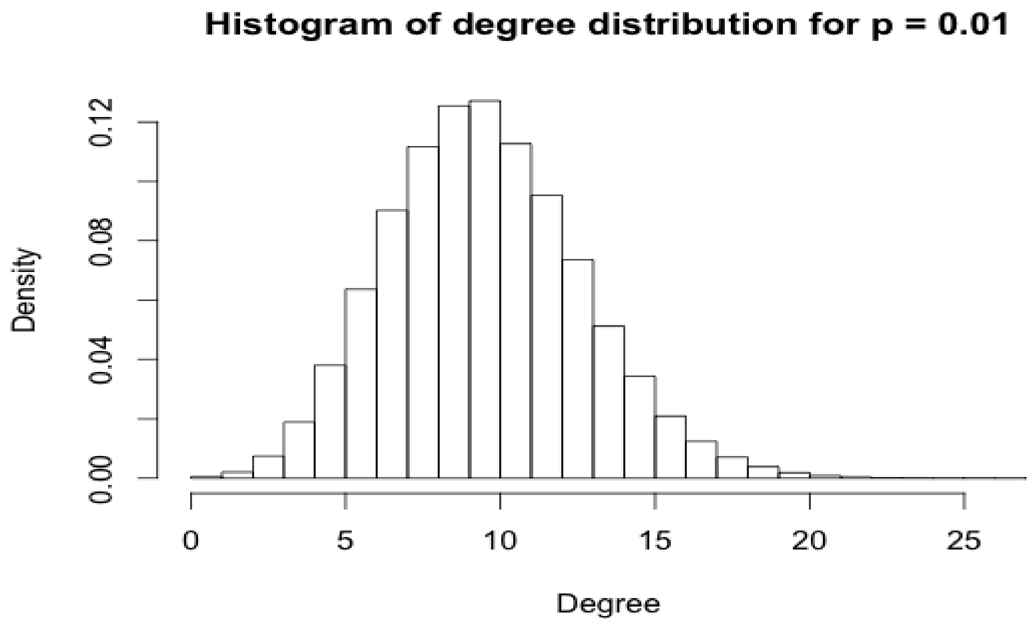
# **Homework 1**

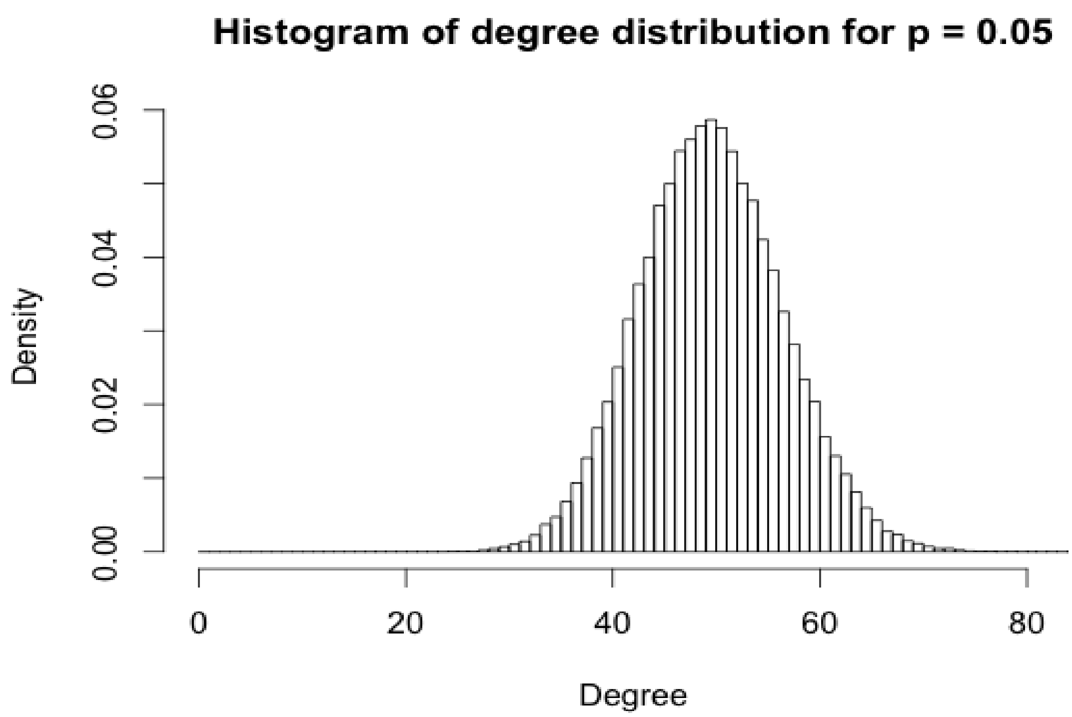
EE232E - Graphs and Network Flows

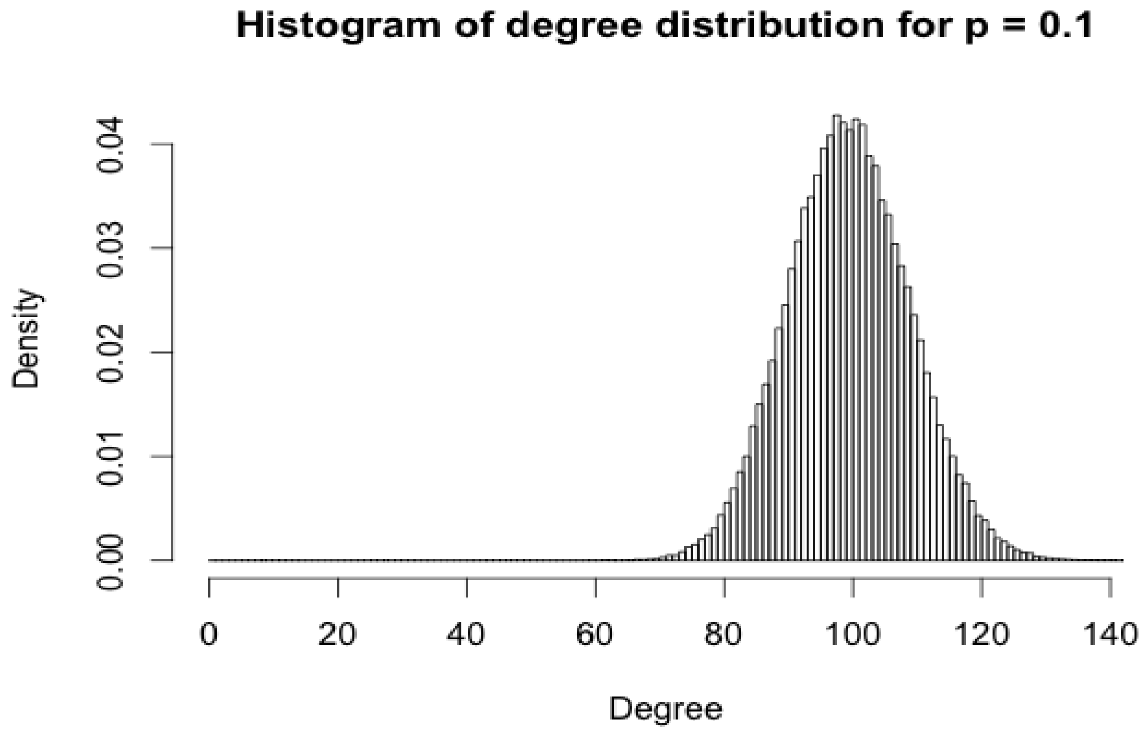
Team:
Rutuja Ubale (UID: 404558257)
Usha Amrutha N (UID: 204590772 )
Pallavi Chakraborty (UID: 404519609)

**Q1. Random networks**

a.) In this section we created undirected random networks using the **random.graph.game** function with 1000 nodes and by varying the probability of drawing an edge between two nodes. The degree distributions of networks generated with probability p for drawing an edge between two arbitrary vertices as 0.01, 0.05 and 0.1 are shown below.



