# Homework 1 Report for EE232E

*Lei Ding*

*Yi Ding*

*Sonu Mishra*

***Question 1: Create random networks***

*(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.*

**Solution:**

The degree distributions of the graph with *p* = 0.01, 0.05 and 0.10 are shown in Figures 1, 2, 3.

Key insights:

1. The distribution in each case looks like Poisson distribution.

   This aligns with the theory: The distribution of a random graph with probability *p* of any of the *N* nodes being connected, is given by the binomial distribution: $P(k) = \binom{N-1}{k} p^k (1-p)^{n-1-k}$. When the value of *N* is large this converges to a Poisson distribution with $\lambda = Np$.

2. The average degree of the graph increases with increase in *p*.

   Mean of any Poisson distribution is given by its parameter $\lambda$. Since $\lambda = Np$, the average degree is directly proportional to *p*. Therefore increase in the average degree with *p* is expected. Intuitively, this means that that when the probability of two nodes being connected increases, the average degree of the graph will increase.

3. Simulation matches well with the theoretical results.

   Theoretical values of average degree are given by $\lambda = Np$. For *N* = 1000 and *p* = 0.01, 0.05 and 0.10, the average degrees can be calculated as 10, 50 and 100, respectively. From Figures 1, 2, 3, we can see that the average degrees of our graphs match well with these theoretical values.

