Network Project: The BA Model

CID: 01743935
NAME: BENCHIZHAO

23rd March, 2020

Abstract: Three kinds of attachments are investigated in network growing model: **1.** pure preferential attachment, **2.** pure random attachment, **3.** random walks and preferential attachment (mixed attachment). The degree distributions $p(k)$ of the three attachments are: **1.** power law decay ($p(k) \propto k^{-3}$), **2.** exponential decay ($p(k) \propto (m/(m+1))^{k-m}$), **3.** power law decay ($p(k) \propto k^{-5}$). The KS test is performed to compare the theory and data, and the results of KS test demonstrate the null hypothesis is false for pure preferential and pure random attachment.

Word count: 2430 words

1 Introduction

The Barabási–Albert (BA) model is an algorithm for generating random scale-free networks using a preferential attachment mechanism[1]. In terms of degree distribution, it is completely equivalent to Yule model [3] and Simon model [4]. Here we are going to investigate the the degree distribution produced by the BA model. There are three kinds of preferential attachments are talked in this report: pure preferential attachment, pure random attachment and mixed attachment.

1.1 Definition of model

The model to be used in the project is defined as follows.

1. Set up initial network at tiem t_0, as graph $\mathcal{G}_0$.
2. Increment time $t \to t+1$ and add one new vertex.
3. Add m edges as follows [2]:
    - Connect one end of the new vertex.
    - Connect the other end of each new to existing vertex chosen with probability Π, which is defined by
      -Pure preferential: $\Pi_{\mathrm{pa}} \propto k$, where k is degree of vertex.
      -Pure random: $\Pi_{\mathrm{rnd}} \propto 1$, the existing vertices are connected with equal probability.
      -Mixed attachment: $\Pi_{\mathrm{mix}} = q\Pi_{\mathrm{pa}} + (1-q)\Pi_{\mathrm{rnd}}$, where q is the probability of choosing Pure preferential.
4. Repeat from 2 until reach final number of N vertices in the network.

1.2 Master equation

The master equation is used to describe how degree k varies when a new vertex is added into system $(n(t+1))$.

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

Let's explain each term in this equation:

- $n(k,t)$: The number of vertices with degree k at time t and this term is the one we start with.
- $m\Pi(k-1,t)n(k-1,t)$: Add one edge to vertices with $k-1$ degree, so that the vertices have k degrees.
- $m\Pi(k,t)n(k,t)$: Add one edge to vertices with k degrees, meaning that these vertices no longer in count.
- $\delta_{k,m}$: Is the new added vertex, if $k=m$, this term is one; Otherwise, is zero.

In long time $t \to \infty$, the number of vertices can be assumed as $n(k,t) \to N(t)p_\infty(k)$. And for each time, only one vertex is added, which means $N(t+1) = N(t) + 1$. Then substitute these equaitons into Eq.(1), we will get

$$p_\infty(k) = mN(t)[\Pi(k-1,t)p_\infty(k-1) - \Pi(k,t)p_\infty(k)] + \delta_{k,m}. \tag{2}$$

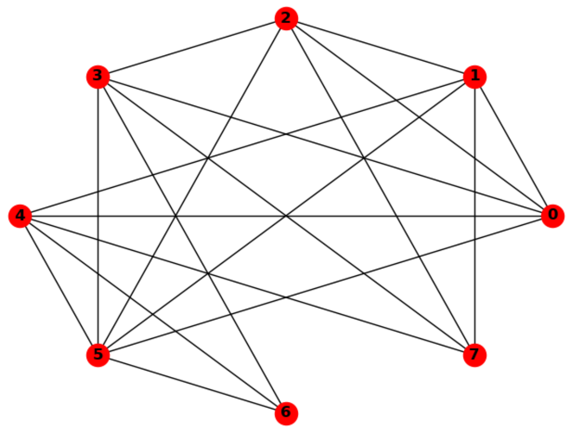
2 Pure Preferential Attachment $\prod_{\mathrm{pa}}$

2.1 Implementation

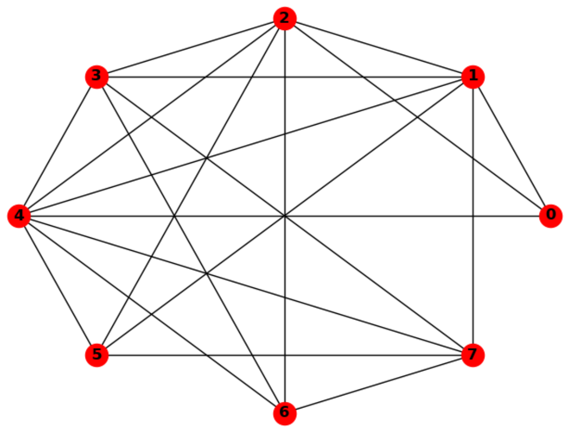
The initial graph was set by the following algorithm:

1. Initialise the system by $\mathcal{N}$ vertices ($\mathcal{N}$ can be any positive integer).
2. Pair vertices up, and $\mathcal{N}(\mathcal{N}-1)$ edges are generated (self-loop is prohibited).
3. Delete the repeated edges and figure out the degree distribution.

For every time the system is initialised, the initial graph should be different. The data generated by this initial graph is more random and more persuasive. The initial graph is shown in Fig.1.



(a)



(b)

Figure 1: (a) and (b) are two different initial graphs, generated by following the algorithm above with $\mathcal{N} = 8$.

After the initialisation, add the first vertex into network, choose m vertices from existing vertices with probability Π_{pa} and connect the new vertex and chosen vertices. At the beginning ($N \sim 1$), the number of edges $E \neq mN(t)$, and the theory in **section1.2** does not satisfied, which means the probability distribution does not follow the $p_\infty(k)$. When $N \to \infty$, the configuration of the initial graph won't affect the results.

