EE232E - Graphs and Network Flows
Homework 2

Swati Arora: 404758379
Vishank Bhatia: 304758488
Anshita Mehrotra: 904743371

Question 1: Random walk on random networks

Part (a)

For part(a), we created an undirected graph using **‘random.graph.game’** of the **igraph** package. The graph created consisted of **1000 nodes** and the probability for drawing an edge between any pair of nodes equal to 0.01

Part (b)

In part (b), we **stimulate a random walker** for the random network graph that was created in part(a). We use the ‘**netrw**’ package for this.

- We first select a node at random (no damping).
- And randomly walk by varying the number of steps as [1-1000]
- The average distance <s<t>> of the walker from the starting node at step t is calculated.
- Also the standard deviation of the this distance is calculated using the following formula

$$\boldsymbol{\sigma 2(t) = \langle (s(t) - \langle s(t) \rangle)2 \rangle}$$

- We then plot the graphs for the values obtained and the step t.
- Results obtained:

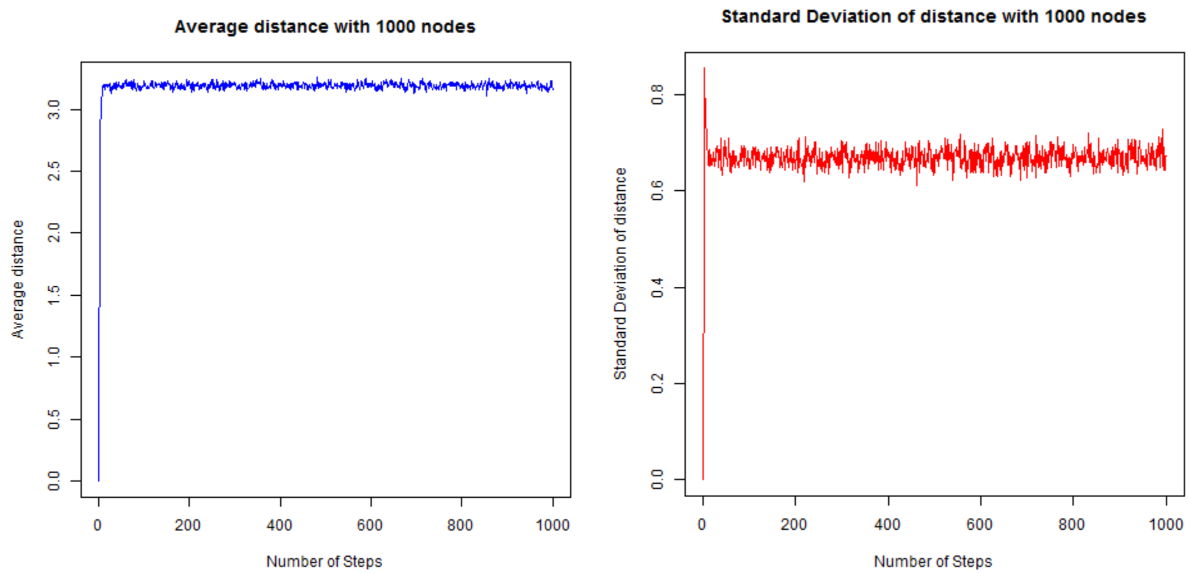

Figure 1 : Average Distance and Standard Deviation v/s Number of Steps

Part (c)

In part (c) we calculate the average distance and the average standard deviation. A random walker in d dimensional has average (signed) distance ⟨s(t)⟩ = 0 and average standard deviation proportional to the number of steps (t). Values observed for our network:

	Value
Average Distance	3
Average standard Deviation	0.5

Table 1: Results for average distance and standard deviation

The reason why our values don't converge to 0 is that **in a d dimensional space the distance can be negative, cancelling out the positive distances**. But in our random graph, all the values are positive so it will not be able to converge to 0.

