Network Project
A Growing Network Model

CID: (01519545)

28th March 2020

Abstract: This report studies the analytical and numerical results for three models of a growing complex network. The initial model is the Barabási-Albert model that utilises preferential attachment, the second model utilised pure random attachment and thirdly a mixture of both attachment methods. In this report network sizes of up to $N = 10^n$ for $n = 2$ to 6 that were averaged over 5x10^{7-n} runs. The numerical results for the degree distributions for each model were found to corroborate the analytical expressions until a cut-off due to the systems finite-size resulting in rapid decay. The largest degree in the Barabási-Albert model was found to analytically scale $\sim \sqrt{N}$ and the numerical results plotted on a log-log axis gave the exponent of N as 0.502 ± 0.015 and also verified that a fat-tail distribution was present. The random attachment was found to not exhibit a fat-tail and thus did not have a power-law decay. However, the mixed attachment model was found to contain a fat-tail distribution.

Word Count: 2490 words excluding font page, figure captions, table captions, acknowledgement and bibliography.

0 Introduction

The project aims to study the degree distribution produced by simple models that describe growing networks[1] and study the resulting fat-tailed distribution to explain the principle of "rich get richer". The initial model we consider is described by Barabási and Albert[2][3]. The objective of the report is to compare the numerical simulations against corresponding mathematical equations that have been simplified for mathematical analysis.

0.1 Definition

The Barabási-Albert model (BA) describes the algorithm for generating and growing scale-free networks based on the Price Model[4] to explain fat-tailed distributions. The BA model studies the principle of cumulative advantage termed as preferential attachment(PA) where the probability of attaching to an existing node is proportional to the number of edges (m) connected to a node, known as the degree (k_i).

1 Phase 1: Pure Preferential Attachment Π_{pa}

1.1 Implementation

1.1.1 Numerical Implementation

The BA model was implemented through the following:

1. Set up initial network with $m + 1$ nodes at time t_0 as an adjacency list represented by a nested list in python.
2. Increment time $t \to t + 1$ and add one node
3. Add m edges to the new node as follows:
   (a) Connect one end of the edge to the new node and update the adjacency list.
   (b) Connect the other end of the edge to an existing node that has not already been connected to the new node with a probability Π_{pa}.
      i. PA, Π_{pa}, is implemented through a list containing the node indices repeated to equal their degree and choosing a value with uniform probability from this list to ensure $\Pi_{\text{pa}} \propto$ degree.
4. Repeat from 2 until there are N nodes in a network

Utilising this method significantly sped up the algorithm allowing for $N = 10^6$ system size to be repeated multiple times within a short time frame.

1.1.2 Initial Graph

The initial graph is a complete and simple graph where all possible edges are present and where there is at most one edge between any two pairs of nodes. The complete graph ensures that the initial nodes all have m edges and to minimise the effect of the initial conditions on the evolution of the network. A simple graph with no self-loops, undirected and unweighted is used to remain consistent with the BA model requirements. The initial graph has $m+1$ nodes so that the minimum number of initial nodes are used to ensure that a new node can always connect with m edges. Also $m > 1$ so that PA works correctly.

1.1.3 Type of Graph

A simple, unweighted and undirected graph is produced as per the requirements of the BA models's properties. The BA model produces a scale-free graph whose distribution follows a power-law asymptotically such that the fraction of nodes, $P(k)$, in the network having k edges becomes $P(k) \sim k^{-3}$. This is due to the stochastic nature of the BA model and the principle of cumulative advantage resulting in a network that contains considerably higher degrees than expected from an exponential graph. The graph is a fat-tailed distribution where $p(k)$ decays slower than an exponential and we expect to see that on a log-log plot, $p(k)$ has a roughly linear fall off for large values of k to obey the above power-law for the tail of the data. The graph should also demonstrate properties independent of the initial graph so that the mathematical equations hold true.

1.1.4 Working Code

The implementation of the BA model was verified by calculating the average degree in a network shown to be $2m$. I also made an adjacency matrix to show that the graph was simple and undirected using the transposition of the matrix. I showed it is unweighted and has no self-loops by taking the sum of the diagonals. General print statements were also used to establish correct coding.

1.1.5 Parameters

The program requires the following parameters; N, which is the total number of nodes in the final network, m which is the number of edges to connect for a new node, p which is the attachment type such as PA and q which is the probability value for Phase 3 of the report. For PA, m > 1 and $N \gg 1$ in the long time limit so that scale-free behaviour can be observed.

1.2 Preferential Attachment Degree Distribution Theory

1.2.1 Theoretical Derivation

The distribution can be determined from the master equation,

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

where Π is the probability that a single new edge is connected to a particular node with degree k. Assuming an asymptotic solution for $p(k, t)$ such that,

$$p_\infty(k) = \lim_{t\to\infty} p(k, t), \tag{2}$$

and substitute into equation (1) resulting in,

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

where different Π are used depending on the type of attachment. We can show that irrespective of the initial network configuration and attachment,

$$\frac{E(t)}{N(t)} \to m, \tag{4}$$

for $t \to \infty$, where $E(t)$ is the number of edges at time t. As the initial network requires the minimum number of initial nodes denoted as $N(0)$ and $E(0)$ for the BA model to work. Thus at time t, $E(t) = E(0) + mt$ and $N(t) = N(0) + t$ and as $N \to \infty$,

$$\lim_{t\to\infty} \frac{E(t)}{N(t)} = \lim_{t\to\infty} \frac{E(0) + mt}{N(0) + t} \tag{5}$$

through induction can be reduced to,

$$\lim_{t\to\infty} \frac{E(t)}{N(t)} = \lim_{t\to\infty} \frac{m}{N(0)/t + 1} \tag{6}$$

resulting in equation (4). For PA and for $m = 1$,

$$\Pi(k,t) = \frac{k}{2E(t)} \tag{7}$$

here in the large N limit the probability approaches $m\Pi(k,t)$ for all m. Substituting equation (4) and (7) adjusted for m, into equation (3) results in,

$$p_\infty(k) = \frac{k-1}{2} p_\infty(k-1) - \frac{k}{2} p_\infty(k) + \delta_{k,m}. \tag{8}$$