2.2 Degree Distribution Theory

From the definition of preferential attachment, we know that $\Pi_{\text{pa}} \propto k$, and probability must be normalised, so

$$\Pi_{\text{pa}}(k,t) = \frac{k}{\sum_{k=0}^{\infty} kn(k,t)}.$$

The number of edges $E(t) = \sum_{k=0}^{\infty} n(k,t)\frac{k}{2}$, then we have

$$\Pi_{\text{pa}} = \frac{k}{2E(t)}.$$

When $N \to \infty$, the number of edges $E(t) \to mN(t)$, for one vertex is added, m edges are added. By taking this limit,

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

Substitute Eq.(3) into Eq.(2), we will get

$$\begin{aligned} p_\infty(k) &= mN(t)\Big[\frac{k-1}{2mN(t)}p_\infty(k-1,t) - \frac{k}{2mN(t)}p_\infty(k,t)\Big] + \delta_{k,m} \\ &= \frac{k-1}{2}p_\infty(k-1,t) - \frac{k}{2}p_\infty(k,t) + \delta_{k,m}. \end{aligned} \tag{4}$$

For $m < k$ in Eq.(4), i.e. $\delta_{k,m} = 0$, we could easily get

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k-1}{k+2}. \tag{5}$$

Since the equation

$$\frac{f(z)}{f(z-1)} = \frac{z+a}{z+b},$$

has solution

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

where A is a constant, $\Gamma(z)$ is the Gamma Function which is defined by $\Gamma(z) = (z-1)!$. Then Eq.(5) has a solution

$$\begin{aligned} p_\infty(k) = \frac{\Gamma(k)}{\Gamma(k+3)} &= A\frac{(k-1)!}{(k+2)!} \\ &= \frac{A}{(k+2)(k+1)k}. \end{aligned} \tag{6}$$

In order to calculate the constant A, the boundary condition should be considered, i.e. $m = k$. The Eq.(4) becomes

$$p_\infty(m) = \frac{m}{2}p_\infty(m) + 1. \tag{7}$$

The first term in Eq.(4) vanishes because $p_\infty(m-1,t)=0$. When a new vertex is added, it should has the minimum value of degree $k=m$, which means there is no way to find a vertex with degree $k=m-1<m$ when $N\to\infty$.

From Eq.(7), we could get

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

Substitute the value A into Eq.(6), the probability of distribution can be achieved

$$p_\infty(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \text{for} k \geq m. \tag{8}$$

To show that Eq.(8) is normalised, the the equation can be rewritten as

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

Then, summing over all possible degree k, we could get

$$\begin{aligned} \sum_{k=m}^{\infty} p_\infty(k) &= m(m+1)\Big[\sum_{k=m}^{\infty}(\frac{1}{k} - \frac{1}{k+1}) - \sum_{k=m}^{\infty}(\frac{1}{k+1} - \frac{1}{k+2})\Big] \\ &= m(m+1)\Big[(\frac{1}{m} - \frac{1}{m+1} + \frac{1}{m+2}\cdots) - (\frac{1}{m+1} - \frac{1}{m+2} + \frac{1}{m+3}\cdots)\Big] \\ &= m(m+1)\Big[\frac{1}{m} - \frac{1}{m+1}\Big] \\ &= 1 \end{aligned} \tag{10}$$

The process above shows the probability $p_\infty(k)$ is normalised. Foe fat-tail distribution $k \gg 1$, an approximation can be derived from Eq.(8)

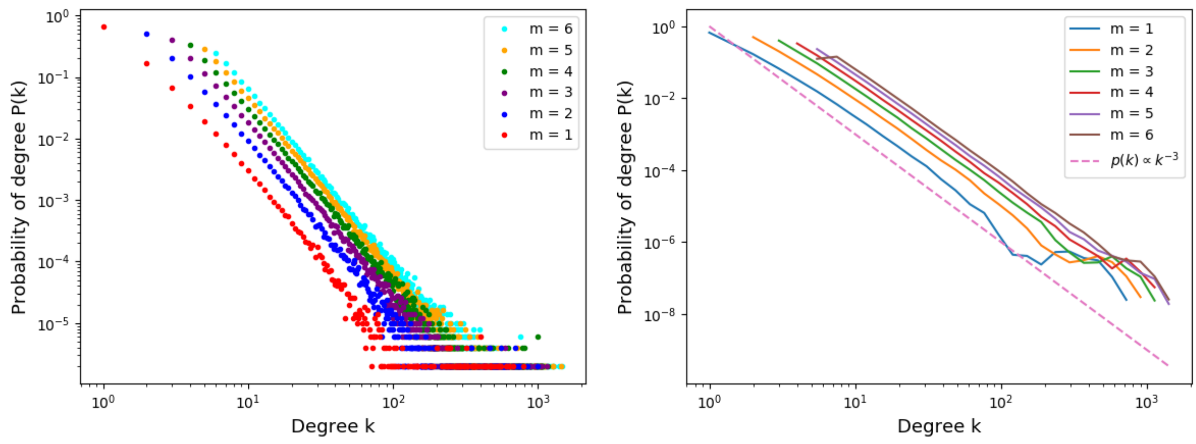
$$p_\infty(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \propto \frac{1}{k^3}. \tag{11}$$

This equation tells us the probability of degree distribution has a power law in k when $k\to\infty$.

2.3 Numerical Degree Distribution

The probability of degree was investigated in a network with $\mathcal{N}=16$ initial vertices and $N=10^5$ added vertices. The m values used are $m=1,2,3,4,5,6$, which is the number of edges formed while a new vertex is added.

In this model, the number of vertices N is a limited value, so the probability p_∞ must be a integer multiple of $1/N$. This effect is call finite-size effect. To minimise the effect, we repeat the simulation 5 times for each m values. Log-binning was made use of to obtain the features that hided by finite-size effect, and the scale factor of log-binning was chosen as $a=1.25$. The probability of degree distribution is shown in Fig.5



(a) Unbinned (b) Binned

Figure 2: (a) and (b) show the probability of degree distribution for $m = 1, 2, 3, 4, 5, 6$. (a) is unbinned, so there is a cutoff when k is large. (b) is binned, and the dash line is the theoretical value.

Fig.2a shows the unbinned data for $m = 1, 2, 3, 4, 5, 6$. When degree $k \gg 1$, the width of data become wider as k increases, and the minimum probability is $1/N$. Fig.2b shows the binned data for all m. The dash line comes from theory (Eq.(11), i.e. $p_\infty(k) \propto k^{-3}$). When k is large ($k \gg 1$), the binned data bump up, this part is known as fat-tail. this effect tells us there are few vertices connect most of edges. Before the bump, the probability decays linearly to the degree k in log-scale, which means the probability decays as a power law of degree k in real-scale. The probability of finding the same degree k increases when m is getting larger. Meanwhile, the larger m value corresponds to larger maximum and maximum degree k. The distributions for all m are parallel to the theory value (dash line), this means the theory is correct.

