EE 232E Graphs and Network Flows

HW1

Weikun Han 804774358
Duzhi Chen 004773782
2017/4/18

Introduction

In this assignment, there are some practices of creating networks and measuring their properties. First, we used R interface to create random undirected networks and checked their connectivity. Then we added a fat-tailed degree distribution to an undirected network. After checking its connectivity, we expanded the network and compared the modularities of the new network and the original one. In the part 3, we created a random graph by simulating its evolution and used fast greedy method to find the community structure. At last, we created a directed network by using the forest fire model and measured its diameter and community structure.

Questions

1(a) Create three undirected random networks with 1000 nodes, and the probability p for drawing an edge between two arbitrary vertices 0.01, 0.05 and 0.1 respectively. Plot the degree distributions.

For every network, we plot a histogram diagram and a scatter-point diagram.

i) For p = 0.01,

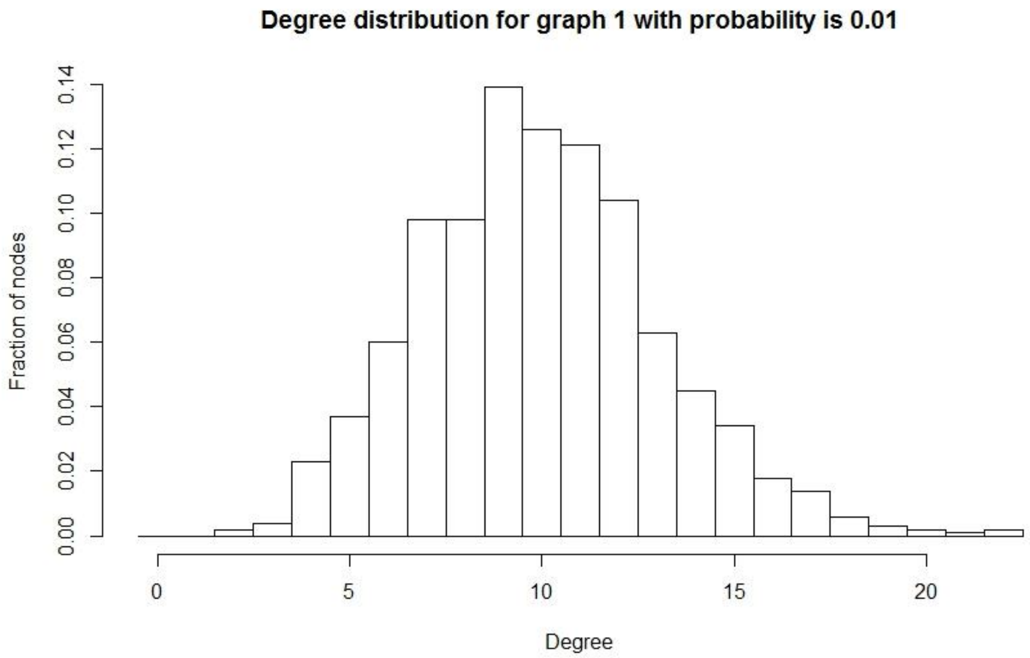

Figure 1, the histogram of the degree distribution for p = 0.01

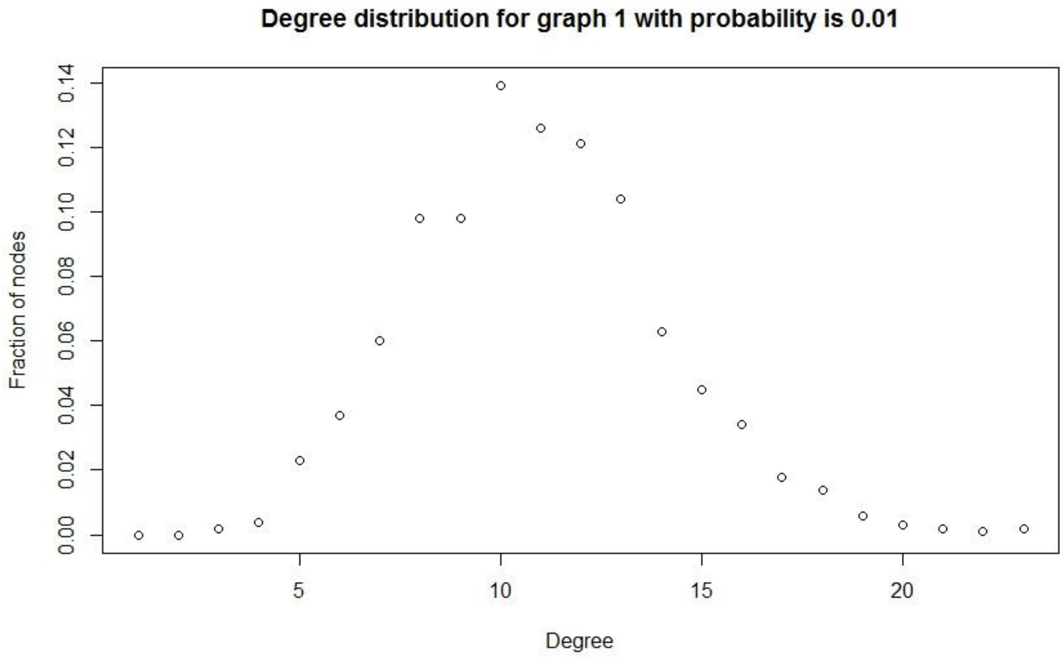

Figure 2, the scatter-point of the degree distribution for p = 0.01

ii) For p = 0.05,

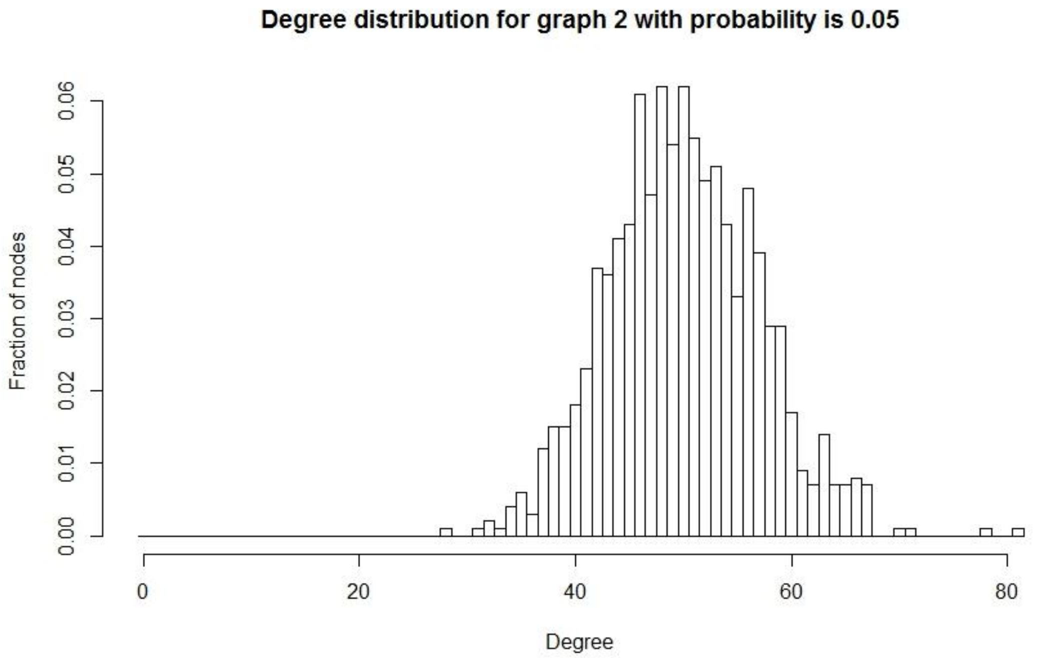

Figure 3, the histogram of the degree distribution for p = 0.05

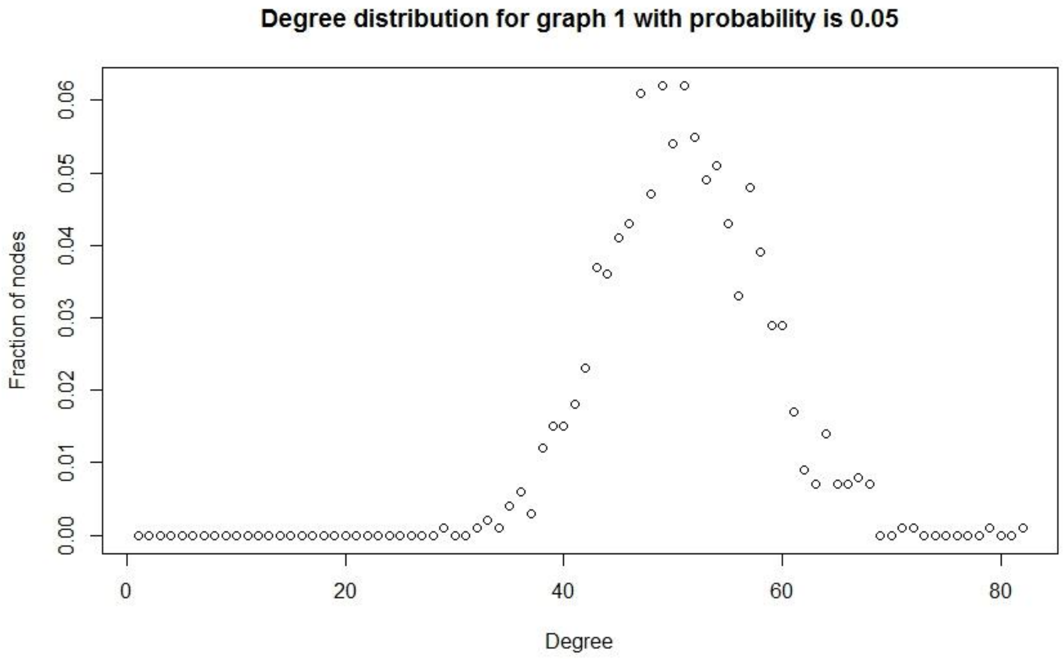

Figure 4, the scatter-point of the degree distribution for p = 0.05

iii) For p = 0.1,

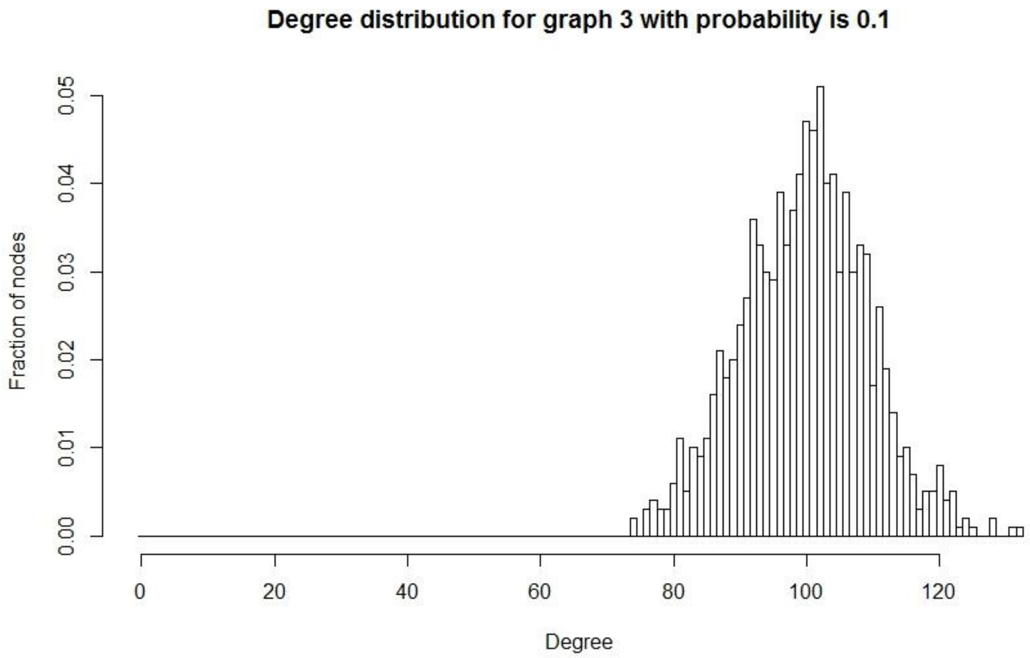

Figure 5, the histogram of the degree distribution for p = 0.1

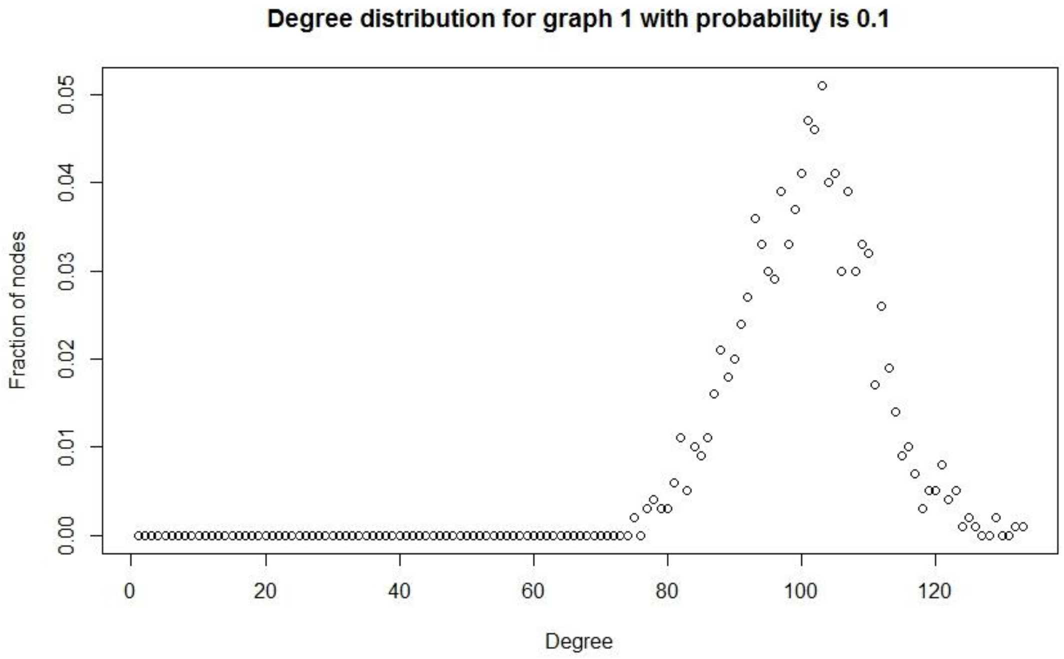

Figure 6, the scatter-point of the degree distribution for p = 0.1

From these 6 figures above, we observe that as the probability p increases, the degree increases.

1(b) Are these networks connected or disconnected? What are the diameters of these networks?

i) For p = 0.01,
   the network is connected;
   the diameter is 5,
   and a path with this diameter is: 2 53 25 114 412 849.
ii) For p = 0.05,
   the network is connected;
   the diameter is 3,
   and a path with this diameter is: 1 7 2 170.
iii) For p = 0.1,
   the network is connected;
   the diameter is 3,
   and a path with this diameter is:29 3 87 875.

1(c) Try to numerically find a value p_c (to three significant figures), so that when $p < p_c$ the generated random networks are disconnected, and when $p > p_c$ the generated random networks are connected.

We create a random unconnected network with initial p = 0.001, and increase p by 0.001 each trial until it is connected. We finally have the threshold pc is about equal to 0.006.

1(d) Can you analytically derive the value of pc?

According to the formula of threshold of connectivity

$$pc = \ln(n)/n$$

where n = 1000. Then we have pc = ln(1000)/1000 = 0.00691, which is close to our result 0.006 from part (c).

2(a) Create an undirected network with 1000 nodes, whose degree distribution is proportional to x^-3. Plot the degree distribution. What is the diameter?

Figure 7 and 8 are the histogram and scatter-point of the degree distribution.

