Investigating methods of edge attachment in the Barabasi-Albert model

Ben R. Amroota

CID: 01508466

26th March 2021

***Abstract*— This project details an investigation into the Barabasi-Albert model for different methods of edge attachment using simple growing graphs. A program is written in Python to implement the BA model numerically and subsequently check the theoretical forms for the degree distributions and largest expected degrees with numerical results. The BA model using preferential attachment is found to produce a fat-tailed degree distribution described by a power-law, and hence models a scale-free network. By performing a data collapse it is shown that random attachment does not exhibit the same fat-tail, thus does not model scale-free systems. By mixing attachment methods, it is found that the largest expected degree obeys a power-law where the exponent depends on the mixing parameter q. For purely random attachment the power-law breaks down and $\langle k_1 \rangle \propto \ln(N)$.**

Word Count – 2498

I. INTRODUCTION

THE Barabasi-Albert model specialises the Price model for undirected graphs and creates random scale-free networks using preferential attachment [1]. A network is a system of discretised units connected by a series of edges, and is described as scale-free if there is no typical degree. This report investigates the consequences on the scale-free nature of a simple network with preferential, random, and mixed attachment.

A. Definition

The Barabasi-Albert model uses preferential attachment to simulate a growing network exhibiting scale-free behaviour, whereby the probability a new node forms an edge with an existing node is proportional to the degree of the existing node. The model is defined as follows:

1. Create an initial graph $\mathcal{G}_0$ at time $t_0 = 0$.
2. Increment time $t \rightarrow t + 1$
3. Add a single node
4. The new node forms m edges with probability Π with existing vertices.
5. Repeat from 2 until a network of N nodes is created.

II. PURE PREFERENTIAL ATTACHMENT Π_{PA}

A. Implementation

1) Numerical Implementation

The network is initialised with an initial graph of $m + 1$ nodes, each degree m. The algorithm creates a final network of $N = t + m + 1$ nodes, where t denotes the number of nodes added. The degree and list of vertex connections each node has is in a list of lists; the list index i labels the node. Once a new node is added, one of the m edges is assigned to an existing node based on the attachment list. The number of times the node appears in the list is equal to the degree of said node, hence choosing randomly from the attachment list is identical to the preferential attachment probability, $\Pi_{PA} \propto k$.

To ensure no edges are repeated, the existing vertex is checked against the list of vertex connections for the new node. Self-loops are avoided by demanding, if the existing vertex chosen is the new node, the existing vertex is rechosen.

2) Initial Graph

The initial graph is a complete simple graph of $m + 1$ nodes with degree m. This satisfies the requirement that no nodes have degree $k < m$ and ensures $N(0) > m$, so no repeated edges form in the initial graph. All nodes in the initial graph have the same degree to prevent a skew in the network. To satisfy theoretical requirements, the initial number of nodes $N(0)$ is chosen such that $N(0) \ll N$, where N is the final number of nodes in the graph.

3) Type of Graph

The BA model produces simple networks. The direction of an edge has no bearing on the degree distribution: there is no difference between the Price and BA model with respect to this project. The graphs simulated are thus unweighted, and do not possess any self-loops or multiple edges. The network evolves into a sparse graph, where a few nodes have considerably larger degrees than the others.

4) Working Code

The algorithm adds one node and m edges per timestep when implemented correctly. This is confirmed by adding individual nodes and manually checking that the list of vertex connections and degrees evolves appropriately for each node. For correct implementation, the $\lim_{t\rightarrow\infty} \frac{E(t)}{N(t)} = m$. This is checked for $t \in \{10, 100, 1000, 10{,}000, 100{,}000\}$ using $m = 2$, and the results are displayed to 4 d.p. in Table. 1. Thus, the limit correctly converges.

t	10	100	1000	10,000	100,000
$E(t)/N(t)$	1.7692	1.9709	1.9970	1.9997	2.0000

Table. 1. The ratio between the total edges and total number of nodes in a network of size $N = t + m + 1$. The ratio converges to m as $t \to \infty$, as required for the theoretical derivations. This demonstrates correct implementation of the BA algorithm.

5) Parameters

The parameters are the degrees m of the nodes added, the number of nodes added t, and the number of simulation repetitions R. The values $m \in \{2, 4, 8, 16, 32, 64\}$ are small compared to the number of nodes in the final graph N to minimise finite-size effects. These values ensure linearly spaced data on a logarithmic scale. The number of added nodes t is large to satisfy the requirement $N \gg E(0), mN(0)$ and reduce finite-size effects. Simulations are repeated $R = 50$ times to improve statistics and estimate errors.

B. Preferential Attachment Degree Distribution Theory

1) Theoretical Derivation

The degree distribution of a growing simple network is described using the master equation. The master equation is

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

where $n(k, t)$ is the number of nodes at time t with degree k, $\Pi(k, t)$ is the probability that the new node forms an edge with an existing node of degree k, and $\delta_{k,m}$ is the Kronecker-delta. The master equation is incompatible with multigraphs, hence the requirement that the graphs contain no multiple edges. The probability a given node at time t having degree k is

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

where $N(t)$ is the number of nodes in the network at time t. Noting $N(t+1) = N(t) + 1$, and substituting, Eq. 1 becomes

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

Assuming $p(k, t)$ approaches some stationary solution

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

Eq. 3 becomes

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

The initial nodes and edges $N(0)$ and $E(0)$ evolve by

$$\begin{gathered} N(t) = N(t-1) + 1 = N(0) + t \\ E(t) = E(t-1) + m = E(0) + mt\,. \end{gathered} \tag{6}$$

where $E(t)$ denotes the number of edges at time t. Taking the limit,

$$\lim_{t\to\infty}\left[\frac{E(t)}{N(t)}\right] = \lim_{t\to\infty}\left[\frac{E(0) + mt}{N(0) + t}\right], \tag{7}$$

$$\lim_{t\to\infty}\left[\frac{E(t)}{N(t)}\right] = \lim_{t\to\infty}\left[\frac{E(0)}{N(0) + t}\right] + \lim_{t\to\infty}\left[\frac{mt}{N(0) + t}\right] = m\,. \tag{8}$$