To investigate how well our theory fits the model, Kolmogorov–Smirnov test (KS test) is performed and the results are shown in Fig.3

The KS test compares the cumulative distribution function of the theoretical distribution and binned distribution and gives two coefficients: D-statistic and p-value. The D-statistic is the maximum absolute difference between theoretical and numerical cumulative distribution function, so the D-statistic value should be minimised. The p-value tells us if the null hypothesis is true. If p-value is large (close to 1), the null hypothesis is true, which means the two distribution are drawn from different sample; If p-value is small (close to 0), the null hypothesis is false, which means the two distribution are drawn from the same sample. Theoretically, when p-value is less than 0.05, it could be considered that the null hypothesis is false.

From Fig.3, the D values are around 0.5, and all p-values $\ll 0.05$. The KS test demonstrates the null hypothesis is false indicating the theory is correct. The D value is incredibly high, because there are only about 25 binned data and the theoretical points are much more than 25. In this case, many theoretical points exist in the interval of binned data points. Since KS test compares the cumulative distribution function of the

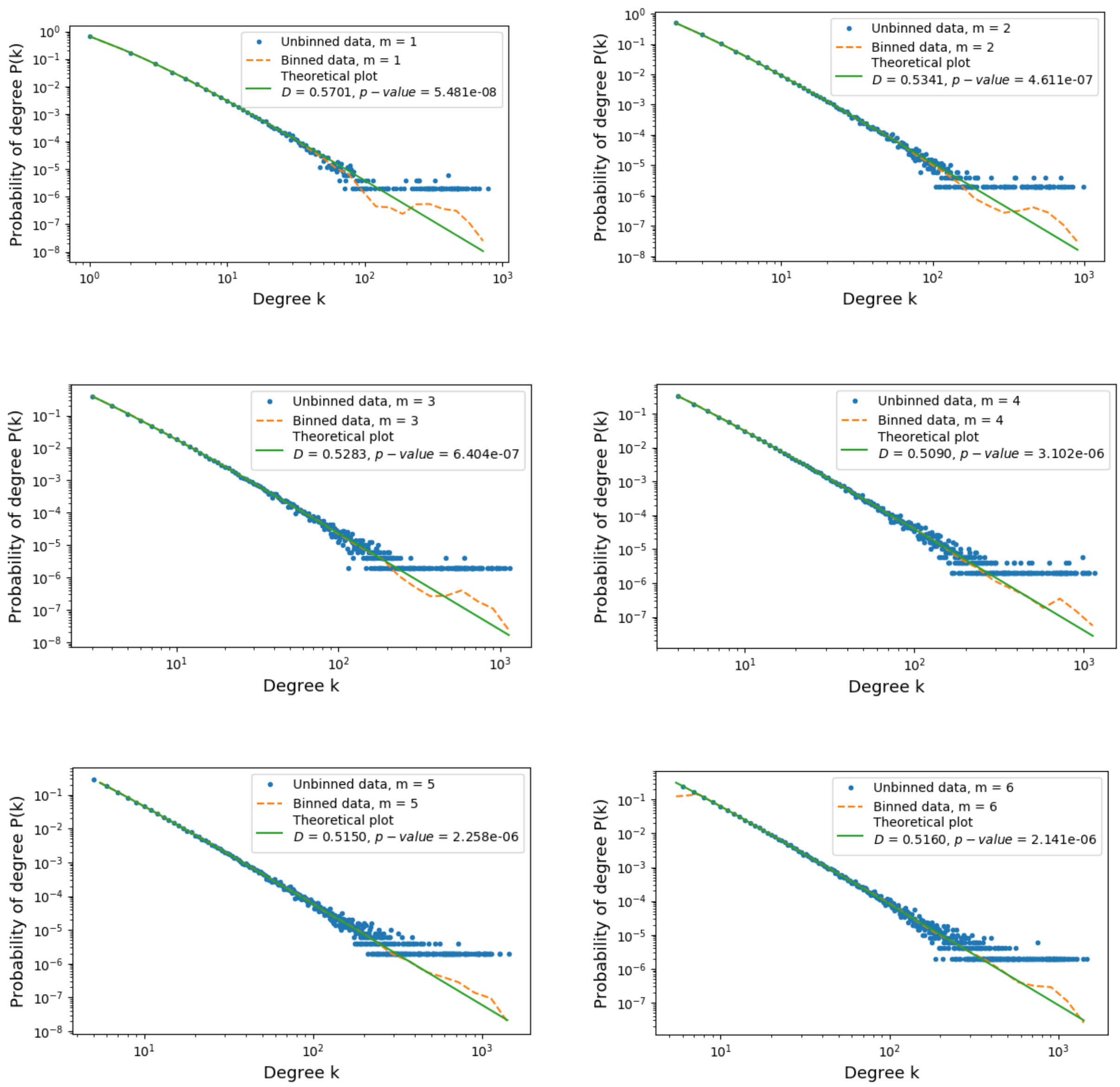


Figure 3: Degree Distribution with $N = 10^5$ vertices for each m value. The cyan dots are unbinned data; The orange dash line is the binned data with scale factor $a = 1.25$; The green line is the theoretical value.

theoretical distribution and binned distribution, that is why the D values are so high. To solve this problem, the second sets of theoretical data are generated on the same position as binned data. The second sets gives reasonable D-values, but the p-values are close to 1 for the number of theoretical points is the same as binned data. Only have about 25 points, which is far from enough to get a precise p-value. The KS test of two sets are shown in Table 1. The p-values form the first set demonstrates the null hypothesis is false; The D-statistic from the second set are about 0.1, indicating the theory fits the numerical data well.