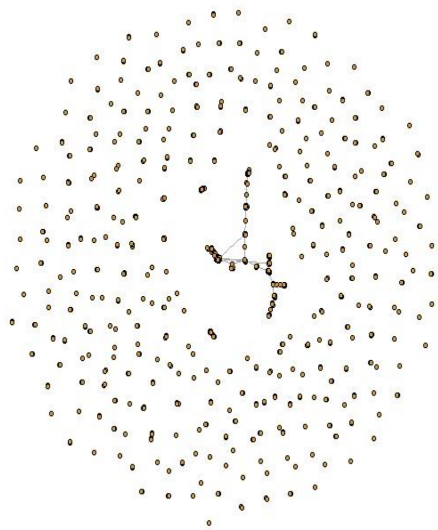



b.) Using **is.connected method** in R we note that all three networks are connected. The value of p is crucial in forming a connected or a disconnected network. In all the three cases the probability p was higher then the threshold value $p_c$ for connectedness of the network. It was seen that as the value of p increases the graph becomes more and more connected. **The average diameters of the networks with p = 0.01, 0.05, 0.1 are 5.38, 3, 3 respectively.**

c.) To find the threshold value $p_c$ for which the graph becomes connected the value of probability p was swept from 0 in steps of 0.001. Since the networks are randomly generated the experiment was repeated 10 times and $p_c$ value which occurred maximum number of times is finally reported. The value of pc such that when $p < p_c$ the generated random networks are disconnected, and when $p > p_c$ the generated random networks are connected was found to be **0.007**.

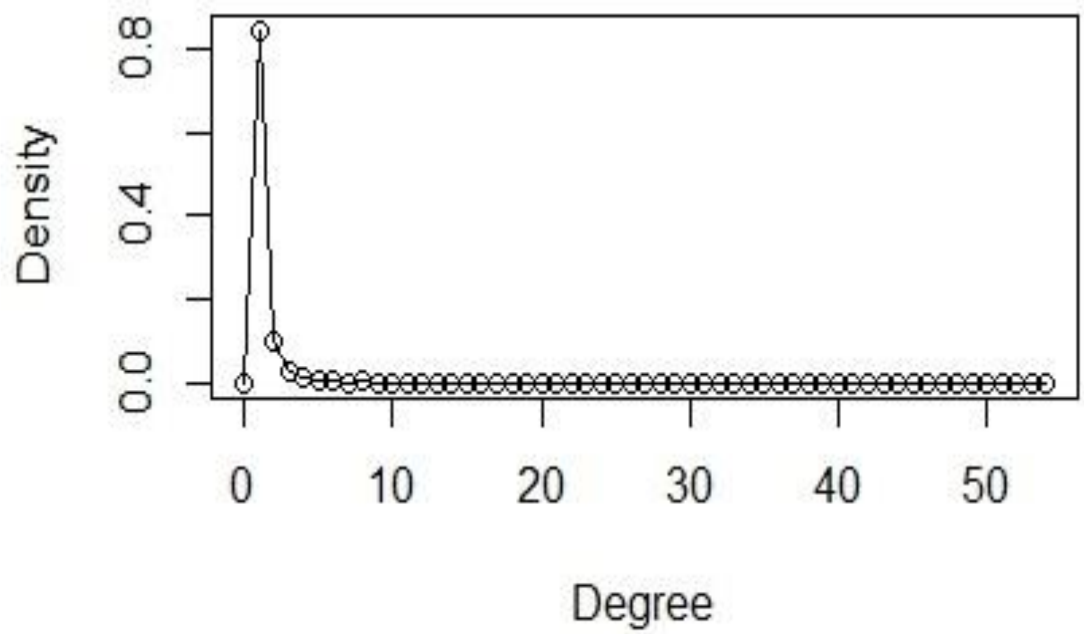
**Question 2. Fat-Tailed Distribution**

a.) In this question, we create a network consisting of 1000 nodes following a power-law distribution. Firstly, we create a distribution following power law, here we have a degree distribution proportional to (x^-3). Next, we sample 1000 points with replacement, which will be considered as nodes in the network. These 1000 nodes are nodes whose degree distribution follows the power law. The nodes are connected to one another through yet another random matching process. This is done using the **sequence.game model in R**. The graph below demonstrates the network generated through power-law distribution. It can be noted that the nodes are mostly disconnected (or sparse). We can also identify the GCC in the middle of the component.

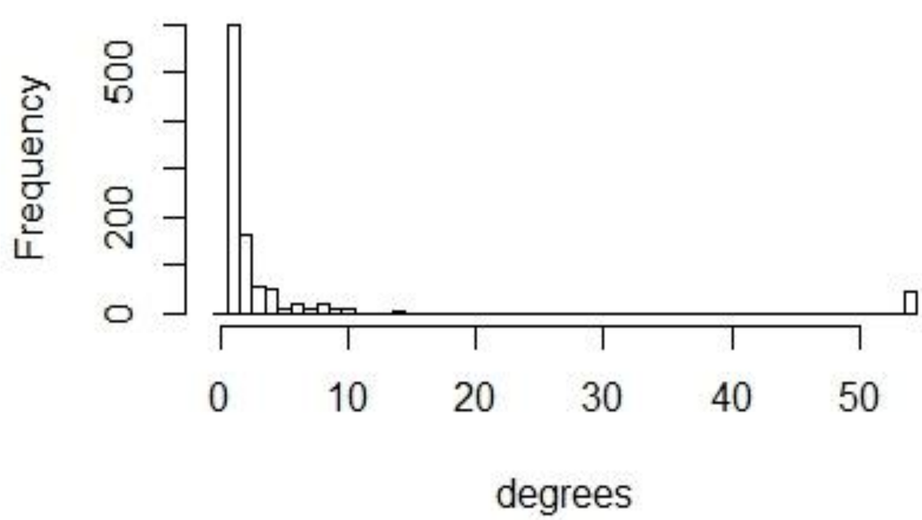


The density vs degrees of distribution and histogram plots are as shown below. We note that the distribution plots follow power law distribution. **The diameter of the above generated network is 17.**