In the limit of $k < m$ for $N \to \infty$, $p_\infty(k) = 0$. To understand the asymptotic degree probability we consider for $k \geq m$. For $m \neq k$ equation (8) rearranges to,

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

which has a solution of the form,

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

where A is a constant. Utilising the following gamma function,

$$\Gamma(z+1) = z\Gamma(z), \tag{11}$$

results in,

$$p_\infty(k) = \frac{A}{k(k+1)(k+2)}. \tag{12}$$

To determine A, we then study the case for $m = k$ so that $\delta_{k,m} = 1$. Substituting this and equation (4) and (7) into equation (3) results in,

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

which can be reduced as for $k < m$, $p_\infty(k) = 0$ and thus $p_\infty(m-1) = 0$ and so,

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

Combining equation (12) and (14) with the condition $k = m$,

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

thus resulting in $A = 2m(m+1)$ and so the degree distribution in the long-time limit for Preferential attachment is,

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

for $k \geq m$.

1.2.2 Theoretical Checks

One theoretical check is normalisation which is satisfied by the condition that $\sum_{k=1}^{\infty} p_\infty(k) = 1$. By substituting equation (16) into the above condition,

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

thus satisfying the condition. Another check is as k becomes large for equation (16), $P_\infty(k) \sim k^{-3}$. This shows that the approximate solution is in line with the expected form for the fat-tailed distribution and the scale free behaviour seen by the BA model.

1.3 Preferential Attachment Degree Distribution Numerics

1.3.1 Fat-Tail

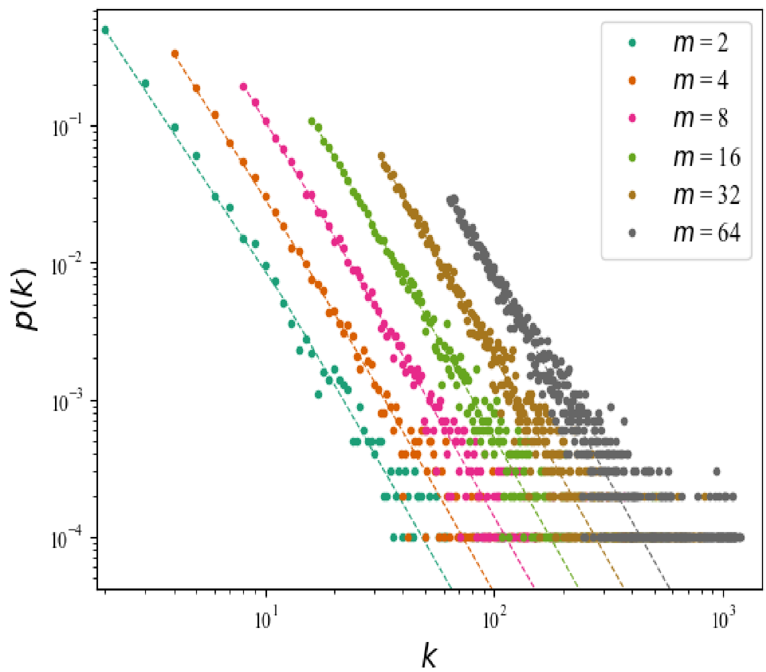

Figure 1: Probability distribution, $p(k)$, as a function of degree k for the BA model for different edges, m, each with a network of size $N = 10^4$. The initial graph was a complete network of size $m + 1$. The respective dashed lines represent the theoretical distribution. The figure shows a fat-tail distribution with significant statistical noise for large k values.

The fat-tailed distribution expected from the asymptotic power-law of the degree distribution for the preferential attachment model contained statistical noise in the tail of the distribution, which is seen in figure (1), plotted on a log-log scale. This is due to large values of k that are not a degree for any of the nodes resulting in zeros that cause problems for some statistical tests. To deal with this issue log-binning was used which scaled the bins so that the size of each bin was progressively incremented by a constant value to deal with the lack of non-zero data points for large k. Bin sizes of 1.1 (i.e. 1.1 times greater than the previous bin) and 1.2 were used in the report.

1.3.2 Numerical Results

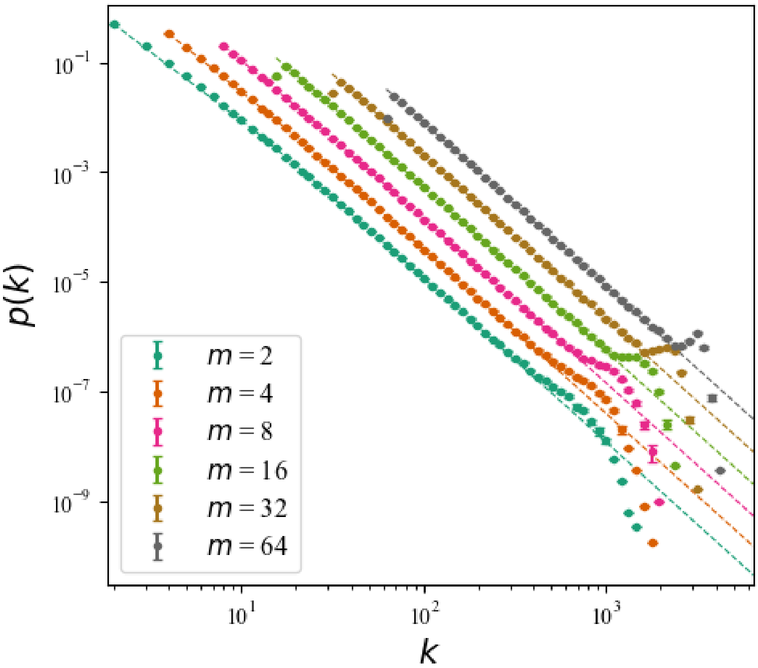

Figure 2: Log binned degree distribution, $p(k)$, as a function of degree k for the BA model. The figure shows networks of size $N = 10^5$ and $m = 2^n$ for $n = 1$ to 6 , averaged over 10x2^{7-n} runs. The initial graph was a complete graph of size $m + 1$. The data points follow the theoretical distribution, dashed lines, until a bump which is followed by a cut-off. The bump is seen to get taller and narrower for increasing m.