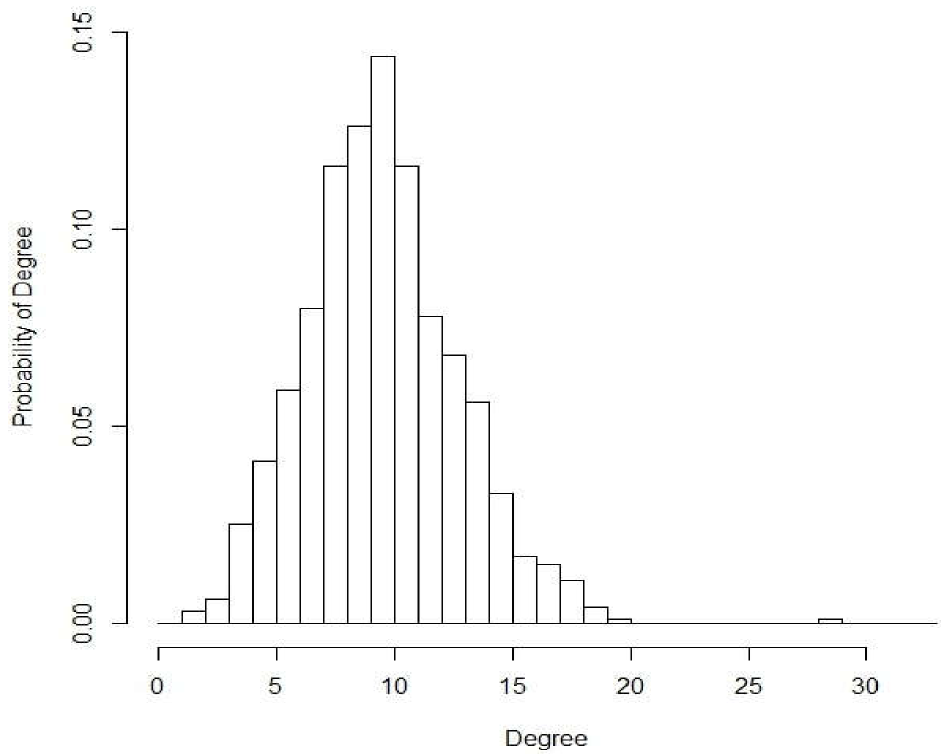


**Figure 1: Degree distribution of the random graph with *p* = 0.01**

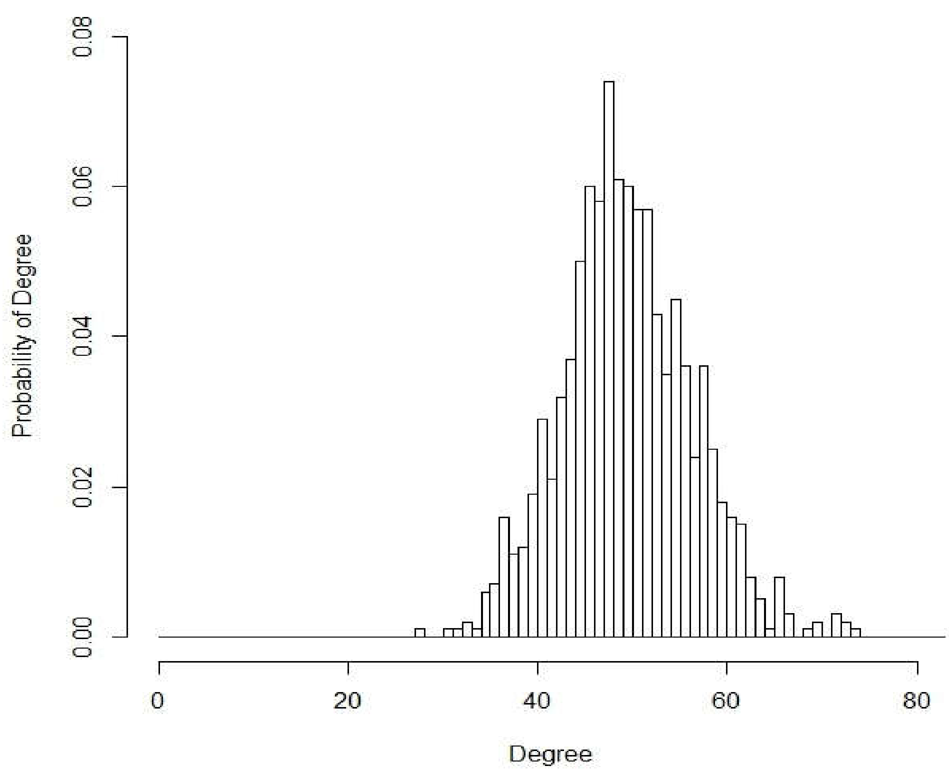


**Figure 2: Degree distribution of the random graph with *p* = 0.05**

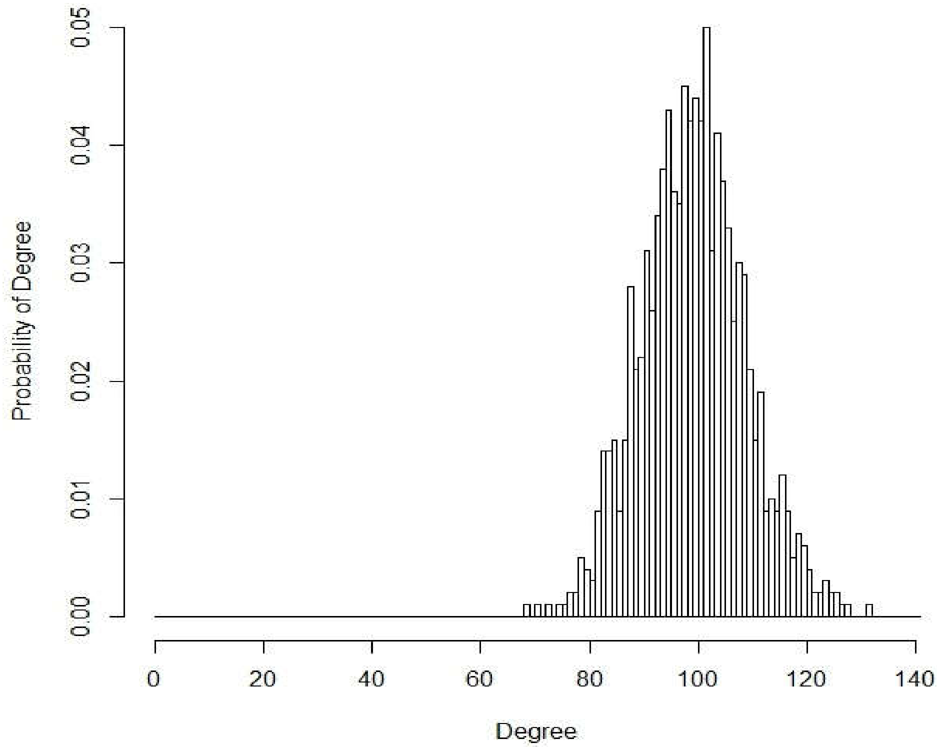


Figure 3: Degree distribution of a random graph with $p$ = 0.10

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

**Solution:**

All three networks were found to be connected.

