The Barabási-Albert model

MATEFY Adam

August 2, 2024

1 Introduction

We have seen in Chapter 10 of the lecture notes [1], that the Gilbert-Erdős-Rényi model ($G(n,p)$ model) is a good model for some applications and has many good properties (such as the fact that many of its properties can be calculated explicitly), but we also saw that this model also has some shortcomings, such as the fact that the distribution of degrees follows a Poisson distribution for large n, while many real life networks have power law degree distribution (such networks are called *scale-free networks*). To this end, in this report a different random graph model, the Barabási-Albert model (BA model), will be presented.

The two main improvements introduced in the BA model are *growth* and *preferential attachment*. While the $G(n,p)$ model has a fixed number of vertices (n), the number of vertices of many real life networks increases. For example, the World Wide Web (WWW) had only one node in 1991, the first webpage (created by Tim Berners-Lee, the creator of WWW), but today there are $\sim 10^{12}$ documents in the Web [2]. Also, the numbers of users in Social Networking Services (SNS), which can be represented by vertices, also increases over time as more and more people register to these services. Hence, we should somehow incorporate in our model that the number of nodes increases. Also, we can observe that celebrities with millions of connections are much more likely to gain more connections than the neighbour Timmy with his 2 followers. Therefore, we should somehow include in our model that new vertices added to the graph are more likely to connect to vertices with bigger degree.

2 The Barabási-Albert model

The simplest model that incorporates these two characteristics (growth and preferential attachment) is the Barabási-Albert model, also referred to as BA model or scale-free model, and is constructed as follows.

Let $G = (V, E)$ be a simple undirected graph with $|G| = m_0$ vertices with $\deg(x) \geq 1\ \forall x \in V$, and let $m \in \mathbb{N}^+$ with $m \leq m_0$. Then, in each step of the iteration we add one new vertex and m new edges to G such that one of the endpoints of each new edges is the new vertex and the other endpoint is selected from the already existing vertices. The probability that a new edge connects to the already existing vertex i is given by

$$P_i = \frac{k_i}{\sum_j k_j} \tag{1}$$

where k_j is the degree of vertex j.

This construction leaves the following question open. In a real life network we usually do not have multiple edges (for example, you can follow someone in SNS only once), but this construction allows the new node to connect to the same node multiple times. So, how do we deal with this? As we will see in Section 3, this situation becomes negligible for huge systems.

I created a Python code which creates a BA graph, the script and its explanation can be found in the Additional resources section (Section 9). A simulation can be seen in Figure 1.

It is easy to see that after t steps the graph has $N(t) = m_0 + t$ nodes and $E(t) = e_0 + mt$ edges (where $e_0 \equiv ||G(t = 0)||$ is the initial number of edges). We can also observe that many nodes have small degrees, but some of them starts to accumulate more and more connections and turns

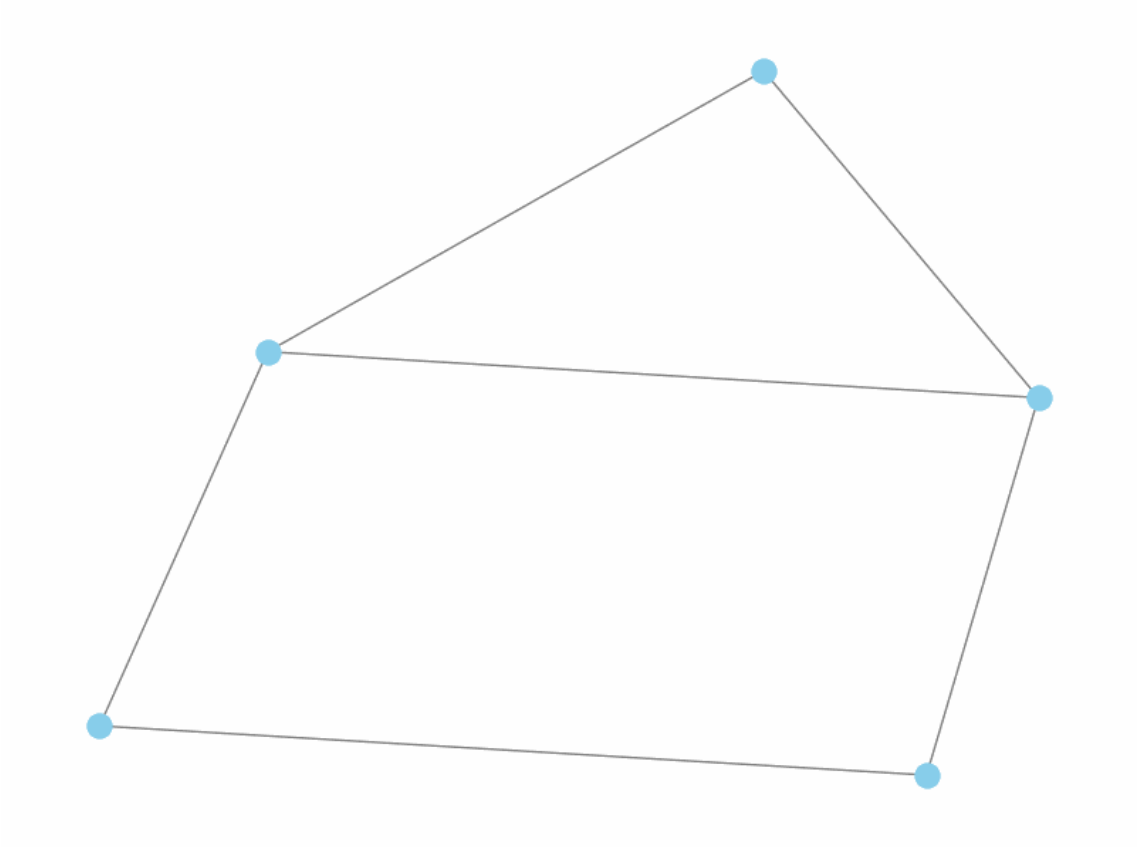


Figure 1: Evolution of the graph over time

into *hubs*. Then, because of preferential attachment, these hubs will gain even more connections while nodes with low degrees will likely stay having low degrees. This is called the *rich-gets-richer phenomenon*, and is a direct consequence of how the graph is constructed.