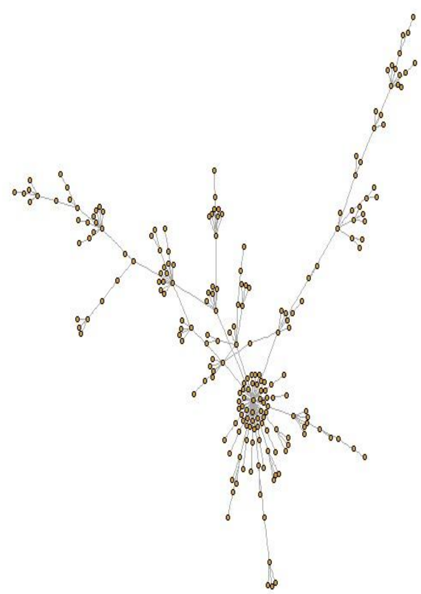
**Degree Distribution with 1000 nodes**



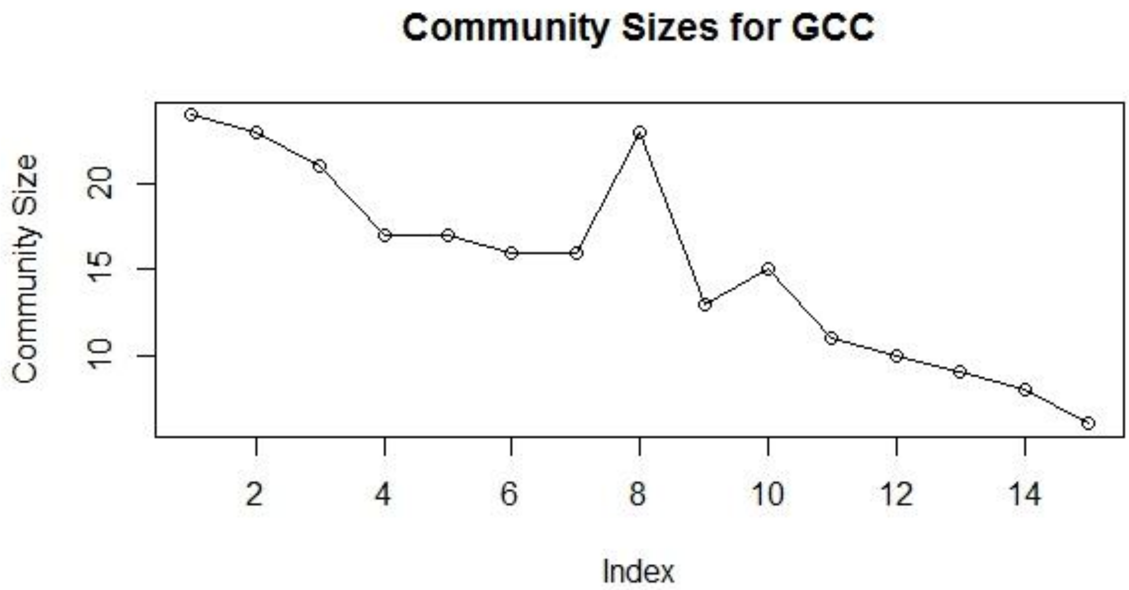

**Histogram of degrees**



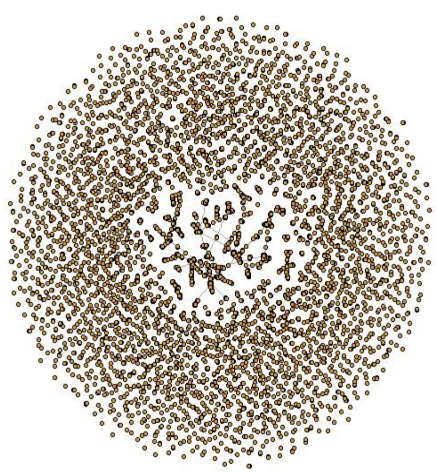
b.) The network is disconnected as we note from the network plot that the there are many nodes which are isolated (or sparse). The giant connected component in the middle is shown in the figure below. There **are 213 nodes in the GCC**. We use the fast greedy method to find the community structure. The community sizes for GCC was found to be 26 communities. The aim of this method is to partition the network into communities. While doing so, we compute the modularity score. Modularity score implies how well the network is partitioned. **The modularity score of the above graph is 0.96428.** High modularity implies that the network is very well clustered (or partitioned) into communities due preferential attachment of nodes to higher degree nodes. Also, High modularity of the GCC denotes that it has dense connections between the nodes within the communities but sparse connections between nodes in different communities, which can be verified from the network graph (above). The GCC is as shown below.



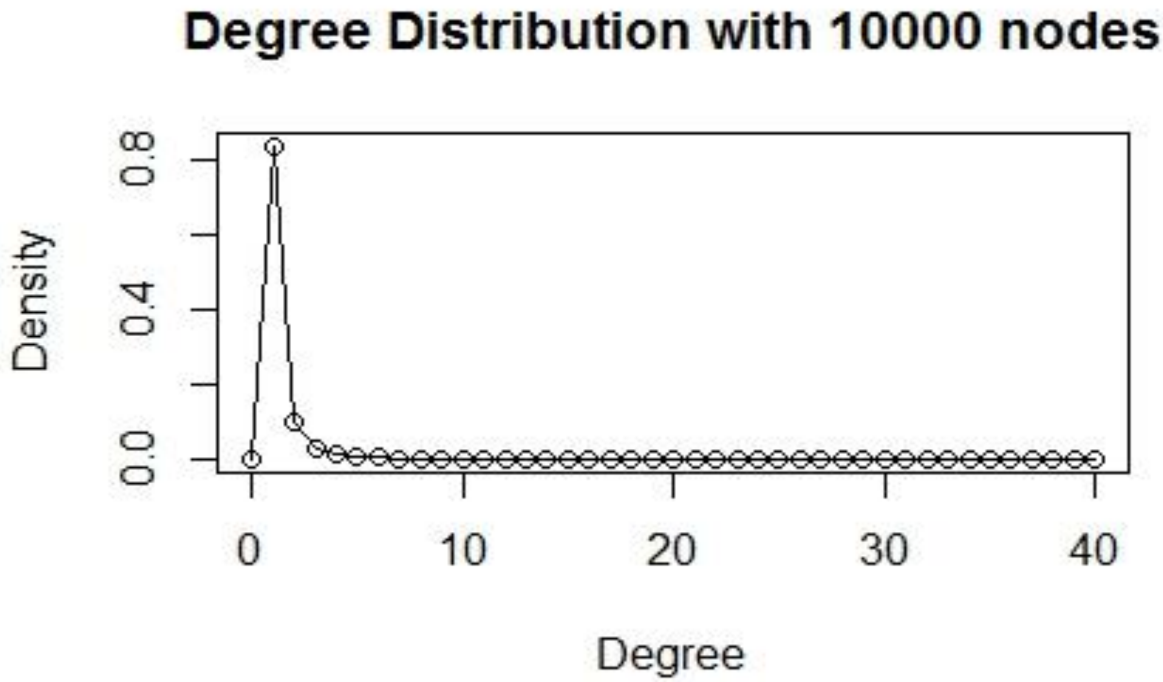
The community size for GCC based on fast greedy algorithm was **26 communities**. And the community structure is plotted as below.

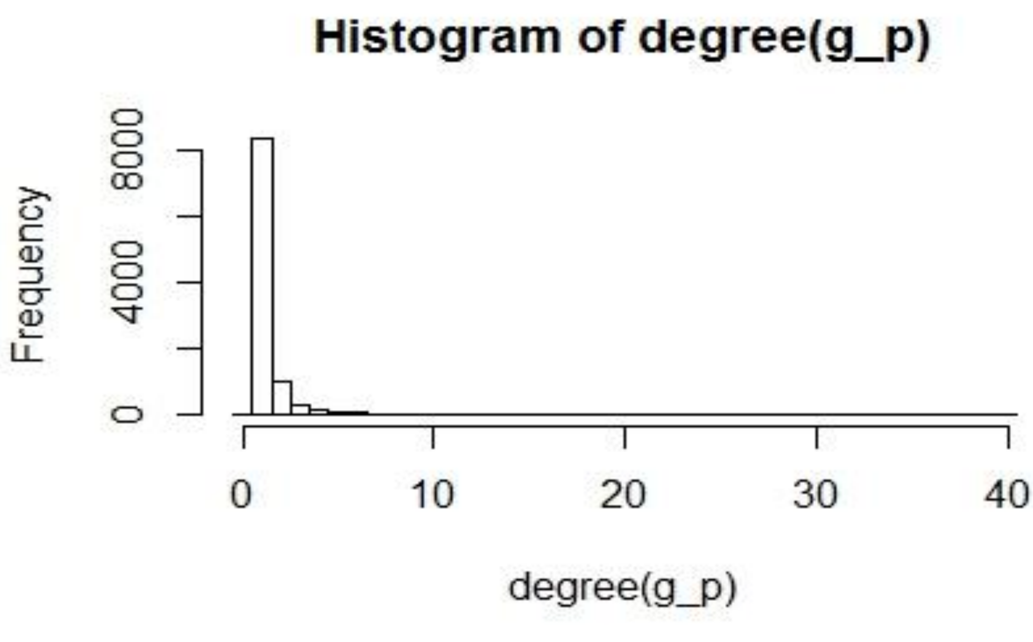



c.) The random network generated with 10000 nodes is as follows.