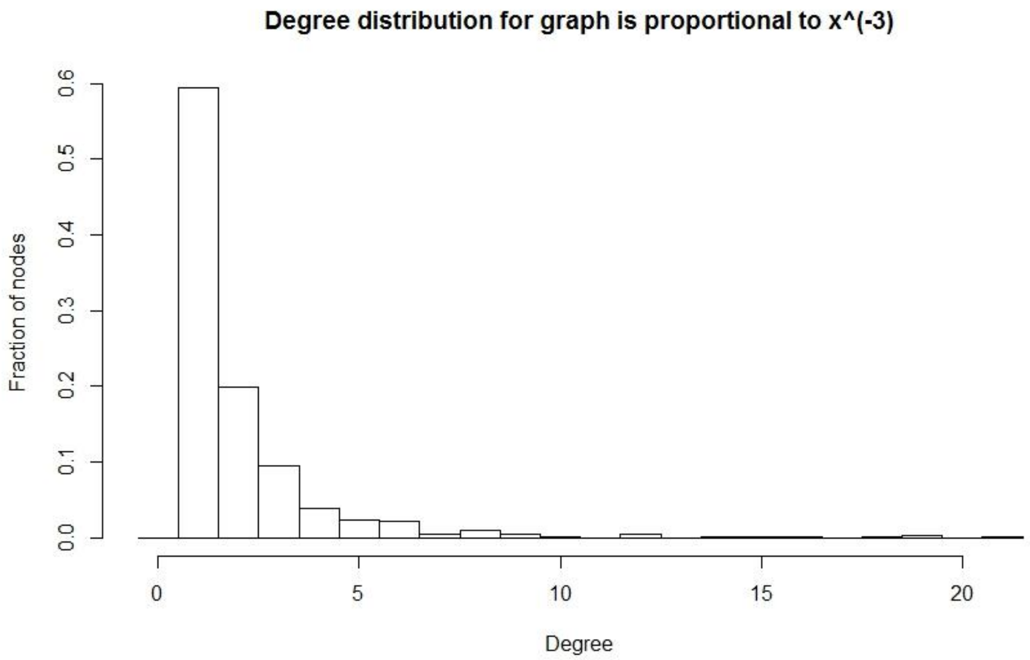

Figure 7, the histogram of the degree distribution

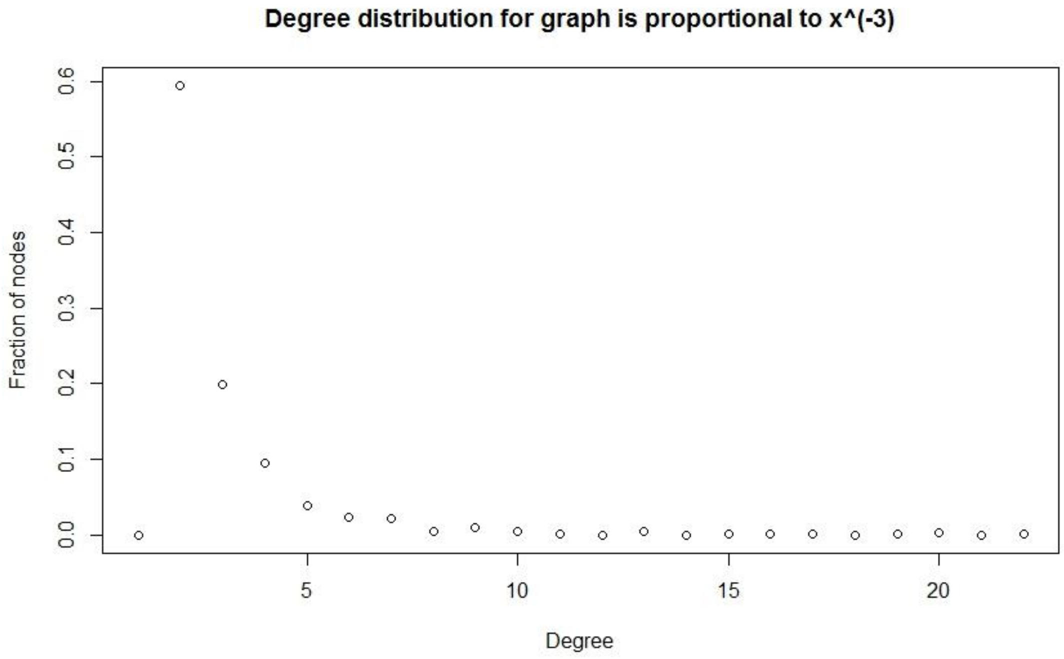

Figure 8, the scatter-point of the degree distribution

The diameter of this network is 20, and a path with this diameter is: 766 568 464 401 272 184 83 19 5 2 1 6 8 18 33 67 210 256 443 892 982

2(b) Is the network connected? Find the giant connected component (GCC) and use fast greedy method to find the community structure. Measure the modularity. Why is the modularity so large?

This network is connected. Figure 9 is the giant connected component (GCC). Then we use fast greedy method to find the community structure. Figure 10 is the community structure. And the modularity is 0.933463. There are three reasons for the modularity being large. The first and main reason is the number of nodes is large; secondly, the graph is undirected; and the third one is that the graph is connected. Which lead to the result that connections between nodes with different modules are sparse and the connections between nodes with the same module are dense.

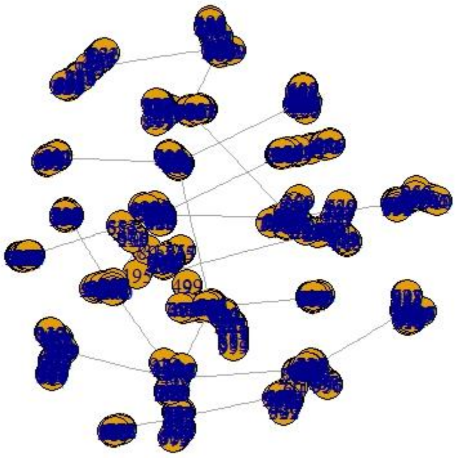

Figure 9, the giant connected component (GCC)

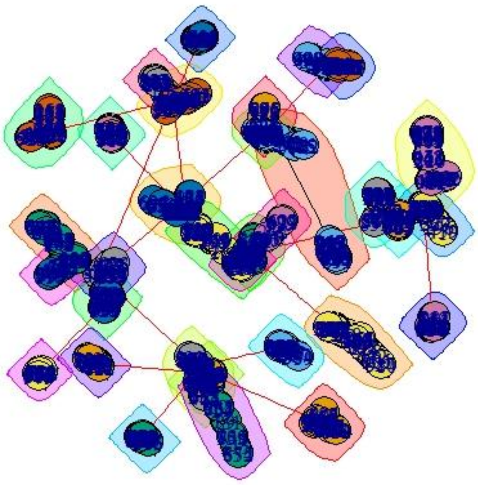

Figure 10, the community structure

2(c) Try to generate a larger network with 10000 nodes whose degree distribution is proportional to x^-3. Compute the modularity. Is it the same as the smaller network's?