	1^{st} set		2^{nd} set	
m	D-statistic	p-value	D-statistic	p-value
1	0.570	5.48×10^{-8}	0.115	0.992
2	0.534	4.66×10^{-7}	0.115	0.992
3	0.528	6.40×10^{-7}	0.077	1
4	0.50	3.10×10^{-6}	0.080	1
5	0.515	2.26×10^{-6}	0.080	1
6	0.516	2.14×10^{-6}	0.080	1

Table 1: KS test for two sets of theoretical data

2.4 The Largest Degree k_1

2.4.1 Theory

Assume only one vertex has the largest degree k_1, then

$$N(t) \sum_{k=k_1}^{\infty} p_\infty(k) = 1.$$

Substitute Eq.(9) into the equation above,

$$\begin{aligned} Nm(m+1) \sum_{k=k_1}^{\infty} \left(\frac{1}{k} - \frac{2}{k+1} + \frac{1}{k+2} \right) &= 1, \\ \Rightarrow \frac{Nm(m+1)}{k_1(k_1+1)} &= 1, \\ \Rightarrow k_1 = \frac{-1 \pm \sqrt{1 + 4Nm(m+1)} - 1}{2}&. \end{aligned}$$

The largest degree k_1 should be positive, so

$$\begin{aligned} k_1 &= \frac{-1 + \sqrt{1 + 4Nm(m+1)} - 1}{2} \\ &\propto \sqrt{N} \text{ for } N \gg 1. \end{aligned} \tag{12}$$

2.4.2 Numerical largest degree

From Eq.(12), the largest degree k_1 is proportional to the number of vertices, i.e. $k_1 \propto \sqrt{N}$. To verify this relation, different size of systems are investigated. Run the programme for $N = 10^2, 10^3, 10^4, 10^5$, and record the maximum degrees. To get better statistical data, repeat 15 times and take the average values. We focus on $m = 6$, which is the maximum value we studied previous, and it can give us a more persuasive data. The maximum degrees are shown in Fig.4.

Fig.4 shows the maximum degree k_1 is linear to the number of vertices with slope 0.498 ± 0.003 in log-scale. So, in real-scale, the largest degree k_1 has a power law of

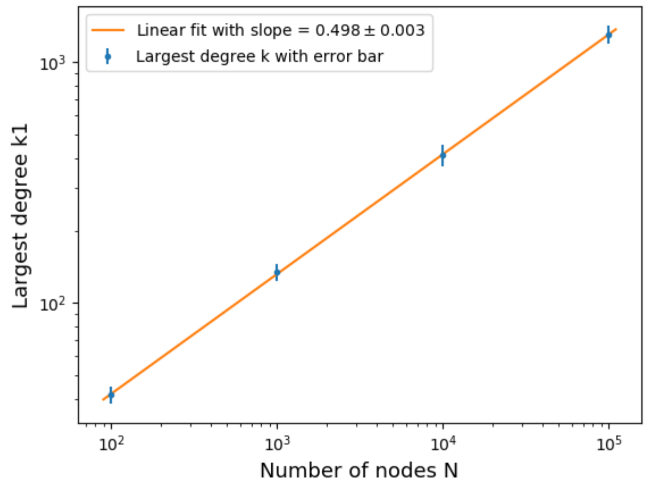
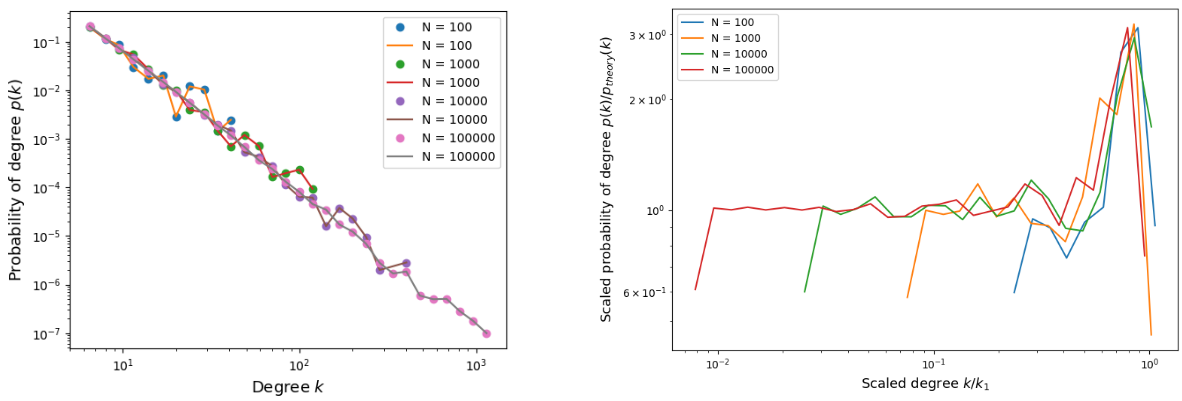


Figure 4: In log-scale, the maximum degree k_1 is linear to the number of vertices with slope 0.498 ± 0.003.

number of nodes, i.e. $k_1 \propto N^{0.498} \sim \sqrt{N}$, which agrees with the theory that derived in **section2**.**4**.**1**.

2.5 Data Collapse

Again, run the programme for $N = 10^2, 10^3, 10^4, 10^5$, and record the degree of every vertices. For getting better statistical data, repeat 15 times. The m value we chose was 6, for the same reason in previous section. The data is shown in Fig.5a.



(a) (b)

Figure 5: (a) Graph of probability distribution respect to degree k for $N = 10^2, 10^3, 10^4, 10^5$ in log-scale for $m = 6$. (b) Graph of scaled probability distribution respect to degree k for $N = 10^2, 10^3, 10^4, 10^5$ in log-scale for $m = 6$.

Fig.5a shows the probability distribution respect to degree k for $N = 10^2, 10^3, 10^4, 10^5$

in log-scale. All data points are basically aligned on a straight line. The y-axis is the probability $p_{\mathrm{data}}(k)$ that is expect to be consistent with $p_{\mathrm{theory}}(k)$, so the y-axis can be collapsed by $p_{\mathrm{data}}(k)/p_{\mathrm{theory}}(k)$, which yields a value of 1. To collapse the x-axis, the k_{data}/k_1 is used. The ansatz we use to collapse the data can be proposed:

$$\frac{p_{\mathrm{data}}(k)}{p_{\mathrm{theory}}(k)} = \mathcal{F}(\frac{k}{k_1}). \tag{13}$$

Fig.5b illustrates the collapsed data. The x-axis is k/k_1 and the y-axis is $p_{\mathrm{data}}(k)/p_{\mathrm{theory}}(k)$. The dramatic increment on the left terminal ($k \ll 1$) is caused by finite-size effect. The 'bump' and cutoff on the right terminal is known as fat tail, corresponding to the initial vertices because they are more likely to be connected with the new added vertices.