Finally, Figure 2 shows the degree distribution of a graph with $n = 100,000$ nodes and $m = 3$. We can observe that the degree distribution follows a power law, just like many real systems, which indicates that this is a promising model. In the following sections, we will also derive this power law.

3 Degree dynamics

Let us approximate the degree k_i of node i as a real continuous random variable. Equation (1) tells us the probability of a new node connecting to node i. Since there are m new edges added with each iteration, the time evolution of k_i is given by

$$\frac{\mathrm{d}k_i}{\mathrm{d}t} = mP_i = \frac{mk_i}{\sum_j^{N-1} k_j} \tag{2}$$

where N is the number of nodes after step t (we have the -1 because we cannot connect to the new (Nth) node). Since each edge has 2 contributions to the total number of edges, the sum of degrees is $2 \times number\ of\ edges$). But, we do not count the degree of the new node. Since there are $e_0 + mt$ edges after step t and the new node has degree m,

$$\sum_j^{N-1} k_j = 2(e_0 + mt) - m = 2e_0 + m(2t - 1).$$

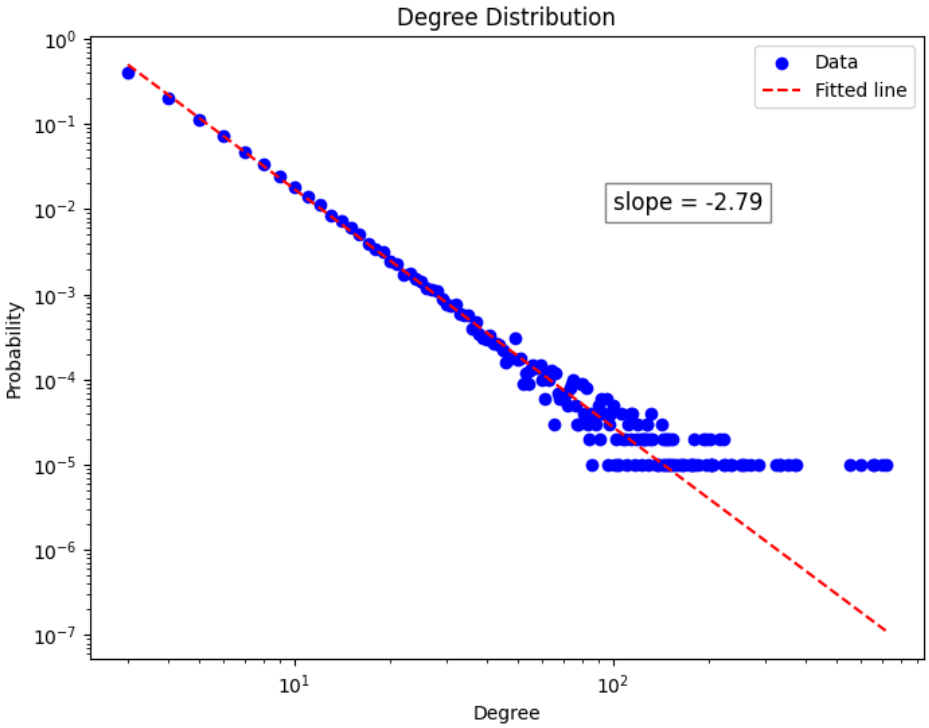


Figure 2: Degree distribution of a BA graph with $N = 100,000$ nodes on a log-log scale. A line was fitted for $k < 40$ and we found that the slope is -2.79, which indicates a power law relation between k and p_k.

Hence, Equation (2) becomes

$$\frac{\mathrm{d}k_i}{\mathrm{d}t} = \frac{mk_i}{2e_0 + m(2t-1)} = \frac{k_i}{\frac{2e_0}{m} + 2t - 1}.$$

Since m and e_0 are constant, $\frac{2e_0}{m}$ and 1 are negligible as t gets large, so we get

$$\begin{aligned} & \frac{\mathrm{d}k_i}{\mathrm{d}t} = \frac{k_i}{2t} \\ \iff & \frac{\mathrm{d}k_i}{k_i} = \frac{\mathrm{d}t}{2t} \\ \iff & \ln(k_i) = \frac{1}{2}\ln(t) + C_1 = \ln\left(C_2 t^{1/2}\right) \\ \iff & k_i = C_2 t^{1/2} \end{aligned}$$

If node i joins the system at time t_i, then $k_i(t_i) = m$, since each node gets m connections when they join the system, so

$$\begin{aligned} & m = C_2 t_i^{1/2} \\ \iff & C_2 = \frac{m}{t_i^{1/2}} \end{aligned}$$

hence

$$k_i = m\left(\frac{t}{t_i}\right)^{\beta} \tag{3}$$

where $\beta = \frac{1}{2}$ is called the *dynamic exponent*. This means that all nodes follow the same power law, as it can be seen in Figure 3. Also, for nodes joining the system earlier the value of t_i is smaller,