Figure (2) shows the log-binned data for the BA model for different m values for $N = 10^5$ and shows their respective asymptotic theoretical distributions predicted by equation (16) as dashed lines. The figure shows that the data follows the theoretical distribution very closely until a bump and then a finite-size cut-off. The standard error on the data points was calculated over multiple runs of the same network and obtaining a standard deviation. The data shows that the bump gets wider and taller for larger m. The data for larger m shows that the initial data point for $m = 16$, 32 and 64 are below the theoretical line. This is due to log binning, as the first bin starts at zero but there are no degrees less than m. As a result, the probability is lower than expected and this difference increases for larger m.

Statistical Test Of Numerical Data For BA Model						
Test	$m=2$	$m=4$	$m=8$	$m=16$	$m=32$	$m=64$
R^2 pvalue	1.0	1.0	1.0	0.841	0.847	0.752
Chi-sqaured pvalue	1.0	1.0	1.0	1.0	1.0	1.0
KS pvalue	1.0	1.0	1.0	0.999	0.964	0.948

Table 1: Statistical tests performed on the numerical data against the theoretical distribution for the BA model for different m and $N = 10^5$. The data was truncated before the bump and finite-size cutoff and the Chi-squared, R^2 and Kolomogrow-Smirnov tests were performed. The KS test pvalues show that the theory and data corroborate. The Chi-squared values show that the theory and data corroborate. The R^2 values decrease as m increases but show that initially the data corroborates very closely with the theory.

1.3.3 Statistics

Table 1 shows the pvalues for the statistical tests performed on numerical data against the theoretical distribution for the BA model. The numerical data was truncated before the bump and cut-off as these were deviating from the expected theory. The KS and Chi-squared tests show that the data follows the expected distribution very closely however, pvalues of 1.0 are significantly high and may not be accurate and would require the consideration of the error bars in the data points to reduce them to reasonable values. The R^2 values show that the pvalues progressively deviate from 1.0. The R^2 test is also invalid to apply to a log-log plot and thus cannot be used to evaluate the goodness of the fit.

1.4 Preferential Attachment Largest Degree and Data Collapse

1.4.1 Largest Degree Theory

The largest expected degree , k_1 is defined as the value of k for which only one node is expected to be found. This is represented as,

$$\sum_{k=k_1}^{\infty} Np_\infty(k) = 1 \tag{18}$$

and substituting equation (16) into equation (18) results in,

$$\frac{2m(m+1)}{2}\sum_{k=k_1}^{\infty}\left(\frac{1}{k+2}-\frac{1}{k+1}\right)-\frac{2m(m+1)}{2}\sum_{k=k_1}^{\infty}\left(\frac{1}{k+1}-\frac{1}{k}\right)=\frac{1}{N} \tag{19}$$

which can be simplified to,

$$\frac{m(m+1)}{k_1^2+k_1}=\frac{1}{N} \tag{20}$$

which can be rearranged to,

$$k_1^2+k_1-Nm(m+1)=0 \tag{21}$$

which is a quadratic equation with two solutions. Considering the positive solution only (physical solution) results in,

$$k_1=\frac{-1+\sqrt{1+4Nm(m+1)}}{2} \tag{22}$$

which is the k_1 for the BA model. In the limit of large N, $k_1 \sim \sqrt{N}$.

1.4.2 Numerical Results for Largest Degree

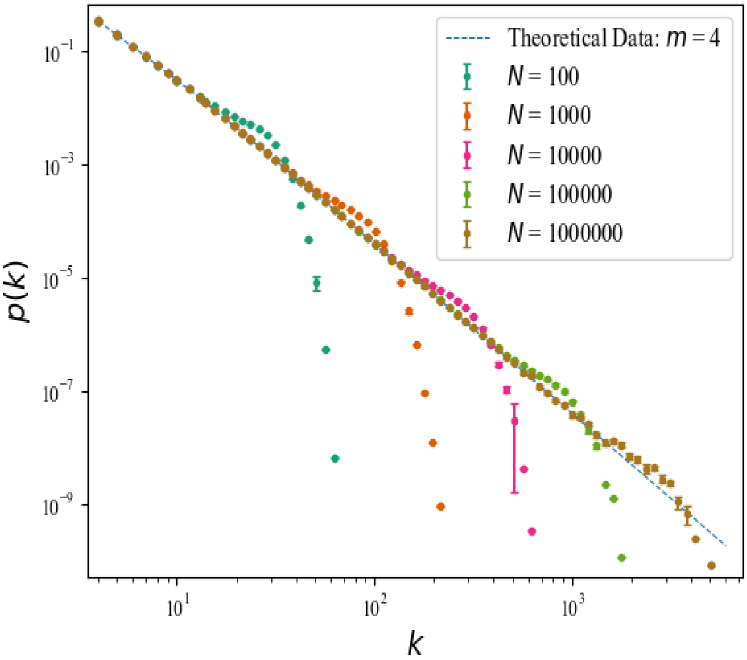

Figure 3: Log binned degree distribution for the BA model for $m = 4$ and different N. The data shows network size's of $N = 10^n$ for $n = 2$ to 6 and was averaged over 5x10^{7-n} and used to calculate the standard error per data point. The dashed line shows the theoretical distribution described by equation (16). The figure shows that higher values of N follow the theoretical distribution for larger values of k preceded by a bump until the cut-off.

Figure (3) shows BA model for $m = 4$ and for different values of N. The figure shows that all N lie on top of the theoretical distribution (dashed line) until the cut-off, with larger N having larger cut-off values. The error on the data points was calculated using the standard error by running each network multiple times. The value of $m = 4$ was used as a reasonable choice to maximise the difference between m and N whilst also having $m > 1$.

Figure (4) shows k_1 as a function of N for $m = 4$. The linear fit with a gradient 0.51 ± 0.015 shows that the scaling relation $k_1 \sim \sqrt{N}$ is true. The off-set of the numerical values from the theoretical relation suggests that there is a slightly different form of equation (22) for finite-size-networks with a different m dependence as the gradient for the theoretical is the same as the numerical data. Also, the values for k_1 was averaged over the individual k_1 values obtained over multiple runs from the data in figure (3) and thus the offset could be explained by the fact that we are not taking the largest value seen across multiple runs but only the average. The error bars were calculated using the standard error.