Part (d)

- In part(d), we perform the steps performed in part(b) for random networks with the number of nodes as 100 and 10000. The number of steps are varied as [1-50]The results obtained are as follows:

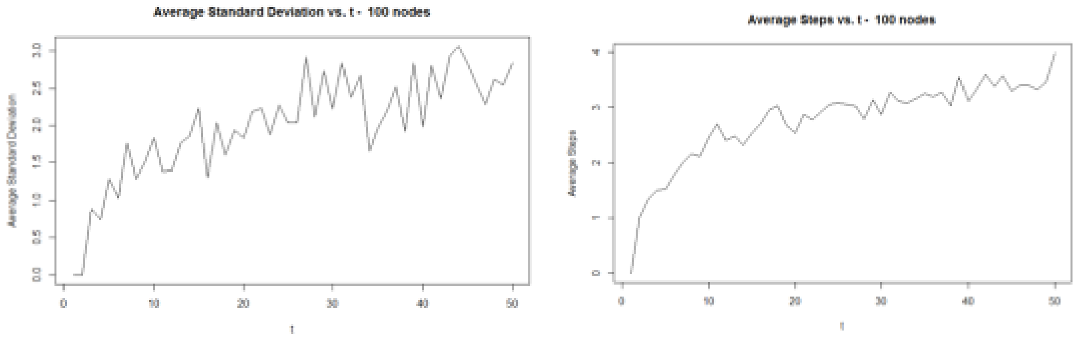

Figure 2 : Average Distance and Standard Deviation v/s Number of Steps with Nodes = 100

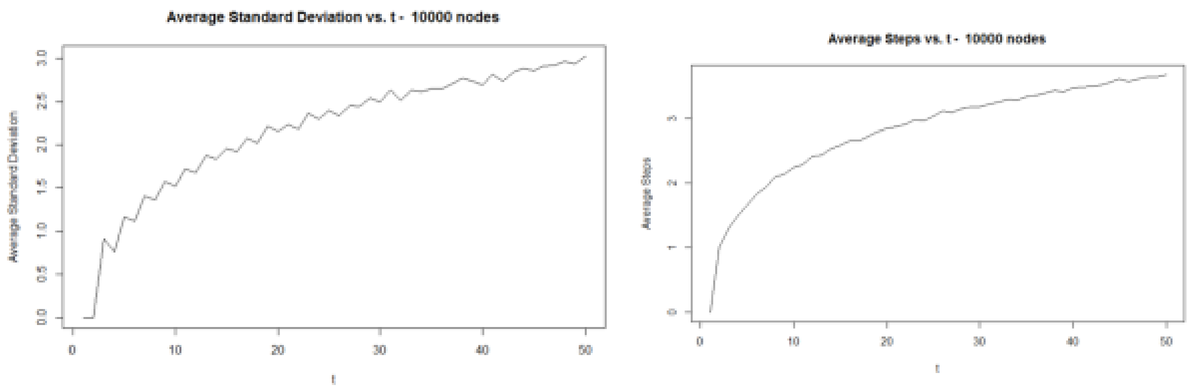

Figure 3 : Average Distance and Standard Deviation v/s Number of Steps with Nodes = 10000

- Next we calculate the diameters for the 3 random network graphs and compare them.

Number of Nodes	Diameter
1000	17
100	10
10000	25

Table 2: Diameters for the random networks

Part (e)

In part (e) we measure the degree distribution of the nodes reached at the end of the random walk on the 1000-node random network. And then compare it with the degree distribution of the graph. The results obtained are as follows:

As we can observe there is not much of a difference between the two graphs.

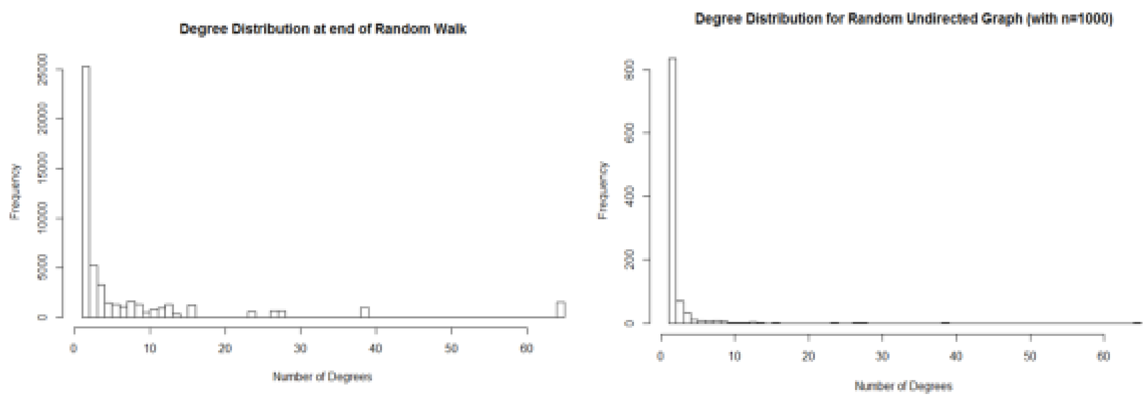

Figure 4 : Degree distribution of random network and nodes at end of random walk

Question 2: Random walk on networks with fat-tailed degree distribution

Part (a)

In part(a) we use **'barabasi.game'** of the igraph package to generate a random network. The degree distribution of the network is proportional to x^{-3}.