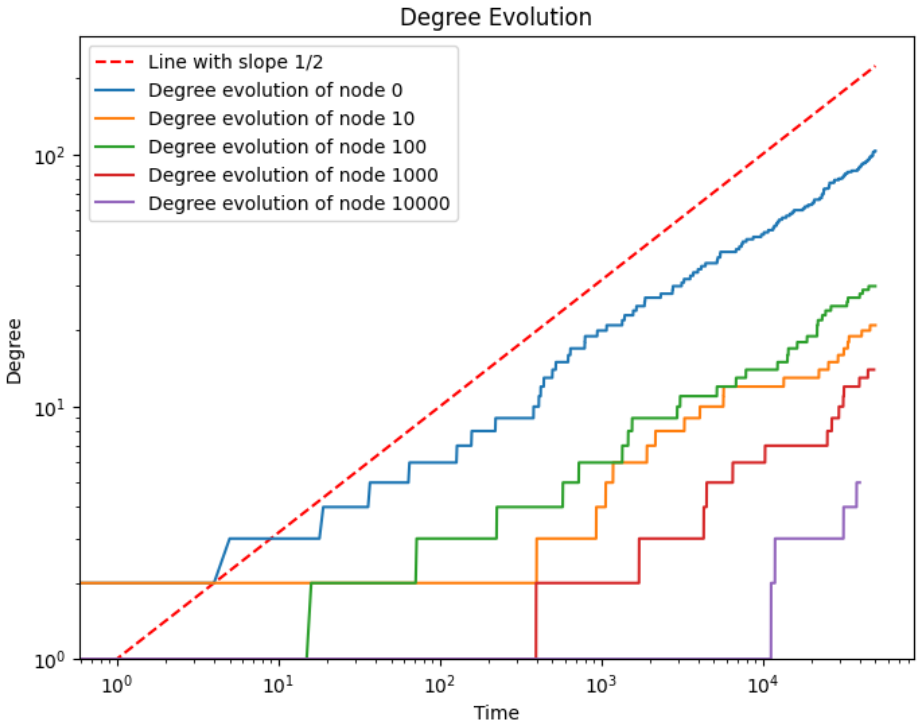


Figure 3: Change of the degree of some nodes joining at different times, as a function of time. We can see that all of the plots have the same slope, which is about 1/2, in accordance with Equation (3)

so k_i is bigger. This means that hubs are big because they arrived earlier to the system. This phenomenon is called *first-mover advantage* in marketing and business. Indeed, we can observe in real life that companies who enter a new market have an advantage over companies which enter later. Next, if we take the derivative of Equation (3), we get

$$\frac{\mathrm{d}k_i}{\mathrm{d}t} = \frac{m}{2\sqrt{t_i t}}.$$

From this we can see that

- nodes which arrived earlier accumulate connections faster (since t_i is smaller),
- since the number of nodes a new node can connect to increases over time, the rate of increase of degree decreases over time ($\sim t^{-1/2}$).

With Eq. (3) we can also answer the question about the construction of the BA model in Section 2. The probability that the new node connects to the same node i two times is

$$P_{double,\ i} = \binom{m}{2} P_i^2 = \frac{m(m-1)}{2}\left(\frac{k_i}{\sum_j^{N-1} k_j}\right)^2 \approx \frac{m(m-1)}{2}\left(\frac{m\left(\frac{t}{t_i}\right)^{1/2}}{2mt}\right)^2 = \frac{m(m-1)}{8t_i t},$$

so the total probability is, using $t_i = i$ (also meaning $t = N$)

$$P_{double} = \sum_{i=1}^{N} P_{double,\ i} = \sum_{i=1}^{N} \frac{m(m-1)}{8t_i t} = \frac{m(m-1)}{8N}\sum_{i=1}^{N}\frac{1}{t_i} = \frac{m(m-1)}{8N}\sum_{i=1}^{N}\frac{1}{i}.$$

If we use a continuous approximation to approximate the sum, we have

$$\sum_{i=1}^{N}\frac{1}{i} \approx \int_1^N \frac{\mathrm{d}x}{x} = \ln(N) - \ln(1) = \ln(N).$$

Using that $\lim_{N\to\infty} \frac{\ln(N)}{N} = 0$, as $N \to \infty$ we have

$$\lim_{N\to\infty} P_{double} \approx \lim_{N\to\infty} \frac{m(m-1)}{8N} \ln(N) = \frac{m(m-1)}{8} \lim_{N\to\infty} \frac{\ln(N)}{N} = 0$$

meaning that the number of double edges is negligible for large N. From this derivation we can also see that we don't necessarily have to require that the new edge must connect to already existing edges, as the number of loops is also negligible for large N (to show this, we just have to replace $\binom{m}{2}$ with $\binom{m}{1}$).

4 Degree distribution

In this section we use the result of the previous section to derive an approximation of the degree distribution. The approach here is based on continuum theory. We saw that the expected degree k_i of node i that joined at time t_i is

$$k_i = m \left(\frac{t}{t_i} \right)^{\beta}. \tag{4}$$

Now we want to find the expected number of nodes with degree smaller than k, that is $k_i(t) < k$. From Equation (4),

$$\begin{aligned} k > k_i(t) &= m \left(\frac{t}{t_i} \right)^{\beta} \\ \iff t_i &> \frac{t m^{1/\beta}}{k^{1/\beta}}. \end{aligned} \tag{5}$$

Since there are $m_0 + t$ nodes at time t and $t - t_0$ of them were added after time t_0 (we add one node at each time step), the probability that a node was added after time t_0 is

$$P(t_i > t_0) = \frac{t - t_0}{m_0 + t}.$$

Hence, the probability that Eq. (5) is satisfied is

$$P\left(t_i > \frac{t m^{1/\beta}}{k^{1/\beta}} \right) = \frac{t - \frac{t m^{1/\beta}}{k^{1/\beta}}}{m_0 + t}.$$

Hence, the cumulative degree distribution is

$$P(k_i(t) < k) = P\left(t_i > \frac{t m^{1/\beta}}{k^{1/\beta}} \right) = \frac{t - \frac{t m^{1/\beta}}{k^{1/\beta}}}{m_0 + t}.$$

The probability distribution is obtained by

$$p_k(t) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{1}{\beta} \frac{t m^{1/\beta}}{m_0 + t} k^{-\frac{1}{\beta} - 1}$$

where for big t the m_0 is negligible, so

$$p_k(t) \approx \frac{1}{\beta} m^{\frac{1}{\beta}} k^{-\frac{1}{\beta} - 1} = \frac{1}{\beta} m^{\frac{1}{\beta}} k^{-\gamma}$$

where γ is the degree exponent, characterizing the network topology. From the derivation we can observe that it is related to the dynamic exponent β by

$$\gamma = \frac{1}{\beta} + 1.$$

Knowing that for the BA model $\beta = \frac{1}{2}$, the degree exponent is $\gamma = 3$, so

$$p_k(t) \approx 2m^2 k^{-3}$$

which gives an approximation of the degree distribution for big k and m, but correctly predicts the power law behaviour of the degree distribution. However, the correct factor is $2m(m+1)$, as we will see in the next section.

5 More precise derivation of degree distribution

In this section, let us derive a more explicit form of the degree distribution. For this, we will use the so called rate equation approach.