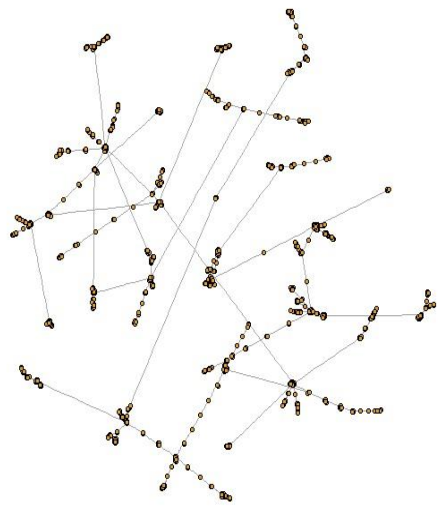


The random network is not connected since there are many isolated (sparse) nodes with a GCC in the middle. **The diameter of the given the network is 33**. The density vs. degree distributions and histograms are given below.

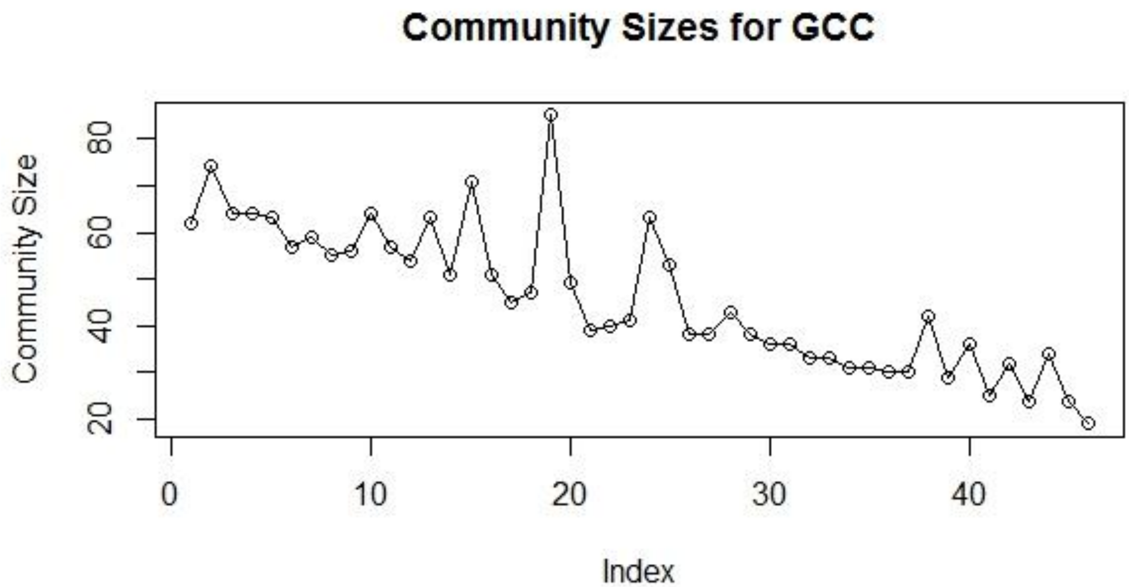


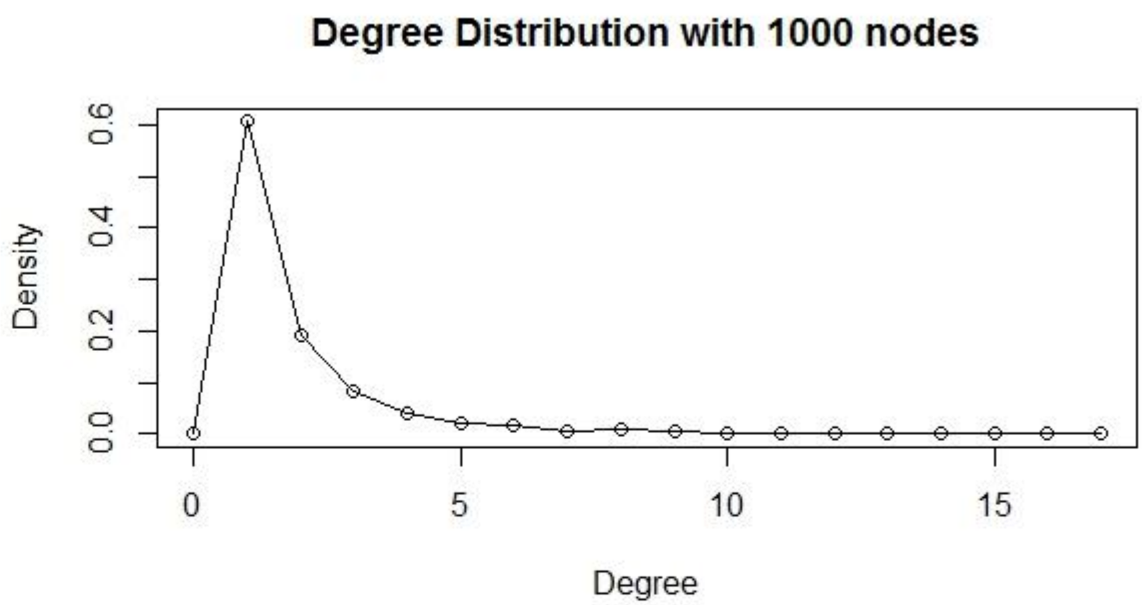

The giant connected component GCC (below) has 1302 nodes

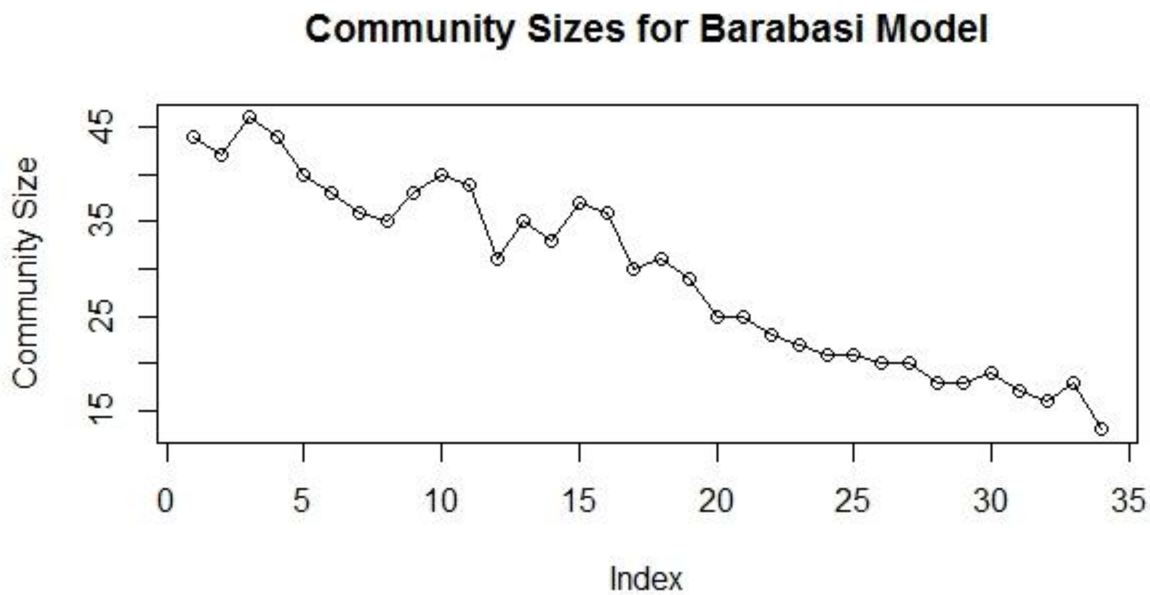


The network is disconnected as we note from the network plot that the there are many nodes which are isolated (or sparse). The giant connected component in the middle is shown in the figure below. There **are 1302 nodes in the GCC**. We use the fast greedy method to find the community structure.  The aim of this method is to partition the network into communities. While doing so, we compute the modularity score. Modularity score implies how well the network is partitioned. **The modularity score of the above graph is 0.9882.