Network Project
The Barabási-Albert Model

CID: 01888850

28 March 2022

Abstract: Networks offer a useful approach to mathematically describe the dynamics of phenomena encountered in many aspects of daily life, such as the distributions of webpages or citations, or the optimal trajectory for crossing bridges. We examine here one example of a network model, the Barabási-Albert model, using three different attachment styles, preferential, random, and existing vertices. The theoretical distributions for such a model inform our understanding, but rely on unrealistic assumptions such as infinite networks. It is therefore necessary to implement the model numerically to see how observed degree distributions compare to theoretical projections; we find, for example, that preferential attachment leads to a fat-tailed distribution, whose effects are lessened once log-binning is applied. In the case of the random attachment network, the largest degree is confirmed to scale with $ln(N)$ for network size N, whereas for preferential attachment the largest degree k_1 goes as $\sqrt{N}$.

Word Count: 2489 words excluding font page, figure captions, table captions, acknowledgement and bibliography.

Introduction

Networks are a useful tool for studying relationships within complex systems. By reducing interactions in a system to a collection of edges and nodes, e.g. the trajectory taken by airplanes between airports, it is possible to rigorously study the arrangements and operations of the components involved. These components can be as varied as friends at a party or shipping routes carrying freight around the world. In this report, we study the degree distributions of networks exhibiting various attachment styles, and label the degree k of a node by counting the number of edges that come out of it. This can used to model, for example, a network of citations in academia in a way that reflects reality well despite its simple nature.

The first phase of this project describes an implementation of the Barabási-Albert model [1] with preferential attachment. In particular, Phase 1 examines the case where the likelihood of introducing a connection between a new and existing node increases when the latter is of higher degree. That is to say, it reflects the "rich get richer" saying and Pareto principle, both of which suggest that resources accumulate based on existing advantages, or $\Pi_{\text{pa}} \propto k$, where Π_{pa} is the probability of a new node connecting to an existing node of degree k. The degree distribution resulting from such an attachment style is then compared to one for random attachment, a model for which Π_{rnd} depends only on the number of existing nodes, in Phase 2 of the report. Finally, in Phase 3, we consider a model with existing vertices, and compare the theoretical and numerical implications of its attachment style.

1 Phase 1: Pure Preferential Attachment Π_{pa}

1.1 Implementation

1.1.1 Numerical Implementation

A preferential list is used to tract the probability Π_{pa} of a new node connecting to an existing node by storing the node information of such a connection. In this way, it is possible to increase the probability that a given node with larger degree appears more often and thus is picked more often for new connections. As self-loops and multigraphs are not considered in this implementation, a list of already-connected vertices is generated to prevent these.

We begin with a complete graph, that is one for which all nodes in the system are connected with each other. It is initialized to contain m_0 nodes and the time starts running at $t = 0$. The algorithm is then as follows:

1. Initialize a preferential list of nodes containing m instances of each
2. Add a node at time t
3. Generate a list of vertices to which the node is already-connected
4. For a set of m links,
   - 4.1. Select an element of the general preferential list
   - 4.2. If the selection is in the subset of already connected vertices, choose a new element

4.3. Connect selected nodes

4.4. Update preferential list and already-connected vertices list

5. Repeat steps (2)-(4) until some t_f

6. Return a table showing the number and final degree of each node

1.1.2 Initial Graph

We consider complete graphs with $m_0 = m$ fully connected nodes. This choice of initial graph is motivated by the uniform degree distribution such graphs have. This initial distribution mitigates bias in the system from artificial hubs [3] and limits initial condition effects.

1.1.3 Type of Graph

We look at simple graphs, meaning ones with undirected, unweighted edges. This has an advantage over weighted graphs in that it scales better with system size. Although the Barabási-Albert model does not itself prohibit self-loops and multigraphs (where the same pair of nodes can share multiple edges), both are prohibited in this implementation. Thus, the minimum degree k of a node is always m.

1.1.4 Working Code

The code is checked by ensuring the degree distribution goes roughly as the power law, $n(k) \propto k^{-3}$, shown in Fig.1. Additionally, we confirm that the average degree satisfies,

$$\lim_{t\to\infty} \langle k \rangle = \frac{mN(t)}{E(t)} = 2m. \tag{1}$$

We fix $m_0 = 2$, and find $\langle k_a \rangle = 3.9994$ with standard mean error of 0.0639 for 10 samples. This compares well to an expected value of $\langle k_e \rangle = 4.0000$.

Another check is conducted to check the number of nodes $N = t + m_0$ by comparing the cardinality of the unique elements of N, to the sum of $m_0 = 2$ and $t = 100$. Both equal 102.