The diameter of a network is defined asthe shortest distance between the two most distant nodes in the network. For the given networks, the diameters are: **5*, 3 and 3** respectively. The drop in the diameter, when $p$ increases, is expected. This is because with richer connectivity, the shortest distance between the two nodes is decreases.

Note: *There was fluctuation in the diameter of the first random graph. Below are the diameter values of the graph with $p$ = 0.01 that we obtained in 10 different runs.

| Runs | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Diameter | 5 | 6 | 5 | 5 | 5 | 5 | 6 | 6 | 5 | 5 |

The diameter values of the other two graphs were stable.

*(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.*

**Solution:**

The value of the percolation threshold, $p_c = 0.008$.

This also matches with our results in part (b). Since the values of $p = 0.01$, 0.05 and 0.10 are greater than $p_c$, our graphs should be connected.

Note: As the network graph is random, there were fluctuations in the value of $p_c$. To get rid of this simulation noise, we took the average over 1K samples of the graph.

***Question 2: Create a network with a fat-tailed degree distribution***

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

**Solution:**

The plot of degree distribution is shown in Figure 4. The diameter of the graph is 19. There were fluctuations in the value of the diameter.

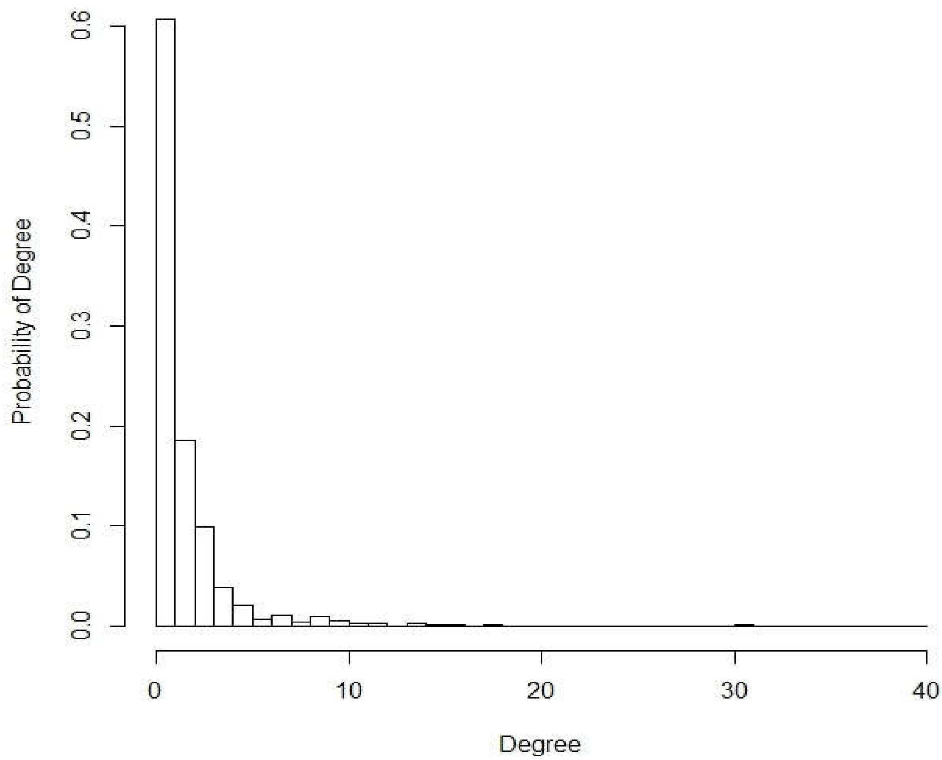


Figure 4: Plot of degree distribution of the graph with degree distribution $x^{-3}$

*(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?*

**Solution:**

Yes, the network is connected.