For a finite network,

$$\lim_{t\to\infty}\left[\frac{E(t)}{N(t)}\right] = \frac{E(0) + m(N - N(0))}{N} = \frac{E(0)}{N} + m - \frac{mN(0)}{N}\,. \tag{9}$$

Hence, provided $E(0) = mN(0)$ or $N \gg E(0), mN(0)$,

$$E(t) = mN(t)\,, \tag{10}$$

For preferential attachment after normalisation the probability $\Pi(k,t)$ is

$$\Pi(k,t) = \frac{k}{\sum_{i=1}^{N} k_i} = \frac{k}{2E(t)}\,. \tag{11}$$

By substituting Eq. 10 and Eq. 11 into Eq. 5,

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

For the case $k > m$, it is evident

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{(k-1)}{(k+2)}\,, \tag{13}$$

The solution to which uses the gamma function,

$$k + \alpha = \frac{\Gamma(k+\alpha+1)}{\Gamma(k+\alpha)}\,. \tag{14}$$

Writing

$$\frac{f(k)}{f(k-1)} = \frac{(k+a)}{(k+b)}\,, \tag{15}$$

it is clear

$$\frac{f(k)}{f(k-1)} = \frac{\Gamma(k+a+1)}{\Gamma(k+b+1)}\frac{\Gamma(k+b)}{\Gamma(k+a)}\,. \tag{16}$$

Defining $g(k)$,

$$g(k) = \frac{\Gamma(k+a+1)}{\Gamma(k+b+1)}\,, \tag{17}$$

Eq. 16 becomes

$$\frac{f(k)}{f(k-1)} = \frac{g(k)}{g(k-1)}\,. \tag{18}$$

In choosing an initial condition $f(1)$, by induction it can be shown that

$$\frac{f(k)}{g(k)} = \frac{f(1)}{g(1)}\,, \tag{19}$$

hence

$$f(k) = A\frac{\Gamma(k+a+1)}{\Gamma(k+b+1)}\,, \tag{20}$$

where A is some arbitrary constant. The solution to Eq. 13 is therefore

$$p_\infty(k) = A\frac{\Gamma(k)}{\Gamma(k+3)}\,, \tag{21}$$

which, using $\Gamma(z+1) = z\Gamma(z)$, is re-expressed as

$$p_\infty(k) = \frac{A}{(k+2)(k+1)k}\,. \tag{22}$$

As no nodes exist with $k < m$, $p_\infty(m-1) = 0$, thus for $k = m$,

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

To determine A, Eq. 23 is equated with Eq. 22 for $p_\infty(k=m)$,

$$p_\infty(k=m) = p_\infty(m) = \frac{A}{(m+2)(m+1)m} = \frac{2}{m+2}\,. \tag{24}$$

Thus, the degree distribution in the long-time limit is

$$p_\infty(k) = \frac{2m(m+1)}{(k+2)(k+1)k}\,, k \geq m\,. \tag{25}$$

2) Theoretical Checks

Taking the limit $k \to \infty$, it is evident from Eq. 25 that $p_\infty(k) \to 0$, consistent with expectation. The smallest degree is m hence $p_\infty(k)$ is largest for $k = m$, satisfying Eq. 27. The values m are positive integers, hence $p_\infty(k) < 1 \; \forall \; k, m$. Equation. 25 is a probability distribution and should be normalised. This is checked by using partial fractions,

$$\sum_m^\infty p_\infty(k) = \frac{2m(m+1)}{2} \sum_m^\infty \left[\frac{1}{(k+2)} - \frac{2}{(k+1)} + \frac{1}{k}\right], \tag{26}$$

which is split into

$$\sum_m^\infty p_\infty(k) = \frac{2m(m+1)}{2} \left(\sum_m^\infty \left[\frac{1}{(k+2)} - \frac{1}{(k+1)}\right] - \sum_m^\infty \left[\frac{1}{(k+1)} - \frac{1}{k}\right] \right). \tag{27}$$

Almost all terms cancel, leaving

$$\sum_m^\infty p_\infty(k) = \frac{2m(m+1)}{2} \left(-\frac{1}{m+1} + \frac{1}{m} \right) = 1\,, \tag{28}$$

confirming that Eq. 25 is normalised. For large k, the degree distribution reduces to a simple power-law of the form $p_\infty(k) \propto k^{-3}$ up to first order in k, consistent with the characteristic form for scale-free distributions $f(k) = ak^b$.

C. Preferential Attachment Degree Distribution Numerics

1) Fat-Tail

The numerical degree distribution is fat-tailed, implying that a few vertices have very large degrees compared to the majority of nodes. This is indicative of scale-free networks which are described using power-laws. Due to poor statistics, there is significant noise in the fat-tail, hence log-binning is used to counter statistical noise and uncover the power-law behaviour. The bins increase logarithmically in width and a scale of 1.2 is chosen.

2) Numerical Results

The network simulated uses $m \in \{2, 4, 8, 16, 32, 64\}$ for $t = 100{,}000$ node additions. The error-bars are the standard errors of the data within each bin. The degree distribution data $p_N(k)$ is log-binned and plotted in Fig. 1. Theoretical models are plotted using dashed lines.

