Network Project:
A Growing Network Model

CID: 01707971

28th March 2022

Abstract: In this paper three versions of the Barabási-Albert network are qualitatively and quantitatively analysed. These are the preferential attachment PA, the random attachment RA and the existing vertices EX models. Their degree distribution $p(k;m)$ and behaviour of average largest degree $k_1(N;m)$ is studied both visually and numerically. It is shown that PA and EX are scale-free networks, with $k_1(N)\sim\sqrt{N}$. Oppositely, RA is not scale-free, and its largest degree scales $\sim \ln N$.

Word count: 2490 words in report (excluding 106 from front page, 321 from figure captions, 371 from table captions, 80 from bibliography and in-text-references).

0 Introduction

As networks stochastically grow, they might be described by an appropriate version of the Barabási-Albert (BA) network.

The aim of this report is to describe the degree distributions $p(k)$ of three of its versions: "original" Preferential Attachment (PA), Random Attachment (RA), and Existing Vertices (EX).

Python-based simulations were used, with m ranging between 1 and 100, and number of nodes from 10^4 to 10^6.

0.1 Definition

The BA network grows adding one node and creating m edges at a time, connecting nodes with probability depending on the chosen model.

In the PA model, the node added to the system is connected by m edges to existing nodes. Here a simple undirected version will be studied, requiring the new node to be connected to m different existing nodes. PA implies the existing nodes to be chosen with probability proportional to their degree k_i. Therefore, for each new edge, a node i has probability of being connected

$$\Pi_{\mathrm{PA}}(k_i; t) = \frac{k_i}{2E(t)}, \tag{0.1}$$

where $E(t)$ is the number of edges in the network at that time. The denominator acts as a normalisation constant, as, in an undirected graph with $N(t)$ nodes

$$2E(t) = \sum_{i=1}^{N(t)} k_i. \tag{0.2}$$

In the RA model the node is connected to m existing nodes with uniform probability

$$\Pi_{RA}(t) = \frac{1}{N(t)}. \tag{0.3}$$

Finally, in the EX model, as the new node is added:

- $r < m$ new edges connect it with a previously existing one with probability Π_1.
- $m - r$ new edges connect two previously existing nodes with probability Π_2.

The case $\Pi_1 = \Pi_{RA}$ and $\Pi_2 = \Pi_{PA}$ only will be studied, with $r = m/2$ only numerically analysed.

1 Phase 1: Pure Preferential Attachment

1.1 Implementation

1.1.1 Numerical Implementation

In the PA network, the probability of an existing node with degree k_i to be among the m ones connected to the newly added node is $m\Pi_{\mathrm{PA}}(k_i)$. Numerically, an attachment list was produced and constantly updated, with every i^{th} node appeared k_i times. Such list had length $2E(t)$. By generating a random number $r \in \{0, 1, \dots, 2E(t) - 1\}$, and choosing the node indexed by r in the attachment list, each i^{th} node had probability proportional to k_i to be chosen, exactly as the PA model demands.

As maximum one edge was allowed between any pair of nodes, every time one was picked, it was checked it had not been already chosen. Finally, as new edges were produced, the occurrence of the newly connected nodes in the attachment list was accordingly increased.

1.1.2 Initial Graph

To prevent the initial graph affecting our understanding of the model, two options were analysed. Being N_0 and E_0 the initial number of nodes and edges, these were:

a) A complete graph with smallest number of nodes required, $N_0 = m$.
b) The smallest complete graph with $E_0 = mN_0$ edges, requiring $N_0 = 2m + 1$, establishing exactly the otherwise asymptotic-only relation

$$E(t) = mN(t). \tag{1.1.1}$$

1.1.3 Type of Graph

The graph was simple and undirected. The graph was initialised complete, but, as the number of edges generated per time step was fixed, it was not for $N > N_0$.

1.1.4 Working Code

The simulation records and updates $N(t), E(t)$, the attachment list, a list of degrees of all nodes and an adjacency list A_i. Data were analysed only after an algorithm checked that length of adjacency list and degrees list both were $N(t)$ length of attachment list and sum of all degrees both were $2E(t)$.

Furthermore, the performance of the algorithms assigning edges between nodes was tested: a graph was grown in a random configuration $k_0(i)$, with $N_0 = 200$ nodes, then expanded "statically", as all nodes added afterwards could only be connected to the original N_0 ones. With PA, the degree of these original nodes is expected to increase proportionally to their original degree, implying the normalised final distribution $k(i)$ should match the original $k_0(i)$. In a simple graph, for $m > 1$, this will not hold exactly, as at every step a node can be picked once. Therefore, the second choice will follow a slightly different distribution, and so the third from the previous two, etc… Fig 1.1 confirms this behaviour: for $m = 1$ in (a), input and output match exactly; for $m = 50$ in (b), the output slightly flattens.

1.1.5 Parameters

Although the evolution of the network is stochastic, its general features are entirely determined by the initial choice of m, the initialisation model, and the number of nodes N at which we want to analyse the graph.

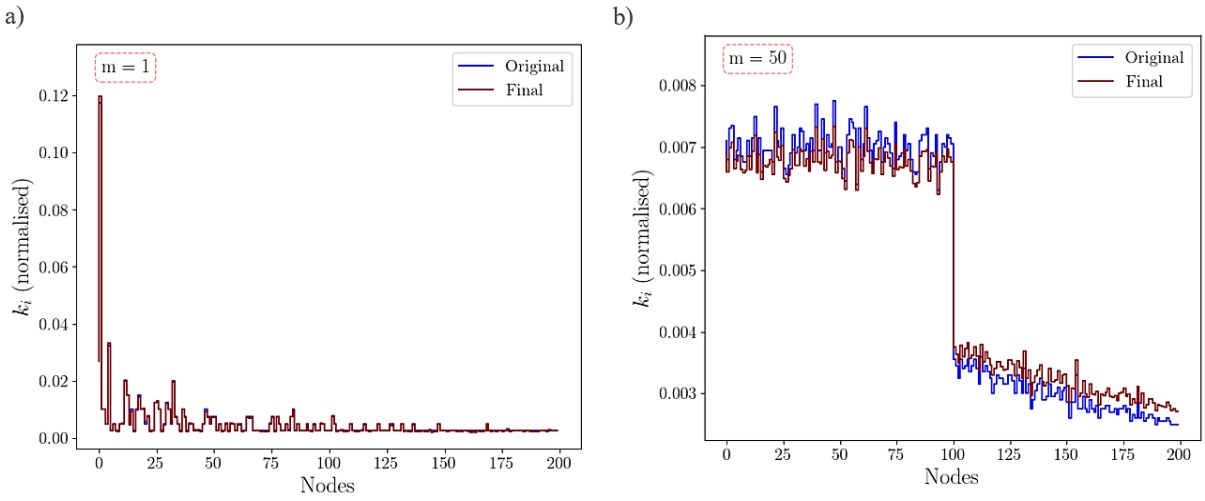

Figure 1.1. Testing results of the PA assignment function. As expected, final and initial $k(i)$ distributions (a) perfectly agree for $m = 1$ and (b) closely resemble each other for bigger m.