To find the giant connected component (GCC), we first found the clusters and got the largest cluster. Then we deleted from graph the nodes that did not have membership in this giant cluster.The resultant graph is the GCC. After this we used fastgreedy.community(), method to find the community structure.

The modularity of the graph was found to be **0.933615.** This is so large because the degree distribution is proportional to $x^{-3}$, due to which the nodes tend to have smaller degrees. This often leads to a large number of small clusters in the graph. These small clusters have rich intra-cluster connectivity but sparse inter-cluster connectivity.

*(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?*

**Solution:**

The modularity of this larger network is found to be **0.977938.**

It is not same as the smaller network; it is slightly higher than the smaller network. The reason behind increase in the modularity is that the larger network will likely have more clusters than the smaller networks.

*(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.*

**Solution:**

The required degree distribution is shown in Figure 5. As desired, the degree distributions of the nodes sampled with this process is fat-tailed as compared to Figure 4.

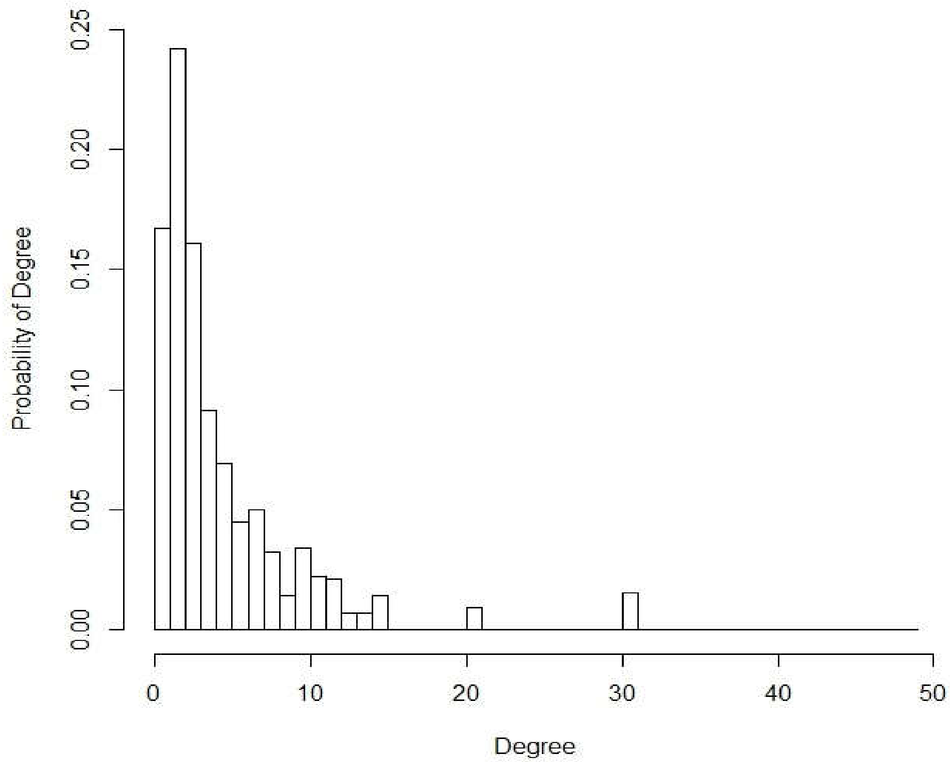


**Figure 5: Degree distribution of node j selected as the 1 hop neighbor of arbitrarily selected nodes.**

***Question 3*: *Creates a random graph by simulating its evolution***

*(a) Each time a new vertex is added it creates a number of linksto old vertices and the probability that anold vertex is citeddepends on its in-degree (preferential attachment) and age.Produce such an undirected network with 1000 nodes. Plotthe degree distribution.*

**Solution:**

In this problem, we use the igraph's library function aging.perfatt.game() to generate an undirected network with 1000 nodes that this problem requires. And the degree distribution of this graph is shown in Figure 6. We can see that most of the degrees are less than 5 and very few nodes have high degrees. The distribution is similar to the right part of Gaussian Distribution.