The data follows the theory but exhibits deviations at large degrees k, attributed to finite-size effects associated with the false assumption that $t \to \infty$. Simulated networks are finite: $N \neq \infty$ in the long-time limit. The theoretical degree distribution $p_\infty(k)$ from Eq. 25 is plotted against the numerical distribution $p_N(k)$ in Fig. 2, highlighting the region where finite-size effects become significant.

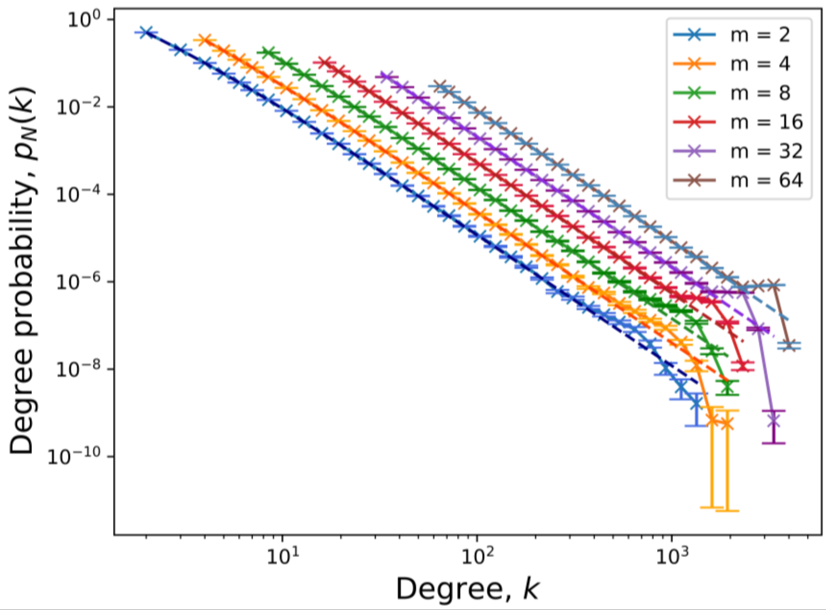

Fig. 1. The degree probability distribution for the BA model using preferential attachment. The theoretical distribution $p_\infty(k)$ as derived in Eq. 25 is plotted using dashed lines and displays a good level of agreement with the numerical data $p_N(k)$. This agreement deteriorates in the tail of the plot and is attributed to finite size effects.

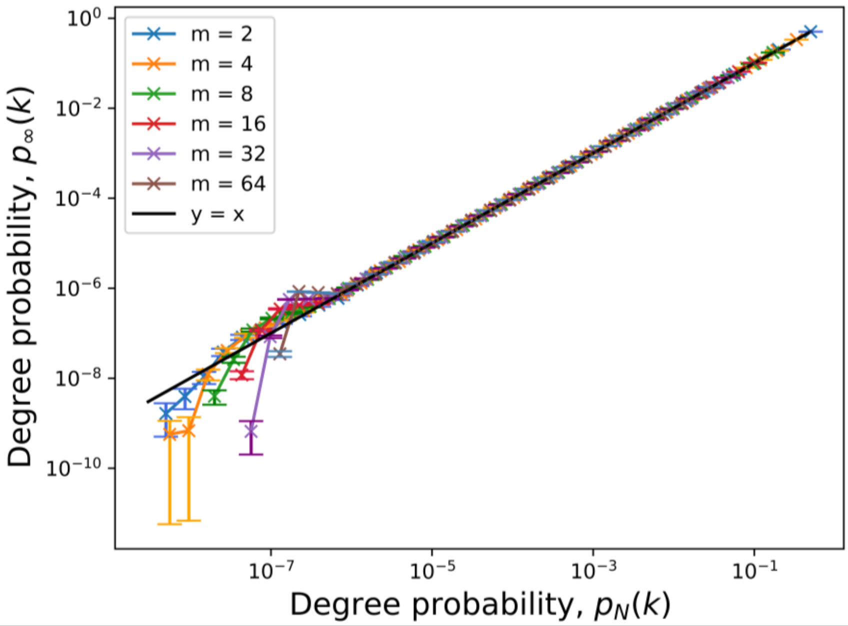

Fig. 2. The theoretical degree probability against the numerical data degree probability. The line $y = x$ is plotted to further illustrate the deterioration in agreement between the numerical data and the theoretical distribution. The agreement breaks down in the limit of large degree k, due to effects associated with the finite size of the system.

3) Statistics

The Kolmogorov-Smirnov test determines whether to reject the null hypothesis that a given data sample comes from some theoretical distribution, and therefore provides a measure of goodness of fit. It is appropriate in this context as the raw data can be approximated as continuous. The two-sample KS test is used, however this assesses the similarity of the two samples without knowing what the theoretical distribution is, hence may not be the best choice.

The KS test compares the numerical data with the theoretical function Eq. 25; the results are detailed in Table 2. The critical statistic at the $\alpha = 0.05$ significance level for the KS test is $D_c = \frac{1.36}{\sqrt{n}}$ for a sample of size n [2, 3]. The p-values are all very close to 1, and the critical statistic D_c exceeds the KS statistic for all m. The null hypothesis that the data samples come from the same distribution cannot be rejected at the $\alpha = 0.05$ significance level, indicating the data provides a good fit with theory. The p-value and KS statistic deviate at higher m due to increased significance of finite-size effects.

m	2	4	8	16	32	64
KS statistic	0.0588	0.0588	0.0667	0.0714	0.1154	0.1250
$p -$ value	0.9999	0.9999	0.9999	0.9999	0.9919	0.9868

Table. 2. The KS statistics and p-values for preferential degree distribution data. The two-sample test is used to compare the numerical data obtained and the theoretical function Eq. 25. The results conclusively show that the null hypothesis cannot be rejected at the 0.05 significance level, hence the data fits well to the theoretical model for all m.

