
50	0.0003405299671514096	1.109035533027885e-12

Table 3: Statistical tests between theoretical prediction and computational implementation for the BA model. Scale = 1.1

mean that the p-value **is** is the probability of the two distributions being the same is not clear.

Nevertheless, there are two important trends to be noticed: Firstly, the $p-$value is getting worst as m increases and secondly, it gets better as scale parameter increases. Hence we will take Table 3 as the main result. Based on the above discussion we can say that the $m = 2$ data is certainly fitting the data, $m = 5$ has probability above 70% which is satisfactory but the remaining two cases do not make the cut. Acknowledging the the discrepancy between theory and data comes from the finite size of the networks at hand, we suggest this finiteness is not simply quantified by the sheer value of N but is also dependent on the N/m ratio.

2.4 Preferential Attachment Largest Degree and Data Collapse

2.4.1 Largest Degree Theory

We expect the largest degree (k_1) to be such so that the expected number of nodes with $k \geq k_1$ is equal to one. Mathematically:

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

We find k_1 by considering the derivation in equations (16)-(19) but replacing m in the summation index with k_1. This gives us the result

$$N \frac{m(m+1)}{k_1(k_1+1)} = 1, \tag{22}$$

and therefore

$$k_1 \approx \sqrt{m(m+1)N}, \tag{23}$$

by approximating $k + 1 \approx k$.

2.4.2 Numerical Results for Largest Degree

Following the discussion in 2.3.3, the value of m chosen for this investigation is 5. This value is large enough to give rise to finite size effects yet not too large as to constitute the computation time too long. This biggest N range that could be obtained with the available means of computation was from 10^2 to 10^7. We plot the results (Fig.4) and produce a linear fit by fitting a straight line to the natural logarithms of N and k_1. The fit coefficients are the slope, $\alpha = 0.490 \pm 0.015$ and the intercept, $\beta = 1.84 \pm 0.16$. According to the theoretical prediction Eq.(23) we expect $\alpha = 0.5$ and $e^\beta = \sqrt{m(m+1)} = 5.48$. We propagate β and its error to get $e^\beta = 6.30 \pm 1.01$ and hence verify that both coefficients point at agreement.

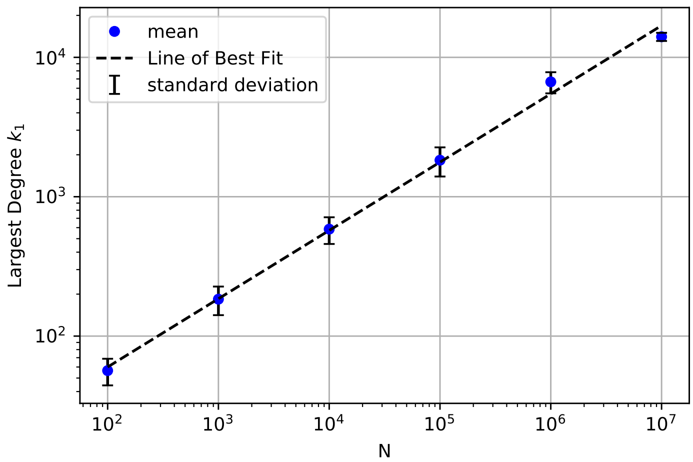


Figure 4: Plot for largest degree k_1 against N for the BA model. A straight line is a good fit as it is included by all errorbars but one.

2.4.3 Data Collapse

It is possible to use the scaling of the largest degree k_1 and the theoretical probability distribution to suggest the following scaling ansatz for the behavior of the BA model:

$$p_{data}(k; m) = p_\infty \mathcal{G}(k/k_1), \tag{24}$$

where $\mathcal{G}(x)$ is the scaling function:

$$\mathcal{G}(x) = \begin{cases} 1 & \text{for } x \leq 1 \\ 0 & \text{for } x \geq 1 \end{cases} \tag{25}$$

Following the above we can produce a data collapse by plotting p_{data}/p_∞ on the $y-$axis and $k/\sqrt{m(m+1)N}$ on the $x-$axis. This is shown in Fig.5.