We expand the network to one with 10000 nodes. This network is connected and the modularity is 0.9782348 which is higher than the modularity of the smaller one.

2(d) You can randomly pick a node i, and then randomly pick a neighbor j of that node. Measure and plot the degree distribution of nodes j that are picked with this process.

Figure 11 and 12 below are the plots of degree distribution of nodes j what are picked with the process described above.

As shown in these two figures, we can see that the degree is very small.

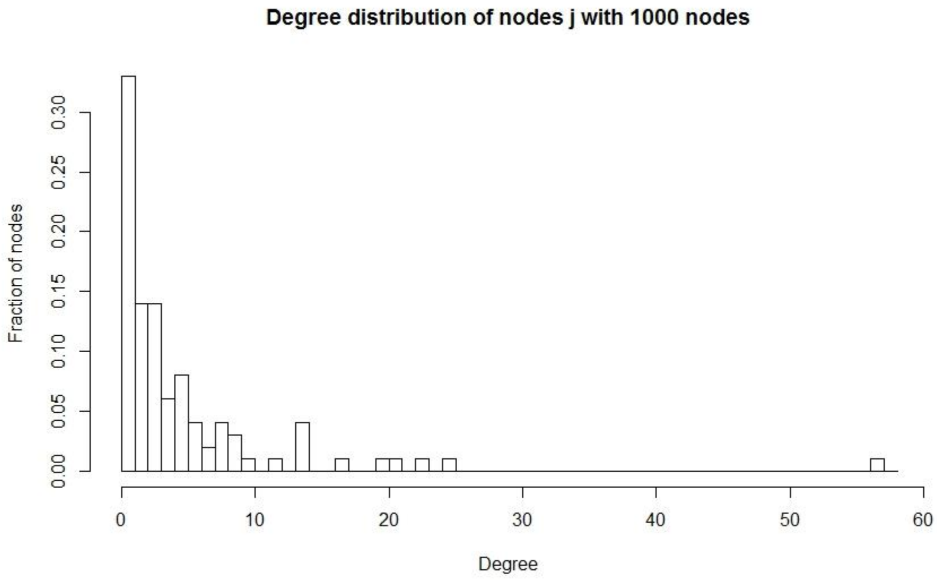

Figure 11, the histogram of the degree distribution of nodes j

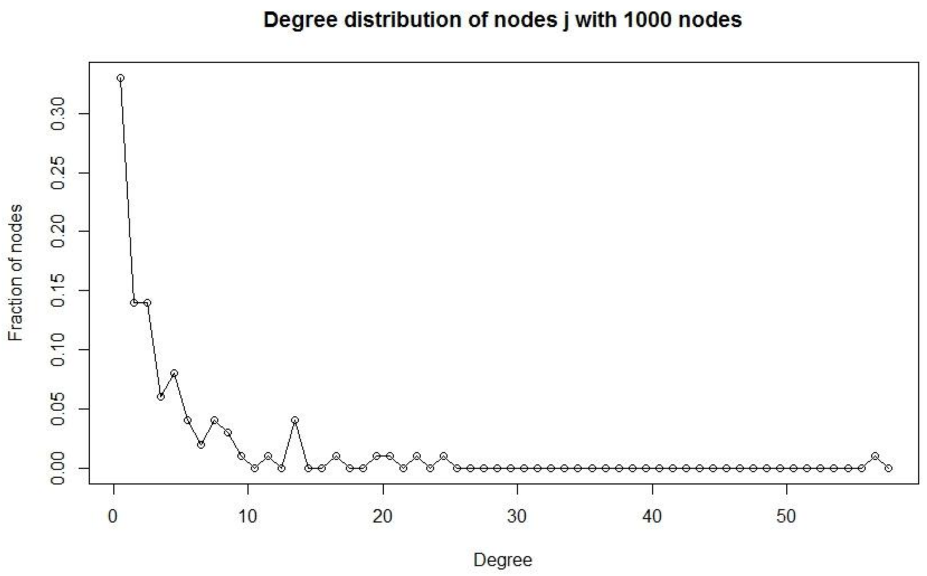

Figure 12, the line chart of the degree distribution of nodes j

3(a) Each time a new vertex is added it creates a few links to old vertices and the probability that an old vertex is cited depends on its in-degree (preferential attachment) and age. Produce such an undirected network with 1000 nodes. Plot the degree distribution.

We create an evolving random undirected graph with 1000 nodes, preferential attachment exponent of 1, aging exponent of -1 and 1000 bins.
Figure 13 and 14 are the plots of the degree distribution of this graph.