D. Preferential Attachment Largest Degree and Data Collapse

1) Largest Degree Theory

The cut-off degree k_1 is the largest expected degree in the network and is the degree after which only one node is anticipated to be found with degree k_1. The sum of expectation values from $k_1 \to \infty$ is hence equal to 1,

$$\sum_{k=k_1}^{\infty} Np_\infty(k) = 2mN(m+1)\sum_{k=k_1}^{\infty}\frac{1}{(k+2)(k+1)k} = 1\,, \tag{29}$$

The sum is evaluated using partial fractions,

$$\sum_{k=k_1}^{\infty} Np_\infty(k) = \frac{2mN(m+1)}{2}\left(-\frac{1}{k+1}+\frac{1}{k}\right) = 1\,, \tag{30}$$

therefore

$$mN(m+1) = k_1(k_1+1)\,. \tag{31}$$

This can be solved quadratically but only the positive solution is physical, therefore

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

hence in the limit of large N, $k_1 \propto \sqrt{N}$.

2) Numerical Results for Largest Degree

Largest degrees are measured for $N = 2^a$ where $a \in \{10, 11, 12, 13, 14, 15, 16, 17\}$ for 100 repetitions. The value $m = 2$ is small satisfies $N \gg mN(0)$. The average largest degree $\langle k_1 \rangle$ is computed by averaging over repeated runs and is plotted against N in Fig. 3. The numerical data agrees within the bounds of error with Eq. 32. Errors are the standard deviations of the largest degrees over different runs, since the degrees have no individual errors. Deviations from theory are attributed to the use of an infinite sum in the derivation of Eq. 32 despite the network being finite.

Pearson's coefficient quantifies the correlation between $\langle k_1 \rangle$ and $\sqrt{N}$. The coefficient is $r = 0.9995$ with associated p-value $p < 0.001$, hence the null hypothesis that there is no linear correlation between variables is rejected with $> 95\%$ confidence.

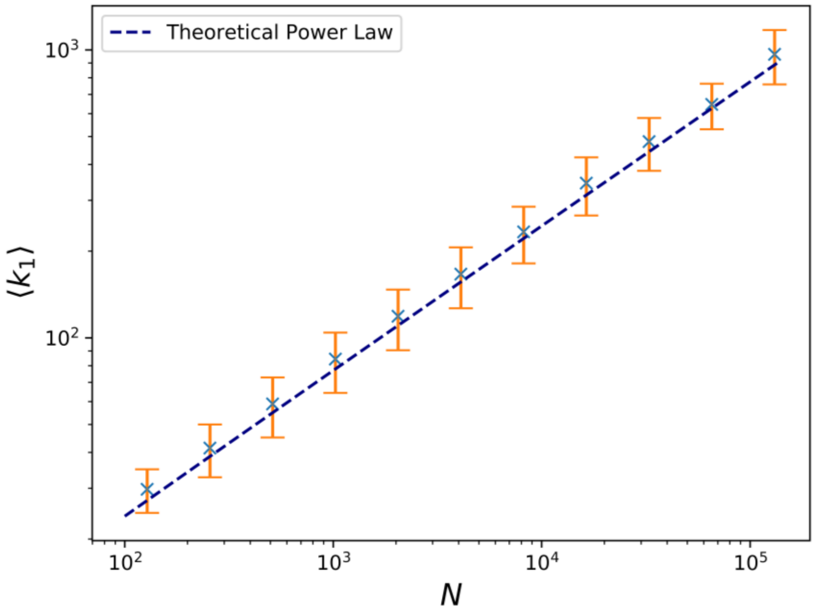

Fig. 3. The largest expected degree averaged over 100 separate runs $\langle k_1 \rangle$ is plotted against the number of nodes in the final graph N. The errors are calculated as the standard deviation of the largest degrees over all 100 runs, since the degrees themselves have no underlying uncertainty. The theoretical power law as determined in Eq. 32 is plotted as the dashed line and agrees with the numerical data within the bounds of error.

3) Data Collapse

The finite-size of the networks affects the degree distribution $p_N(k)$, thus a finite-size scaling ansatz is proposed,

$$p_N(k) = p_\infty(k)\mathcal{F}\left(\frac{k}{k_c}\right), \tag{33}$$

where k_c is the cut-off degree and $p_\infty(k)$ is the theoretical degree distribution. The cut-off degree is the largest degree in the network before finite-size effects become significant and causes nodes in the region $k \to k_c$ to bunch-up, creating the bumps in Fig. 1.

It is proposed that the cut-off degree and largest degree are affected by scaling and assumed that cut-off degree will scale like $k_c \sim k_1$. Therefore, the data is collapsed by plotting $p_N(k)/p_\infty(k)$ against $k/\sqrt{N}$, as seen in Fig. 4.

For $k/\sqrt{N} \ll 1$, the finite-size effects are insignificant: the numerical and theoretical degree distributions align. For $k/\sqrt{N} \geq 1$, the finite-size becomes significant and the numerical data deviates from predictions; theory is only exact for infinite networks.