1.2 Preferential Attachment Degree Distribution Theory

1.2.1 Theoretical Derivation

The evolution of the BA model is stochastic. Nevertheless, a mean field master equation is here derived with the aim of describing the average behaviour of its degree distribution $p(k)$.

To improve readability, the following notation is adopted: the degree k and time t dependence of a variable will be represented by subscripts and superscripts respectively.

The number of nodes in the network with degree k at time $t+1$, n_k^{t+1} is

$$n_k^{t+1} = n_k^t + m\Pi_{PA_{k-1}}^t n_{k-1}^t - m\Pi_{PA_k}^t n_k^t + \delta_{km}, \tag{1.2.1}$$

Substituting the probability expression given by Eq.(0.1) and $n_k^t = N^t p_k^t$ and rearranging,

$$N^{t+1}p_k^{t+1} - N^t p_k^t = \frac{m(k-1)N^t p_{k-1}^t}{2E^t} - \frac{mkN^t p_k^t}{2E^t} + \delta_{km}. \tag{1.2.2}$$

In the long-time limit, $p(k)$ can be assumed not to change between time steps, hence $p_k^{t+1} = p_k^t = p_k$. Furthermore, E^t reduces to Eq.(1.1.1) for any initial graph, resulting in

$$p_k = \frac{1}{2}[(k-1)p_{k-1} - kp_k] + \delta_{km}. \tag{1.2.3}$$

For $k > m$, it rearranges as

$$\frac{p_k}{p_{k-1}} = \frac{k-1}{2+k}, \tag{1.2.4}$$

with exact solution

$$p_k = A\frac{\Gamma(k-1+1)}{\Gamma(k+2+1)} = A(m)\frac{\Gamma(k)}{\Gamma(\mathrm{k}+3)}, \tag{1.2.5}$$

for some constant $A(m)$, with Γ being the Gamma function, defined, for positive integers n, as (Weisstein, 2002)

$$\begin{aligned}\Gamma(1) &= 1\\ \Gamma(n) &= (n-1)!.\end{aligned} \tag{1.2.6}$$

Then

$$p_k = \frac{A(m)}{k(k+1)(k+2)}. \tag{1.2.7}$$

$A(m)$ is determined by solving Eq.(1.2.4) for $k = m$ and imposing continuity with Eq.(1.2.8). As $k \geq m$, $p_{m-1} = 0$, and it yields

$$\frac{2}{m+2} = \frac{A(m)}{m(m+1)(m+2)}, \tag{1.2.8}$$

finally leading to

$$p_k = \frac{2m^2+2m}{k(k+1)(k+2)}. \tag{1.2.9}$$

1.2.2 Theoretical Checks

Clearly $p(k) > 0\ \forall m, k \in \mathbb{N}$, as required. Besides, it is normalised, as

$$\begin{aligned}\sum_{k=m}^{\infty} p_k &= \sum_{k=m}^{\infty} \frac{2m^2+2m}{k(k+1)(k+2)}\\ &= (2m^2+2m)\sum_{k=m}^{\infty}\left[\frac{1}{2k} - \frac{1}{k+1} + \frac{1}{2(k+2)}\right]\\ &= (2m^2+2m)\left[\frac{1}{2m} + \frac{1}{2(m+1)} - \frac{1}{m+1} + \sum_{k=m+2}^{\infty}\left(\frac{1}{2}\frac{1}{k+2} - \frac{1}{k+2} + \frac{1}{2}\frac{1}{k+2}\right)\right]\\ &= (2m^2+2m)\left[\frac{1}{2m} - \frac{1}{2}\frac{1}{m+1}\right] = 1.\end{aligned} \tag{1.2.10}$$

The cumulative density function (CDF) is

$$\begin{aligned}F_{k_i}(k) = P(k_i \leq k) &= \sum_{k_i=m}^{k} p_{k_i}\\ &= (2m^2+2m)\sum_{k_i=m}^{k}\left[\frac{1}{2k_i} - \frac{1}{k_i+1} + \frac{1}{2(k_i+2)}\right]\\ &= 1 + (2m^2+2m)\left[\sum_{k_i=m+2}^{k}\frac{1}{2k_i} - \sum_{k_i=m+2}^{k+1}\frac{1}{k_i} + \sum_{k_i=m+2}^{k+2}\frac{1}{2k_i}\right]\end{aligned}$$

$$= 1 + (2m^2 + 2m)\left[-\frac{1}{k+1} + \frac{1}{2}\frac{1}{k+1} + \frac{1}{2}\frac{1}{k+2}\right]$$

$$= 1 - \frac{m^2 + m}{(k+1)(k+2)}. \tag{1.2.11}$$

Finally, note this is a scale-free distribution, with infinite variance as

$$V \sim \langle k^2 \rangle = \sum_{k=m}^{\infty} \frac{k^2}{k(k+1)(k+2)} = \sum_{k=m}^{\infty} -\frac{1}{k+1} + \frac{2}{k+2} =$$

$$= -\frac{1}{m+1} + \sum_{k=m+1}^{\infty} \frac{1}{k+1} \sim \zeta(1) \tag{1.2.12}$$

where ζ is Riemann Zeta function, which diverges at 1.

1.3 Preferential Attachment Degree Distribution Numerical Results

1.3.1 Fat-Tail

This is a fat tailed distribution, as decaying faster than an exponential. To deal with the problem this presents, i.e. many large k not being the degree of any vertex, two approaches were used. The first was a log-binning, with values of bin edges growing exponentially as e^{Δ}, with scaling $\Delta > 1$. The bins' increasing size corrects the otherwise non-uniform sampling, allowing all bins to have a comparable number of nodes, retrieving the distribution's true behaviour.