1.1.5 Parameters

The parameters needed are:

- m, the number of nodes added at time t_i, where $t_i \in (t_0, t_f)$
- m_0, the initial number of nodes (in our case, $m_0 = m$ is used)
- t, the time
- t_f, the cutoff time for a finite system

1.2 Preferential Attachment Degree Distribution Theory

1.2.1 Theoretical Derivation

We begin with the master equation for a network of any attachment style [2],

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

where Π is the attachment probability, $N(k,t)$ is the number of nodes of degree k at time t, and $\delta_{k,m}$ is a Dirac delta function that accounts for the new node added at $k = m$. We can convince ourselves this equation makes sense by noting the left hand side describes the number of nodes of degree k at one step after time t, and the right hand side accounts for how the quantity of nodes of degree k at time t have degrees changes with given probability.

In the case of preferential attachment, the probability of a new node attaching to an existing node depends on the degree of the latter. This is normalized with respect to system size,

$$\Pi_{\text{pa}}(k) \approx \frac{k}{2mN(t)}, \tag{3}$$

and $p_k(t)$, the probability of picking a node of degree k, is given by,

$$p_k(t) = \frac{N(k,t)}{N(t)}, \tag{4}$$

where N(t) is the total number of nodes at time t.

We can combine the master equation with the expressions found in Eqs. (3) and (4),

$$(N+1)p_k(t+1) = \begin{cases} Np_k(t) + \frac{k-1}{2}p_{k-1}(t) - \frac{k}{2}p_k(t), & \text{if } k > m \\ Np_m(t) + 1 - \frac{m}{2}p_m(t), & \text{if } k = m \\ 0, & \text{if } k < m \end{cases}, \tag{5}$$

Note $k < m$ is not allowed by construction. Working in the limit where $t \to \infty$, we assume additional incremental increases in t are diminishing and can be neglected, and so $\lim_{t\to\infty} p_k(t+1) = \lim_{t\to\infty} p_k(t) = p_k(\infty) = p_k$. This allows us to simplify Eq.(5) for positive t,

$$p_k = \begin{cases} \frac{k-1}{k+2}p_{k-1}, & \text{if } k > m \\ \frac{2}{m+2}, & \text{if } k = m \end{cases} \tag{6}$$

Next, we develop a recursive relation by combining Eqs.(5) and (6):

$$\begin{aligned} p_{m+1} &= \left[\frac{m}{m+3}\right][p_m] = \left(\frac{m}{m+3}\right)\left(\frac{2}{m+2}\right) = \frac{2m}{(m+2)(m+3)}, \\ p_{m+2} &= \left[\frac{m+1}{m+4}\right][p_{m+1}] = \left(\frac{m+1}{m+4}\right)\left(\frac{2m}{(m+2)(m+3)}\right) = \frac{2m(m+1)}{k(k+1)(k+2)}, \\ p_{m+3} &= \left[\frac{m+2}{m+5}\right][p_{m+2}] = \left(\frac{m+2}{m+5}\right)\left(\frac{2m(m+1)}{(m+2)(m+3)(m+4)}\right) \\ &= \frac{2m(m+1)}{(m+3)(m+4)(m+5)} = \frac{2m(m+1)}{k(k+1)(k+2)}, \end{aligned} \tag{7}$$

implying the theoretical degree distribution for preferential attachment is given by,

$$p_k\left(\Pi_{\text{pa}}\right) = \frac{2m(m+1)}{k(k+1)(k+2)} \text{ for } k \geq m. \tag{8}$$

1.2.2 Theoretical Checks

We can confirm the approximate theoretical solution for $p(k)$ behaves as expected by verifying (a) the probability is normalized, (b) the continuous limit for sufficiently large k exhibits power law behavior, and that it is consistent with Eq.(1).

(a) We consider a sum over k of Eq.(8), factor out the relevant m terms, and decompose the sum using partial fractions,

$$\begin{aligned}\sum_{k=m}^{\infty} p_k\left(\Pi_{\mathrm{pa}}\right) &= \sum_{k=m}^{\infty} \frac{2m(m+1)}{k(k+1)(k+2)} = m(m+1)\sum_{k=m}^{\infty} \frac{2}{k(k+1)(k+2)} \\ &= m(m+1)\sum_{k=m}^{\infty} -\frac{2}{k+1} + \frac{1}{k+2} + \frac{1}{k}, \\ &= m(m+1)\left[\sum_{k=m}^{\infty}\left(\frac{1}{k} - \frac{1}{k+1}\right) - \sum_{k=m}^{\infty}\left(\frac{1}{k+1} - \frac{1}{k+2}\right)\right],\end{aligned} \tag{9}$$

and simplified where terms in the first and second component sums cancel,

$$\begin{aligned}\sum_{k=m}^{\infty} p_k &= m(m+1)\left[\frac{1}{m} - \frac{1}{m+1}\right] = m(m+1)\frac{(m+1)-m}{m(m+1)} \\ &= m+1-m = 1.\end{aligned} \tag{10}$$

Thus we find $\sum_{k=m}^{\infty} p_k = 1$, as desired.