Part(b)

In part (b), we **stimulate a random walker** for the random network graph that was created in part(a). We use the '**netrw**' package for this.

- We first select a node at random (no damping).
- And randomly walk by varying the number of steps as [1-1000]
- The average distance <s<t>> of the walker from the starting node at step t is calculated.
- Also the standard deviation of the this distance is calculated using the following formula

$$\sigma 2(t) = \langle (s(t) - \langle s(t) \rangle) 2 \rangle$$

- We then plot the graphs for the values obtained and the step t.
- Results obtained:

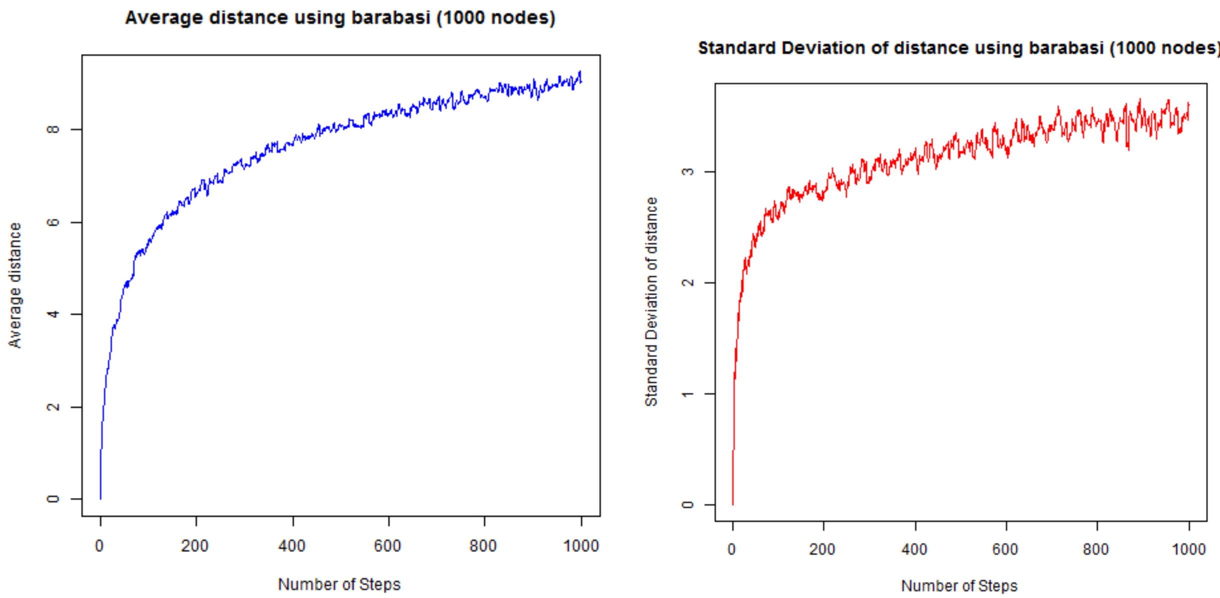

Figure 5 : Average Distance and Standard Deviation v/s Number of Steps

Part (c)

In part (c) we calculate the average distance and the average standard deviation. A random walker in d dimensional has average (signed) distance $\langle s(t) \rangle = 0$ and average standard deviation proportional to the number of steps (t). Values observed for our network:

	Value
Average Distance	3
Average standard Deviation	3

Table 3: Results for average distance and standard deviation (fat tailed network)

The reason why our values don't converge to 0 is that **in a d dimensional space the distance can be negative, cancelling out the positive distances**. But in our random graph, all the values are positive so it will not be able to converge to 0.