3 Pure Random Attachment Π_{rnd}

The network growth mechanism is basically the same as the one that talked in previous section, but the attachment assumption is different: all existing vertices are equal when attaching new edges, i.e. $\Pi_{\mathrm{rnd}} \propto 1$.

3.1 Degree Distribution Theory

The normalisation condition for $\Pi_{\mathrm{rnd}} \propto 1$ is $\sum_{n=0}^{\infty} \Pi_{\mathrm{rnd}} = 1$, so the probability $\Pi_{\mathrm{rnd}} = 1/N$. Substitute the probability into the master equation (Eq.(2)), we will get

$$\begin{aligned} p_\infty(k) &= mN(t)\Big[\frac{1}{N(t)}p_\infty(k-1) - \frac{1}{N(t)}p_\infty(k)\Big] + \delta_{k,m} \\ &= m\big[p_\infty(k-1) - p_\infty(k)\big] + \delta_{k,m}. \end{aligned} \tag{14}$$

$$p_\infty(k) = \frac{m}{m+1}p_\infty(k-1) + \delta_{k,m}. \tag{15}$$

For $k > m$, i.e. $\delta_{k,m} = 0$, so

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

By induction, we could get

$$p_\infty(k) = \Big(\frac{m}{m+1}\Big)^{k-m} p_\infty(m). \tag{17}$$

Let's focus on $k = m$, the term $p_\infty(m-1)$ in the Eq.(14) vanishes for $k = m-1 < m$, and $k = m$ is the minimum degree for each added vertices, then we will get

$$p_\infty(m) = m[0 - p_\infty(m)] + 1, \tag{18}$$

$$p_\infty(m) = \frac{1}{m+1}. \tag{19}$$

Substitute Eq.(19) into Eq.(17), the distribution of degree k is obtained:

$$p_\infty(k) = \frac{1}{m+1}\Big(\frac{m}{m+1}\Big)^{k-m}. \tag{20}$$

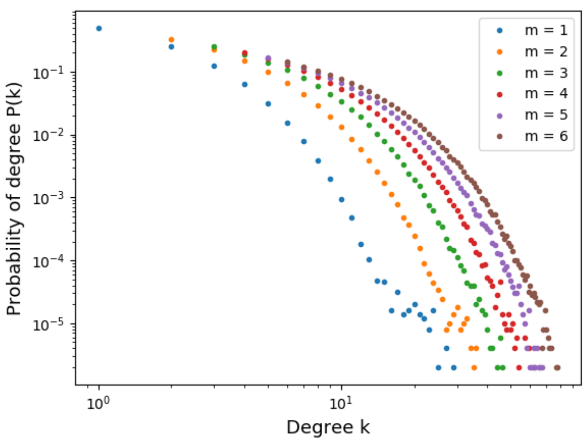
To check the normalisation of the degree distribution $p_\infty(k)$, we should sum over all possible values of the equation:

$$\begin{aligned}
\sum_{k=m}^{\infty} p_\infty(k) &= \frac{1}{m+1}\sum_{k=m}^{\infty}\Big(\frac{m}{m+1}\Big)^{k-m} \\
&= \frac{1}{m+1}\Big(\frac{m+1}{m}\Big)^{m}\sum_{k=m}^{\infty}\Big(\frac{m}{m+1}\Big)^{k} \\
&= \frac{1}{m+1}\Big(\frac{m+1}{m}\Big)^{m}\Big(\frac{(\frac{m}{m+1})^m}{1-\frac{m}{m+1}}\Big) \text{ Using Taylor expansion} \\
&= 1.
\end{aligned} \tag{21}$$

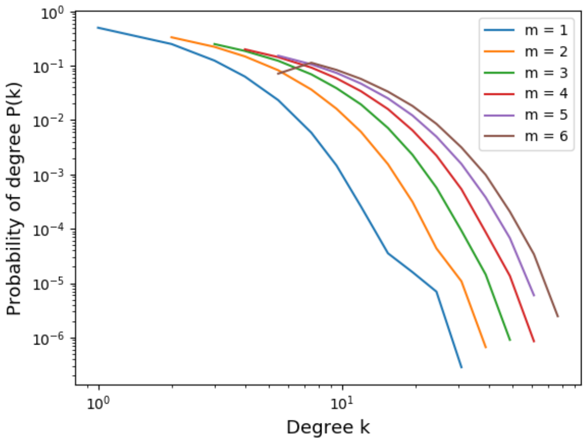
The degree distribution $p_\infty(k)$ is normalised.

3.2 Degree Distribution Numerical Results

Similar to what we have done in **section 2.3**: the network system has $\mathcal{N} = 16$ initial vertices, $N = 10^5$ added vertices, and $m = 1, 2, 3, 4, 5, 6$. The degree distribution in log-scale is shown in Fig.6. We can conclude from the curve that the fat-tail is not obvious and the probability decay exponentially, which consistent with our expectation.



(a) Unbinned



(b) Binned

Figure 6: (a) and (b) show the probability of degree distribution for $m = 1, 2, 3, 4, 5, 6$.

The KS test was also performed to check the goodness of fitness. The results are shown in Fig.7. This KS test has the same problem talked previously. But in this case, the p-value is too large (p-values are around $0.2 \gg 0.05$, where 0.05 is widely used as threshold). This indicate that the null hypothesis is true.