1.4.3 Data Collapse

Figure (5) shows a data collapse for the data seen in figure (3). The results show that k_1 is the dominant scale that is associated with a system and thus controls its behaviour. The bump seen in figure (5) is a finite-size bump due to the excess probability in that region and is followed by a cut-off. The cut-off represents a decay due to the finite-size of the system and represents the effects seen in the tail of the distribution.

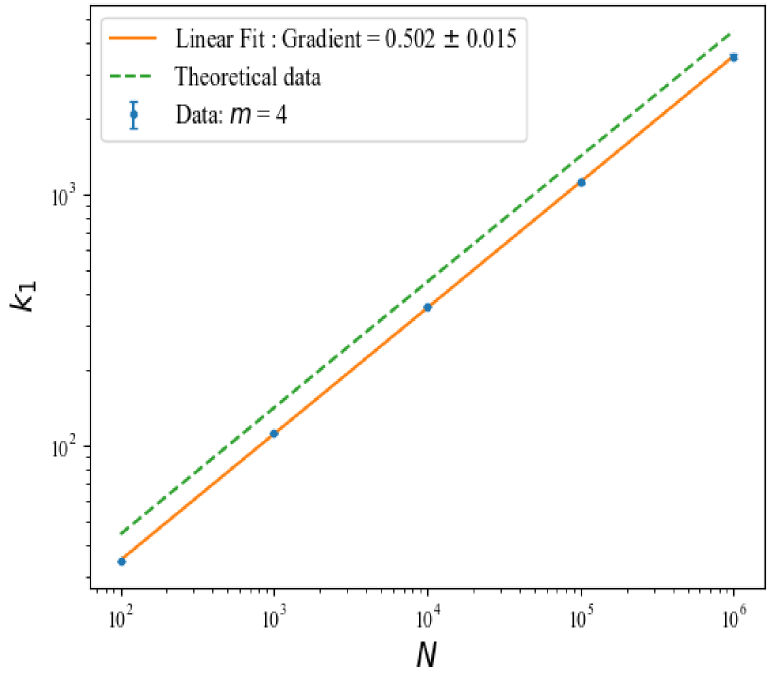

Figure 4: Largest degree, (k_1), for a node plotted against system size $N = 10^n$ for $n =$ 2 to 6. This is derived from the data in figure (3). k_1 values represent the mean of the largest degree calculated across multiple runs of each N with the error bars representing the standard error. The solid line represents a linear fit with a gradient of 0.502 ± 0.015. The dashed line represents the theoretical scaling relation predicted by equation (22). The figure shows that the two lines have a similar gradient and thus scales as predicted however there is an offset from the theoretical relation.

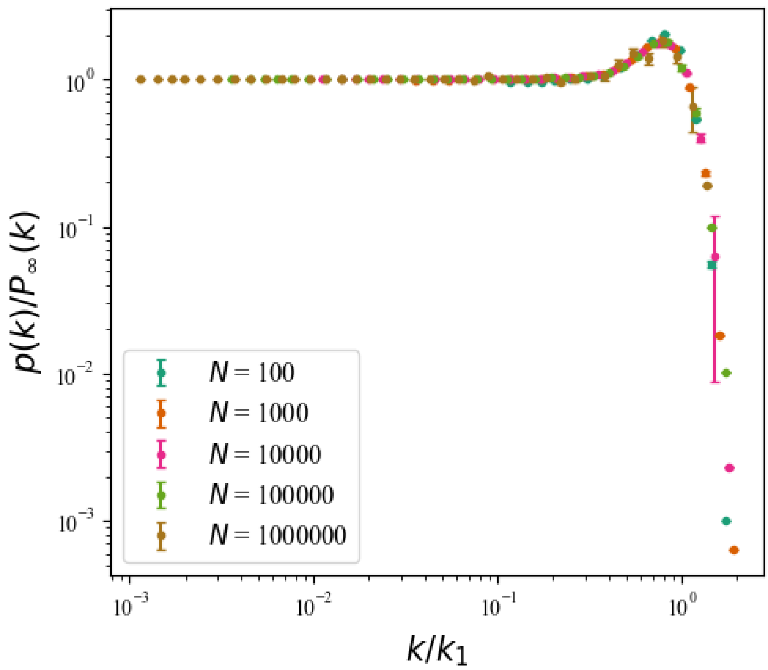

Figure 5: Data collapse for the data seen in figure (3). The y-axis was scaled by dividing $p(k)$ by the theoretical distribution seen in equation (16) and the x-axis was scaled by dividing k by the largest average degree k_1. The figure shows a cut-off seen by a sharp decay that is proceeded by a characteristic bump.

2 Phase 2: Pure Random Attachment Π_{rnd}

2.1 Random Attachment Theoretical Derivations

2.1.1 Degree Distribution Theory

In pure random attachment for $m = 1$ the probability is,

$$\Pi(k, t) = \frac{1}{N(t)}, \tag{23}$$

which can be extended to $m \geq 1$ as $N \to \infty$ which can now be substituted into the masters equation (3) resulting in,

$$p_\infty(k)(1 + m) = mp_\infty(k - 1) + \delta_{k,m} \tag{24}$$

for the case $k > m$ equation (24) becomes,

$$p_\infty(k)(1 + m) = mp_\infty(k - 1) + \delta_{k,m} \tag{25}$$

which through induction can be reduced to.

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

In the case where $k = m$, $p_\infty(k) = 0$ and thus,

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

thus the degree distribution in the long-time limit for pure random attachment is,

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

for $k \geq m$.