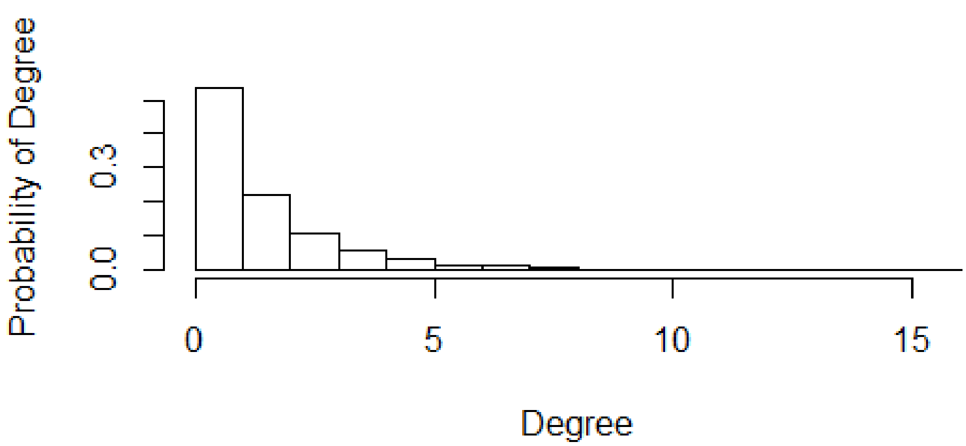


Figure 6 Random Graph Degree Distribution

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

**Solution:**

In this problem, we use the igraph's library function fastgreedy.community() to generate the community structure. And then we use modularity() to calculate the modularity of the generated graph.

I run the R program for ten times and the result is shown in table 1.

Table 1: Modularity values

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Modularity | 0.9354 | 0.9354 | 0.9358 | 0.9354 | 0.9357 | 0.9350 | 0.9352 | 0.9351 | 0.9359 | 0.9346 |

As we know, the modularity can measure the strength of division of a network into modules and networks with high modularity have dense connections between nodes within modules but sparse connections between nodes in different modules. And the range of modularity is [-1/2,1]. So we can see that our table shows that the modularity is pretty high, which means the graphs we generate are highly clustered as communities, which means dense connections between nodes within modules but sparse connections between nodes in different modules.

***Question 4: Use the forest fire model to create a directed network***

*(a) This is a growing network model, which resembles how theforest fire spreads by igniting trees close by. Plot the in andout degree distributions.*

**Solution:**

The forest fire model has following four rules which are executed simultaneously:

- A burning cell turns into an empty cell.
- A tree will burn if at least one neighbor is burning.
- A tree ignites with probability *f* even if no neighbor is burning.
- An empty space fills with a tree with probability *p*.

Based on these rules, a network created using forest fire model will have following features:

- Heavy-tailed in-degree distribution.
- Heavy-tailed out-degree distribution.
- Shrinking diameter: The diameter of the network decreases in time.
- Densification power-law.
- Communities.

In the problem, we choose to build a forest fire network with 1000 nodes with the forward burning probability equals to 0.37 and the backward burning ratio equals to 0.32/0.37. The in-degree distribution and out-degree distribution are as follows.