Part (d)

- In part(d), we perform the steps performed in part(b) for random networks with the number of nodes as 100 and 10000. The results obtained are as follows:

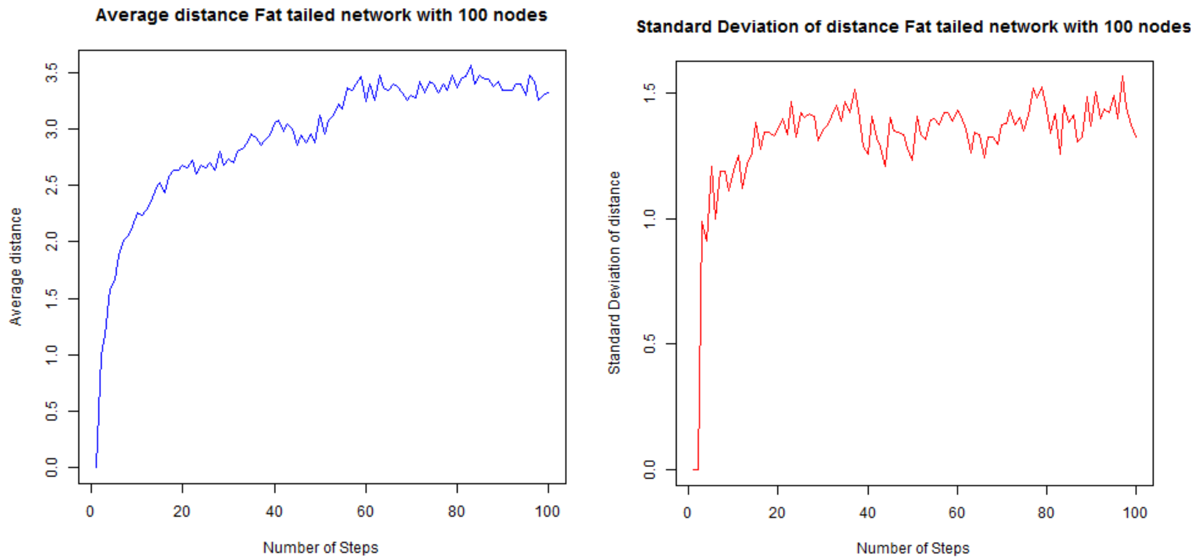

Figure 6 : Average Distance and Standard Deviation v/s Number of Steps with nodes = 100

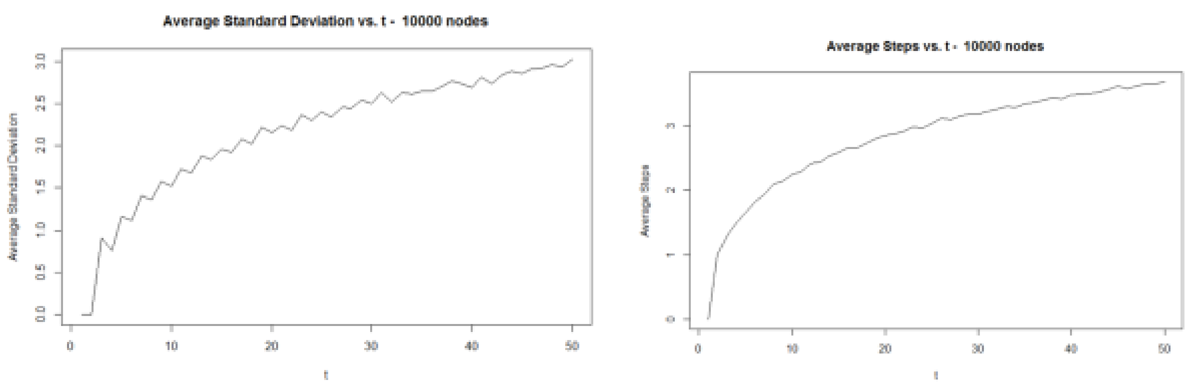

Figure 7: Average Distance and Standard Deviation v/s Number of Steps with nodes = 10000

- Next we calculate the diameters for the 3 random network graphs and compare them.

Number of Nodes	Diameter
1000	17
100	10
10000	24

Table 4: Diameters for the random networks

- <u>Observations:</u>
  - It can be clearly seen that as the **diameter for the network increases, the time for the distribution to converge is slower**.
  - The distribution converges at a much faster rate when the number of nodes is 100 than when the number of nodes is 10000, since the diameter has increased two-folds.