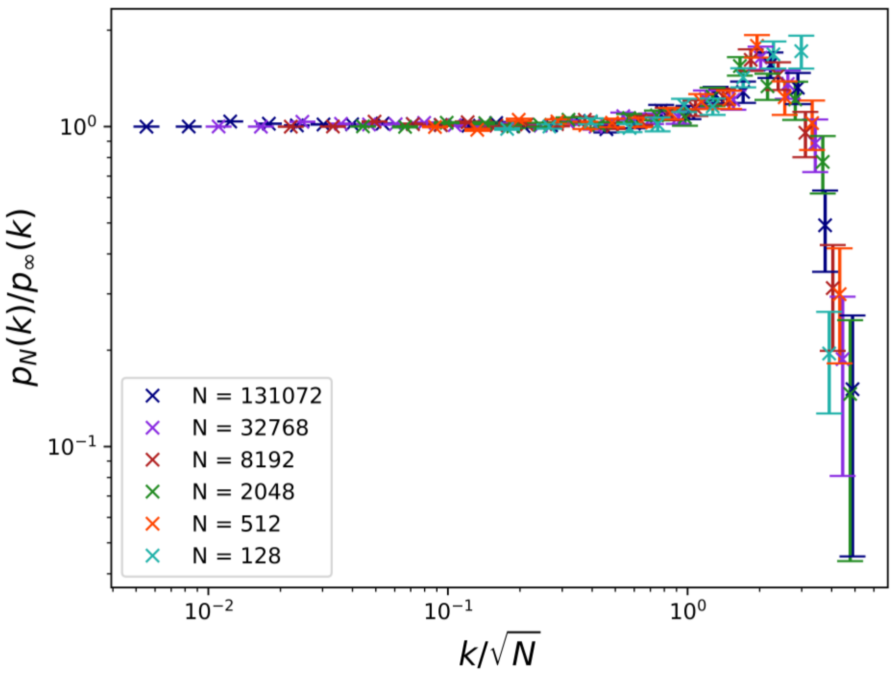

Fig. 4. The data collapse of the degree probability distributions as a function of degree for different final graph sizes N. The data collapses onto a single function which is a constant value of 1 for small arguments $k/\sqrt{N} \ll 1$. For arguments $k/\sqrt{N} \approx 1$ the distribution begins to be affected by finite-size effects, and for larger arguments $k/\sqrt{N} \gg 1$ the scaling function rapidly decays.

III. PURE RANDOM ATTACHMENT Π_R

A. *Random Attachment Theoretical Derivations*

1) *Degree Distribution Theory*

For random attachment the attachment probability is independent of degree, $\Pi_R(t) = \frac{1}{N(t)}$. Equation 5 becomes

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

Therefore, for $k > m$,

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

and for $k = m$,

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

since $p_\infty(m-1) = 0$. Rewriting Eq. 35 as

$$p_\infty(k) = \left(\frac{m}{(1+m)}\right)^{k-m} p_\infty(m) \,, \tag{37}$$

allows Eq. 36 and Eq. 37 to be combined,

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

The degree always satisfies $k > m$ and $m > 0$, therefore $p_\infty(m), p_\infty(k) < 0 \, \forall \, m, k$ as required. Equation. 38 is normalised using standard results for geometric progressions,

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

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

Taking $k \to \infty$, $p_\infty(k) \to 0$, as expected. Thus $p_\infty(k)$ exhibits expected behaviour, however the degree distribution now evolves non-uniformly with the addition of nodes: the distribution no longer describes a scale-free network.

2) Largest Degree Theory

By analogy with preferential attachment, the sum of the expectation values from $k_1 \to \infty$ must equal one,

$$\sum_{k=k_1}^{\infty} p_\infty(k) = \sum_{k=k_1}^{\infty} \left(\frac{m}{m+1}\right)^{k} \frac{m^{-m}}{(m+1)^{1-m}} = \frac{1}{N}\,, \tag{41}$$

therefore

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

Performing a change of summation indices $k_1 + j = k$ gives,

$$\sum_{j=0}^{\infty} \left(\frac{m}{m+1}\right)^{k_1+j} = \left(\frac{m}{m+1}\right)^{k_1} \sum_{j=0}^{\infty} \left(\frac{m}{m+1}\right)^{j} = \left(\frac{m}{m+1}\right)^{k_1} (1+m)\,, \tag{43}$$

where the last step comes from the standard formula for a geometric sum. This evaluates to

$$\frac{m^m}{N(m+1)^m} = \left(\frac{m}{m+1}\right)^{k_1}\,, \tag{44}$$

and taking the logarithm returns

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

Therefore,

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

and hence $k_1 \propto \ln N$ rather than scaling as a power-law.

B. Random Attachment Numerical Results

1) Degree Distribution Numerical Results

Random attachment is simulated using identical parameters as for preferential attachment. Error-bars are the standard errors. The degree distribution $p_N(k)$ in Fig. 5 closely agrees with theory for low degrees. The distribution is not fat-tailed because the network is not scale-free, therefore the data is un-binned. The finite-size effects become more prominent for increasing m, hence the deviation from the theoretical model at larger degrees.

The results of the Kolmogorov-Smirnov tests are detailed in Table 3. The KS statistics remain low for all m, and p-values deviate further from one as m increases. Consequently, the null hypothesis can only not be rejected at the 0.05 significance level for $m \leq 8$. This is caused by finite-size effects, which are more significant here because the data is un-binned, hence the statistics are poorer.