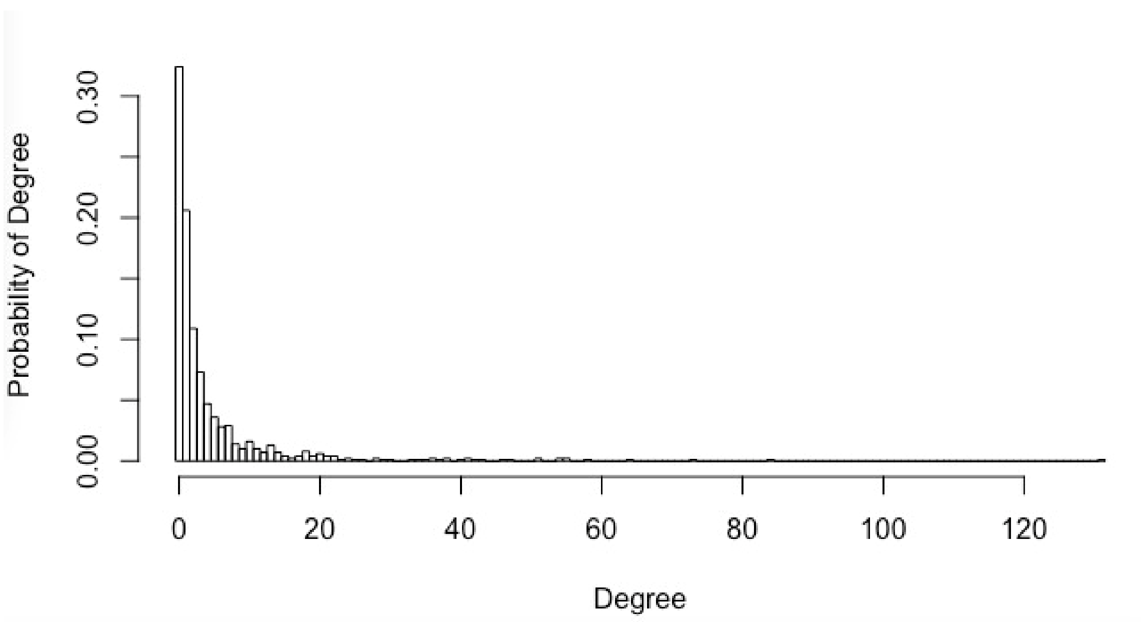


Figure 7: In Degree Distribution of a Forest Fire Graph.

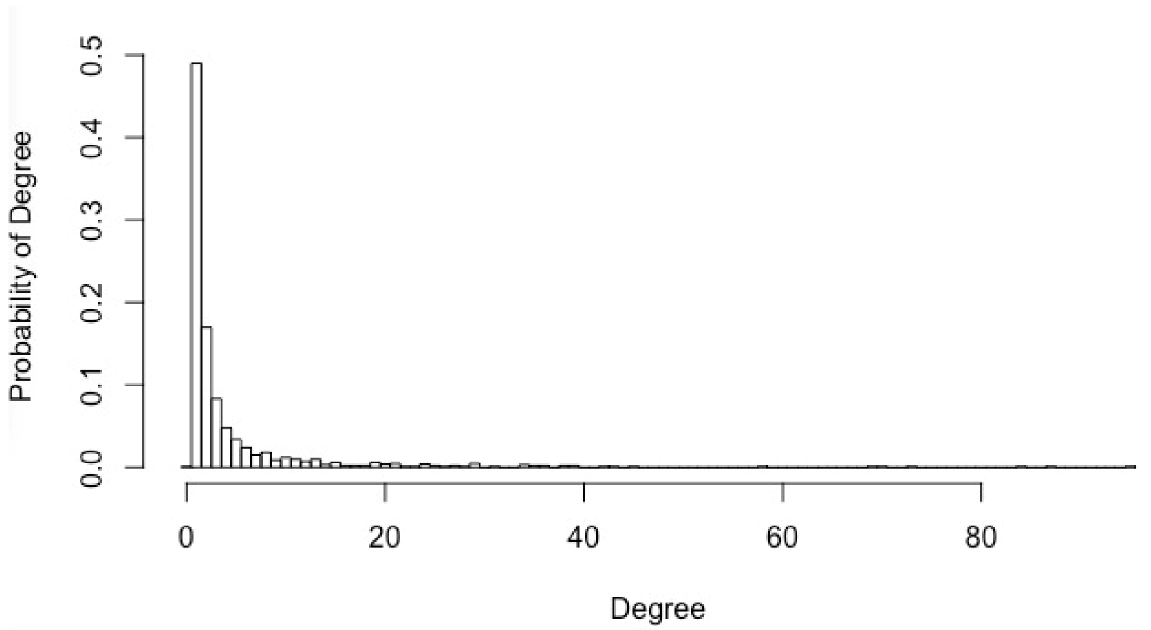

Figure 8: Out Degree Distribution of a Forest Fire Graph.

As shown in Figure 7 and Figure 8, both in-degree and out-degree distributions are heavy tailed.

(b) *Measure the diameter.*

**Solution:**

Since the diameter of a forest fire graph is shrinking with the time, thus we measure the final diameter of the graph. We run the graph 10 times and measure the final diameter.