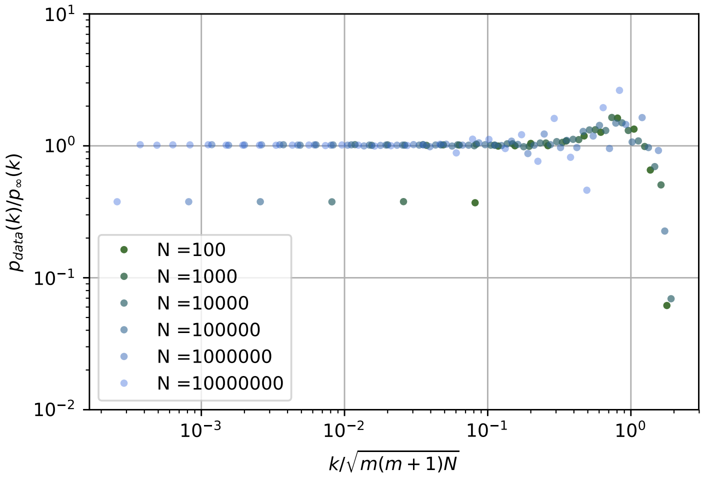


Figure 5: Data collapse of degree distribution of BA Model of Preferential Attachment for different network sizes N and parameter $m = 5$. We see that the plot equals 1 until finite size effects start to show close to $x = 1$. The characteristic bump shows an excess of measured nodes with a certain degree which are not predicted by the theoretical distribution. These are the nodes that would have had larger degree had the network been larger but instead pile up at lower k.

2.5 Phase 2: Pure Random Attachment Π_{rnd}

2.6 Random Attachment Theoretical Derivations

2.6.1 Degree Distribution Theory

To derive the theoretical form of the degree distribution we start from the master equation

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

but now replace Π_{pa} with $\Pi_{\text{rnd}} = \frac{1}{N}$ which is the normalised solution for adding edges with uniform probability. Again, replacing $n(k,t)$ with $p_\infty(k)N$ and recalling $N(t+1) = N(t) + 1$ we arrive at:

$$p_\infty(k) = m[p_\infty(k-1) - p_\infty(k)] + \delta_{m,k}. \tag{26}$$

For $k > m$ this is rearranged to

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

which we recognise as a geometric series. As before, in the long-time limit $p_\infty(k < m) = 0$. We can therefore start with $p_\infty(m)$ and write:

$$p_\infty(m+x) = \left(\frac{m}{m+1}\right)^x p_\infty(m). \tag{28}$$

For $k = m$

$$p_\infty(m) = -mp_\infty(m) + 1, \tag{29}$$

giving $p_\infty(k) = \frac{1}{m+1}$. Finally the complete probability distribution is

$$p_\infty(k) = \frac{1}{m+1}\left(\frac{m}{m+1}\right)^{k-m}, \tag{30}$$

by substituting $k = m + x$.

Checking for the correct theoretical properties we see that

$$\lim_{k\to\infty} p_\infty(k) = \frac{1}{m+1}\left(\frac{m}{m+1}\right)^{\infty} = 0, \tag{31}$$

as $m/(m+1)$ is always smaller than one. To check whether it is normalised we compute the sum:

$$\sum_{k=m}^{\infty} p_\infty(k) = \frac{1}{m+1}\sum_{k=m}^{\infty}\left(\frac{m}{m+1}\right)^{k-m}, \tag{32}$$

which we recognise as

$$\frac{1}{m+1}\sum_{0}^{\infty} r, \tag{33}$$

with $r = \frac{m}{m+1}$. Taking the standard result

$$\sum_{0}^{\infty} r = \frac{1}{1-r} = \frac{1}{1-\frac{m}{m+1}} = m+1. \tag{34}$$

We see that the probability distribution is normalised as

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

Thus, the derived probability distribution has the correct properties. Namely, it tends to zero at large k, it is normalised and it equals zero for $k < m$ (by construction).