The second was the use of the CDF $F(k)$, necessarily being non-zero for large values of k.

1.3.2 Numerical Results

To test the model previously derived, which only holds for $N \gg 1$, networks were simulated until achieving a large number of nodes, $N = 10^6$, for different values of m ranging between 1 and 100.

Results are shown in Fig 1.3, for $m = 1, 10, 100$. For each m, 50 networks were simulated and their distributions averaged.

Such average perfectly matched theoretical expectations but for the largest values of k. The tail, in fact, displays a different behaviour. For both initial graphs, it presents a bump followed by a steep fall. This is induced by the finite size of the system, not allowing the distribution to infinitely extend, hence causing an "accumulation of probability" (bump) followed by the fall.

Fig 1.3(b) also shows, for the (b) choice of initial graph, the bump is preceded by a drop. This originates as the first N_0 nodes are initialised with higher degree $(2m)$ than those following (m): due to PA, this gap in their initial degree turns into the reduction of nodes with intermediate degree that the drop represents.

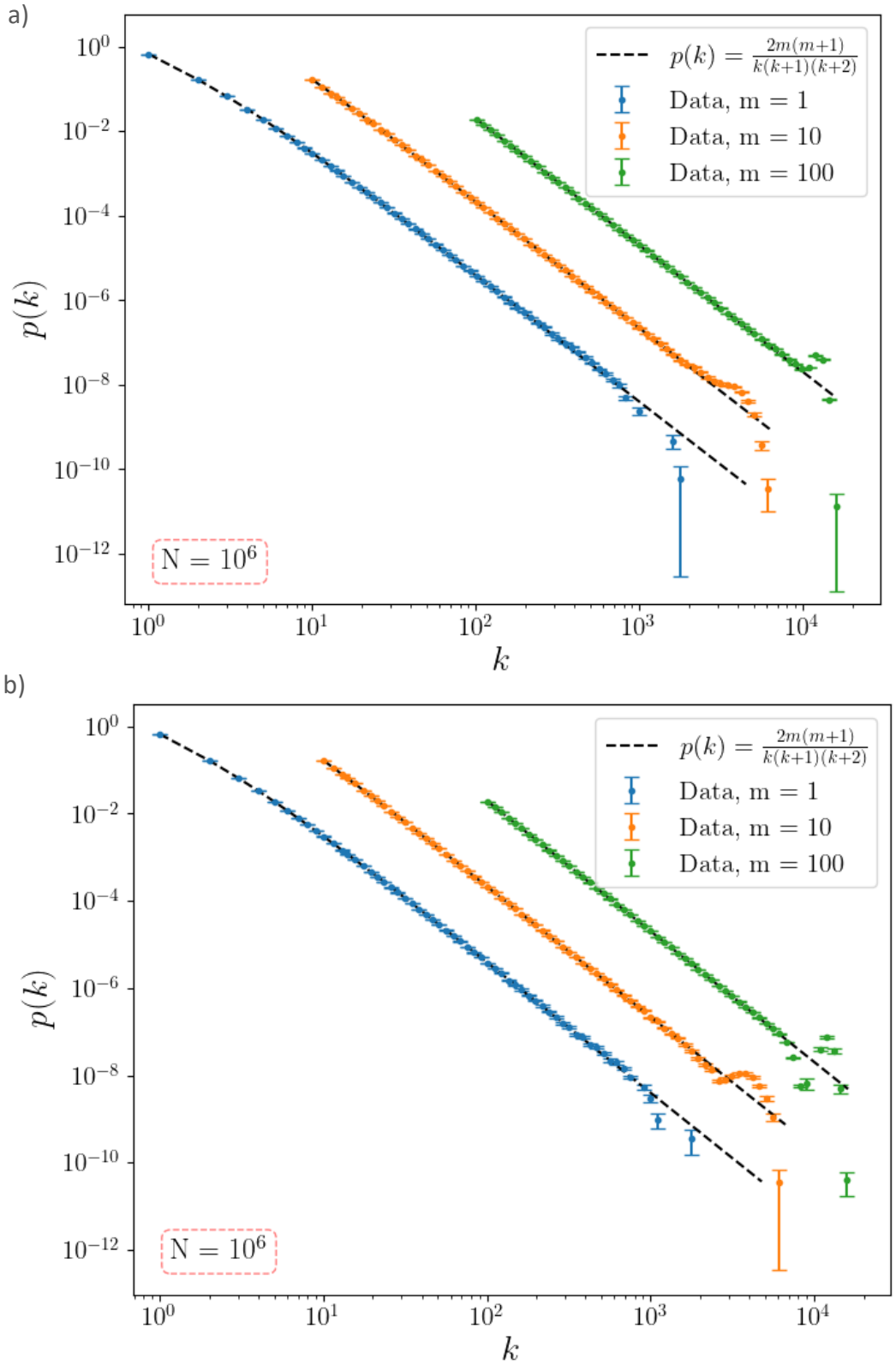

Figure 1.3. Observed and theoretical log-binned $p(k)$ distributions compared for a) $N_0 = m + 1$ initialisation and b) $N_0 = 2m + 1$ model. Both match perfectly but for large k, as the graph's finiteness induces an accumulation followed by a cutoff. Model (b) also presents a drop before the bump, due to the large gap in the degree of the original node, and those following. Error-bars display standard error (SE), as in all future $p(k)$ plots, unless explicitly stated otherwise.

1.3.3 Statistics

To test the model, Kolmogorov-Smirnov (KS) D statistic was used. Oppositely, Pearson's χ^2 was discarded as requiring independence and finite variance, both violated by PA networks.

Given theoretical and observed CDFs $F_{th}(x)$ and $F_{obs}(x)$, the KS statistics is

$$D = \sup_x |F_{th}(x) - F_{obs}(x)|. \tag{1.3.1}$$

The test was applied onto the whole distribution first, then separately on main body and tail, treating them as full CDFs.

The results of the tests are shown in Table 1.1.

Table 1.1: Statistical Validity of PA model

The table displays the results of the KS tests applied onto the whole distribution, its main body and its tail. There is an extremely high overall agreement between theory and observation, and an extremely low one in the tail.