Let $N(k,t)$ denote the number of nodes with degree k at time t. Then, the probability of a node having degree k at time t is given by

$$p_k(t) = \frac{N(k,t)}{N(t)} \tag{6}$$

where $N(t)$ is the total number of nodes at t. Also, the preferential attachment says

$$P(k) = \frac{k}{\sum_j k_j}. \tag{7}$$

We saw that at time t the number of nodes is $N(t) = m_0 + t$ and the number of edges is $E(t) = e_0 + mt$. When t is big, m_0 and e_0 are negligible, so $N(t) \approx t$ and $E(t) \approx mt$. Then, since the total number of degrees is $2E$, Equation (7) becomes

$$P(k) \approx \frac{k}{2mt}. \tag{8}$$

Next, let us calculate the expected number of nodes with degree k that gain a new connection when we add a new node. From Equation (8) the probability that the new node connects to a node with degree k is $\frac{k}{2mt}$, from Equation (6) there are $N(k,t) = p_k(t)N(t) \approx p_k(t)t$ nodes with degree k, and the new node connects to m nodes, so the expected value of the new links gained by degree k nodes when a new node arrives is

$$\frac{k}{2mt} \cdot p_k(t)t \cdot m = \frac{k}{2}p_k(t). \tag{9}$$

Now, let us observe what happens to $N(k,t)$ and $p_k(t)$ when we add a new node.

- If the new node connects to a node with degree k, then the degree of this node becomes $k+1$, which decreases $N(k,t)$. From Equation (9) the expected value of this change is

$$\frac{k}{2}p_k(t).$$

- If the new node connects to a node with degree $k-1$, then the degree of this node becomes k, which increases $N(k,t)$. From Equation (9) the expected value of this change is

$$\frac{k-1}{2}p_{k-1}(t).$$

At time $t+1$ the expected number of nodes with degree k is $N(k,t+1) = N(t+1)p_k(t+1) = (t+1)p_k(t+1)$. Also, by the previous two points,

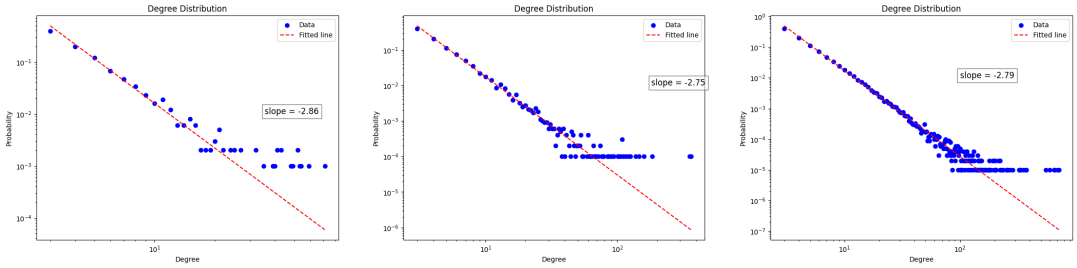
$$N(k,t+1) = N(k,t) + \frac{k-1}{2}p_{k-1}(t) - \frac{k}{2}p_k(t). \tag{10}$$

where if we use that $N(k,t+1) = (t+1)p_k(t+1)$ and that $N(k,t) = tp_k(t)$,

$$\begin{aligned} (t+1)p_k(t+1) &= tp_k(t) + \frac{k-1}{2}p_{k-1}(t) - \frac{k}{2}p_k(t) \\ \iff (t+1)p_k(t+1) - tp_k(t) &= \frac{k-1}{2}p_{k-1}(t) - \frac{k}{2}p_k(t). \end{aligned} \tag{11}$$

However, one should note that every node has at least degree m. This means that $p_k(t) = 0$ when $k < m$. Also, the new node has degree m. Hence, for $k = m$ Eq. (11) becomes

$$(t+1)p_m(t+1) - tp_m(t) = 1 - \frac{m}{2}p_m(t). \tag{12}$$



(a) Simulation with maximum nodes $N = 1000$ (b) Simulation with maximum nodes $N = 10,000$ (c) Simulation with maximum nodes $N = 100,000$

Figure 4: Simulations for 10^3, 10^4, 10^5 nodes with $m = 3$. We can observe that all three distributions are described by the same power law, thus these simulations suggest that the degree distribution is independent of time, i.e. the degree distribution is stationary.

Now let us look at Figure 4, which shows the degree distribution of graphs with different length of simulations. We can observe that the relationship between the degrees and their probabilities is given by the same power law, so we expect a stationary degree distribution, that is a degree distribution independent of time t. With this assumption, if we take the limit $t \to \infty$, we expect to get $p_k(\infty) = p_k$. Applying this to Equation (11) and (12) we get on the left hand side

$$(t+1)p_k(t+1) - tp_k(t) \to tp_k(\infty) + p_k(\infty) - tp_k(\infty) = p_k(\infty) = p_k \tag{13}$$

$$(t+1)p_m(t+1) - tp_m(t) \to p_m \tag{14}$$

using that $t + 1 \approx t$ for large t. And, on the right hand side we get

$$\frac{k-1}{2}p_{k-1}(t) - \frac{k}{2}p_k(t) \to \frac{k-1}{2}p_{k-1}(\infty) - \frac{k}{2}p_k(\infty) = \frac{k-1}{2}p_{k-1} - \frac{k}{2}p_k \tag{15}$$

$$1 - \frac{m}{2}p_m(t) \to 1 - \frac{m}{2}p_m(\infty) = 1 - \frac{m}{2}p_m. \tag{16}$$

Hence, from Equations (13) and (15) we get

$$\begin{aligned} p_k &= \frac{k-1}{2}p_{k-1} - \frac{k}{2}p_k \\ \iff p_k &= \frac{k-1}{k+2}p_{k-1} \qquad (17) \\ \iff p_{k+1} &= \frac{k}{k+3}p_k \qquad (18) \end{aligned}$$

where we performed a shift of index $k \to k+1$ in Eq. (17). And, from Eq. (14) and (16)

$$\begin{aligned} 1 - \frac{m}{2}p_m &= p_m \\ \iff p_m &= \frac{2}{m+2}. \qquad (19) \end{aligned}$$