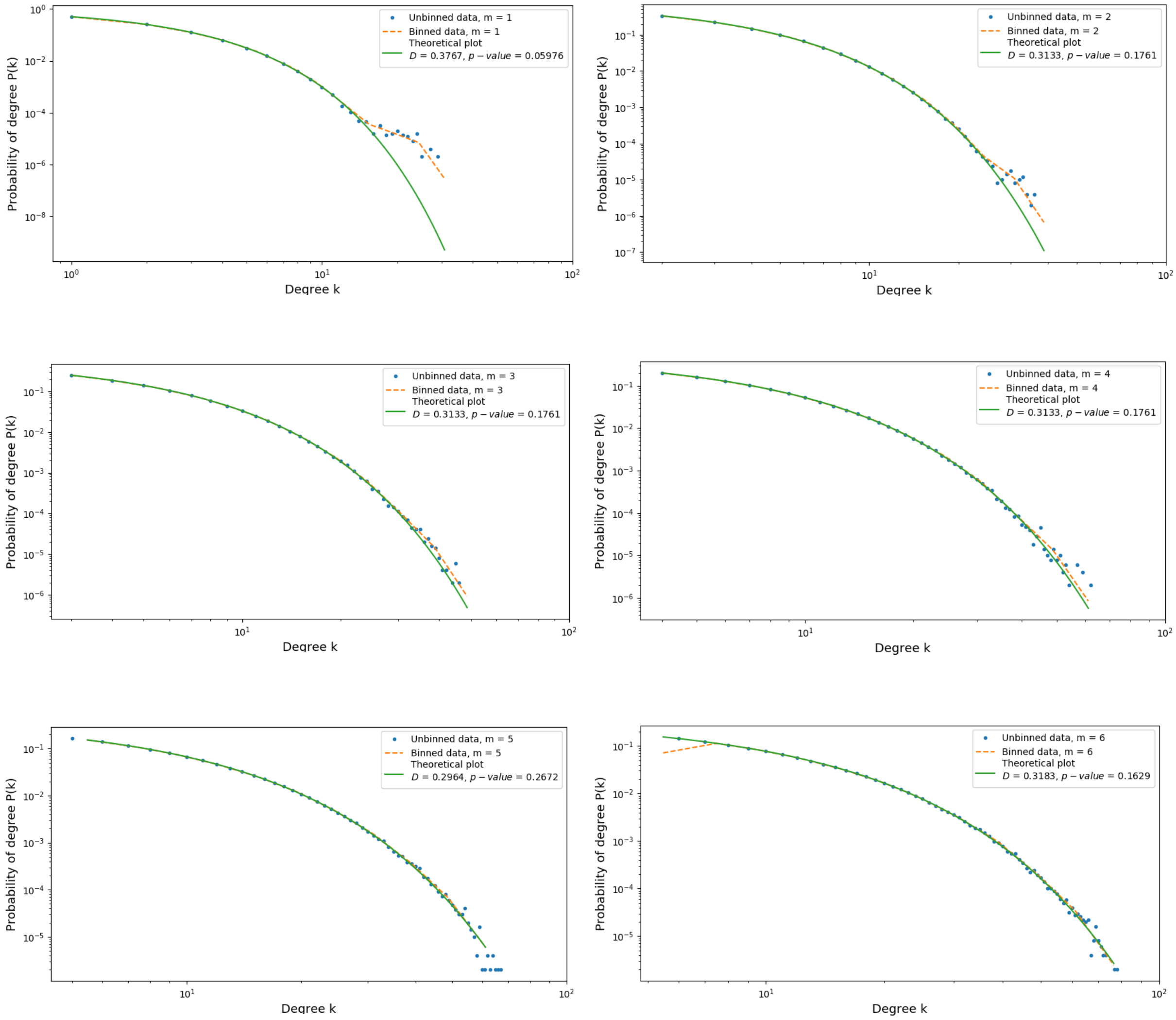


Figure 7: Degree Distribution with $N = 10^5$ vertices for each m value. The blue dots are unbinned data; The orange dash line is the binned data with scale factor $a = 1.25$; The green line is the theoretical value.

3.3 Largest Degree Theory

Using the same logic that used in the preferential largest degree. From Eq.(20), we could get

$$\frac{1}{m+1}\sum_{k=k_1}^{\infty}\left(\frac{m}{m+1}\right)^{k-m} = \frac{1}{N}, \tag{22}$$

$$\left(\frac{m}{m+1}\right)^{k_1-m}\left[\frac{1}{m+1}\sum_{k=k_1}^{\infty}\left(\frac{m}{m+1}\right)^{k-k_1}\right] = \frac{1}{N}. \tag{23}$$

Because of Eq.(21), term inside the square bracket equals to 1, then

$$\left(\frac{m}{m+1}\right)^{k_1-m} = \frac{1}{N}. \tag{24}$$

Taking logarithm on both side,

$$(k-m)\ln\left(\frac{m}{m+1}\right) = -\ln N. \tag{25}$$

$$\Rightarrow k_1 = m - \frac{\ln N}{\ln m - \ln m + 1}, \tag{26}$$

$$\Rightarrow k_1 = -\frac{\ln N}{\ln m - \ln m + 1}, \text{ for } k_1 \gg m, (N \gg 1). \tag{27}$$

Since m is a constant value, when $N \gg 1$, $k_1 \propto \ln N$

3.4 Largest Degree Numerical Results

Run the programme for $N = 10^2, 10^3, 10^4, 10^5$, and record the maximum degrees. To get better statistical data, repeat 15 times and take the average value. We focus on $m = 6$, which is the maximum value we studied previous, and it can give us a more persuasive data. The maximum degrees are shown in Fig.8. The theoretical curve line passes through all experiment data points with error bars, which proves the theory is correct.

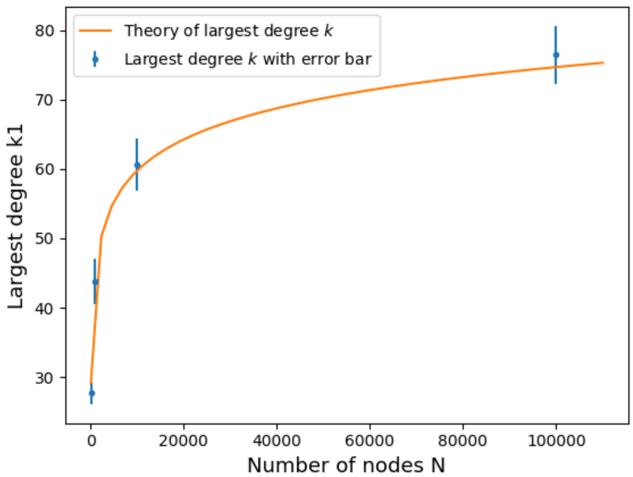


Figure 8: In real-scale, the maximum degree k_1 for $m = 6$ is represented by blue dots with error bars. The orange curve is the theory for largest degree.