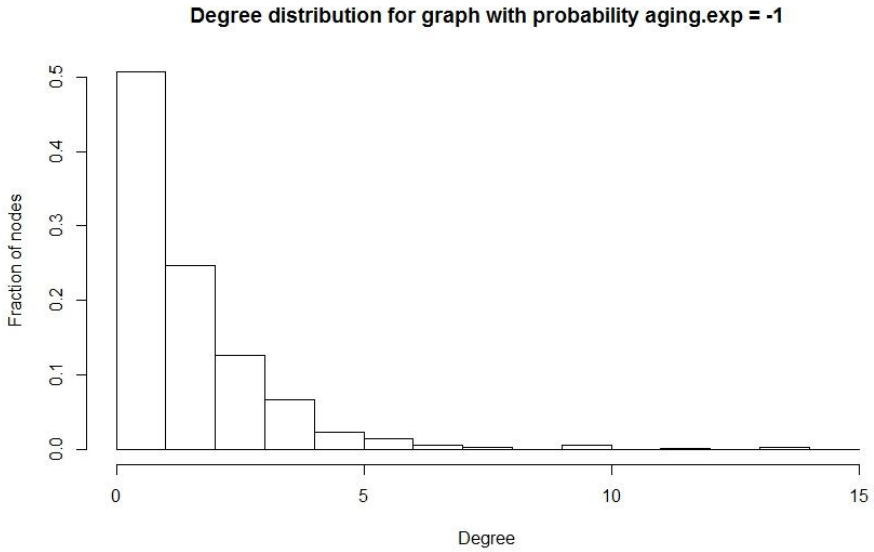

Figure 13, the histogram of the degree distribution of the random graph

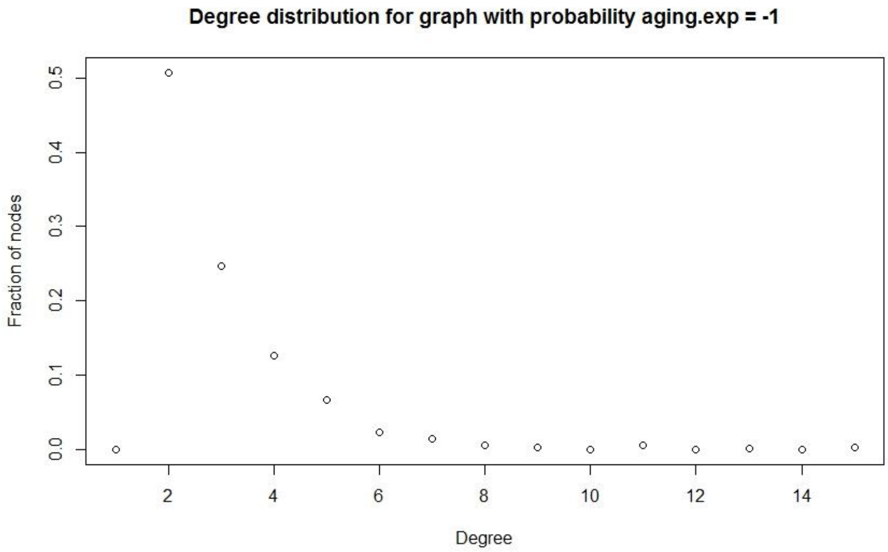

Figure 14, the scatter-point of the degree distribution of the random graph

3(b) Use fast greedy method to find the community structure. What is the modularity?

We use the fast-greedy algorithm to find the community structure and compute the modularity.

Figure 15 is the GCC and Figure 16 is the community structure. The modularity is 0.9347871.

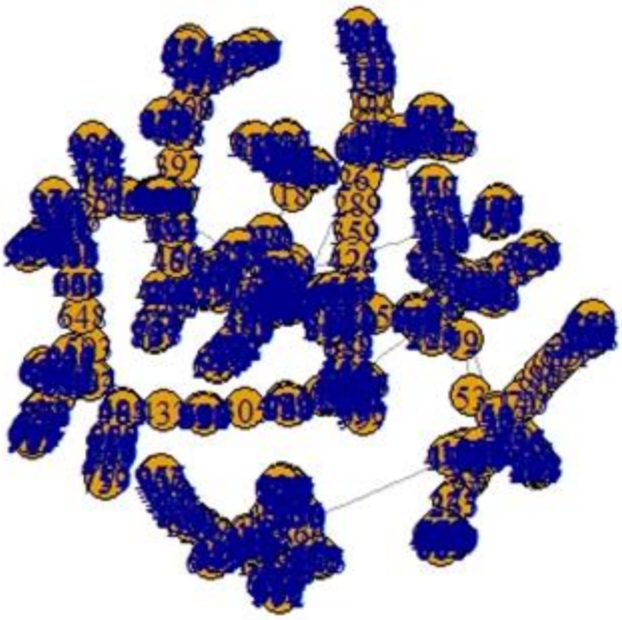

Figure 15, the giant connected component (GCC)

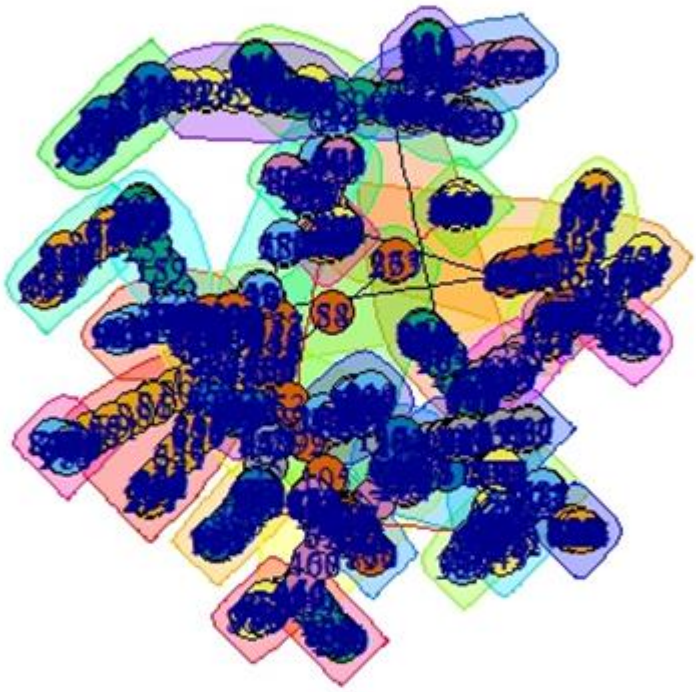

Figure 16, the community structure

4(a) This is a growing network model, which resembles how the forest fire spreads by igniting trees close by. Plot the in and out degree distributions.