Statistics \ m		1	10	25	50	100
Body-Tail Cut k_{bt}		800	2000	3000	4000	6000
Whole's:	D	$3.3 \cdot 10^{-5}$	$6.5 \cdot 10^{-5}$	$9.5 \cdot 10^{-5}$	$8.9 \cdot 10^{-5}$	$1.3 \cdot 10^{-4}$
	p-value	0.999	0.980	0.758	0.820	0.381
Body's:	D	$3.3 \cdot 10^{-5}$	$6.5 \cdot 10^{-5}$	$9.5 \cdot 10^{-5}$	$8.9 \cdot 10^{-5}$	$1.3 \cdot 10^{-4}$
	p-value	0.999	0.980	0.758	0.820	0.381
Tail's:	D	0.410	0.225	0.322	0.330	0.405
	p-value	0.0	10^{-34}	10^{-162}	0.0	0.0

There is high agreement between the whole function and expectation, and extremely low agreement between the tail and expectation.

1.4 Preferential Attachment Largest Degree and Data Collapse

1.4.1 Largest Degree Theory

The largest degree k_1 can be considered such that contributions from the tail of the distribution are negligible, i.e. the average number of nodes with degree $k_i > k_1$ in the N-nodes network should be 1. In terms of CDF this is

$$N\left[1 - F_{k_i}(k_1 - 1)\right] = 1, \tag{1.4.1}$$

leading, for for $k_1 \gg 1$, to

$$k_1(N; m) = \sqrt{Nm(m+1)}. \tag{1.4.2}$$

Alternatively, consider the approximate average rate at which the degree of a node increases with every new node (Barabási 2016)

$$\frac{dk_i}{dN} = m\Pi_{PA}(k) = \frac{mk_i}{2E(t)}. \tag{1.4.3}$$

Using Eq.(1.1.1) for $E(t)$ it yields

$$\frac{dk_i}{mk_i} = \frac{1}{2}\frac{dN}{mN}. \tag{1.4.4}$$

Integrating, as the node was introduced with degree k_{i0} when there were $N_{i0} - 1$ nodes already,

$$k_i(N) = k_{i0}\sqrt{\frac{N}{N_{i0}}}, \tag{1.4.5}$$

Being k_1 the maximum possible k_i, it is obtained for the firstly initialised nodes. Requiring $N_{i0} = N_0$,

$$k_1(N) = k_{i0}\sqrt{N/N_0}, \tag{1.4.6}$$

where $k_{i0} = m$ for initialisation (a) and $2m$ for initialisation (b).

1.4.2 Numerical Results for Largest Degree

The behaviour of $k_1(N; m)$ was inspected for large values of m, so that $N_0 \propto m$ and $m + 1 \approx m$. A power law

$$k_1 = a\, m^{D_m} N^{D_N}$$

was fitted for N ranging between 10^4 and 10^6, and $m = 25, 50, 100$. The best fit were

- for initialisation (a): $a = 2.00 \pm 0.13$, $D_m = 0.44 \pm 0.01$, $D_N = 0.497 \pm 0.003$, with $\chi^2_{red} = 0.17$ and p-value 0.99.
- for initialisation (b): $a = 1.91 \pm 0.14$, $D_m = 0.45 \pm 0.01$, $D_N = 0.498 \pm 0.005$, with $\chi^2_{red} = 0.12$ and p-value 0.99.

These two agree with each other, with, most importantly, $D_N \sim 0.5$, as expected from both theoretical methods. Besides, the similarity of a suggests the dependence on the initial configuration to be negligible as N grows larger, oppositely to what expected from the second model in Eq.(1.4.6). Instead, Eq.(1.4.2) has a completely wrong scaling in m.

The behaviour of $k_1(N;m)$ is shown in Fig.1.4 for initial graph (a).

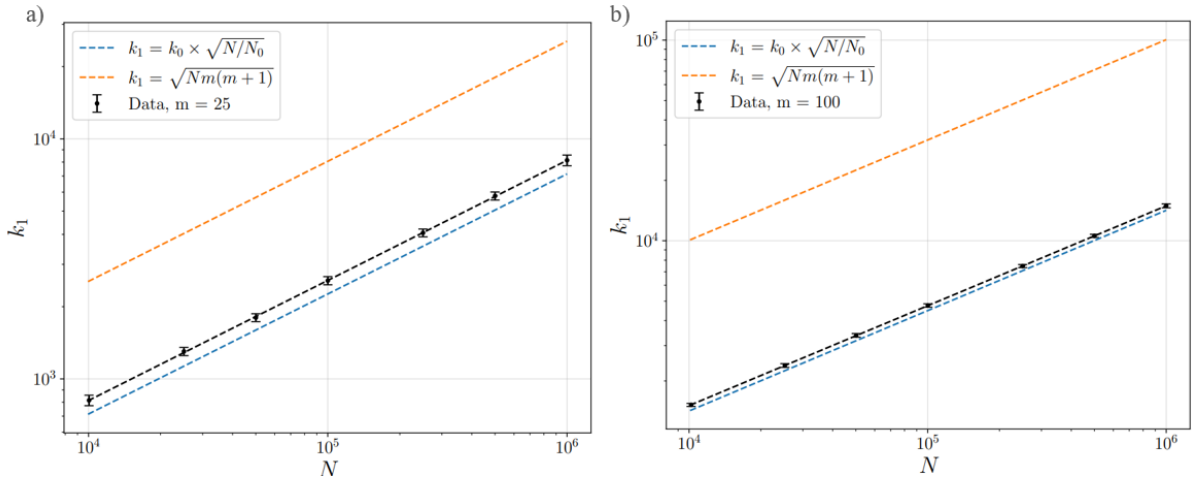

Figure 1.4. The largest degree, for large m, is better approximated by Eq.(1.4.6), even though not quite the same. It grows approximately with $\sqrt{N}$.

1.4.3 Data Collapse

By scaling the x-axis by N^{D_N}, which k_1 depends on, and the y-axis by the theoretical $p_{th}(k)$ data are expected to collapse onto one curve. The results are shown in Fig 1.5, for both models (a) and (b). As expected, data matches perfectly. Furthermore, the collapse highlights the deviations in the tail, i.e. the bump, the steep fall, and, for initialisation (b), the drop, confirming previous suggestions.