4 Random Walks and Preferential Attachment Π_{mix}

To simplify the code, we set the initial graph fully connected. If we have $\mathcal{N}$ vertices at the beginning, then the degree of each vertex is $\mathcal{N} - 1$. In this case, the random walks will stop at any existing vertices with equal probability. In fact, for $N \to \infty$, the initial graph won't affect the statistical results.

4.1 Degree Distribution Theory

The network growth mechanism is basically the same, but the probability of attachment changes to Π_{mix},

$$\begin{aligned}\Pi_{\text{mix}}(k,t) &= q\Pi_{\text{pa}}(k,t) + (1-q)\Pi_{\text{rnd}}(k,t) \\ &= q\frac{k}{2mN(t)} + (1-q)\frac{1}{N(t)},\end{aligned} \tag{28}$$

where q is the probability of choosing preferential attachment, which means the probability of choosing random attachment is $1-q$. If $q=0$, the model collapses to pure random attachment; If $q=1$, the model collapses to pure preferential attachment. To get rid of these extreme situation, we will focus on $q=0.5$. Substitute Eq.(28) into the master equation (Eq.(2)),

$$\begin{aligned}p_\infty(k) &= mN(t)\Big[\Big(\frac{q(k-1)}{2mN(t)} + \frac{1-q}{N(t)}\Big)p_\infty(k-1) - mN(t)\Big(\frac{qk}{2mN(t)} + \frac{1-q}{N(t)}\Big)p_\infty(k)\Big] + \delta_{k,m} \\ &= \Big[\frac{q(k-1)}{2} + m(1-q)\Big]p_\infty(k-1) - \Big[\frac{qk}{2} + m(1-q)\Big]p_\infty(k) + \delta_{k,m}.\end{aligned} \tag{29}$$

For $k > m$, $\delta_{k,m} = 0$, and substitute $q = 0.5$ into equation, we can get

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k + \frac{2m}{q} - 2m}{k + \frac{2m}{q} - 2m + \frac{2}{q} + 1} = \frac{k+2m}{k+2m+5}, \tag{30}$$

This equation has a solution in form of Gamma Function:

$$p_\infty(k) = A\frac{\Gamma(k + \frac{2m}{q} - 2m)}{\Gamma(k + \frac{2m}{q} - 2m + \frac{2}{q} + 1)} = A\frac{\Gamma(k+2m)}{\Gamma(k+2m+5)}. \tag{31}$$

To find out the coefficient A, let $q = 0.5$ and $k = m+1$,

$$\begin{aligned}p_\infty(m+1) &= A\frac{\Gamma(3m+1)}{\Gamma(3m+6)} \\ &= \frac{A}{(3m+5)(3m+4)(3m+3)(3m+2)(3m+1)}.\end{aligned} \tag{32}$$

Let's consider the boundary condition, i.e. $k = m$, and again assume $p_\infty(k < m) = 0$, as $N \to \infty$,

$$p_\infty(m) = -\Big[\frac{qk}{2} + m(1-q)\Big]p_\infty(k) + 1 = \frac{4}{3m+4}. \tag{33}$$

Substitute $k = m+1$ into Eq.(30),

$$\begin{aligned}p_\infty(m+1) &= \frac{3m}{3m+5}p_\infty(m) \\ &= \frac{3m}{3m+5}\frac{4}{3m+4}.\end{aligned} \tag{34}$$

The coefficient A can be easily obtained by comparing Eq.(32) and Eq.(34),

$$A = 12(3m+3)(3m+2)(3m+1)m. \tag{35}$$

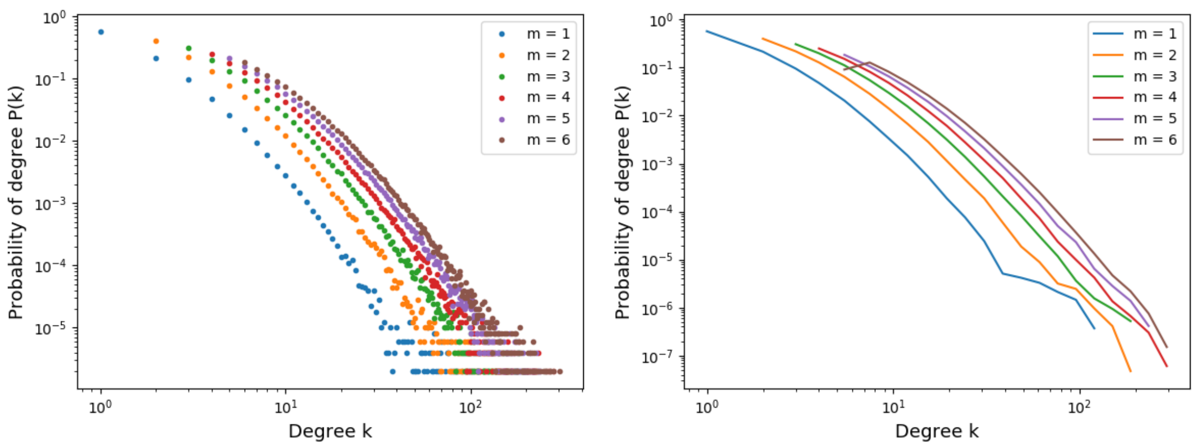
s A is known, the Eq.(31) becomes

$$p_\infty(k) = \frac{12(3m+3)(3m+2)(3m+1)m}{(k+2m+4)(k+2m+3)(k+2m+2)(k+2m+1)(k+2m)} \tag{36}$$

For $k \to \infty$, the degree distribution has a power law of degree, i.e. $p_\infty(k) \propto k^{-5}$. The distribution of degree p_∞ is normalised, for it is derived from Π_{mix}, which is the combination of two normalised probabilities Π_{pa} and Π_{rnd}.

4.2 Numerical results

The setting and factors are exactly the same as the one we set in **section2**. The initial vertices $\mathcal{N} = 16$, the added number $N = 10^5$, added edges $m = 1, 2, 3, 4, 5, 6$, the binned factor $a = 1.25$. The visual illustrations of distribution of degree is shown in Fig.9. The figure on the left shows the unbinned distribution, and the figure on the right shows the binned data. The curves in Fig.9b can be considered a straight line (log-scale) when degree k is large, which means the distribution of degree has a power law of degree k. When k is small, the distribution curve follows an exponential decay. This is consistent with our theory. The fat-tail is observable but not very obvious.