** High modularity implies that the network is very well clustered (or partitioned) into communities due preferential attachment of nodes to higher degree nodes. Also, High modularity of the GCC denotes that it has dense connections between the nodes within the communities but sparse connections between nodes in different communities, which can be verified from the network graph (above).
The community size for GCC based on fast greedy algorithm was **46 communities**. And the community structure is plotted as below

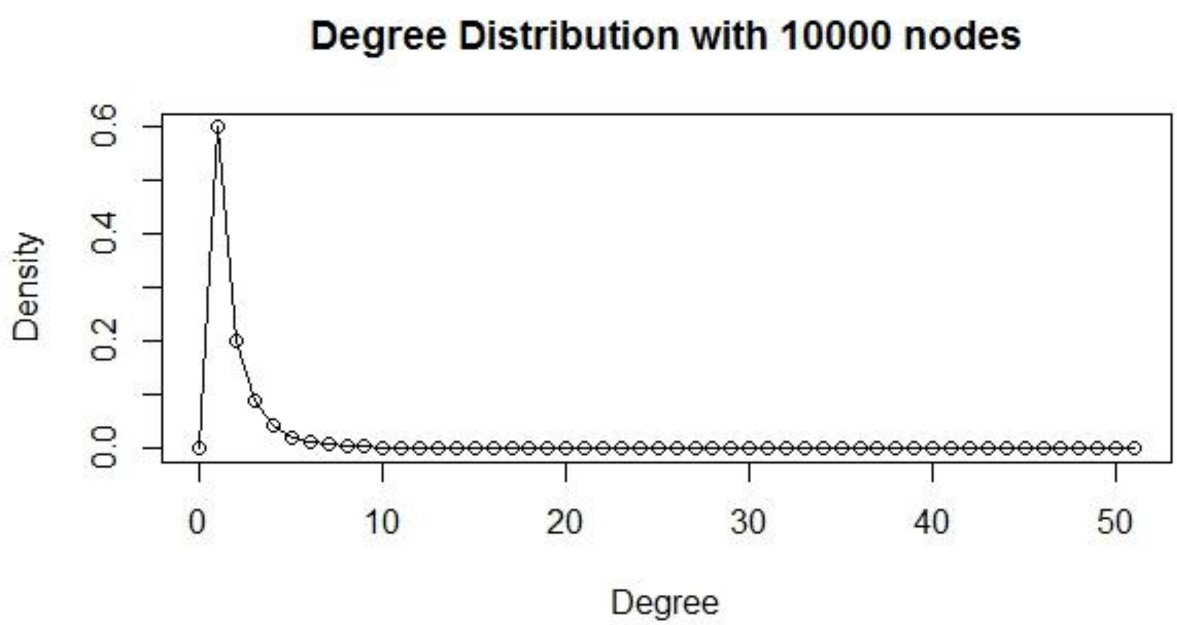


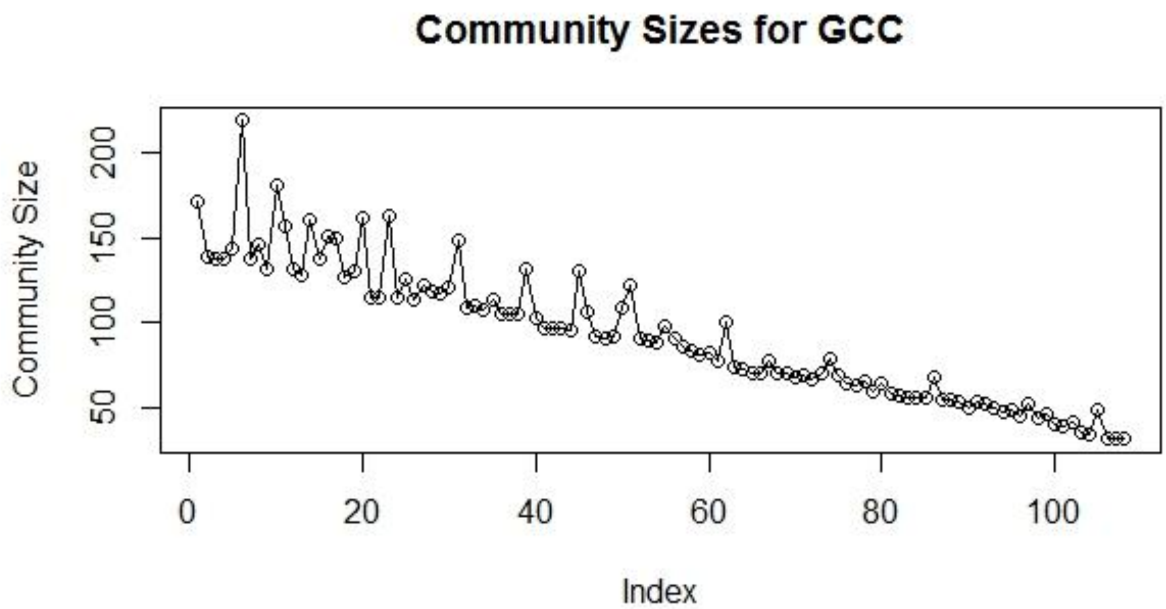

**Note: We have conducted experiments using Barabasi network model as well. Th which follows power law distribution. The random networks for this model were 100% connected, and hence GCC was the entire network. The modularity for 1000 and 10000 node random network were 0.934 with 34 communities and 0.977 with 105 communities respectively. The diameter that we got using this model was 17. The degree distributions for 1000 and 10000 nodes and the community structure plots using Barabasi model are shown below.**