2.6.2 Largest Degree Theory

Similar to 2.4.1 the condition for the largest degree k_1 is

$$N\sum_{k=k_1}^{\infty} p_\infty(k) = 1. \tag{36}$$

By substituting the expression for $p_\infty(k)$ we have

$$\frac{N}{m+1}\sum_{k=k_1}^{\infty}\left(\frac{m}{m+1}\right)^{k-m} = \frac{N}{m+1}\left(\frac{m}{m+1}\right)^{-m}\sum_{k=k_1}^{\infty}\left(\frac{m}{m+1}\right)^{k}. \tag{37}$$

Here, we make the substitution $x = k - k_1$ in order to obtain

$$\frac{N}{m+1}\left(\frac{m}{m+1}\right)^{k_1-m}\sum_{x=0}^{\infty}\left(\frac{m}{m+1}\right)^{x} = N\left(\frac{m}{m+1}\right)^{k_1-m} = 1, \tag{38}$$

where we have made use of the standard result

$$\sum_{r=0}^{\infty} = \frac{1}{1-r} \tag{39}$$

with $r = m/(m+1)$. By taking the natural logarithm of this result we rearrange for

$$k_1 = \frac{\ln N}{\ln\left[(m+1)/m\right]} + m. \tag{40}$$

2.7 Random Attachment Numerical Results

2.7.1 Degree Distribution Numerical Results

As before we first plot the produced probability distribution in comparison to theory (Fig.6) and then proceed with the more instructive fractional difference plots (Fig.7). We see that apart from the large k region the model seems to be a good fit for the data as the fractional difference is equally distributed about zero pointing at noise instead of any systematic flaw in the theoretical prediction. In contrast, as k grows larger the observations are skewed towards positive y which indicate a systematic variation from theory; evidence of finite size effects.

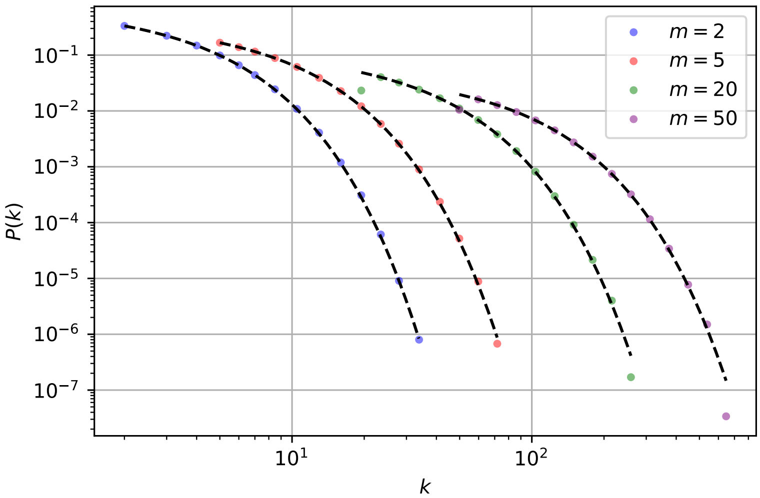
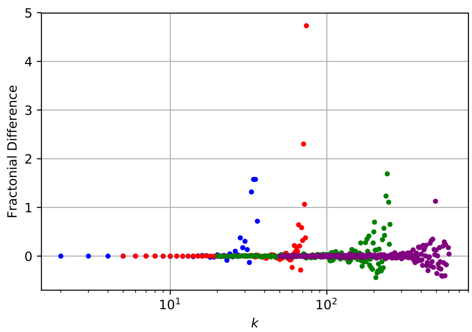
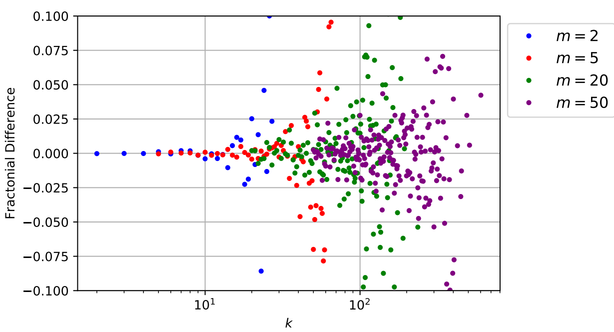


Figure 6: Numerical data for different values of m along the corresponding theoretical prediction (black lines). We see that the colour markers fit the lines well.



(a) We notice that the fractional difference is minimum at low k but grows as k approaches k_1. This is a manifestation of the finite size of the network but also of the fact that probability itself is becoming smaller.