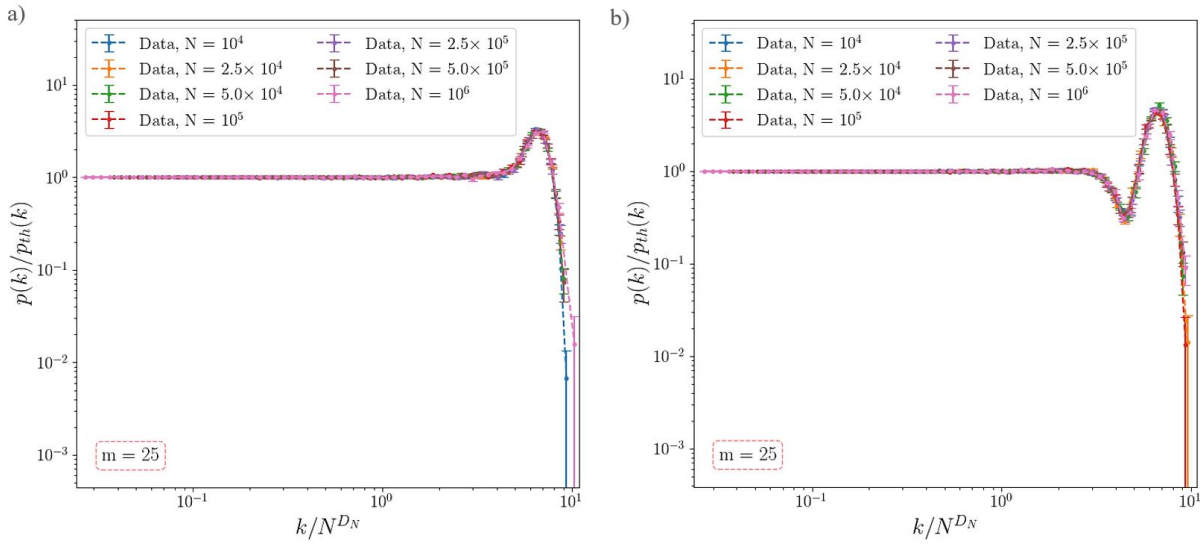

Figure 1.5. Collapse for different N perfectly matches. While the expectation is the straight line $y = 1$, the collapse highlights the deviations in the tail. Both models, (a) above and (b) below, present a clear bump and steep cutoff, caused by the finite size of the system, "cutting" the distribution and generating an "accumulation" of probability. (b) also presents a small drop, caused by the larger degree of the firstly initialised nodes ($2m$) with respect to the following (m). As it can be seen, this enlarges the bump even more.

2 Phase 2: Pure Random Attachment

2.1 Random Attachment Theoretical Derivations

2.1.1 Degree Distribution Theory

With RA, existing nodes have equal probability of being linked to the new one. As in section 1.2, we derive a mean field master equation of form

$$n_k^{t+1} = n_k^t + m\Pi^{\mathrm{t}}_{RA_{k-1}} n_{k-1}^t - m\Pi^t_{RA_k} n_k^t + \delta_{km}. \tag{2.1.1}$$

By substituting Π_{RA} from Eq.(0.3), as $n_k^t = N^t p_k^t$, it yields

$$N^{t+1}p_k^{t+1} - N^t p_k^t = mp_{k-1}^t - mp_k^t + \delta_{km}. \tag{2.1.2}$$

In the long-time limit, this is rearranged into

$$p_k = \frac{m}{1+m} p_{k-1} + \frac{\delta_{km}}{1+m}. \tag{2.1.3}$$

For $k = m$, it gives $p_m = 1/(1+m)$. For $k > m$ we have

$$p_k = \frac{m}{1+m} p_{k-1} = \left(\frac{m}{1+m}\right)^2 p_{k-2} = \cdots = \left(\frac{m}{1+m}\right)^{k-m} p_m, \tag{2.1.4}$$

yielding

$$p_k = \frac{1}{1+m}\left(\frac{m}{1+m}\right)^{k-m}. \tag{2.1.5}$$

As required, $p(k) > 0\ \forall m, k \in \mathbb{N}$ and it is normalised:

$$\sum_{k=m}^{\infty} p_k = \sum_{k=m}^{\infty} \frac{1}{1+m}\left(\frac{m}{1+m}\right)^{k-m} = \frac{1}{1+m}\sum_{k'=0}^{\infty}\left(\frac{m}{1+m}\right)^{k'}$$

$$= \frac{1}{1+m}\frac{1}{1-\frac{m}{1+m}} = \frac{1}{1+m}(1+m) = 1. \tag{2.1.7}$$

The cumulative distribution $F_{k_i}(k)$, found from the related truncated series, is

$$F_{k_i}(k) = P(k_i \leq k) = 1 - \left(\frac{m}{1+m}\right)^{k-m}. \tag{2.1.8}$$

As expected (Barabási 2016), this system is not scale-free, since (WolframAlpha)

$$\langle k^2 \rangle = m(5m+1). \tag{2.1.9}$$

2.1.2 Largest Degree Theory

In the $k_1 \gg 1$ limit, as in section 1.4.1, from the CDF we deduce

$$k_1 = m + \frac{\ln N}{\ln(1+1/m)}. \tag{2.1.10}$$

Alternatively, the average rate at which an existing node increases its degree is

$$\frac{dk_i}{dN} = m\Pi_{RA}(k) = \frac{m}{N(t)}. \tag{2.1.11}$$

Integrating for an initial node, with $N_{i0} = N_0$, yields

$$k_1 = k_0 + m\ln(N/N_0), \tag{2.1.12}$$

where k_0 is the degree initial nodes are initialised with. Both methods have $\ln(N)$ dependence. However, it will turn out that, for large m, the closest solution to simulated data is

$$k_1 = k_0 + m\ln(2N/N_0). \tag{2.1.13}$$

2.2 Random Attachment Numerical Results

2.2.1 Degree Distribution Numerical Result

A simulation was written and tested. As for PA, consistency between features was checked. and the assignment function was tested by analysis of its "static" expansion.

Oppositely to PA, the RA algorithm, as confirmed in Fig 2.1, ignores the initial distribution, so that, all nodes tend to the same degree, with no differences between choices of m.