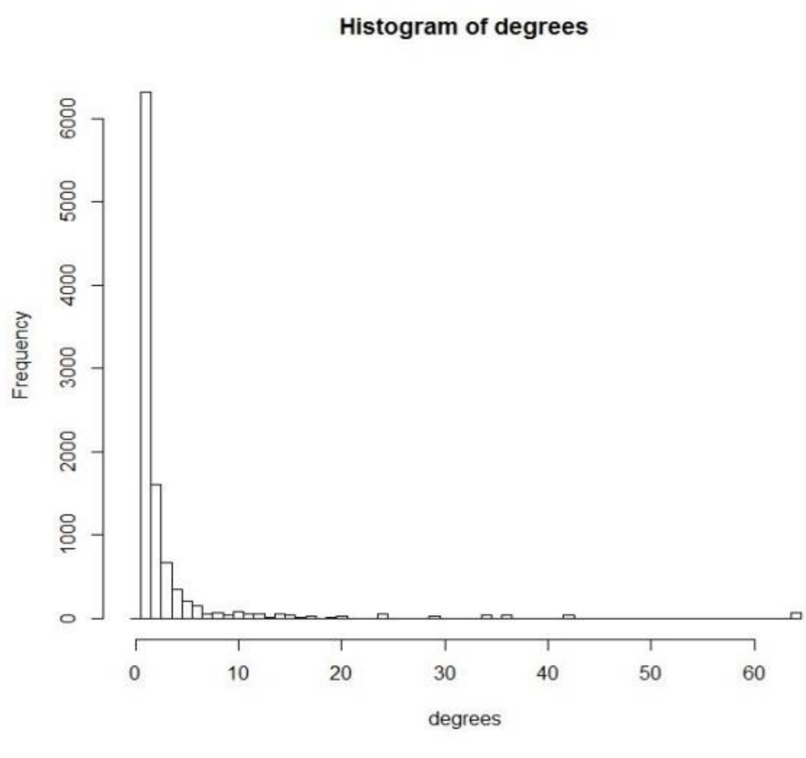



**Plots for 10000 node random network using Barabasi model**



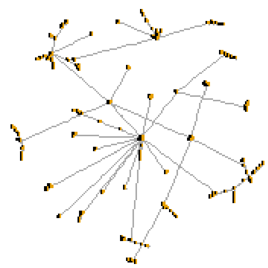


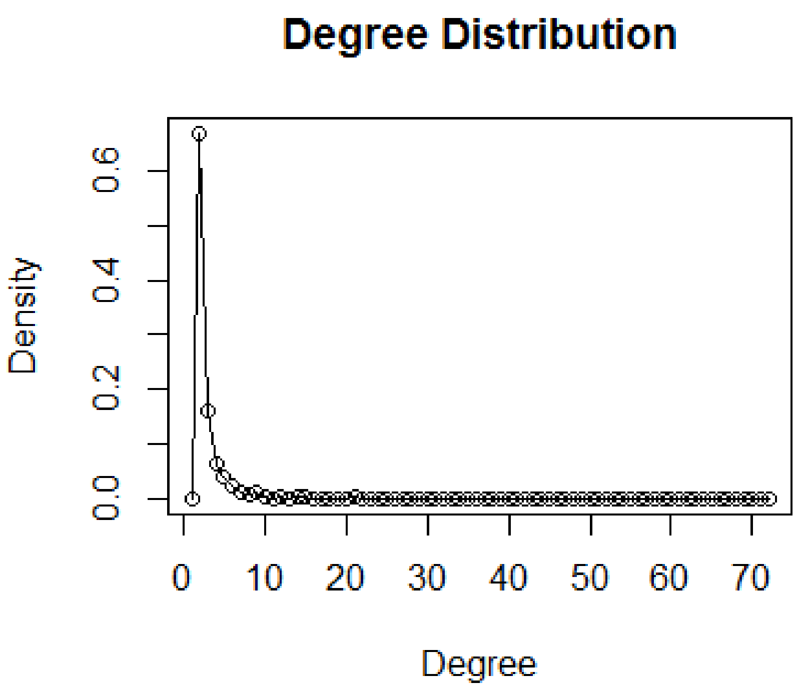

**Question 2d**
On the network with 1000 vertices a random walk was done to find the degree distribution. A vertex was picked up randomly and then the degree of a neighbor node (picked randomly) was found. This experiment was repeated 10000 times. The result of the experiment and the degree distribution of the network is shown below. We can see that it **follows the power law form (X^-3).**



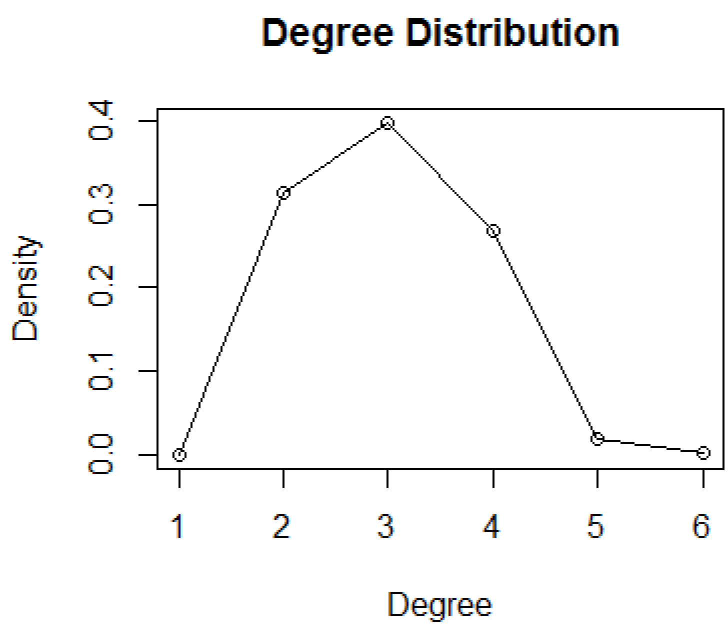
**Q3. Random graph by simulating its evolution**

a) We created a random graph with 1000 nodes by simulating its evolution. The probability that an old vertex is cited depends on its in-degree (preferential attachment) and age. The following results were for the aging exponent 0.



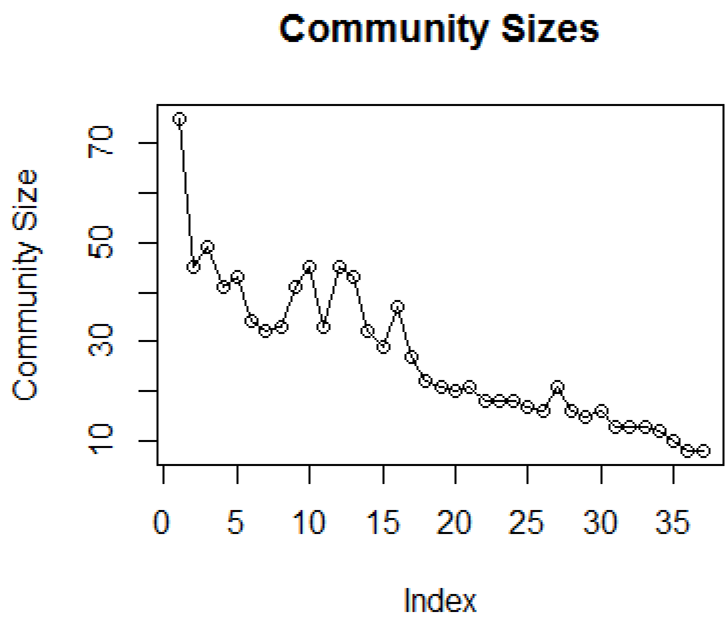



The following results were for the aging exponent -4.