(b) Now we work in the continuous limit where $t \to \infty$ and $k \to \infty$. We have,

$$\lim_{k\to\infty} p_k\left(\Pi_{pa}\right) = \lim_{k\to\infty} \frac{2m(m+1)}{k(k+1)(k+2)} \propto \frac{1}{k^{-3}}, \text{ for } k \gg m. \tag{11}$$

This can be verified visually in Fig.(1).

(c) We consider the average degree $\langle k\rangle$ and recall the number of edges should satisfy $E = mN$. Once the simulation has been run, that quantity should reach $E_{tot} = 2mN = 2E$, resulting in $\langle k\rangle = 2m$. We can perform this check using the expression in Eq.(8),

$$\begin{aligned}\langle k\rangle &= \sum_{k=m}^{\infty} p_k(k) = \sum_{k=m}^{\infty}\left[\frac{2m(m+1)}{k(k+1)(k+2)}(k)\right] \\ &= m(m+1)\sum_{k=m}^{\infty}\left[\frac{2}{k+1} - \frac{2}{k+2}\right] \\ &= m(m+1)(\frac{2}{m+1}) = 2m.\end{aligned} \tag{12}$$

Therefore, $\langle k\rangle = 2m$ as required, and our theoretical check is consistent with Sec.(1.1.4).

1.3 Preferential Attachment Degree Distribution Numerics

1.3.1 Fat-Tail

The main method used for dealing with fat-tail distributions was log-binning. Fat-tailed distributions are those for which there is a significant probability of finding large values k

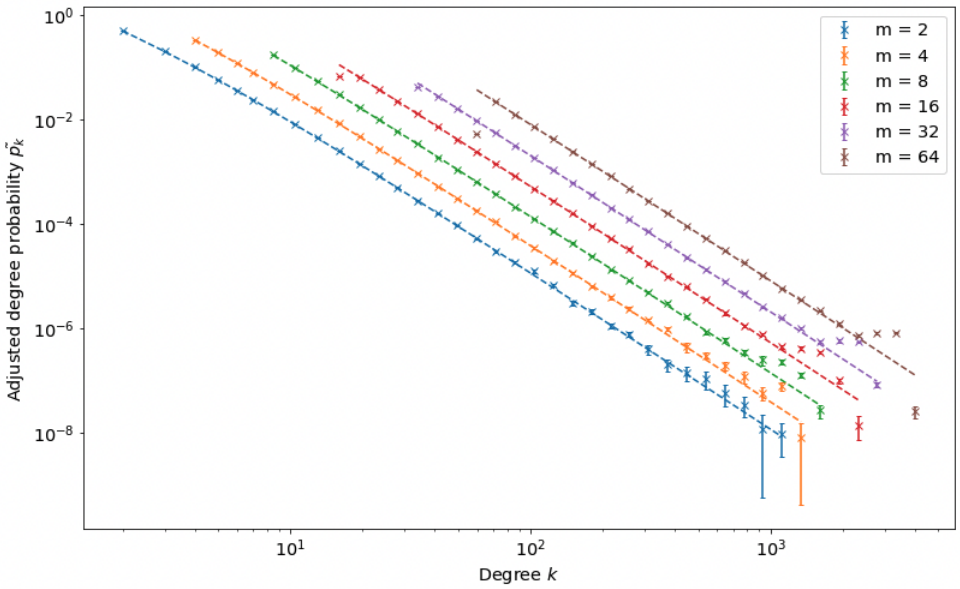

Figure 1: A log-binned degree probability distribution $\tilde{p}_k$ is plotted against degree k for several values of system-size m. The theoretical dashed lines, which track Eq. (10) for each system show close correspondence with measured data points, plotted as crosses of matching color. The behavior is less promising toward the tail-end of the distribution, where large error bars indicate it is not due to random error, but rather attributable to numerical results approximating the infinite limit of theory.

[2]. This means that the probability, p_k, of the fat-tail has a decay slower than exponential, which results in a "fat tail" for large values of k. Because that part of the distribution does not have enough data scattered throughout, it is not possible to gauge the true underlying behavior, and so the little data there is spans orders of magnitude in k.

Log-binning is useful for extracting a clearer sense of the true behavior because the bins generated scale exponentially in k and so the data is separated into bins divided by bin width. This can be seen from the relationship between lower and upper bounds of each bin, given by b_{i-1} and b_{i+1} respectively [2]:

$$\frac{b_{i+1}}{b_i} = \exp\{\Delta\} \text{ where } \Delta > 0. \tag{13}$$

In the case of preferential attachment, we use the value $\Delta = 1.2$ to make the plot shown in Fig.(2) sufficiently smooth and still maintain the separations necessary for it to be useful.