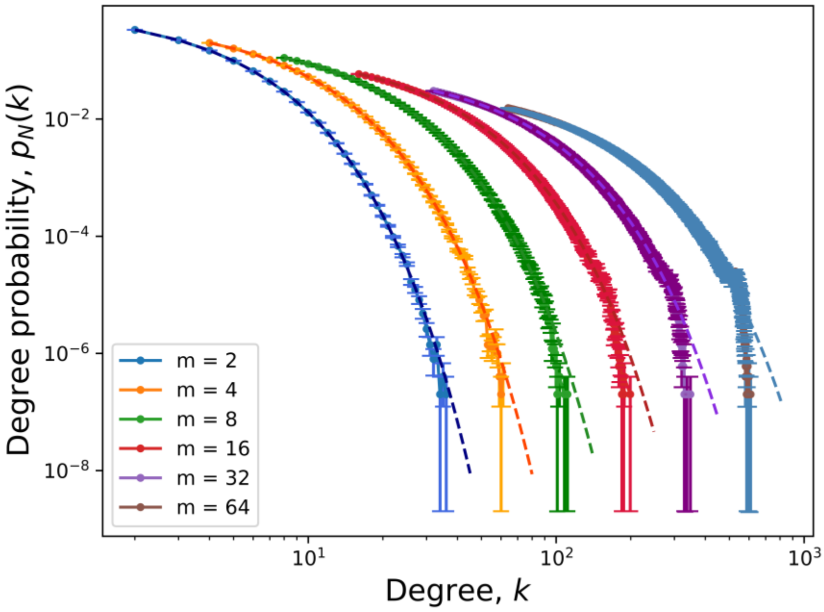

Fig. 5. The degree probability distribution for random attachment in a growing network. The theoretical distributions as derived in Eq. 38 are plotted as the dashed lines and exhibit reasonably close agreement with the data until finite-size effects become relevant. The distributions do not display a fat-tail since they are not scale-free. Hence, the data has no need for log-binning.

m	2	4	8	16	32	64
KS statistic	0.0571	0.0517	0.0714	0.0693	0.0657	0.0670
$p-$value	0.9999	0.9999	0.9574	0.7854	0.5126	0.1723

Table. 3. The results from performing the Kolmogorov-Smirnov test on the degree distributions. The p-values deviate significantly from one for increasing m, due to the increased significance of finite-size effects, hence the null hypothesis must be rejected for $m = 16, 32, 64$ respectively.

2) Largest Degree Numerical Results

Largest degrees are measured for $N = 2^a$ where $a \in \{10, 11, 12, 13, 14, 15, 16, 17\}$ using $m = 2$ and are averaged over 100 repetitions. Standard deviations are the error-bars. The average largest degree is plotted in Fig. 6 against the natural logarithm of the number of nodes and displays a linear fit. The theoretical prediction from Eq. 46 is plotted in blue and demonstrates an increasingly good fit with the data. The deviation is attributed again to the use of an infinite summation in deriving Eq. 46.

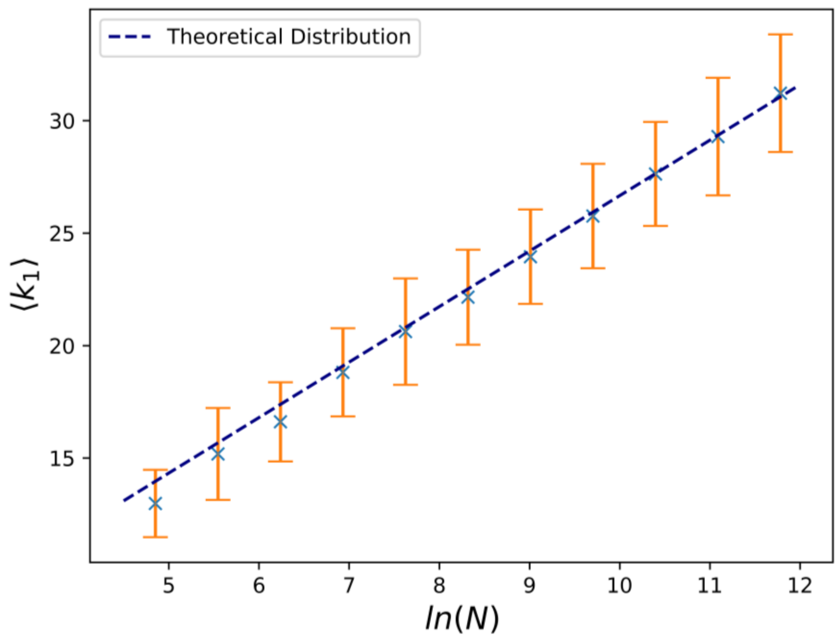

Fig. 6. The largest degree averaged over 100 runs plotted against the natural logarithm of the number of nodes in the system $\ln(N)$ for random attachment in the BA model. The theoretical distribution Eq. 46 is plotted as the dashed blue line. There is increasingly closer agreement as N increases, and the theoretical predictions agree within the bounds of error for all points.

The degree distributions for different N are collapsed by proposing a similar finite-size scaling ansatz as before. Plotting $p_N(k)/p_\infty(k)$ against $k/\ln N$ produces the collapse seen in Fig. 7 and Fig. 8.

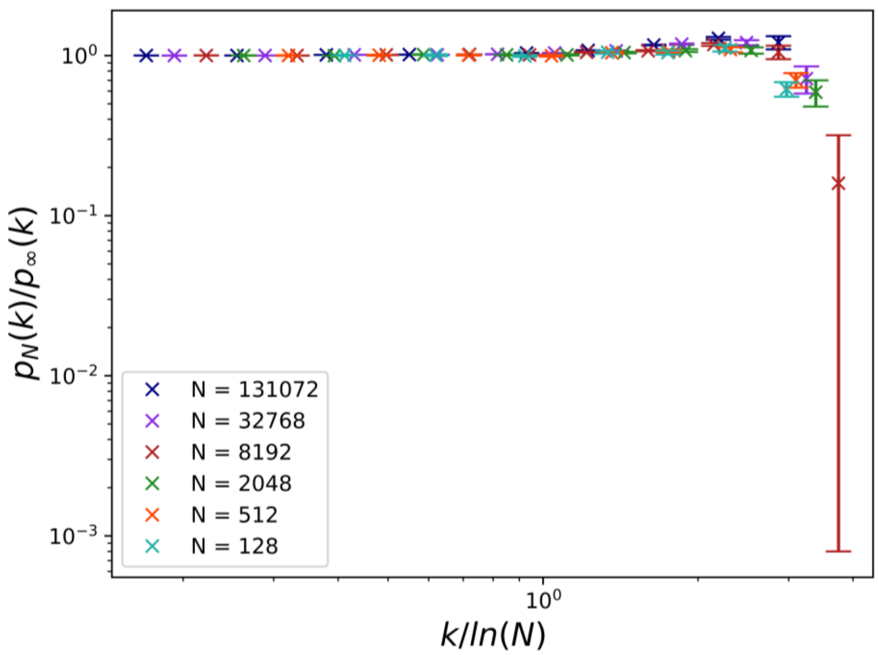

Fig. 7. The data collapse of the degree probability distribution. Unlike for preferential attachment, random attachment networks are not scale free. Hence, the peak is not nearly as defined, and the decay at large arguments $k/\ln N$ is due to the finite nature of the network.