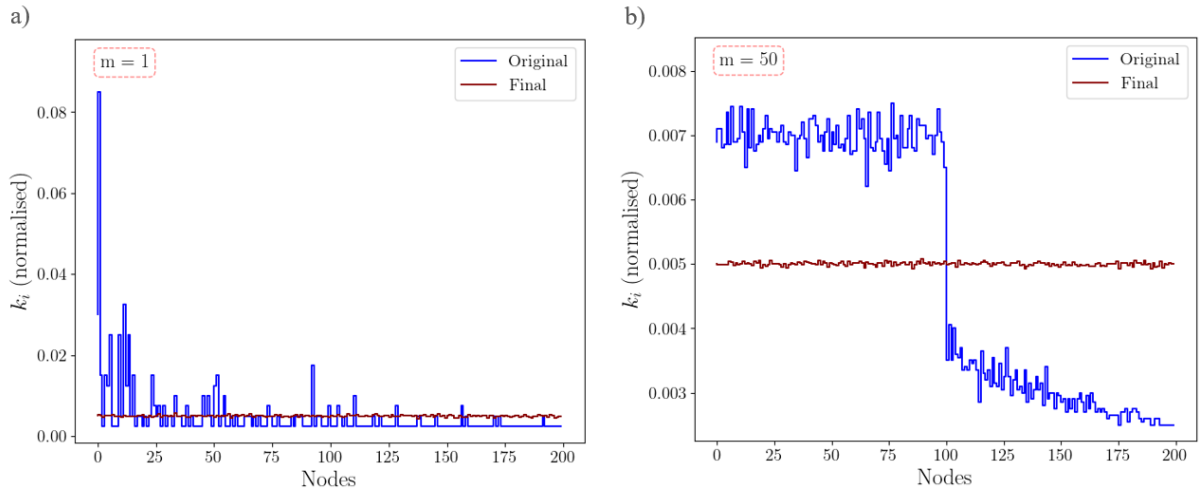

Figure 2.1 RA assignment ignores the initial distribution, with all k_i tending to the same value.

The validity of the theoretical model was tested with same procedure as in section 1.3.2. Data for $N = 10^6$ is plotted in Fig 2.2 for both initial graph models. Data and theory perfectly match but, due to the finiteness of the system, for the tail, where bump, fall and, for (b) drop, occur.

For RA, nodes do not influence each other's growth and variance is finite, thus both χ^2 and KS were performed, with results in Table 2.1.

Both tests confirm the distribution's body perfectly matches the expectations, and both reject the fit for the tail. For the distribution in its whole, the χ^2 test on the flatten $\{k_i\}$ observations strongly rejects the theoretical model, while the KS test doesn't. This is due the different weight given to tail events, in turn caused by the different assumptions behind the statistics.

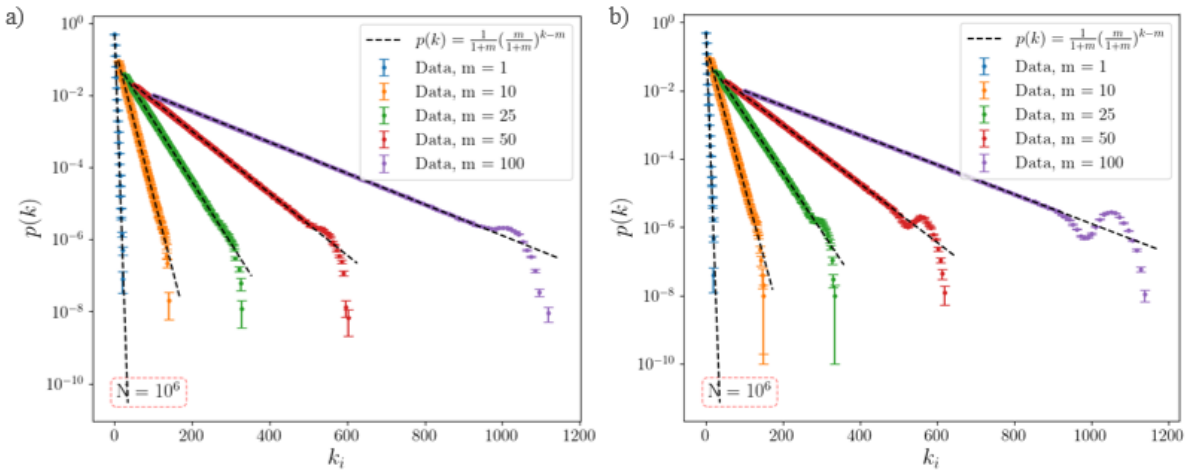

Figure 2.2. Data for both initialisations perfectly match theoretical expectations but for the tail, where the "usual" bump and fall occur. Again, (b) also presents a small drop before the bump.

Table 2.1: Statistical Test of RA model

Test were performed for $m = 1, 10, 25, 50, 100$. The $\chi^2$was applied to the set of all degree values of all realisations, not their mean, to ensure a high enough statistics, with each bin count bigger than 5 (Cochran, 1952). For the same reason, theory and observations were binned. All tests confirm the distribution's body perfectly matches the expectations and the tail totally diverges. However, KS and χ^2 disagree on the distribution as a whole, probably due to the different weight given to observations in the tail.

Statistics \ m	1	10	25	50	100
Body-Tail Cut k_{bt}	16	120	260	480	920
Whole's: χ^2_{red}	0.938	2.004	12.17	49.67	156
p-value	0.495	10^{-5}	10^{-59}	0.0	0.0
Body's: χ^2_{red}	0.737	0.717	0.799	0.912	0.924
p-value	0.682	0.907	0.811	0.605	0.577
Tail's: χ^2_{red}	1.214	10.5	85.73	293	1614
p-value	0.205	10^{-14}	10^{-74}	0.0	0.0
Whole's: D	$1.2 \cdot 10^{-4}$	$5.9 \cdot 10^{-5}$	$6.56 \cdot 10^{-5}$	$7.2 \cdot 10^{-5}$	$7.3 \cdot 10^{-5}$
p-value	0.485	0.995	0.981	0.959	0.954
Body's: D	$1.2 \cdot 10^{-4}$	$5.9 \cdot 10^{-5}$	$6.56 \cdot 10^{-5}$	$7.2 \cdot 10^{-5}$	$7.3 \cdot 10^{-5}$
p-value	0.485	0.995	0.981	0.959	0.999
Tail's: D	0.051	0.136	0.100	0.131	0.249
p-value	0.037	10^{-22}	10^{-43}	10^{-148}	0.0

2.2.2 Largest Degree Numerical Results

Figure 2.3 shows $k_1(N)$, for $m = 1, 10, 50, 100$ for the (a) initialisation model. Interestingly, Eq.(2.1.11) best fits $m = 1$, while for $m = 50$ and 100, the best fit is Eq.(2.1.13), with $m = 10$ in between. Same was obtained for initialisation model (b).