1.3.2 Numerical Results

The numerical results shown in Fig.(1) use data for several values of m, namely $m = 2, 4, 8 \ldots, 128$. The log-binned degree probability distribution p_k is plotted against degree k and uses data from 100 runs of the model. As for any probabilistic process, there will be statistical variance t o account for. By using larger data sets, we can better gauge the true, underlying distribution hidden by noise in smaller runs. This allows for a comprehensive assessment of the standard deviations for each m. The standard deviations are normalized by the standard error of the mean and represented on the plot in the form

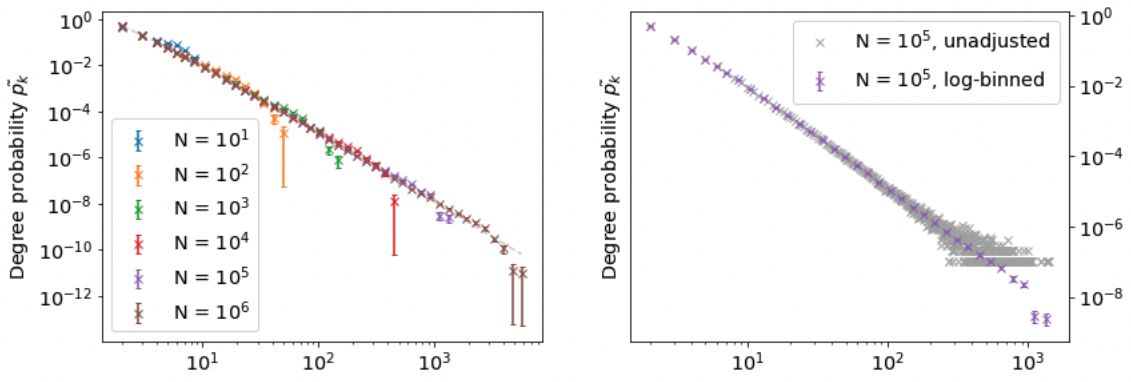

Figure 2: **Left.** Degree distribution of $m = 2$ for different N of nodes added. The theoretical distribution given in Eq. (10) is tracked by the grey line and shows good correspondence. **Right.** Plot shows the benefits of applying log-binning to a fat-tail distribution.

of error bars. These are an objective assessment of how well the numerical results reflect theoretically-predicted distributions, and we find good correspondence once log-binning is instituted.

Indeed, the numerical results only seem to exhibit large error around k^3, or very small p_k. This can be attributed to finite size system effects, which cut off nodes that in an infinite limit would accumulate a higher degree count. This corresponds to higher information loss and is more likely to happen for large k, but once understood can help explain deviations from theoretical results.

Lastly, though steps are taken to minimize the initial condition effects (addressed in Sec. (1.1.2)), these are entirely removed. Because we work with a finite system, the effects are more pronounced, since there is never enough time for the incremental changes to reduce initial effects to negligible levels.

1.3.3 Statistics

We assess goodness of fit using a Chi-Squared test by assuming the observed degree values can be treated as a normal random variable,

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i}. \tag{14}$$

This relates the observed degree count O_i to the expected quantity E_i from theoretical calculations. We fix $m = 2$ such that we are in a region where initial condition effects are not as pronounced, and use 10 bins for significance. Using our data, we find $\chi^2 = 20.3$, which corresponds to a p-value of greater than 0.05, and thus is statistically significant. In the case of this test, the null hypothesis was that the observed value fits the expected, and so we would not want to reject it.

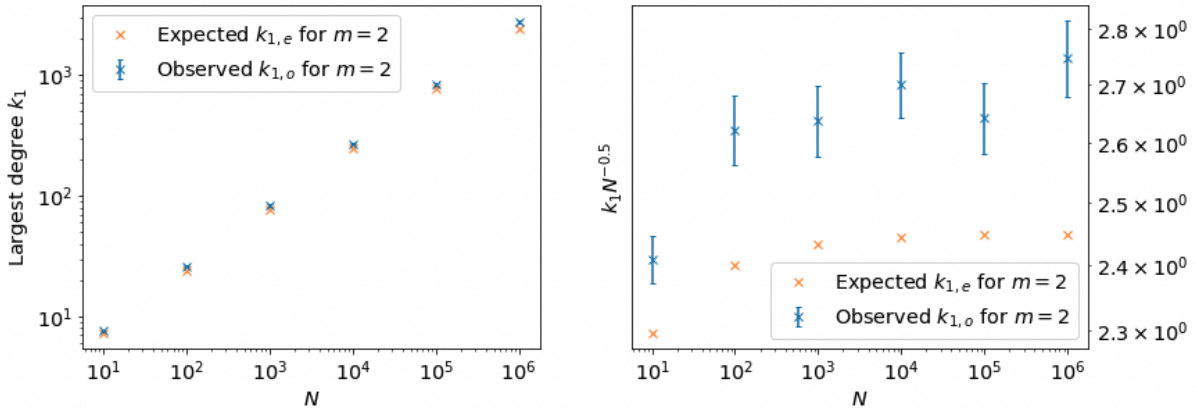

Figure 3: **Left.** Plot of largest degree distribution, k_1, against N given to highlight good agreement. **Right.** Plot shows ratio of largest degree and leading order N to draw attention to the scaled differences.