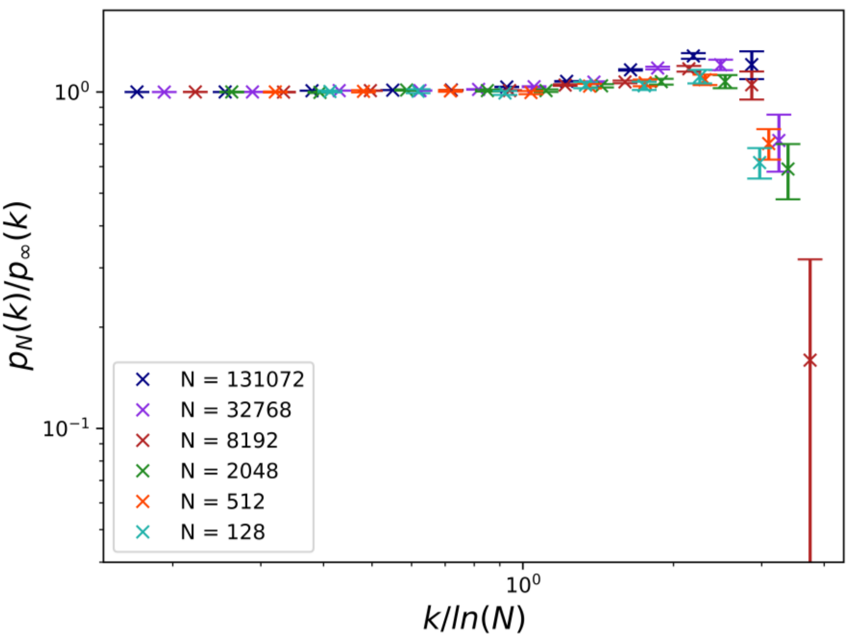

Fig. 8. A zoomed in version of the data collapse of the degree probability distribution in Fig 7. This is to highlight the peak of the data collapse, which is not nearly as defined as for preferential attachment. This is because randomly attached networks are not scale-free and hence are not fat-tailed.

IV. MIXED PREFERENTIAL AND RANDOM ATTACHMENT

A. *Mixed Attachment Model Theoretical Derivations*

For mixed attachment, where q is the mixing parameter

$$\Pi(k) = q\Pi_{PA}(k) + (1-q)\Pi_R(k) = \frac{qk}{2E(t)} + \frac{(1-q)}{N(t)}. \tag{47}$$

Substituting into Eq. 5 and using earlier relations,

$$p_\infty(k) = \left[\frac{q(k-1)}{2} + (1-q)\mathrm{m}\right] p_\infty(k-1) - \left[\frac{qk}{2} + (1-q)\mathrm{m}\right] + \delta_{k,m}\,, \tag{48}$$

which simplifies to

$$\left[\frac{2+qk+2m(1-q)}{2}\right] p_\infty(k) = \left[\frac{q(k-1)+2m(1-q)}{2}\right] p_\infty(k-1) + \delta_{k,m}\,. \tag{49}$$

For $k \neq m$,

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \left[\frac{k + \frac{2m}{q} - 2m - 1}{k + \frac{2m}{q} - 2m + \frac{2}{q}}\right]. \tag{50}$$

By use of the gamma function,

$$p_\infty(k) = A\frac{\Gamma\left(k + \frac{2m}{q} - 2m\right)}{\Gamma\left(k + \frac{2m}{q} - 2m + \frac{2}{q} + 1\right)}. \tag{51}$$

For the specific case $q = 2/3$, using $\Gamma(z+1) = z\Gamma(z)$,

$$p_\infty(k) = A\frac{\Gamma(k+m)}{\Gamma(k+m+4)} = \frac{A}{(k+m+3)(k+m+2)(k+m+1)(k+m)}. \tag{52}$$

The constant A is found by finding $p_\infty(m)$ and equating this to $p_\infty(k=m)$. This gives

$$p_\infty(k) = \frac{6m(2m+2)(2m+1)}{(k+m+3)(k+m+2)(k+m+1)(k+m)}. \tag{53}$$

For large k, the distribution again reduces to a power-law $p_\infty(k) \propto k^{-4}$ up to leading order in k, consistent with the characteristic form for scale-free distributions. The probability of finding a degree in an infinite system with infinite degree is hence zero, as expected.

For $q = 1/2$, the result is

$$p_\infty(k) = \frac{12m(3m+3)(3m+2)(3m+1)}{(k+2m+4)(k+2m+3)(k+2m+2)(k+2m+1)(k+2m)}. \tag{54}$$

B. Mixed Attachment Model Numerical Results

Simulated networks for $q = 2/3$ use identical parameters as before. The degree distribution is plotted in Fig. 9, and error-bars are the standard errors. The distribution is fat-tailed and decays faster than purely preferential attachment, as expected. The KS results are shown in Table. 3 and show that the null hypothesis cannot be rejected at the 0.05 significance level.

m	2	4	8	16	32	64
KS statistic	0.0526	0.0555	0.0588	0.0588	0.0667	0.0667
$p-$value	1.0	1.0	0.9999	0.9999	0.9999	0.9999

Table. 3. The KS statistics and p-values for mixed attachment; the null hypothesis cannot be rejected at the 0.05% significance level: the data fits well to Eq. 53.