2.1.2 Largest Degree Theory

Substituting equation (28) into equation (18) as done for PA results in,

$$\sum_{k=k_1}^{\infty} \left(\frac{m}{m+1}\right)^{k-m} \frac{1}{1+m} = \frac{1}{N} \tag{29}$$

rearranging the sum and then changing the index of the sum results in,

$$\frac{m^m(m+1)^{1-m}}{N} = \sum_{k=k_1}^{\infty} \left(\frac{m}{m+1}\right)^{k} = \left(\frac{m}{m+1}\right)^{k_1} \sum_{j=0}^{\infty} \left(\frac{m}{m+1}\right)^{j}. \tag{30}$$

Utilising the result for the sum of an infinite geometric series results in ,

$$\left(\frac{m}{m+1}\right)^{k_1} = \frac{m^m(m+1)^{1-m}}{N(m+1)} \tag{31}$$

and taking the logarithm of both sides and rearranging, the expected cut-off degree for random attachment is,

$$k_1 = m - \frac{\ln N}{\ln m - \ln(m+1)} \tag{32}$$

2.2 Random Attachment Numerical Results

2.2.1 Degree Distribution Numerical Results

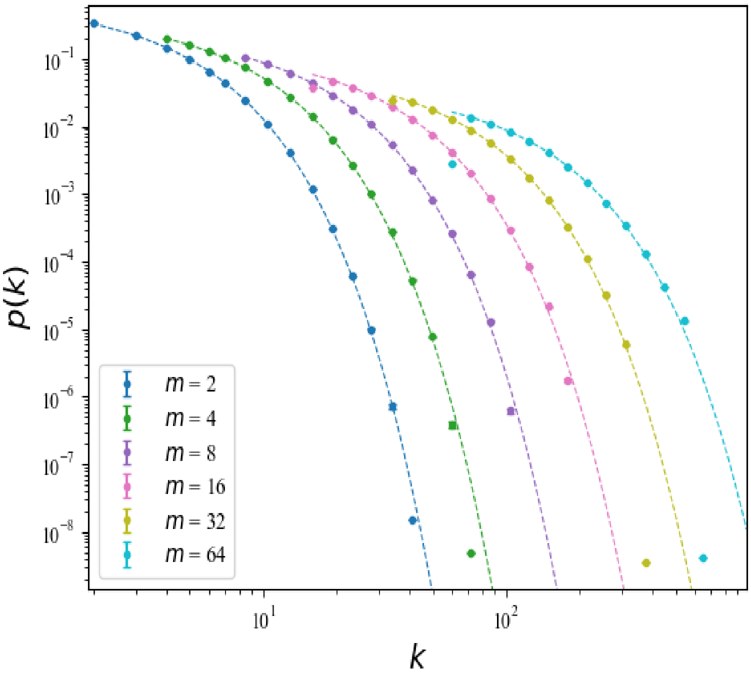

Figure 6: Log binned degree distribution, $p(k)$, as a function of degree k for pure random attachment. The figure shows networks of size $N = 10^5$ and $m = 2^n$ for $n = 1$ to 6 , averaged over $10\text{x}2^{7-n}$ runs. The initial graph was a complete graph of size $m + 1$. The data points follow the theoretical distribution described by equation (28).

Figure (6) shows the degree distribution for pure random attachment for different m and $N = 10^5$. The error bars on the data is the standard error calculated by averaging over multiple runs. Equation (28) shows that pure random attachment does not give a power-law behaviour and therefore does not produce a scale-free network. This is seen in the results as the theoretical and numerical results follow closely with no visible bump. The numerical results do deviate at large k and large m. Numerical tests as described in PA section were performed on the data set above, the results are shown in Table (2), which show that the numerical results follow the theoretical distribution very closely however, the same issue regarding high pvalues remain.

Statistical Test For Pure Random Attachment						
Test	$m = 2$	$m = 4$	$m = 8$	$m = 16$	$m = 32$	$m = 64$
R^2 pvalue	1.0	1.0	1.0	0.915	0.983	0.542
Chi-sqaured pvalue	1.0	1.0	1.0	1.0	1.0	1.0
KS pvalue	1.0	1.0	1.0	1.0	1.0	0.948

Table 2: Statistical tests performed on the numerical data against the theoretical distribution for the pure random attachment model for different m and $N = 10^5$. The KS test pvalues show that the theory and data corroborate however the same problem of pvalues being too high remains true. The Chi-squared values show that the theory and data corroborate. The R^2 values decrease as m increases but show that initially, the data corroborates very closely with the theory.

2.2.2 Largest Degree Numerical Results

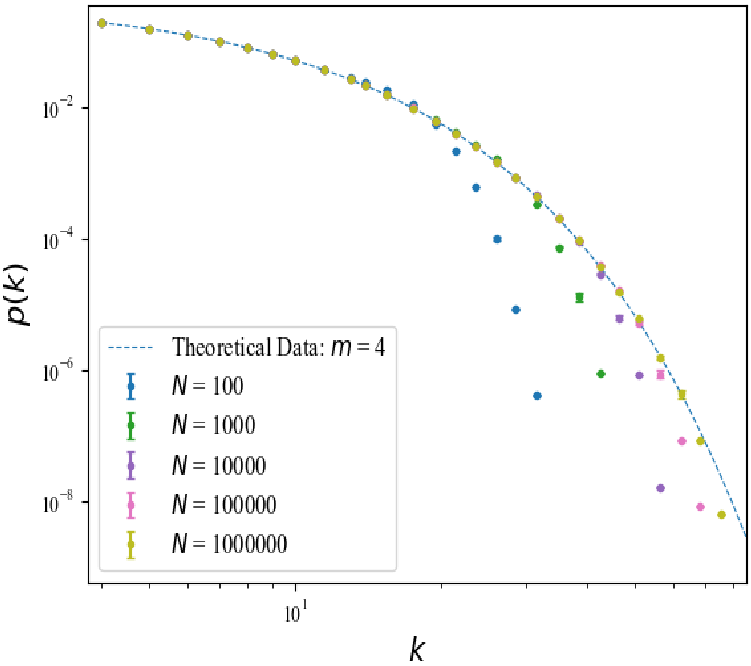

Figure 7: Log binned degree distribution for the pure random attachment model for $m = 4$ and different N. The data shows network size's of $N = 10^n$ for $n = 2$ to 6 and was averaged over 5x10^{7-n} and used to calculate the standard error per data point. The dashed line shows the theoretical distribution described by equation (28). The figure shows that the data only follows the theoretical relation for 1.5 orders of magnitude greater for each N until it deviates.