(b) Magnified plot to show observations in the 0.1 range. We see that the points in the range are symmetrically scattered about $y = 0$ which indicates random error but a good fit.

Figure 7: Fractional difference between data and theoretical distribution as defined in Eq.(20) against degree k.

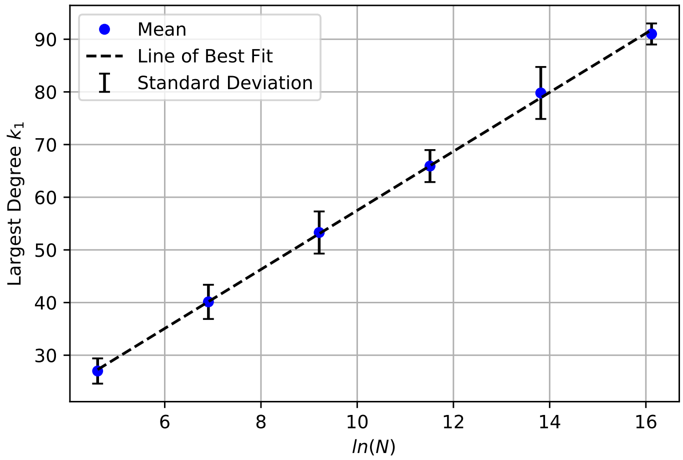


Figure 8: Straight line fit to largest degree k and $ln(N)$. This is a good fit as the line is included in all error bars and not points are systematically above or below the line.

2.7.2 Largest Degree Numerical Results

For the same reasons as for the corresponding section for the BA case (2.4.1) the value of m was set to 5. The results are shown in Fig.8 where a straight line is fitted to k_1 and the natural logarithm of N. The straight line is a good fit thereby corroborating the theoretical prediction of $k_1 \sim \ln N$. More specifically, the fit coefficients are the slope $\alpha = 5.60 \pm 0.72$ and the intercept $\beta = 1.5 \pm 0.7$. Looking at the predictions of Eq.(40) we expect a slope of $1/[\ln{(m+1)}/m] = 5.48$ and the intercept to equal the choice of m. We see that the theoretical prediction is verified with respect to the slope but not the intercept. This is explained by the fact that the slope is much less sensitive to the errors at small system size than the intercept. Fitting a line to the largest three sizes gives $\beta = 3.7 \pm 4.7$, a very uncertain value which does nevertheless include and has a closest mean to m.

3 Phase 3: Existing Vertices Model

3.1 Existing Vertices Model Theoretical Derivations

As for the previous to versions we start with the master equation

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

which we now update as

$$\begin{aligned} n(k,t+1) = n(k,t) + \frac{m}{2}\Pi_{\text{pa}}(k-1,t)n(k-1,t) - \frac{m}{2}\Pi_{\text{pa}}(k,t)n(k,t) \\ \frac{m}{2}\Pi_{\text{rnd}}(k-1,t)n(k-1,t) - \frac{m}{2}\Pi_{\text{rnd}}(k,t)n(k,t) + \delta_{k,m} \end{aligned} \tag{41}$$

to reflect the current problem. Recalling $\Pi_{\text{pa}} = \frac{k}{2E}$ and $\Pi_{\text{rnd}} = \frac{1}{N}$,using $E = mN$ and $N(t+1) = N(t) + 1$ and substituting for the probability distribution such that $n(k) = p_\infty(k)N$ we arrive at

$$p_\infty(k) = \frac{1}{4}\left[(k-1+2m)p_\infty(k-1) - (k+2m)p_\infty(k)\right] + \delta_{k,m}. \tag{42}$$