(a) Unbinned (b) Binned

Figure 9: (a) and (b) show the probability of degree distribution for $m = 1, 2, 3, 4, 5, 6$.

The KS test was performed for each m value. The results are shown in Fig.10, making comparison between theory and data. The D-values and p-values are given as well. Again, the results have the same problem we came across in **section2.3**, and the same methods can be applies to solve it. In Table.2, the p-values of the first set are all smaller than 0.05, demonstrating the null hypothesis is false; The D-statistic from the second set are about 0.1, indicating the theory fits the numerical data well.

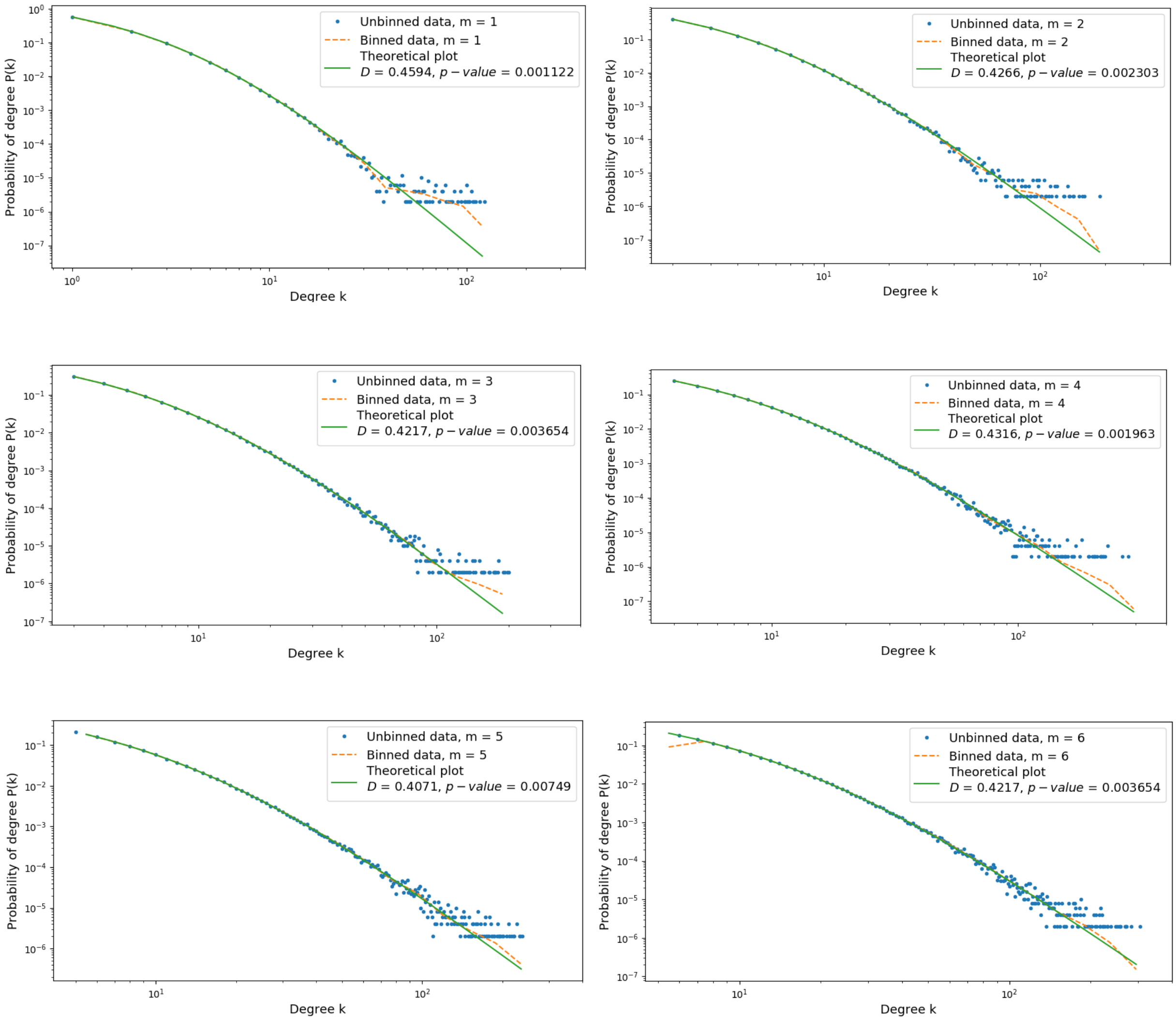


Figure 10: Degree Distribution with $N = 10^5$ vertices for each m value. The blue dots are unbinned data; The orange dash line is the binned data with scale factor $a = 1.25$; The green line is the theoretical value.

	1^{st} set		2^{nd} set	
m	D-statistic	p-value	D-statistic	p-value
1	0.459	0.0011	0.166	0.9447
2	0.427	0.0023	0.105	0.9998
3	0.422	0.0036	0.111	0.9997
4	0.432	0.0020	0.053	1
5	0.407	0.0007	0.059	1
6	0.421	0.0037	0.056	1

Table 2: KS test for two sets of theoretical data

5 Conclusion

Three kinds of attachments are investigated in network growing model: **1.** pure preferential attachment, **2.** pure random attachment, **3.** random walks and preferential attachment (mixed attachment). The degree distribution for the three attachments $p(k)$ are: **1.** power law decay ($p(k) \propto k^{-3}$), **2.** exponential decay ($p(k) \propto (m/(m+1))^{k-m}$), **3.** power law decay ($p(k) \propto k^{-5}$). The fat-tail is obvious in pure preferential attachment and random walks and preferential attachment. There is no fat-tail in pure random attachment. The model is limited by finite-size effect, and all theories are based on the assumption: $N \to \infty$. The N value we use is 10^5. If a larger N is applied, for example $N = 10^6$, the statistical results could be better, and the features of the BA model are more observes.

References

[1] A.-L.Barabás and R.Albert, *Emergence of scaling in random networks*, Science. 286, 173(1999)

[2] Tim Evans *Network project student notes*, Imperial College London, 2020.

[3] G.U.Yule, *A mathematical theory of evolution based on the conclusions of Dr. J.C. Wills*, F.R.S. Phil. Trans. B, **21-87** (1925)

[4] H.A. Simon, *On a class of skew distribution function*, Biometrica, **42** 425 (1955)