Figure (7) shows that for $m = 4$, as N increases the numerical data follows the theoretical relation for larger k before rapidly declining implying that as N increases the results become more accurate. This is further corroborated by figure (8). The results in figure (8) show that k_1 is consistent with the theoretical predictions and as N increases the data points approach the theoretical prediction thus behaving as expected. Figure (9) corroborates the finding as it shows the lack of a bump indicating that utilising pure random attachment does not result in a scale-free network as seen by PA.

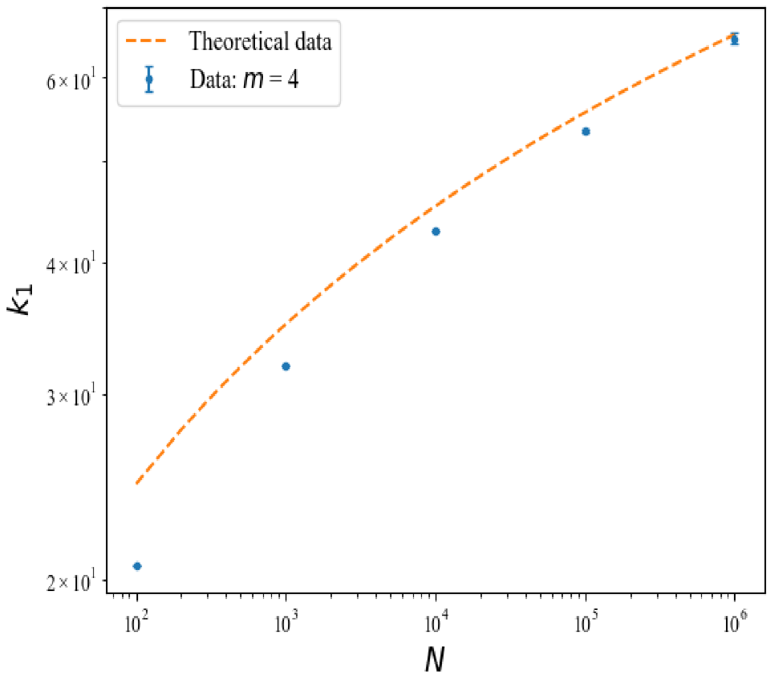

Figure 8: Largest degree, (k_1), for a node plotted against system size $N = 10^n$ for $n = 2$ to 6 for pure random attachment. This is derived from the data in figure (7). k_1 values represent the mean of the largest degree calculated across multiple runs of the N with the error bar. The dashed line represents the theoretical scaling relation predicted by equation (32). The figure shows that as N increases k_1 the data points approach the theoretical value which is expected.

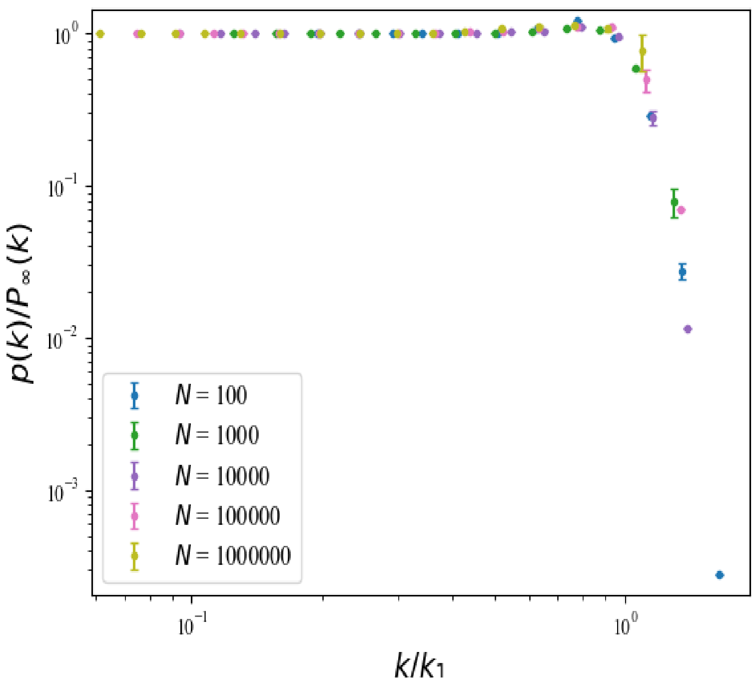

Figure 9: Data collapse for the data seen in figure (7). The y axis was scaled by dividing $p(k)$ by the theoretical distribution seen in equation (28) and the x-axis was scaled by dividing k by the largest average degree k_1. The figure shows a rapid decline without any bump showing that it is not a scale-free network.

3 Phase 3: Mixed Preferential and Random Attachment

3.1 Mixed Attachment Model Theoretical Derivations

For the mixed attachment model (MA),

$$\Pi(k,t) = q\Pi_{BA} + (1-q)\Pi_{rand}, \tag{33}$$

here q is the probability that an edge will be attached using PA, Π_{BA}, and $1-q$, Π_{rand}, is for random attachment. Substituting equation (7) and equation (23) we get,

$$\Pi(k,t) = \frac{qk}{2E(t)} + \frac{1-q}{N(t)} \tag{34}$$

utilising this and substituting into the master equation we find that,

$$p_\infty(k) = m\Pi(k-1,t)p_\infty(k-1)N(t) - m\Pi(k,t)p_\infty(k)N(t) + \delta_{k,m}, \tag{35}$$

and using the same procedure as described for PA and random attachment,

$$p_\infty(k) = \frac{2}{2m+2-mq} \tag{36}$$

for $q \neq 0$ and for $k = m$. For $k > m$,

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

and utilising the gamma function similarly as equation (10) results in

$$p_\infty(k) = \frac{A\Gamma(k + \frac{2m}{q} - 2m)}{\Gamma(k + \frac{2m}{q} + \frac{2}{q} - 2m + 1)}. \tag{38}$$

Utilising a value of $q = \frac{2}{3}$ and applying the same method as discussed in PA, results in,

$$p_\infty(k) = \frac{12m(2m+1)(2m+2)(2m+3)}{(k+m)(k+m+1)(k+m+2)(k+m+3)(4m+6)} \tag{39}$$

which in limit of large k the power-law behaviour is $p_\infty(k) \sim k^{-4}$. Using a value of $q = \frac{1}{2}$ results in,

$$p_\infty(k) = \frac{12m(3m+1)(3m+2)(3m+3)}{(k+2m+4)(k+2m+3)(k+2m+2)(k+2m+1)(k+2m)} \tag{40}$$

which in the limit of large k the power-law behaviour is $p_\infty(k) \sim k{-}5$