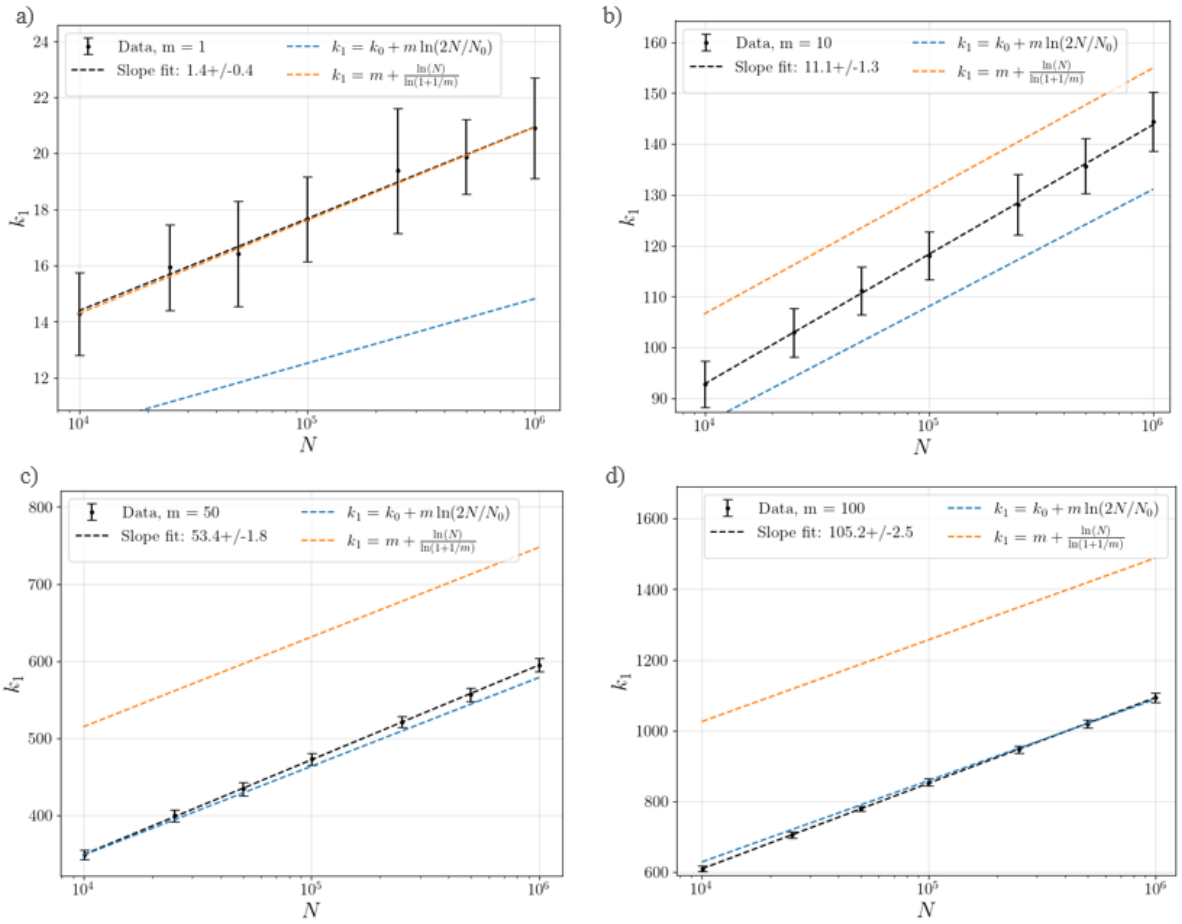

Figure 2.3. $k_1(N)$ is plotted for $m = 1, 10, 50, 100$. It grows approximately with $\ln(N)$, and it varies with m with respect to the proposed fits.

3 Phase 3: Existing Vertices Model

3.1 Theoretical Derivations

In the EX model

- $r < m$ new edges connect the new node with a previously existing one with probability Π_1.
- $m - r$ new edges connect two previously existing nodes with probability Π_2.

In the following section $\Pi_1 = \Pi_{RA}$ and $\Pi_2 = \Pi_{PA}$ is assumed. Multiple connections between edges were allowed due to the complexity of enforcing the otherwise simple graph. The mean field master equation reads

$$n_k^{t+1} = n_k^t + r\Pi^t_{RA_{k-1}} n^t_{k-1} - r\Pi^t_{RA_k} n^t_k + 2(m-r)\Pi^t_{PA_{k-1}} n^t_{k-1} - 2(m-r)\Pi^t_{PA_k} n^t_k + \delta_{kr}. \tag{3.1.1}$$

This does not account for nodes being connected more than once in the same step, which, however, scale as $\sim\Pi^n$, hence were considered negligible in the large N limit studied.

Substituting probabilities, as $n_k^t = N^t p_k^t$, in the long-time limit, it yields

$$p_k = rp_{k-1} - rp_k + \frac{(m-r)(k-1)p_{k-1}}{m} - \frac{(m-r)kp_k}{m} + \delta_{kr}. \tag{3.1.2}$$

For $r < m$, it is rearranged as

$$\left[1 + r + (m-r)\frac{k}{m}\right] p_k = \left[r + \frac{(m-r)(k-1)}{m}\right] p_{k-1} + \delta_{kr}. \tag{3.1.3}$$

For $k > m$ and $r = m/2$, similarly to section 1.2, the solution is

$$p_k = \frac{A(m, m/2)}{(k+m)(k+m+1)(k+m+2)}. \tag{3.1.4}$$

Considering continuity at $k = r$, as $k \geq r$ imposes $p_{k<r} = 0$, we require

$$\frac{A(m, m/2)}{(m+r)(m+r+1)(m+r+2)} = \frac{1}{1 + r + r^2/m}. \tag{3.1.5}$$

As $r = m/2$, it yields

$$A(m, m/2) = \frac{3}{2} m(3m+2), \tag{3.1.6}$$

hence

$$p_k\left(m, \frac{m}{2}\right) = \frac{3}{2} \frac{m(3m+2)}{(k+m)(k+m+1)(k+m+2)}. \tag{3.1.7}$$