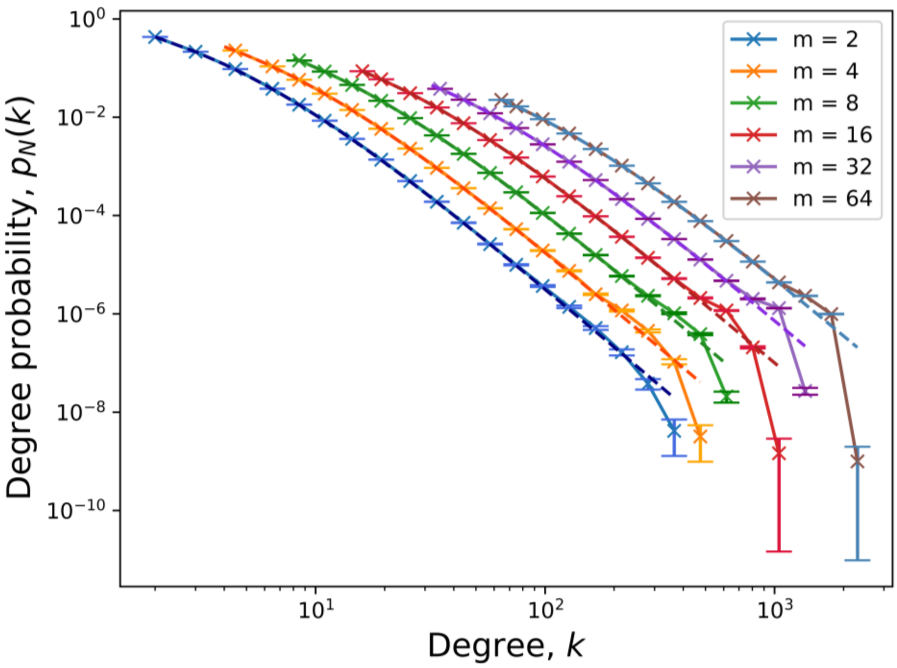

Fig. 9. The degree probability distribution for a mixed network with $q = 2/3$. The data displays a fat-tail as anticipated and the finite-size effects are preceded by a small peak, as for preferential attachment. Both are less pronounced than for preferential, due to the random attachment contribution.

As can be seen in Fig. 10, the average largest degree follows a power-law for $q \neq 0$, the exponent of which reduces with q, hence it is proposed that the power-law exponent is determined by q. This form breaks down for random attachment and becomes logarithmic.

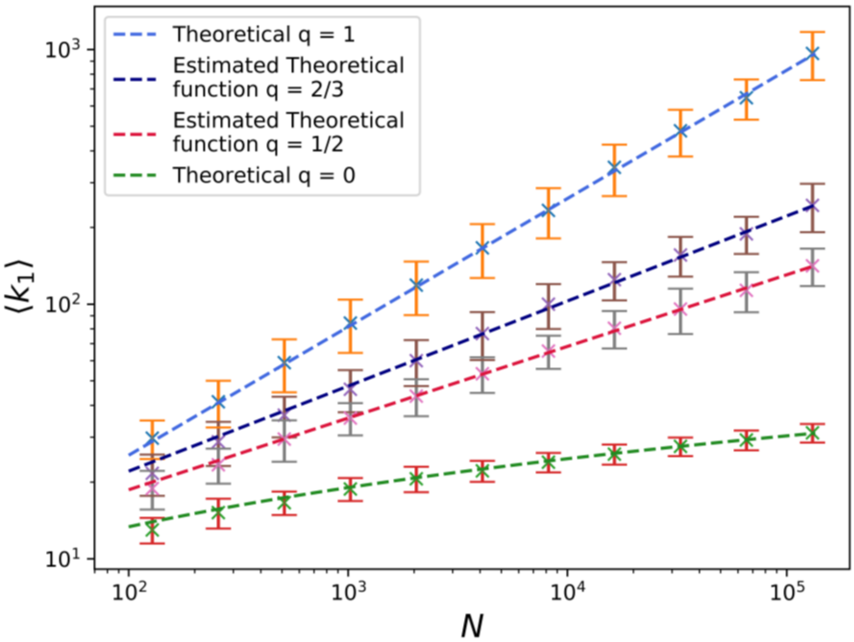

Fig. 10. The average largest degrees for all methods of attachment considered. For all cases with $q \neq 0$, a power-law relation is observed. The exponent of the power law thus depends on the parameter q, and by extension the amount of preferential attachment in mixed attachment.

V. CONCLUSION

This project implements the BA algorithm for three methods of edge attachment: preferential, random, and a mixture of the two. Comparing theoretical and numerical results, it is found preferential attachment describes scale-free networks whilst random does not. The results for mixed attachment depend on the parameter q; the largest expected degree follows a power-law with the exponent determined by q, however this breaks down for random attachment.

VI. REFERENCES

[1] Barabási AL. Network science. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2013 Mar 28;371(1987):20120375.

[2] S. Rick Wicklin, “Critical values of the kolmogorov-smirnov test,” 20/05/19 https://blogs.sas.com/content/iml/2019/05/20/critical-values-kolmogorovtest.html [Accessed: 29.03.2021].

[3] Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. SIAM review. 2009;51(4):661-703.

VII. ACKNOWLEDGEMENTS

I would like to thank Max Falkenberg McGillivray for their log-binning code applied to the relevant data in the report.