We use the forest fire model to simulate a random growing network with 1000 nodes. Figure 13 and 14 are the plots of the input degree distribution; Figure 15 and 16 are the plots of the output degree distribution. From these images, we can see that the in and out degree distributions are similar.

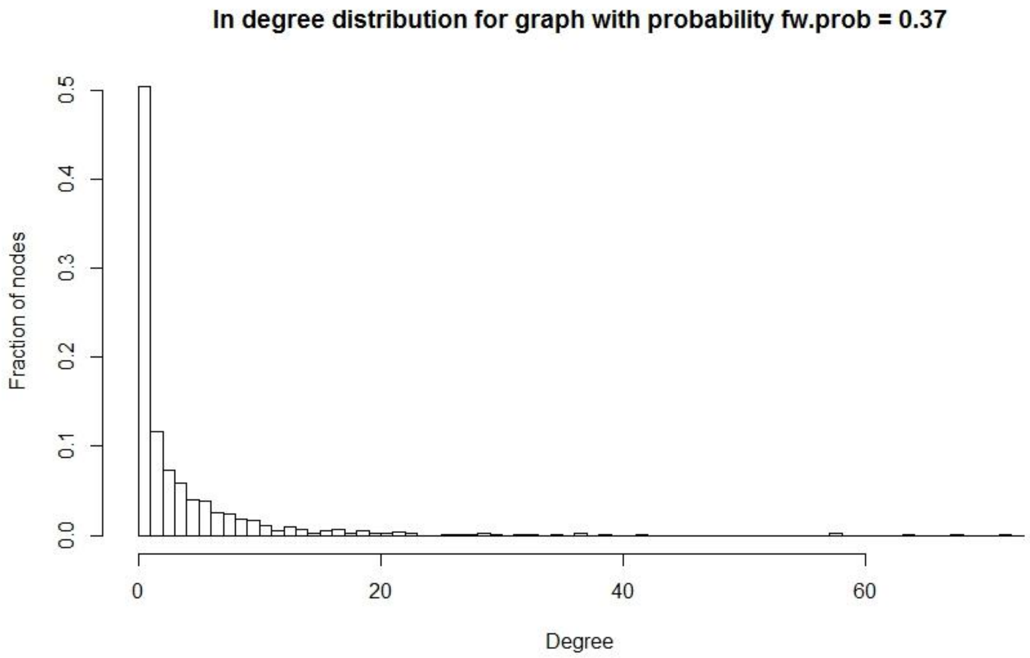

Figure 13, the histogram of the in-degree-distribution

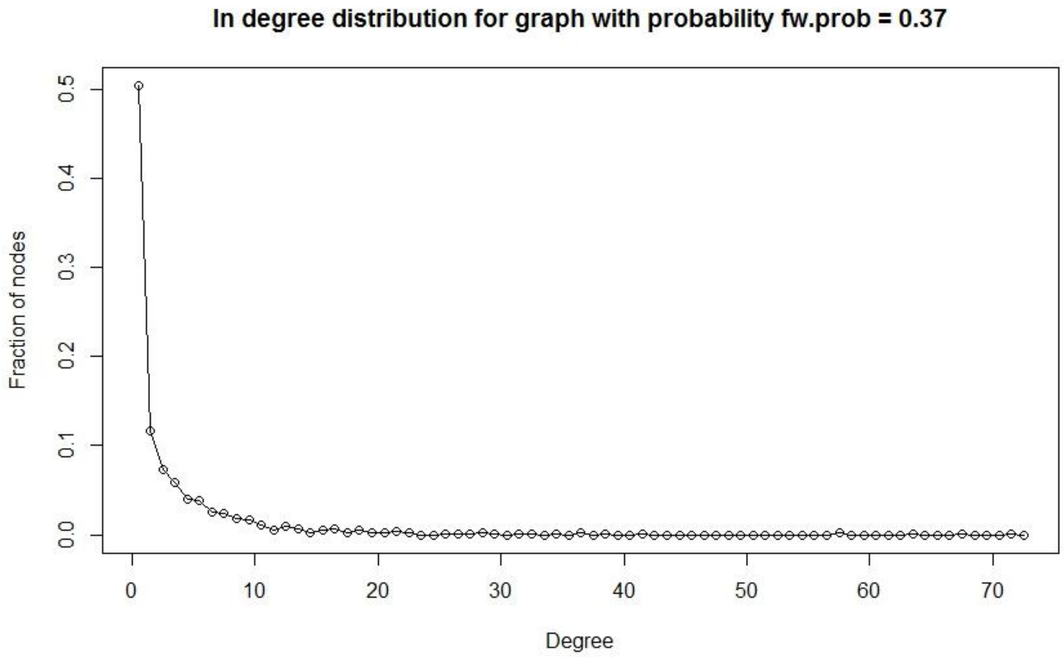

Figure 14, the scatter-point of the in-degree-distribution

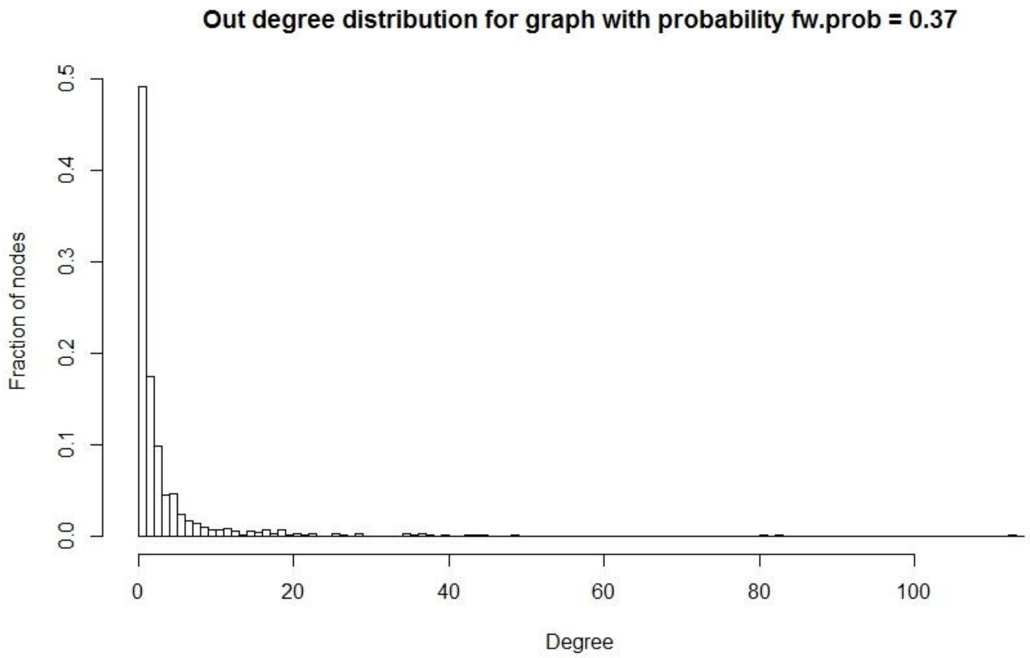

Figure 15, the histogram of the out-degree-distribution

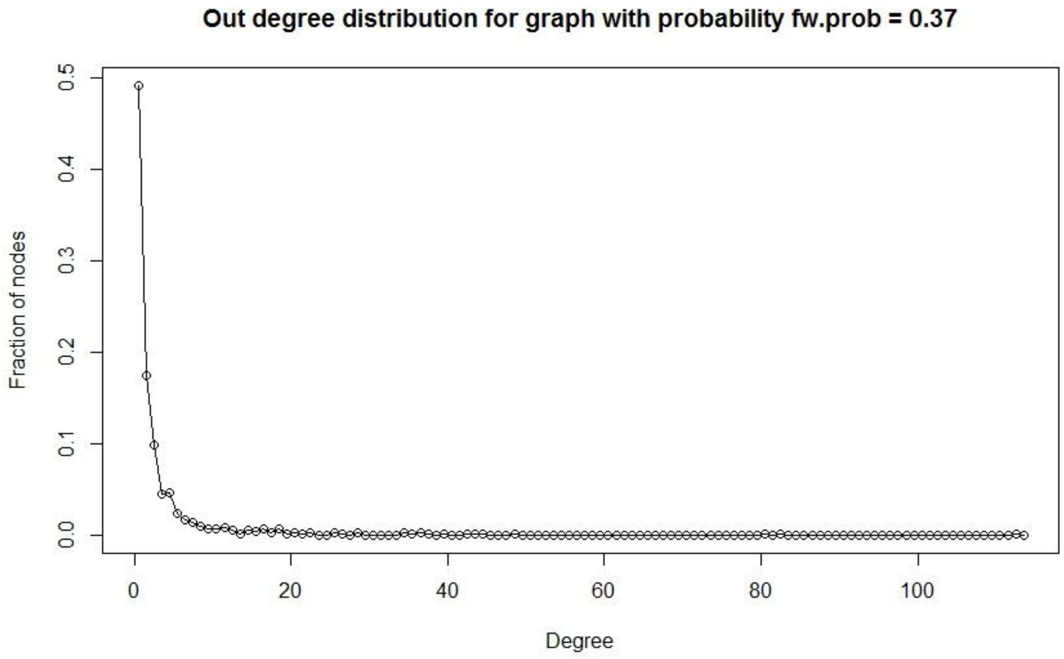