This gives us a recursive expression for calculating p_k. Let us write down p_k for the smallest possible nonzero degrees using Eq. (18) and (19)

$$\begin{aligned} p_m &= \frac{2}{m+2} = \frac{2m(m+1)}{m(m+1)(m+2)}, \\ p_{m+1} &= \frac{m}{m+3}p_m = \frac{2m}{(m+2)(m+3)} = \frac{2m(m+1)}{(m+1)(m+2)(m+3)}, \\ p_{m+2} &= \frac{m+1}{m+4}p_{m+1} = \frac{2m(m+1)}{(m+2)(m+3)(m+4)}, \\ p_{m+3} &= \frac{m+2}{m+5}p_{m+2} = \frac{2m(m+1)}{(m+3)(m+4)(m+5)}. \end{aligned}$$

We can see a pattern in the denominator in the above sequence. If we substitute $k = m + 3$ in the last line, we get the general formula:

$$p_k = \frac{2m(m+1)}{k(k+1)(k+2)}. \tag{20}$$

Indeed, for $k + 1$

$$p_{k+1} = \frac{k}{k+3} p_k = \frac{2m(m+1)}{(k+1)(k+2)(k+3)} = \frac{2m(m+1)}{(k+1)\big((k+1)+1\big)\big((k+1)+2\big)}.$$

which has the same form as Eq. (20). Hence, by mathematical induction, we can conclude that Eq. (20) is the correct formula.

Finally, let us see a few remarks about this result.

- Since we approximated $N \approx t$ and $E \approx mt$, this result is explicit only when we start with 1 node.
- For high degree, $k \approx k + 1 \approx k + 2$, which gives us

  $$p_k \approx \frac{2m(m+1)}{k^3}$$

  that is a power law relation of the form $p_k \sim k^{-3}$, which is very close to the results of simulations of Figure 2 and 4.
- When m also is big, we get $p_k \approx 2m^2 k^{-3}$, so we recover the approximation in Section 4.
- The exponent of the degree distribution is independent of m, which is in agreement with the simulations in Figure 5.
- The degree distribution is indeed stationary (independent of t). This implies that systems with different size, age, and history develop the same degree distribution.
- The derivation method explained here (the rate equation approach) is also applicable to many other growing networks.

6 The absence of growth and preferential attachment

One can ask the question if both growth and preferential attachment are necessary to produce a power law degree distribution. To answer this question, we will remove one of these two characteristics from the graph and see what happens.

6.1 No preferential attachment

If we remove preferential attachment, that is a new node connects to any node with the sane probability, then the chance that the new node connects to node i is given by

$$P_i = \frac{1}{m_0 + t - 1} \tag{21}$$

since the graph has $m_0 + t$ nodes at time t but the new node cannot connect to itself. We can then use the derivation method used in Sections 3 and 4 to derive the degree distribution. We have

$$\frac{\mathrm{d}k_i}{\mathrm{d}t} = mP_i = \frac{m}{m_0 + t - 1}$$

The solution of this ODE is

$$\begin{aligned} \mathrm{d}k_i &= \frac{m}{m_0 + t - 1}\mathrm{d}t \\ \iff k_i &= m \ln(m_0 + t - 1) + C \end{aligned}$$

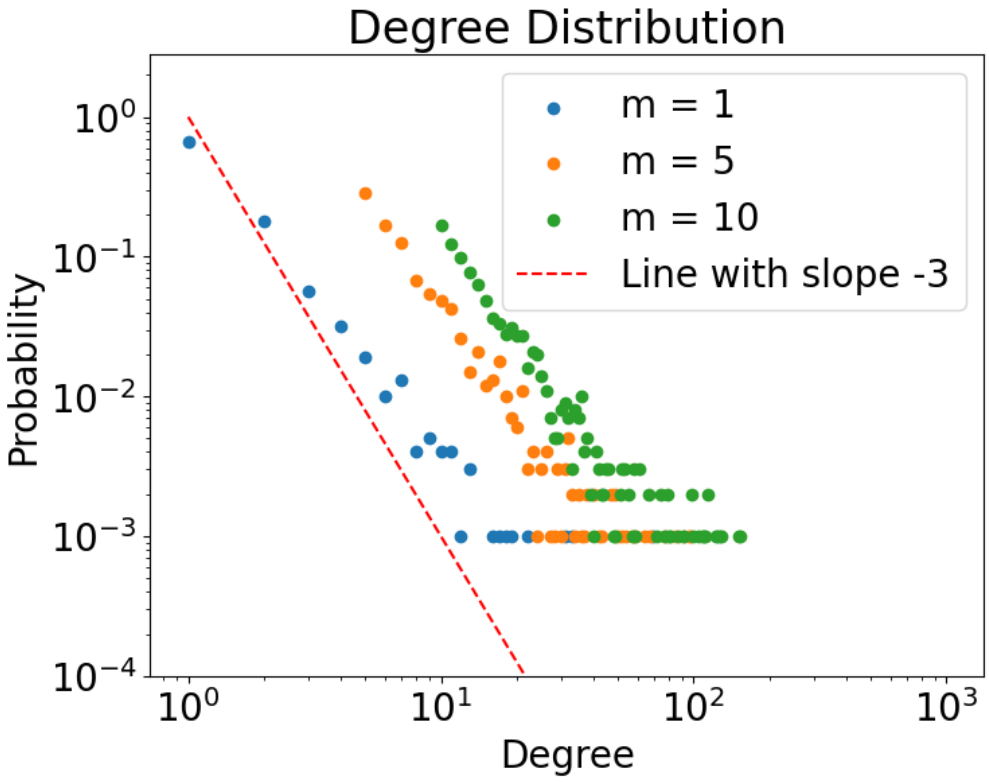


Figure 5: Simulations for 10^4 nodes and new connections 1, 5, 10. We can observe that all three distributions are described by the same power law, thus these simulations suggest that the degree distribution is independent of the number of new connections, as predicted by our model.