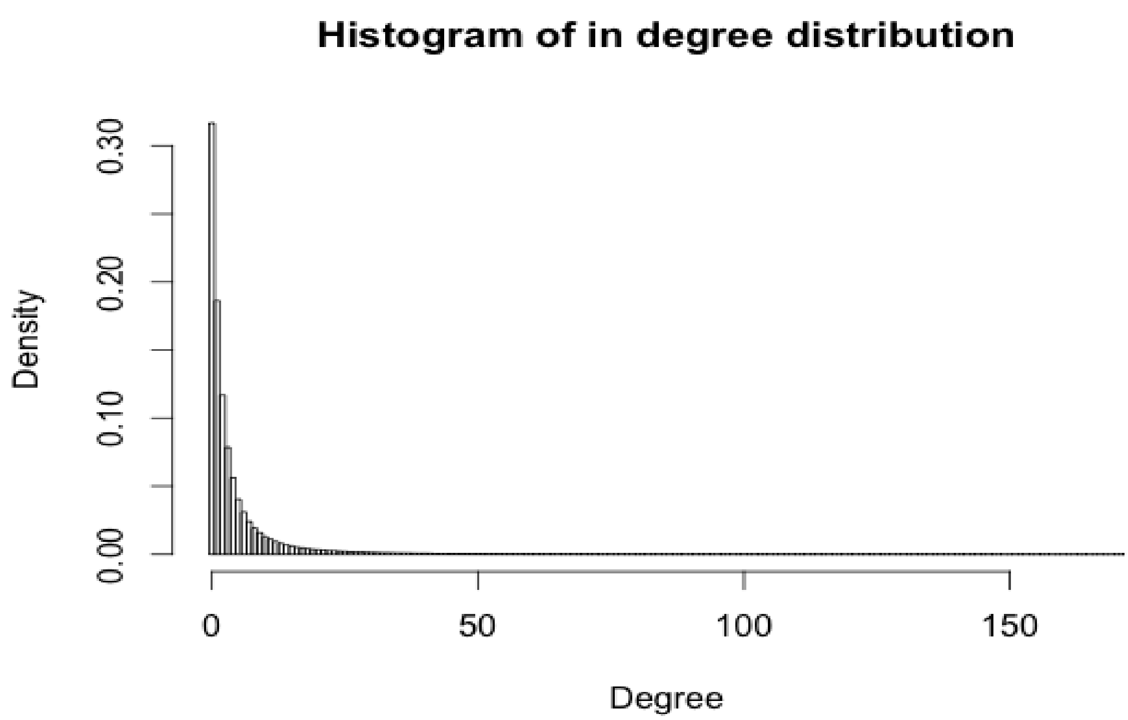

The maximum degree of the network is much higher for aging exponent = 0. However, for aging exponent = - 3 we see that the degree distribution dies out very quickly.

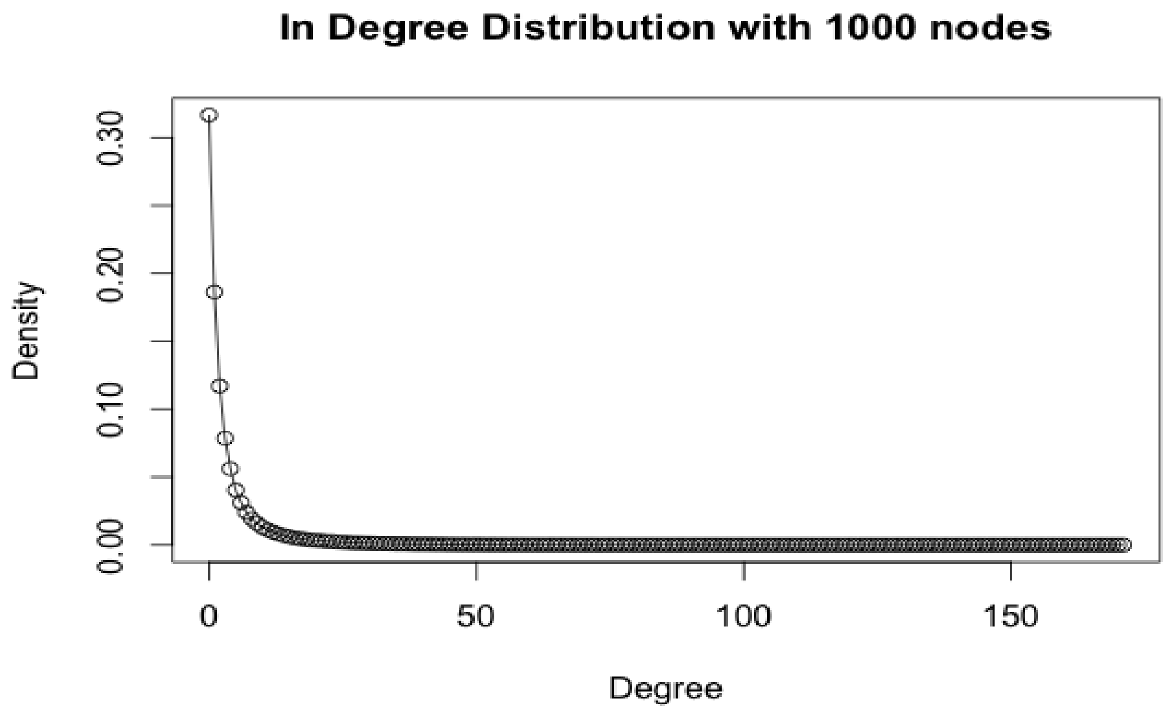
b) **The modularity for when aging exponent is 0 was found to be 0.928**. We found the community structure of the network using the fast greedy method and plotted the community sizes, as shown below. We can tell that a network is densely populated within the community but sparsely across communities.