Table 2: Diameter values

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Diameter | 12 | 10 | 10 | 13 | 9 | 10 | 9 | 11 | 8 | 10 |

From the table above, we can conclude that the diameter of a forest fire graph is shrinking and the final diameter is always very small.

(c) *Measure the community structure and modularity.*

**Solution:**

To compute the community and modularity, We use random walk and run 10 times, the following table shows the modularity results

Table 3: Madularity values

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Modularity | 0.56 | 0.24 | 0.21 | 0.52 | 0.44 | 0.43 | 0.39 | 0.52 | 0.26 | 0.58 |

As shown in the table above, each time, the modularity will fluctuate due to the graph is created randomly. The average of the modularity is 0.41, which is not very high.

Finally, we plot the community structure of a forest fire graph. To illustrate it more clearly, we choose the node size 200. Following is the community structure of a forest fire graph with 200 nodes.

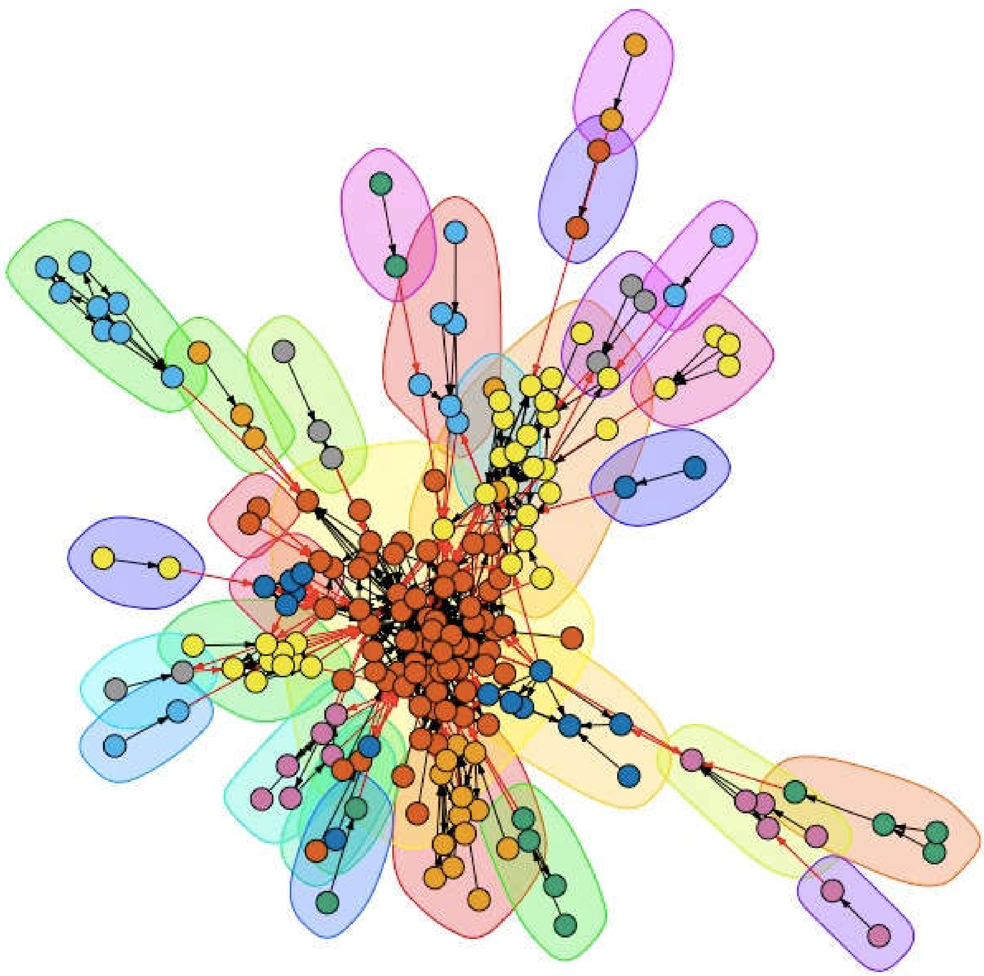

Figure 9: Community Structure of Forest Fire Graph with 200 nodes.