Here we will need the identity that we have proven before, i.e. that

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

has solution

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

Guided by this, we rearrange Eq.(42) into

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k+2m-1}{k+2m+4}, \tag{43}$$

where we recognise $a = 2m-1$ and $b = k+2m+4$. This is valid for $k > m$. We now see that

$$p_\infty(k) = A\frac{\Gamma(k+2m)}{\Gamma(k+2m+5)} \tag{44}$$

$$= \frac{A}{(k+2m+4)(k+2m+3)(k+2m+2)(k+2m+1)(k+2m)}. \tag{45}$$

We can obtain the value of the constant A by considering the case $k = m$. Then Eq.(42) becomes

$$p_\infty(m) = -\frac{3}{4}mp_\infty(m) + 1. \tag{46}$$

We rearrange for $p_\infty(m)$ and equate to Eq.(45) to obtain

$$\frac{4}{3m+4} = \frac{A}{3m(3m+4)(3m+3)(3m+2)(3m+1)} \tag{47}$$

and hence

$$A = 12m(3m+3)(3m+2)(3m+1). \tag{48}$$

Finally,

$$p_\infty(k) = \frac{12m(3m+3)(3m+2)(3m+1)}{(k+2m+4)(k+2m+3)(k+2m+2)(k+2m+1)(k+2m)}. \tag{49}$$

As before, we will now check whether the function has the correct properties. It is trivial to see that it tends to 0 as $k \to \infty$. Below we will check the normalisation. We start from the normalisation condition

$$\sum_{k=m}^{\infty} p_\infty(k) = 1, \tag{50}$$

and substitute the expression for $p_\infty(k)$ after having breaking it into partial fractions:

$$A\sum_{k=m}^{\infty} -\frac{1}{6(k+2m+1)} + \frac{1}{4(k+2m+2} - \frac{1}{6(k+2m+3)} + \frac{1}{24(k+2m+4} + \frac{1}{24(k+2m)}. \tag{51}$$

We now carry the summation through to every fraction and take out terms so that we have the five sums look the same (As we did in the BA case). Doing the algebra shows that the summation equals $1/A$ and therefore

$$\sum_{k=m}^{\infty} p_\infty(k) = \frac{A}{A} = 1. \tag{52}$$

3.2 Existing Vertices Model Numerical Results

Data was collected for networks of $m = 2$, 8, and 16(We used 100, 10 and 10 networks respectively). The result is shown in Fig.9. This is evidently a very bad fit. Both the theoretical derivation and the computational implementation where checked for errors but none were found.

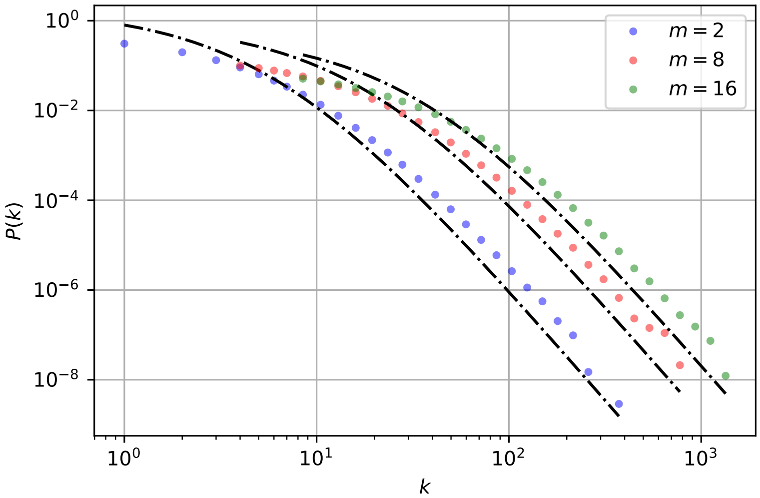


Figure 9: Numerical data for different values of m along the corresponding theoretical predictions (black lines) for the Existing Vertices model. We see that this is a very bad fit. The data points are systematically either above or below the predictions giving now room for attributing this to noise.

4 Conclusions

It was generally showed that analytical prediction agreed with numerical data apart from the existing vertices case were the possibility of a flaw in either the numerical model or the prediction remains open. The theoretical probability distributions modelled the data very well; while finite size effects made their appearance at the high k region they were also confined there. Investigations of largest degree k_1 where successful and so was the performed data collapse.

References

[1] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999. [Online]. Available: `http://science.sciencemag.org/content/286/5439/509`

[2] A. D. Broido and A. Clauset, "Scale-free networks are rare," *Nature Communications*, vol. 10, no. 1, p. 1017, 2019. [Online]. Available: `https://doi.org/10.1038/s41467-019-08746-5`