where we can find $C = m[1 - \ln(m_0 + t_i - 1)]$ by using the condition $k_i(t_i) = m$. So,

$$k_i(t) = m[\ln(m_0 + t - 1) - \ln(m_0 + t_i - 1) + 1] = m \ln\left(e\frac{m_0 + t - 1}{m_0 + t_i - 1}\right).$$

We can again find the expected number of nodes with degree smaller than k

$$\begin{aligned} & k > m \ln\left(e\frac{m_0 + t - 1}{m_0 + t_i - 1}\right) \\ \iff & t_i > 1 - m_0 + \frac{e(m_0 + t - 1)}{e^{k/m}} = 1 - m_0 + (m_0 + t - 1)\exp\left(1 - \frac{k}{m}\right) \end{aligned}$$

then find the cumulative degree distribution

$$\begin{aligned} P(k_i(t) < k) &= P\left(t_i > 1 - m_0 + (m_0 + t - 1)\exp\left(1 - \frac{k}{m}\right)\right) = \frac{t - \left(1 - m_0 + (m_0 + t - 1)\exp\left(1 - \frac{k}{m}\right)\right)}{m_0 + t} \\ &= \frac{(m_0 + t - 1)\left(1 - \exp\left(1 - \frac{k}{m}\right)\right)}{m_0 + t} \end{aligned}$$

and finally the degree distribution

$$p_k(t) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{(m_0 + t - 1)\exp\left(1 - \frac{k}{m}\right)}{m(m_0 + t)}$$

where for t big (i.e. $t - 1 \approx t$)

$$p_k(t) \approx \frac{\exp\left(1 - \frac{k}{m}\right)}{m} = \frac{e}{m}\exp\left(-\frac{k}{m}\right).$$

Hence, the degrees accumulate connections logarithmically, which is much slower than the power law increase, and the degree distribution decays exponentially, that is much faster than the power law decay, so hubs are not likely to form. We can also see this difference in Figure 6. Therefore, we can conclude that without the preferential attachment the network is not scale free and no hubs are formed, so preferential attachment is an essential part of the model.

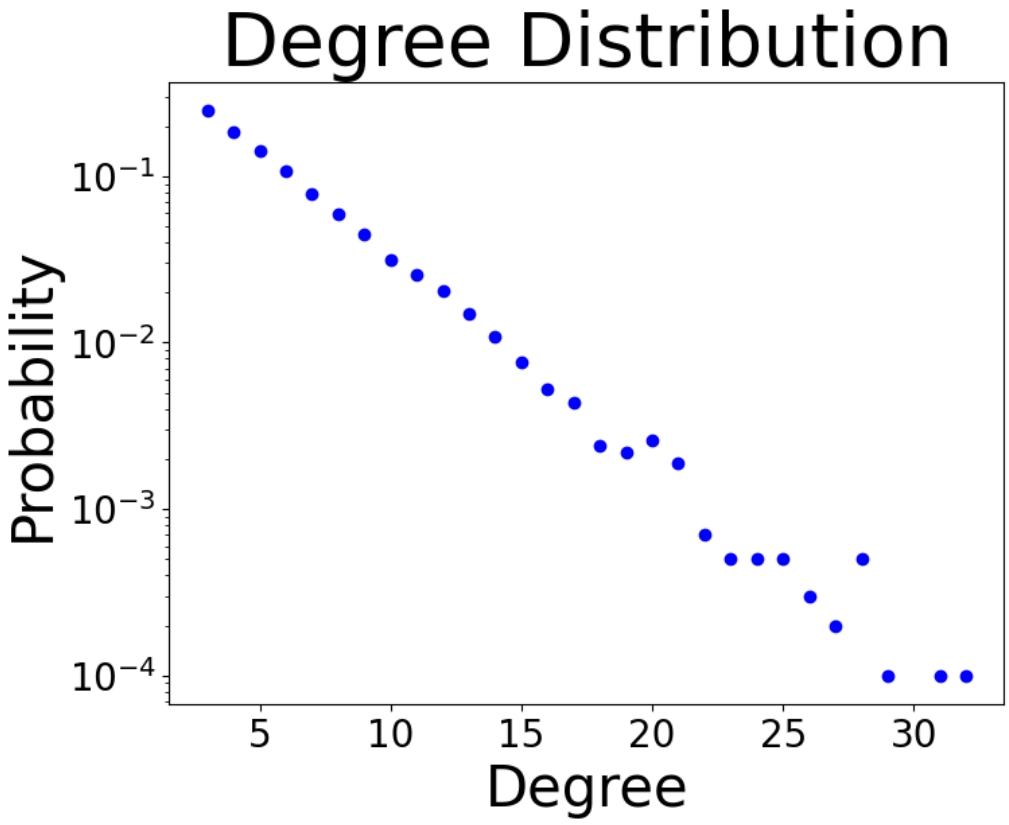


Figure 6: Simulations for 10^4 nodes and new connections 3, without preferential attachment. We can see that the degree distribution is exponential (note that this time only the y axis has logarithmic scale).

6.2 No growth

Now let us see what happens when we have preferential attachment but no growth. For this we will modify the model as follows. We fix the number of nodes N and at each step we choose a random node (each node has the same probability to be chosen), and connect it to another node i with probability

$$P_i = \frac{k_i}{\sum_j k_j}.$$

Then, at each step there is a $1/N$ probability that the node i is chosen as one end of a new edge, and $\frac{N-1}{N} P_i$ probability that it gets a connection from another node (we choose a different node with probability $(N-1)/N$, and this node connects to i with probability P_i). For t big the total number of edges is t, so the total number of degrees is $\sum_j k_j = 2t$. Hence, we have

$$\frac{\mathrm{d}k_i}{\mathrm{d}t} = \frac{N-1}{N}\frac{k_i}{2t} + \frac{1}{N}$$

which can be written as

$$\frac{\mathrm{d}k_i}{\mathrm{d}t} - \frac{N-1}{N}\frac{k_i}{2t} = \frac{1}{N}$$