1.4 Preferential Attachment Largest Degree and Data Collapse

1.4.1 Largest Degree Theory

We next examine the theoretical dependence of the largest expected degree, k_1, on the number of vertices N in a finite size system. We consider the sum $\sum_{k=k_1}^{\infty} p_k$ and note exactly one node will have the largest degree, implying the sum should equal $\frac{1}{N}$,

$$\frac{1}{N} = \sum_{k=k_1}^{\infty} p_{k_1} = m(m+1)\left[\sum_{k=k_1}^{\infty}\left(\frac{1}{k} - \frac{1}{k+1}\right) - \sum_{k=k_1}^{\infty}\left(\frac{1}{k+1} - \frac{1}{k+2}\right)\right], \tag{15}$$

which can be simplified where the second terms cancel in the infinite limit of the sums,

$$\begin{aligned} \frac{1}{N} &= m(m+1)\left[\frac{1}{k_1} - \frac{1}{k_1+1}\right] \\ &\Rightarrow k_1^2 + k_1 - Nm(m+1) = 0. \end{aligned} \tag{16}$$

This quadratic is solved for k_1 to mathematically describe the dependence of largest expected degree on number of vertices,

$$k_1 = -\frac{1}{2} \pm \frac{\sqrt{1+4Nm(m+1)}}{2}. \tag{17}$$

Since we are dealing with a real system, the negative solution is discarded and we are left with,

$$k_1 = -\frac{1}{2} + \frac{\sqrt{1+4Nm(m+1)}}{2} \text{ for } N \gg 1. \tag{18}$$

1.4.2 Numerical Results for Largest Degree

We pick $m = 2$ so that initial condition effects can be reduced using a reasonably small network (a smaller value would not work as the network would either not grow or be a tree). Using $N = 10^1, 10^2, \ldots, 10^6$, we plot the maximum degree distribution shown in

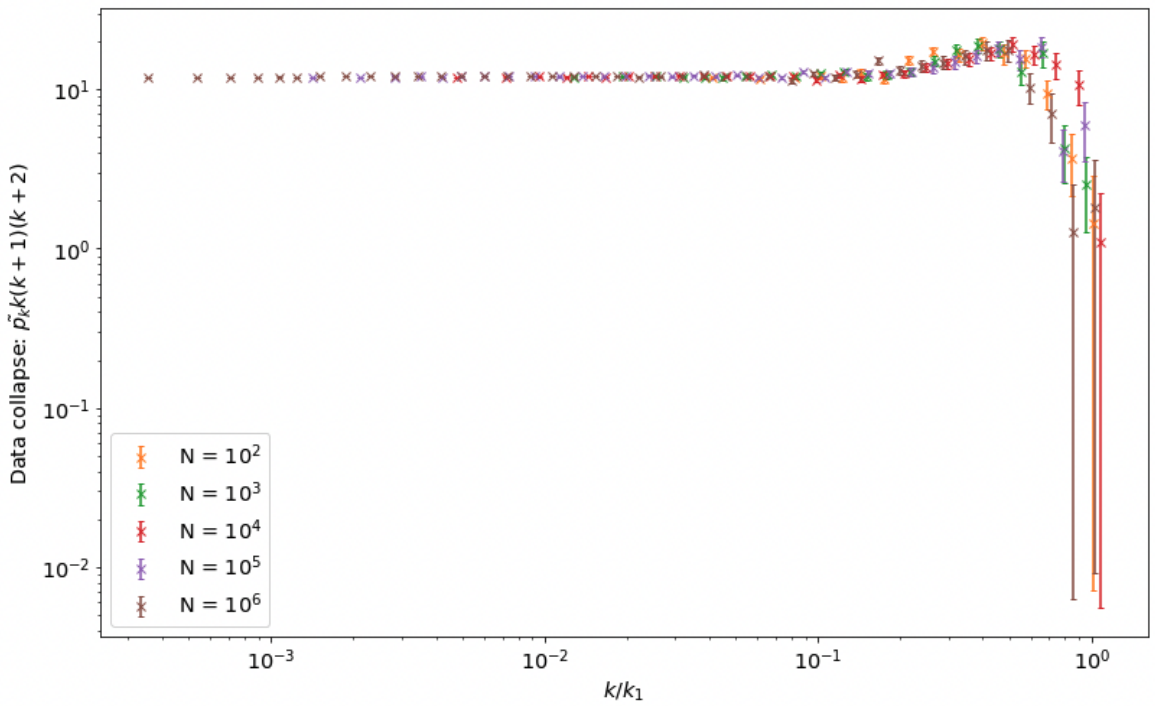

Figure 4: Data collapse for the system is achieved for $N = 10^2, 10^3, \ldots, 10^6$ using a reasonable fixed value $m = 2$. Errors are computed from 100 runs and error bars generated as using the standard mean error.