3.2 Mixed Attachment Model Numerical Results

The degree distribution seen in figure (10) is for $q = \frac{2}{3}$ for the mixed attachment model. The numerical data follows the theoretical distribution described by equation (39) very closely until a bump and then a cut-off. The bump becomes wider and narrower as m

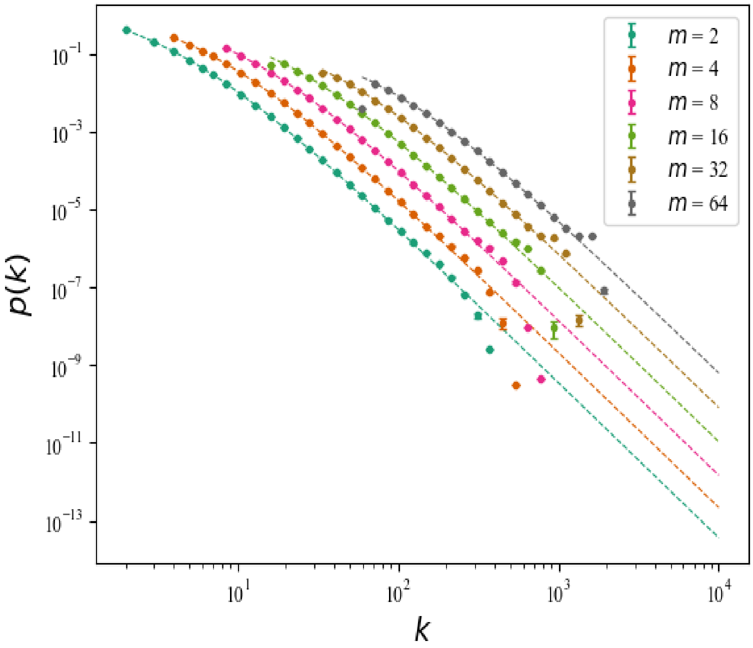

Figure 10: Log binned degree distribution, $p(k)$, as a function of degree k for $q = \frac{2}{3}$ for the mixed attachment model. The figure shows networks of size $N = 10^5$ and $m = 2^n$ for $n = 1$ to 6 , averaged over $10\text{x}2^{7-n}$ runs. The initial graph was a complete graph of size $m+1$. The data points follow the theoretical distribution described by equation (39) until a bump followed by a cut-off. The data indicates that it has a fat-tail distribution.

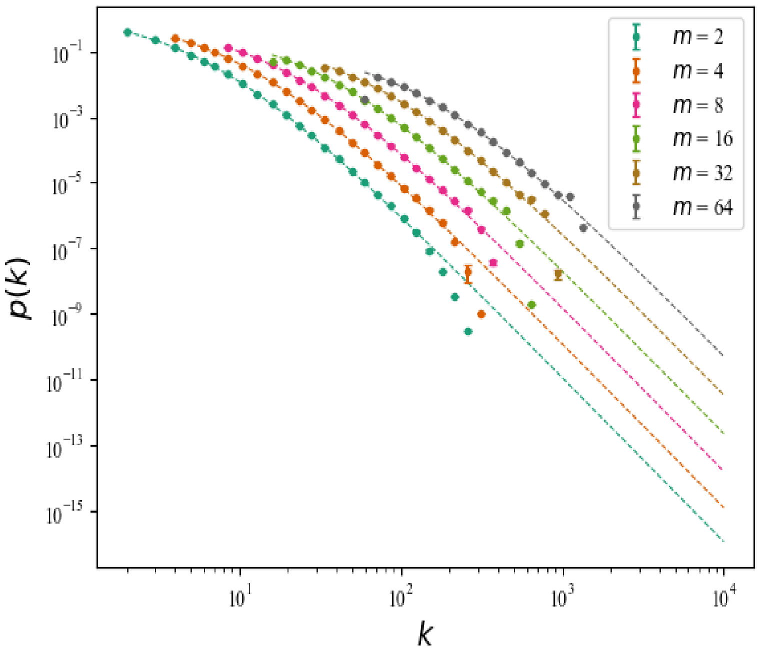

Figure 11: Log binned degree distribution, $p(k)$, as a function of degree k for $q = \frac{1}{2}$ for the mixed attachment model. The figure shows networks of size $N = 10^5$ and $m = 2^n$ for $n = 1$ to 6 , averaged over $10\text{x}2^{7-n}$ runs. The initial graph was a complete graph of size $m+1$. The data points follow the theoretical distribution described by equation (40) until a bump followed by a cut-off. The data indicates that it has a fat-tail distribution.

increases. The results show that for $q = \frac{2}{3}$ it has a fat-tail but decays faster than the BA model due to $\frac{1}{3}$ of the component being random attachment, this result is also seen in figure (12). Similarly for $q = \frac{1}{2}$ in figure (11) whose distribution is described by equation (40) shows similar properties. The fat-tail seen here decays faster than $q = \frac{2}{3}$ as expected. It is therefore understood that as long as $q > 0$ a fat- tail distribution will emerge due to a power-law as seen in figure (13) and the smaller the q the larger the k will have to be to see the finite-size effect. These results are further verified by figure (14) which is the data collapse for $q = \frac{2}{3}$ and shows a distance bump similar to PA.