To solve this ODE first we find the integrating factor $\mu(t)$

$$\mu(t) = \exp\left(\int -\frac{N-1}{N}\frac{\mathrm{d}t}{2t}\right) = \exp\left(-\frac{N-1}{2N}\ln(t)\right) = t^{-(N-1)/2N}$$

from which the solution of the ODE is

$$\begin{aligned} k_i(t) &= \frac{\int \mu(t)/N \mathrm{d}t + C}{\mu(t)} = t^{(N-1)/2N}\left(\int \frac{t^{-(N-1)/2N}}{N}\mathrm{d}t + C\right) \\ &= t^{(N-1)/2N}\left(\frac{1}{N}\frac{t^{-(N-1)/2N+1}}{1-(N-1)/2N} + C\right) = \frac{2N}{N(N+1)}t + Ct^{(N-1)/2N} \\ &= \frac{2}{N+1}t + Ct^{(N-1)/2N} \end{aligned}$$

which reduces to

$$k_i(t) = \frac{2}{N}t + Ct^{1/2}$$

for N big. We saw that $\sum_j k_j = 2t$, so

$$\begin{aligned} 2t = \sum_j^N \frac{2}{N}t + Ct^{1/2} = N\left(\frac{2}{N}t + Ct^{1/2}\right) = 2t + CNt^{1/2} \\ \Rightarrow 0 = CNt^{1/2} \\ \Rightarrow C = 0. \end{aligned}$$

Hence,

$$k_i(t) = \frac{2}{N}t. \tag{22}$$

At early stages, when all nodes connect to nodes it does not have a connection with yet, the evolution of the graph is very similar to the BA graph, and the degree distribution follows a power law. However, Eq. (22) says that all nodes have the same expected degrees, which means that as t increases the degree distribution approaches a Gauss-distribution. We can also observe these in Figure 7. This change also implies that the distribution is not stationary. Hence, growth is also an essential part of the BA model.

7 Diameter and local clustering coefficient

The average distance of two nodes in the BA model for $m > 1$ and large N is given by [4]

$$d_{avg} \approx \frac{\ln(N)}{\ln(\ln(N))}$$

and it can be shown that the diameter (the maximum distance in the graph) scales the same way [4,5]

$$d \sim \frac{\ln(N)}{\ln(\ln(N))}.$$

Also, for $m = 1$ we have [4,5]

$$d \sim \ln(N)$$

This means, that the diameter of the BA graph is the same or even smaller than that of the $G(n,p)$ graph (which scales as $d \sim \ln(N)$).

Let us continue by calculating the local clustering coefficient. The local clustering coefficient is defined as follows.

Definition 7.1 (Local clustering coefficient) *For a finite simple graph $G = (V, E)$ and for any $x \in V$, the local clustering coefficient C_x is defined as*

$$C_x := 2\frac{Nr_x(\triangleleft)}{k(k-1)}$$

where $Nr_x(\triangleleft)$ is the number of closed triangles containing x, and k is the degree of x.

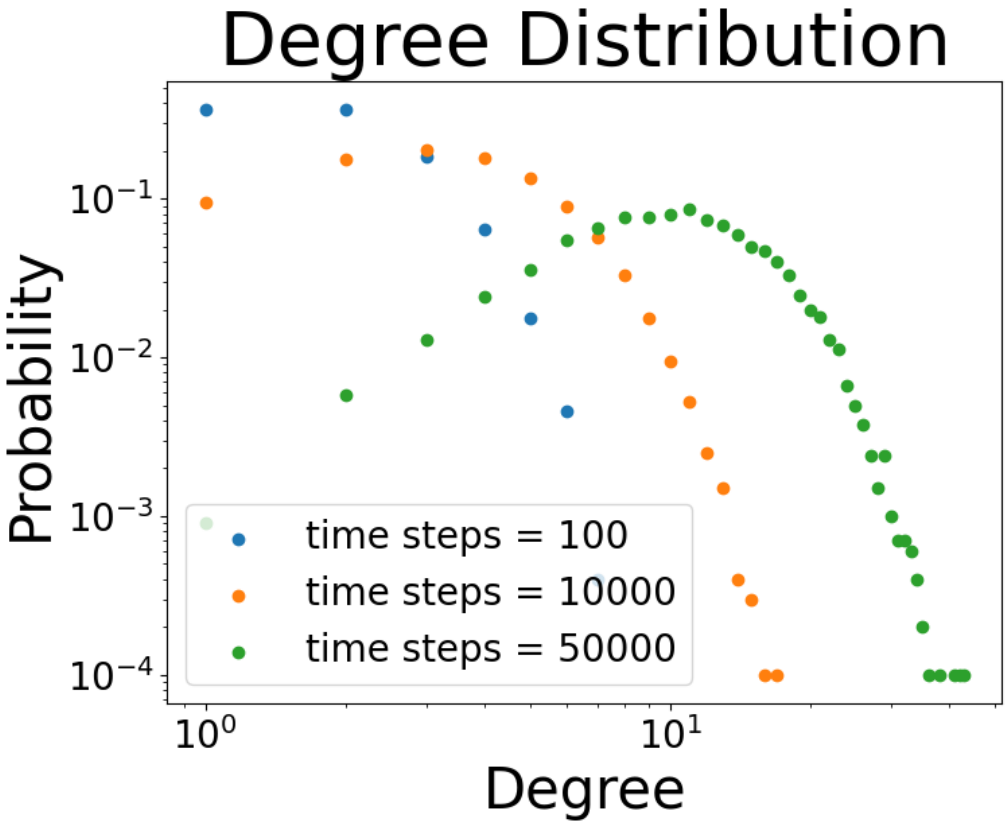


Figure 7: Simulations for 10^4 nodes without growth, with 1 new connection and 100, 10,000, 50,000 time steps. We can see that the degree distribution is very similar to a power law for low number of steps, but later develops a peak and approaches a Gaussian distribution.