Fig.(2). We expect to see $k_1 \propto N^{0.5}$ for $N \gg 1$ following our expression for k_1 in Eq. (22). We can compute the slope using a linear relationship,

$$\log k_1 = (0.505 \pm 0.002) \log N. \tag{19}$$

This is a deviation of roughly 1% from expectation. If we divide the leading order term by k_1, we can calculate the difference between predicted and observed k_1, which we expect to be approximately constant.

1.4.3 Data Collapse

If we take the maximum degree k_1 as a normalization for the system we find evidence of finite size effects in the tail of the distribution. Specifically, we look at the ratio $\frac{k}{k_1}$ to align the distributions and multiply p_k by k such that each distribution height is a function exclusively of m. This manipulation of Eq.(8) highlights the complex nature of the system.

We observe the data collapse results in a constant function until a bump followed by a decay. During this decay, the error bars are very large, whereas for the majority of the function prior to that point, errors remain small. The bump is due to finite system size effects (which were minimized with the choice of initial conditions, but cannot be removed entirely). We note this occurs around the point where errors increase for k_1, and infer the reason is errors scaling roughly obeys data scaling. In this way, an understanding of the maximum degree scale provides valuable insight for the behavior of the data collapse.

2 Phase 2: Pure Random Attachment Π_{rnd}

2.1 Random Attachment Theoretical Derivations

2.1.1 Degree Distribution Theory

We return to the master equation described in Eq.(2), this time to study random attachment. Now, the probability a new node will attach to an existing node is given by,

$$\Pi_{\text{rnd}} = \frac{1}{N(t)}. \tag{20}$$

This implies any existing node, regardless of its degree k, is as likely a candidate for the new node to attach to as any other. Using this new probability in combination with Eq.(2) and Eq.(4) leads to,

$$(N+1)p_k(t+1) = \begin{cases} Np_k(t) + mp_{k-1}(t) - mp_k(t), & \text{if } k > m \\ Np_m(t) - mp_m(t) + 1, & \text{if } k = m \end{cases}, \tag{21}$$

where $k < m$ is again forbidden by construction.

We work in the limit where $t \to \infty$ allowing us to use the notation $p_k(\infty) = p_k$ since for sufficiently large t the incremental changes diminish. Then we can rewrite Eq.(19) for $k \geq m$,

$$p_k = \begin{cases} \frac{1}{m+1}, & \text{if } k > m \\ \frac{m}{m+1}p_{k-1}, & \text{if } k = m, \end{cases} \tag{22}$$

and develop a recursive relation similarly to how it was done in Eq.(7),

$$\begin{aligned} p_{m+1} &= \left[\frac{m}{m+1}\right] p_m = \left(\frac{m}{m+1}\right)^1 \frac{1}{(m+1)} \\ p_{m+2} &= \left[\frac{m}{m+1}\right] p_{m+1} = \left(\frac{m}{m+1}\right)^2 \frac{1}{(m+1)}, \\ p_{m+3} &= \left[\frac{m}{m+1}\right] p_{m+2} = \left(\frac{m}{m+1}\right)^3 \frac{1}{(m+1)}, \end{aligned} \tag{23}$$

and expressing these in the general form,

$$p_k\left(\Pi_{\text{rnd}}\right) = \left(\frac{m}{m+1}\right)^{k-m} \frac{1}{m+1} \text{ for } k \geq m. \tag{24}$$

This recursive relation describes the theoretical degree distribution for random attachment. We note there is a more pronounced dependence on m as compared to the case of the preferential attachment, and this makes sense given Eq.(20) has no $N(k,t)$ dependence where Eq.(3) did.

We can perform a simple check to ensure the result is reasonable. In particular, we consider Eq.(24) summed over k and rearrange terms,

$$\sum_{k=m}^{\infty} p_k\left(\Pi_{\text{rnd}}\right) = \sum_{k=m}^{\infty} \left[\left(\frac{m}{m+1}\right)^{k-m} \frac{1}{m+1}\right] = \sum_{k=m}^{\infty} \frac{m^{k-m}}{(m+1)^{k-m+1}}, \tag{25}$$