For the normalisation, with $\tilde{k} = k + m$ and same procedure as for PA, we have

$$\begin{aligned} \sum_{k=m/2}^{\infty} p_k &= \frac{3}{2} m(3m+2) \sum_{\tilde{k}=3m/2}^{\infty} \left[\frac{1}{2\tilde{k}} - \frac{1}{\tilde{k}+1} + \frac{1}{2(\tilde{k}+2)}\right] \\ &= \frac{3}{2} m(3m+2) \left[\frac{1}{3m} + \frac{1}{3m+2} - \frac{2}{3m+2}\right] = 1, \end{aligned} \tag{3.1.8}$$

as required. Finally, as in previous sections, the CDF is derived as :

$$\begin{aligned} F_{k_i}\left(k; r = \frac{m}{2}\right) = P(k_i \leq k) &= \frac{3}{2} m(3m+2) \sum_{\tilde{k}_i=3m/2}^{k+m} \left[\frac{1}{2\tilde{k}_i} - \frac{1}{\tilde{k}_i+1} + \frac{1}{2(\tilde{k}_i+2)}\right] \\ &= 1 + \frac{3}{2} m(3m+2) \left[-\frac{1}{k+m+1} + \frac{1}{2}\frac{1}{k+m+1} + \frac{1}{2}\frac{1}{k+m+2}\right] \\ &= 1 - \frac{3}{4} \frac{m(3m+2)}{(k+m+1)(k+m+2)}. \end{aligned} \tag{3.1.9}$$

3.2 Numerical results

As for previous models, simulation's features were checked to be consistent with each other, whereas the functions involved in the stubs' assignment had already been tested in previous sections.

Fig. 3.1 shows the behaviour of the EX model, which was studied only for the complete $N_0 = 2m + 1$ initialisation graph. Simulated and expected distributions perfectly match but for large k, where the finite size of the system induces a bump and a steep fall, preceded by a small drop, as in previous sections for this initialisation mode.

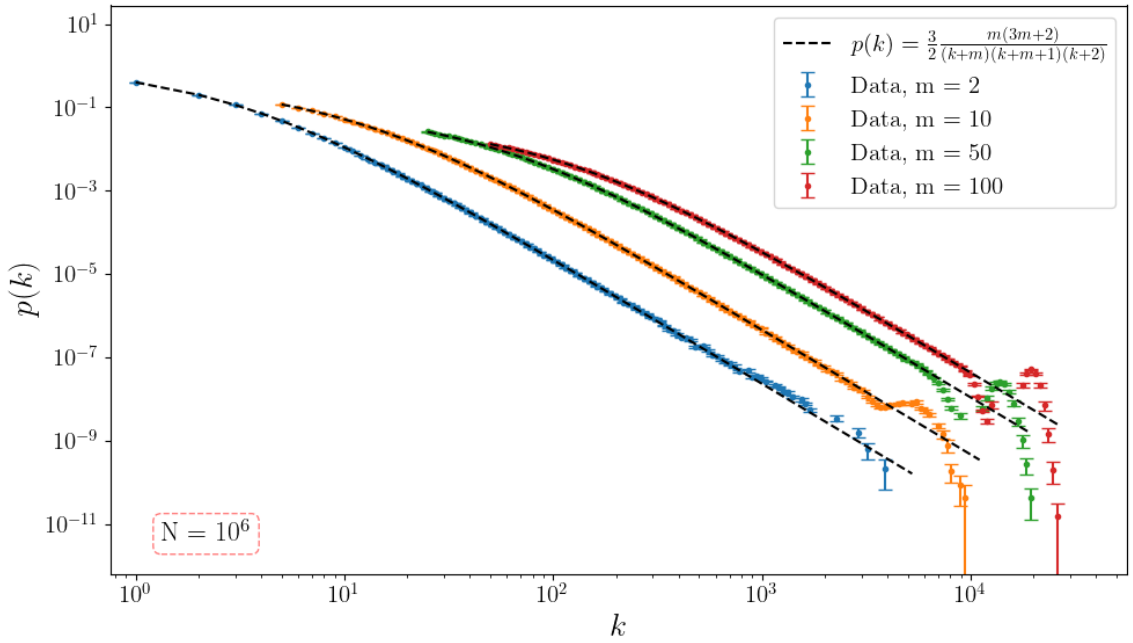

Figure 3.1. As in previous sections, data matches the theoretical model but for the tail of the distribution, where drop, bump and fall occur.

To evaluate the fit, as for PA, the χ^2 test could not be used, oppositely to the KS. Results are shown in Table 3.1. Although the tail of the obtained distribution clearly diverges from theoretical expectations, KS does not reject the model.

Table 3.1: Statistical Validity of EX model

The table shows the results of the statistical tests on $m = 2, 10, 50, 100$. The results of the test highlight, once again, that the distribution's body perfectly matches whereas the tail does not at all.

Statistics \ m	2	10	50	100
Body-Tail Cut k_{bt}	1000	2000	4000	6000
Whole's: D	$1.1 \cdot 10^{-4}$	$1.1 \cdot 10^{-4}$	$9.9 \cdot 10^{-5}$	$1.3 \cdot 10^{-4}$
p-value	0.891	0.919	0.968	0.643
Body's: D	$1.1 \cdot 10^{-4}$	$1.1 \cdot 10^{-4}$	$1.0 \cdot 10^{-4}$	$9.9 \cdot 10^{-5}$
p-value	0.897	0.919	0.962	0.975
Tail's: D	0.24	0.155	0.179	0.214
p-value	0.0	10^{-15}	0.0	0.0

4 Conclusions

In this report, the degree distribution of three versions of the Barabási-Albert network were studied: preferential attachment, random attachment and existing vertices model.

By visual and numerical means, the first and the last were shown being scale free networks, with degree following a quasi-power law, and the largest degree scaling approximately with the square root of the system's size. Oppositely the random attachment graph was shown not being scale free, with its largest degree growing with logarithm of its size.

5 References

Barabási, A.L. 2016. Network Science. Cambridge University Press. Available at http://networksciencebook.com/. Last visited 28/03/2022.

Cochran, W.G. 1952. The χ^2 Test of Goodness of Fit. Ann. Math. Statist. 23 (3). Available at https://doi.org/10.1214/aoms/1177729380. Last visited 28/03/2022.

Evans, T.S.2022. Network Notes. Imperial College London.

Weisstein, E.W. 2002. Gamma Function. From *MathWorld*--A Wolfram Web Resource. Available at: https://mathworld.wolfram.com/GammaFunction.html. Last visited 28/03/2022.

Wolfram Alpha. Available at: https://www.wolframalpha.com/input?i=1%2F%28m%2B1%29+*+sum+%28k%5E2+*+%28m%2F%281%2Bm%29%29%5E%28k-m%29%29+from+k%3Dm+to+infty. Last visited on 28/03/2022.