Figure 16, the scatter-point of the out-degree-distribution

4(b) Measure the diameter.

The diameter is 11 and a path with this diameter is: 767 299 277 156 134 32 12 10 8 3 2 1.

4(c) Measure the community structure and modularity.

Figure 17 is the GCC and Figure 18 is the community structure.
The length of the community structure is 24, and the modularity is 0.4862787.

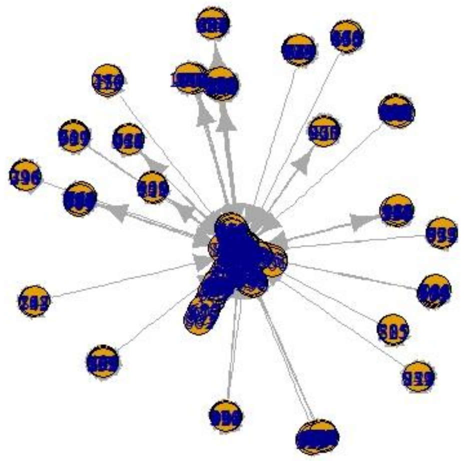

Figure 17, the giant connected component (GCC)

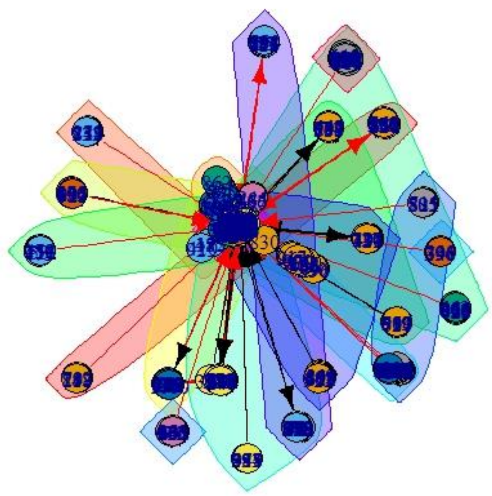

Figure 18, the community structure

Conclusion

In general, the results meet our expectations.