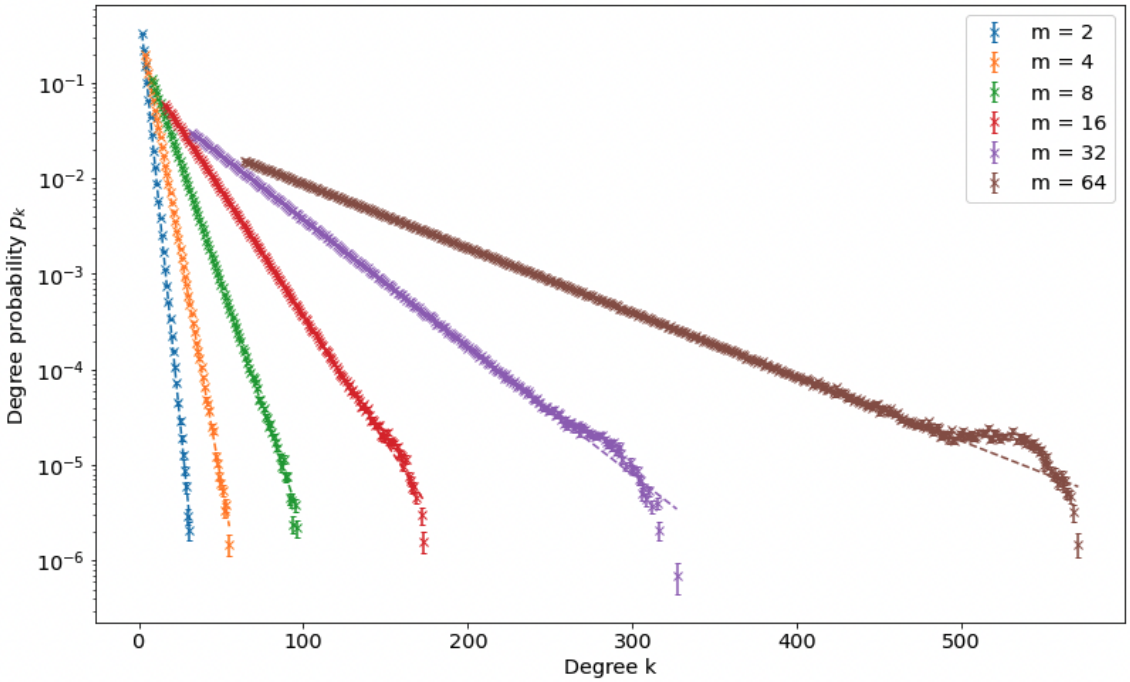

Figure 5: The logbinned-degree probability distribution $\tilde{p}_k$ is plotted against degree k for the random attachment run of the Barabási-Albert model. The dashed lines track the theoretical result of Eq. (31) and the measured data is graphed as crosses of matching color.

This can be simplified,

$$\begin{aligned}\sum_{k=m}^{\infty} p_k &= \sum_{k=m}^{\infty} \frac{m^{k-m}}{(1+m)^{1+k-m}} = \sum_{k=m}^{\infty} \frac{m^k m^{-m}}{(m+1)^k (m+1)^{-m+1}} \\ &= \frac{m^{-m}}{(m+1)^{-m+1}} \sum_{k=m}^{\infty} \left[\frac{m}{(m+1)}\right]^k .\end{aligned} \tag{26}$$

We relabel the index to $k = i + m$, and identify the last term as a geometric sum. Using its series representation, $\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$, we write,

$$\begin{aligned}\sum_{k=m}^{\infty} p_k &= \frac{m^{-m}}{(m+1)^{-m+1}} \left(\frac{m}{m+1}\right)^m \sum_{i=0}^{\infty} \left(\frac{m}{m+1}\right)^i \\ &= \frac{m^{-m}(m+1)^m}{(m+1)^{-m+1}} \left(\frac{m}{m+1}\right)^m (m+1) \\ &= m^{-m} \left(\frac{m}{m+1}\right)^m (m+1)^m \\ &= \frac{m^{-m} m^m (m+1)^m}{(m+1)^m} = 1.\end{aligned} \tag{27}$$

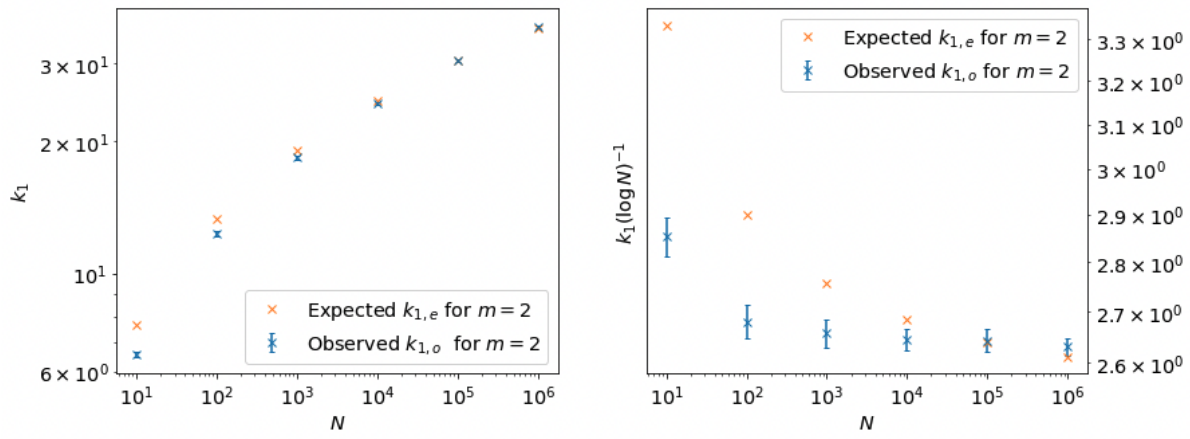

Figure 6: Visual results described in Eq.(35) are here shown for expected $k_{1,e}$ and observed $k_{1,o}$ largest degree values. The latter tends to the former for large values of N, but the approximation is not as good for smaller N.