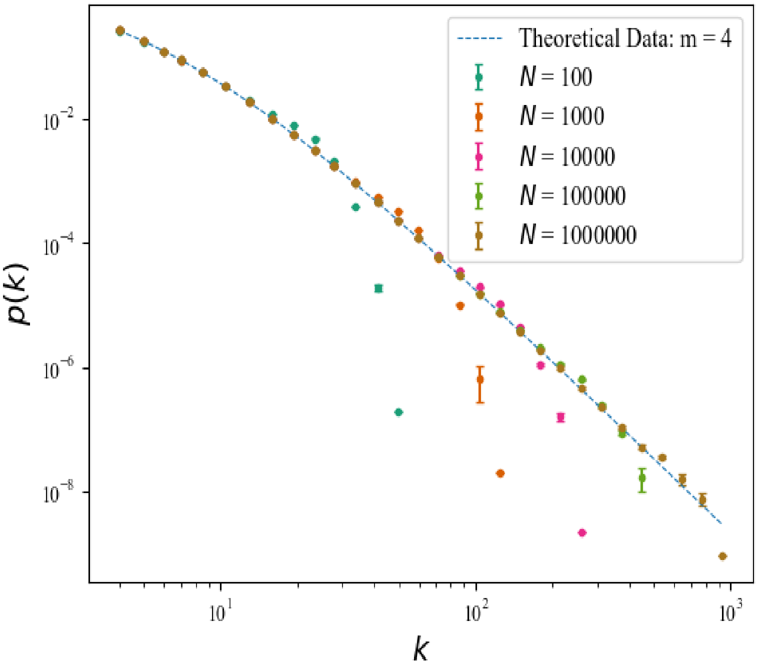

Figure 12: Log binned degree distribution for the mixed attachment model for $q = \frac{2}{3}$, $m = 4$ and different N. The data shows network size's of $N = 10^n$ for $n = 2$ to 6 and was averaged over $5\text{x}10^{7-n}$ and used to calculate the standard error per data point. The dashed line shows the theoretical distribution described by equation (39). The figure shows that higher values of N follow the theoretical distribution for larger values of k until the cut-off, which is preceded by a bump.

4 Conclusions

In this report, we saw overall that for all values of N the data would follow the asymptotic theoretical distribution in the long time limit until a certain point which would be followed by a rapid decay and in the case of preferential attachment an initial bump. The BA model showed that k_1 was the dominant scale for a system and controlled its behaviour. The results produced in the report also show that the random attachment model did not have any fat-tail distribution, however combining this with PA resulted in a power-law and thus a fat-tail as seen by mixed attachment.

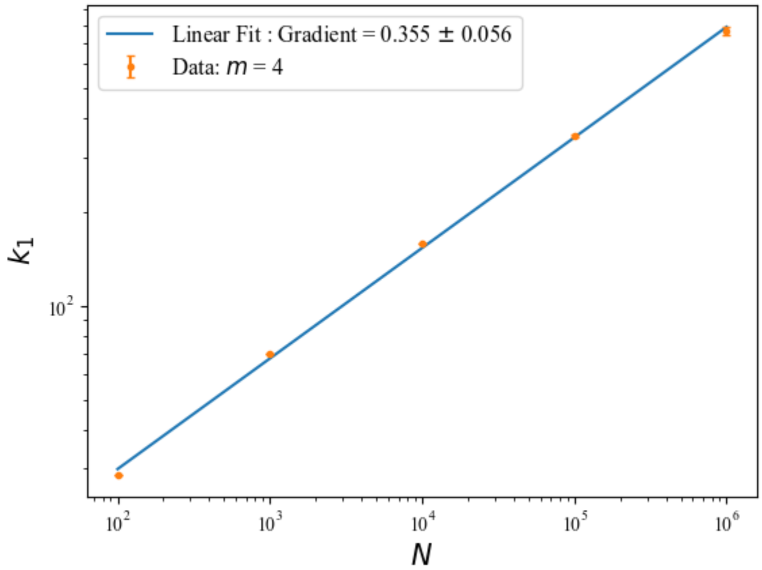

Figure 13: Largest degree, (k_1), for a node plotted against system size $N = 10^n$ for $n =$ 2 to 6 for mixed attachment with $q = \frac{2}{3}$. This is derived from the data in figure (12). k_1 values represent the mean of the largest degree calculated across multiple runs of each N with the error bars representing the standard error. The solid line represents a linear fit with a gradient of 0.355 ± 0.056. The line fit indicates a power-law decay that indicates a fat-tail is present.

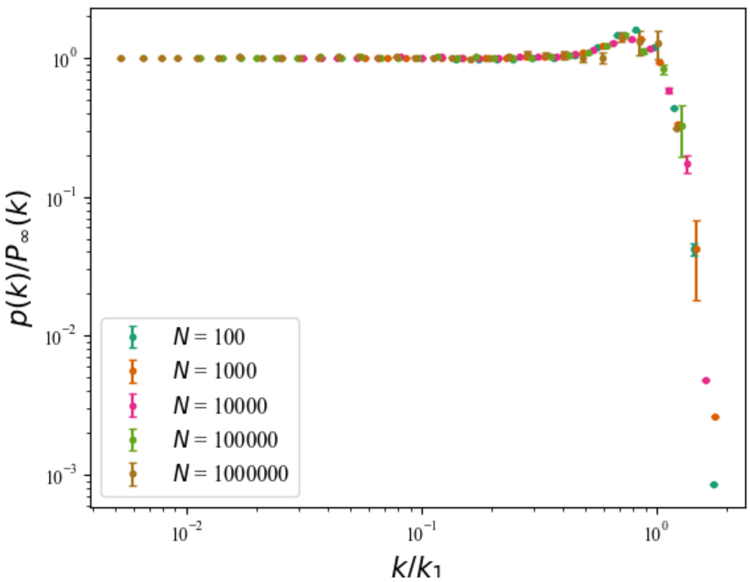

Figure 14: Data collapse for the data seen in figure (12) for mixed attachment for $q = \frac{2}{3}$. The y-axis was scaled by dividing $p(k)$ by the theoretical distribution seen in equation (16) and the x-axis was scaled by dividing k by the largest average degree k_1. The figure shows a cut-off seen by a sharp decay that is proceeded by a characteristic bump. The bump indicates a fat-tail similar to PA figure (5)

References

[1] T.S. Evans, *Amazing Paper*, Journal of Astounding Results, **7** (2019) 1234.

[2] K.Christensen and N.Maloney, *Complexity and Criticality*, Imperial College Press, London, 2005.

[3] A.-L. Barabási and R. Albert, Science 286, 509 (1999).

[4] J. de S. Price, A general theory of bibliometric and other cumulative advantage processes J.Amer.Soc.Inform.Sci. 27 292–306 (1976).