As we will see, the local clustering coefficient depends on the time, so let us calculate it at time N. Let the probability that an edge exists between nodes i and j be $P(i,j)$. Then, the probability that $i,\ j,\ \ell$ form a triangle is $P(i,j)P(i,\ell)P(j,\ell)$.
Let us use continuous degree approximation. To find the number of triangles containing ℓ we sum the probabilities for all N nodes of the graph

$$Nr_\ell(\triangleleft) = \int_{i=1}^{N}\int_{j=1}^{N} P(i,j)P(i,\ell)P(j,\ell)\ \mathrm{d}i\ \mathrm{d}j \tag{23}$$

Next, let us calculate $P(i,j)$. If node j was added at time $t_j = j$ (we can use this notation since we add one node at each time), and node i has degree k_i, then

$$P(i,j) = mP_i(k_i(t_j)) \approx m\frac{k_i(j)}{\sum_{a=1}^{j} k_a(j)} \approx m\frac{k_i(j)}{2mj} = \frac{k_i(j)}{2j}$$

where we neglected the initial number of nodes and edges. From Eq. (3)

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2} = m\left(\frac{j}{i}\right)^{1/2}$$

if node i was added at $t_i = i$. Hence

$$P(i,j) = \frac{m}{2}(ij)^{-1/2}.$$

Substituting this into Eq. (23)

$$Nr_\ell(\triangleleft) = \int_{i=1}^{N}\int_{j=1}^{N} P(i,j)P(i,\ell)P(j,\ell)\ \mathrm{d}i\ \mathrm{d}j = \frac{m^3}{8}\int_{i=1}^{N}\int_{j=1}^{N}(ij)^{-1/2}(i\ell)^{-1/2}(j\ell)^{-1/2}\ \mathrm{d}i\ \mathrm{d}j$$
$$= \frac{m^3}{8\ell}\int_{i=1}^{N}\frac{\mathrm{d}i}{i}\int_{j=1}^{N}\frac{\mathrm{d}j}{j} = \frac{m^3}{8\ell}\left(\ln(N)\right)^2$$

Next, from Eq. (3) the degree of node ℓ at time $t = N$ is

$$k_\ell(N) = m\left(\frac{N}{\ell}\right)^{1/2}$$

so we have for N large

$$k_\ell(N)\left(k_\ell(N)-1\right) \approx k_\ell^2(N) = m^2\frac{N}{\ell}$$

so the clustering coefficient is

$$C_\ell = 2\frac{Nr_\ell(\triangleleft)}{k_\ell(N)(k_\ell(N)-1)} = 2\frac{\frac{m^3}{8\ell}\left(\ln(N)\right)^2}{m^2\frac{N}{\ell}} = \frac{m}{4}\frac{\left(\ln(N)\right)^2}{N}.$$

If we compare this with the clustering coefficient of the $G(n,p)$ model, given by $C_x \sim \frac{1}{N}$, we can see that the BA model is locally more clustered.

8 Summary

The Gilbert-Erdős-Rényi model fails to predict some characteristics of some real life systems. One of these systems are the scale-free networks, whose degree distribution follows a power law. To model these kind of systems, a new model was introduced. This new model is the Barabási-Albert model (BA model), and has two important characteristics: growth (i.e. the number of nodes increases over time) and preferential attachment (i.e. the probability that an edge is formed between two nodes depends on the degree of the nodes). As we saw in Section 6, both of these are necessary to produce a scale-free network.
The basic properties of this model are summarized in Table 1. An important property of the BA model is that its degree distribution indeed follows a power law, so it can model scale-free network. However, this model is also not perfect, as it fails to predict the following properties of real life systems.

- In the BA model the degree exponent of the degree distribution is always $\gamma = 3$, but for real systems it varies between 2 and 5.
- The BA model produces an undirected graph while many real life networks are directed (for example, following someone on social media happens in one direction).
- In real life networks, links and nodes can disappear (for example, if you unfollow someone on social media or if you delete your account).
- The BA model does not distinguish between nodes, while in real life some nodes might be more likely to get attraction (e.g. in the network of research papers, certain research areas are more popular than others, so paper related to these have a higher chance to get more views).

Therefore, the main purpose of the BA model is to capture the main mechanisms of real systems, investigate the emergence of the scale-free property, and serve as a foundation for more advanced models where we incorporate the above mentioned properties.
For further interest, the reader is encouraged to read the sections of [2] not covered in this report (namely Sections 5.6-5.8 and 5.14).

Number of nodes	$N = t$
Number of links	$E = mt$
Average degree	$\langle k \rangle = 2m$
Degree dynamics	$k_i = m \left(\frac{t}{t_i} \right)^{1/2}$
Degree distribution	$p_k \sim k^{-3}$
Average distance	$d_{avg} \approx \frac{\ln(N)}{\ln(\ln(N))}$
Local clustering coefficient	$C = \frac{m}{4} \frac{(\ln(N))^2}{N}$

Table 1: Basic properties of the BA model, for large t.

9 Additional resources

In this section the general structure of the program used for the simulations will be explained. More details can be found in the comments of the script. The full script can be accessed from the following link (Google account needed):
`https://colab.research.google.com/drive/12AgmVxuopMboy6AUnNmOw4hIyPrBedEi?usp=sharing`.

The program is organized into blocks. The first block contains the libraries necessary for the program, the second block can generate and show a BA graph given the initial graph, the number of new connections m, and the maximum number of nodes, and the rest of the blocks were used to create the plots in this report. These blocks are structures the following way. Firstly, we give the characteristics of the graph (initial graph, the number of new connections m, and the maximum number of nodes), and define a few functions which describe how the graph evolves. Then, we create the graph using these functions, and finally we create a plot of some measured parameters of the graph.

10 References

1. SML 2024 Spring, Graph theory lecture notes
2. Albert-László Barabási, Network Science, `https://networksciencebook.com/`
3. Albert-László Barabási, Réka Albert, Hawoong Jeong, Mean-field theory for scale-free random networks, Physica A: Statistical Mechanics and its Applications, Volume 272, Issues 1–2, 1999, Pages 173-187, ISSN 0378-4371, `https://doi.org/10.1016/S0378-4371(99)00291-5` (`https://www.sciencedirect.com/science/article/pii/S0378437199002915`)
4. Cohen, Reuven & Havlin, Shlomo. (2003). Scale-Free Networks Are Ultrasmall. Physical review letters. 90. 058701. 10.1103/PhysRevLett.90.058701.
5. B. Bollobas and O. Riordan, in Handbook of graphs andnetworks, eds. S. Bornholdt and H.G. Schuster (Wiley-VCH, Berlin, 2002).