2.1.2 Largest Degree Theory

We next examine the theoretical dependence of the largest expected degree, k_1, on the number of vertices N in a finite size system. We consider the sum $\sum_{k=k_1}^{\infty} p_k$ and note some node must have the largest degree, meaning the sum should equal $\frac{1}{N}$,

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

we identify, as in Sec. (2.1.1), a geometric sum, and use the same procedure as before,

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

This can be solved for k_1 by taking the logarithm of both sides and rearranging,

$$k_1 = \frac{\log\left|\left(\frac{m}{m+1}\right)^m \frac{1}{N}\right|}{\log\left|\frac{m}{m+1}\right|}. \tag{30}$$

Eq.(30) gives the relationship between the number of vertices N and the largest expected degree, k_1, for the case of random attachment.

2.2 Random Attachment Numerical Results

2.2.1 Degree Distribution Numerical Results

The numerical results here reflect a similar pattern to those of Sec.(1.3.2), namely, they agree with theoretical degree probability distributions until a point determined by system size. After that point, we observe finite system size effects. Errors were calculated the same as in Sec.(1.3.2), allowing us to compare results to those of the run for preferential attachment. These can also be seen in Fig.(5), where p_k is plotted against k for $N = 100,000$ and $m = 2, 4, 8, \ldots, 128$.

2.2.2 Largest Degree Numerical Results

Numerical results for largest degree k_1 are consistent with theoretical results, as can be seen in Fig. (5), which plots the ratio of Eq.(30) over N. This allows us to visually assess the correspondence between expected $k_{1,e}$ and observed $k_{1,o}$,

$$\lim_{N\to\infty} \frac{k_{1,o}}{k_{1,e}} = 0. \tag{31}$$

3 Phase 3: Existing Vertices Model

3.1 Existing Vertices Model Theoretical Derivations

We now investigate an attachment where, for each time step, half the new edges attach to a new vertex and half to both ends of an existing vertex. We choose, for simplicity, to make Π_{ra} the probability of the first connection for r edges and Π_{pa} the probability of the second for $r-m$ edges. We choose $r = m/2$. Returning to the master equation fisrt seen in Eq.(2), we now have,

$$\Pi_{\text{ev}} = r\Pi_{\text{ra}}(k) + (r-m)\Pi_{\text{pa}} = r\frac{k}{2mN(t)} + (r-m)\frac{1}{N(t)} \tag{32}$$

Then we borrow from the results of Eqs. (8) and (24) and note the remaining $m-r$ edges connect both ends to existing vertices, so for $k \geq m$,

$$\frac{m}{2}\left[p_k\left(\Pi_{\text{pa}}\right) + p_k\left(\Pi_{\text{rnd}}\right)\right] = \frac{m^2(m+1)}{k(k+1)(k+2)} + \left(\frac{m}{m+1}\right)^{k-m}\frac{m}{m+1}. \tag{33}$$

We simplify the second term to find,

$$p_k\left(\Pi_{\text{ev}}\right) = \frac{m^2(m+1)}{k(k+1)(k+2)} + \left(\frac{m}{m+1}\right)^{k-m+1} \text{ for } k \geq m. \tag{34}$$

3.2 Existing Vertices Model Numerical Results

The results here produce a mix of the preferential and random attachments found in previous phases. We fix $N = 10000$ and use $m = 2$ and $m = 4$. The $p(k)$ follows a fat-tail distribution similar to its behavior for preferential attachment. Again, there is a cut-off in degree k after which the theoretical distribution is not followed closely by the numerical results.

Given that in the real world, we cannot expect networks to neatly follow exactly one attachment style, it is useful to consider combinations of the two for a system. This will reflect numeric problems found in either of the system, and will not be as straightforward, but will better inform our understanding of a given network.

4 Conclusions

The Barabási-Albert model was implemented using three different attachment styles: preferential, random, and existing vertices. We find good correspondence between theoretical descriptions of degree distributions $p(k)$ for small values of k, however there was a clear

cutoff in the case of preferential attachment after which finite networks show large deviation from theoretical results for infinite networks. This was not so much the case for the random attachment mechanism, where the deviation for a finite system size approximation was not as pronounced. Finally, the largest degree scaled as $\sqrt{N}$ for preferential attachment and as $log(N)$ for random attachment.

References

[1] A. Barabási and R. Albert. "Emergence of scaling in random networks", *Science*. 1999.

[2] T.S. Evans, *Network Notes*, March 2022.

[3] M. Pósfai. "Network science: The Barabási-Albert model.