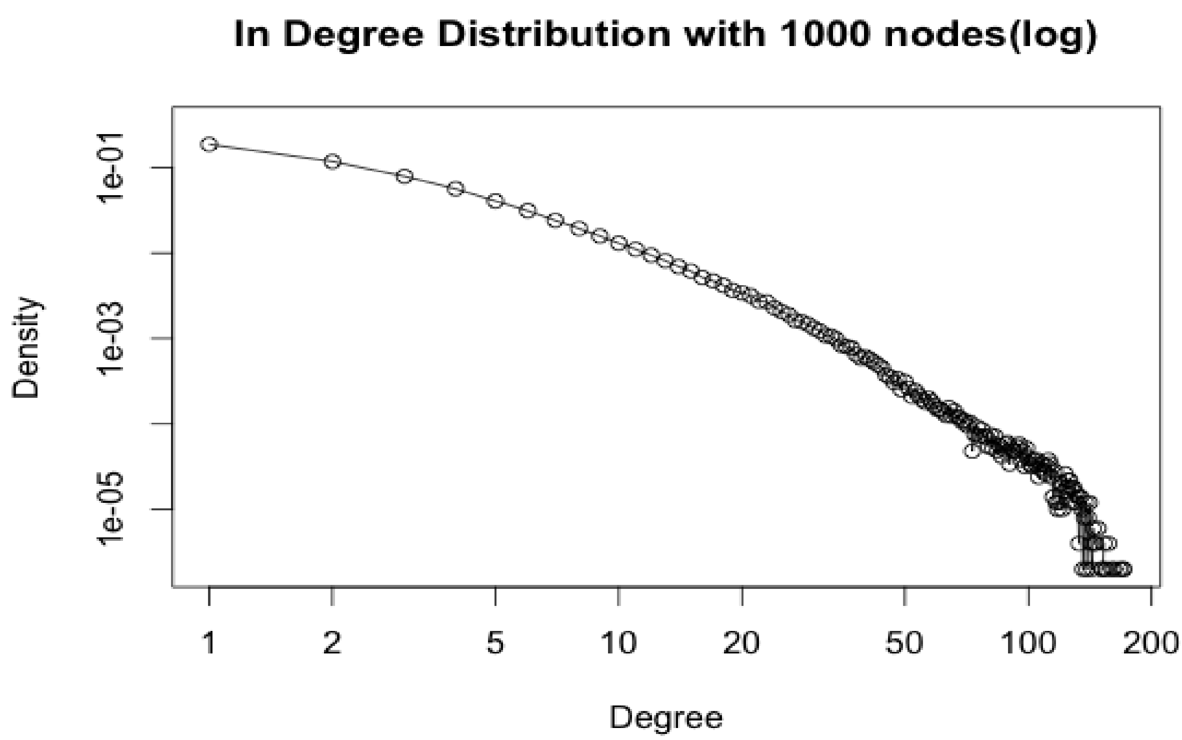


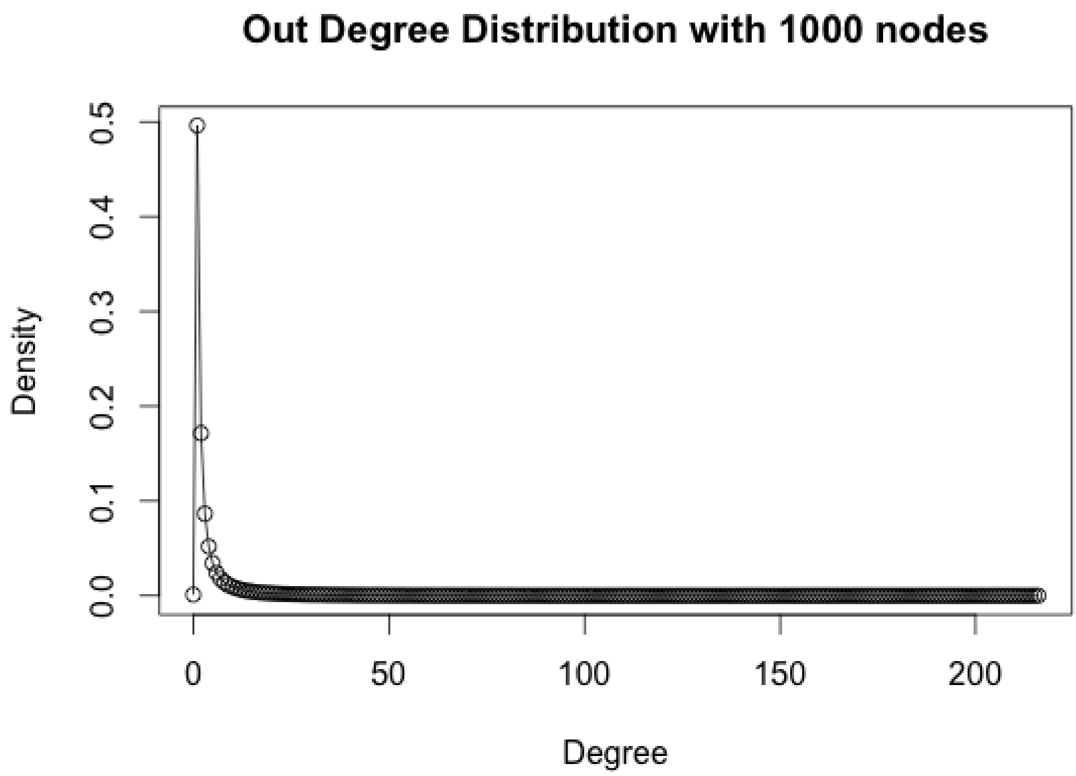

**Q4. Forest fire model**

a.) A directed forest fire model of network was created using the **forward burning probability as 0.37 and the backward burning ratio as 0.87**. The in and out degree distributions for a directed network with 1000 nodes using forest fire model are shown below. Both the distributions are seen to decay quite fast.

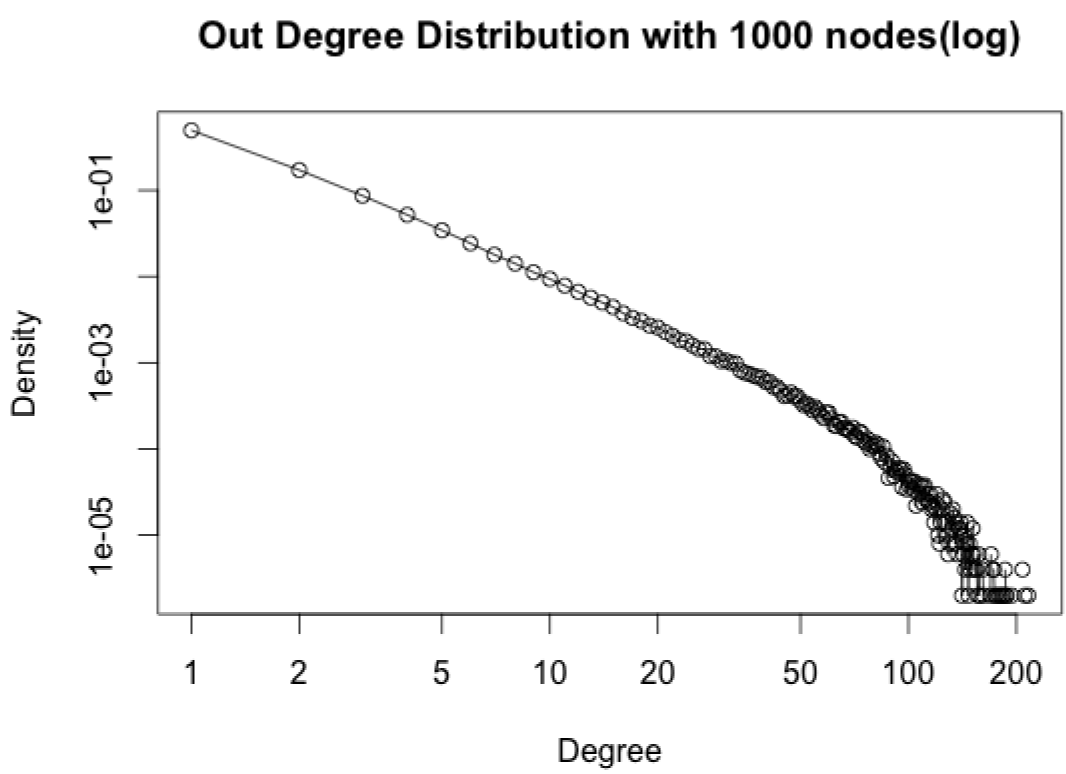



In Degree Distribution with 1000 nodes
Density
0.00
0.10
0.20
0.30
0
50
100
150
Degree


In Degree Distribution with 1000 nodes(log)
Density
1e-05
1e-03
1e-01
1
2
5
10
20
50
100
200
Degree



Out Degree Distribution with 1000 nodes
Density
0.0
0.1
0.2
0.3
0.4
0.5
0
50
100
150
200
Degree

Out Degree Distribution with 1000 nodes(log)
Density
1e-05
1e-03
1e-01
1
2
5
10
20
50
100
200
Degree

b.) **The diameter of the network was found to be 10**.
c.) We have used the **edge betweenness algorithm** to find the community structure of this directed network. The network had only **1 GCC**. We note that there were **314 communities** in the whole network. The community sizes are plotted below. Also, it can be seen that there is a single community with large number of vertices and the rest are communities with small number of vertices. The network is also shown below.
**The modularity is 0.198796**.