Part (e)

In part (e) we measure the degree distribution of the nodes reached at the end of the random walk on the 1000-node random network. And then compare it with the degree distribution of the graph. The results obtained are as follows:

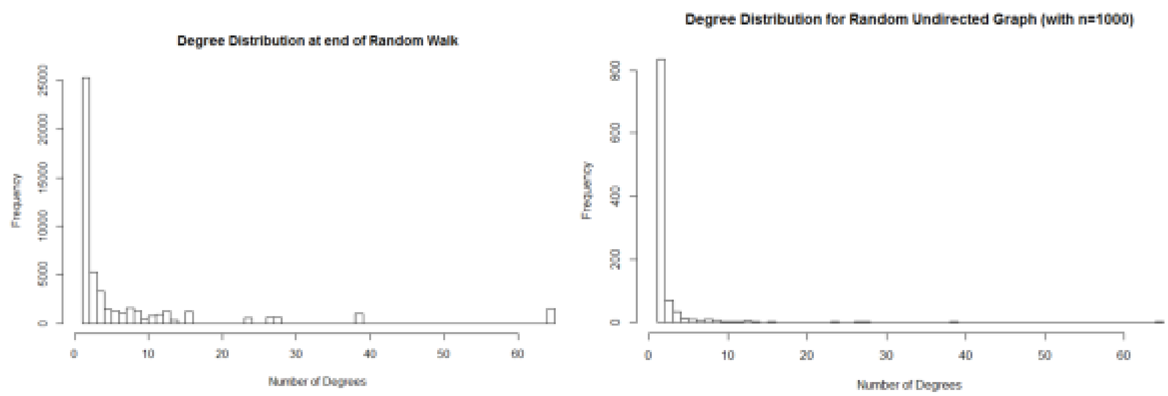

Figure 8 : Degree Distribution of Random Network and Nodes at end of Random Walk

As we can observe there is not much of a difference between the two graphs.

Question 3: Page Rank

Part (a)

Here we use the random walk to stimulate PageRank. For random walks on the network created in 1(a), we measure the probability that the walker visits each node. We then calculated the **correlation between the number of degree and the visit probability**. The value obtained was: **0.976**

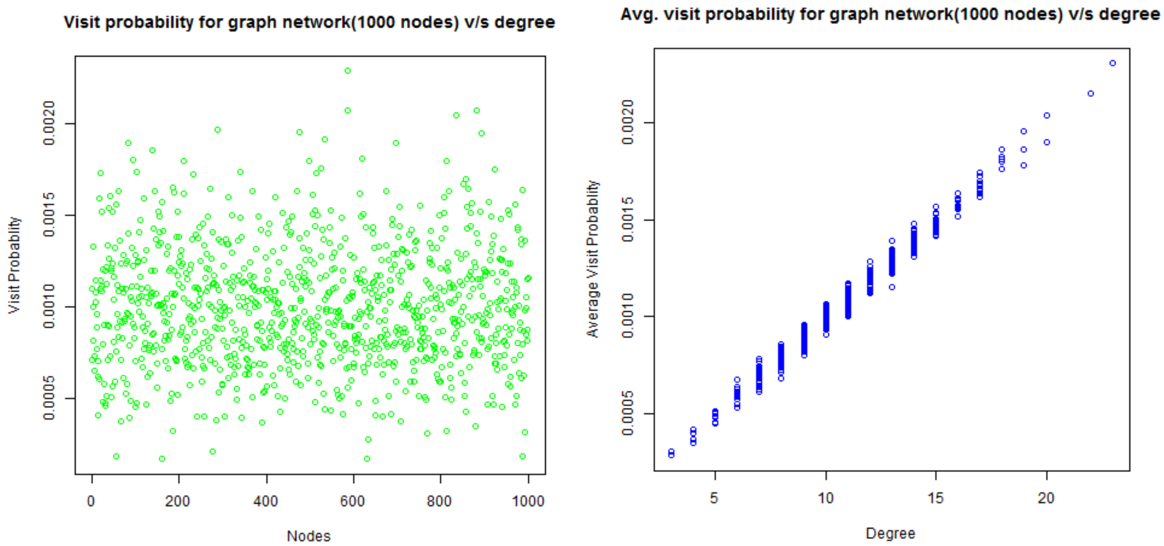

Figure 9 : Number of nodes v/s Visit Probability and Average visit probability v/s number of degree

As we can observe, **the degree of nodes and their visit probability are linearly dependent as they have high a correlation value**.

Part (b)

In part(b), we again stimulate a random walker. We create a directed network with the number of nodes as 1000. The probability p for drawing an edge between any pair of nodes is 0.01. We the calculated the degree of each node in the graph and then the probability of visiting each node within vertex sequence. The result obtained are as follows:

Visit Probability of nodes	Correlation
Total degree	0.621
In degree	0.911
Out degree	-0.008

Table 5: Results for Correlation between nodes of network and their visit probability

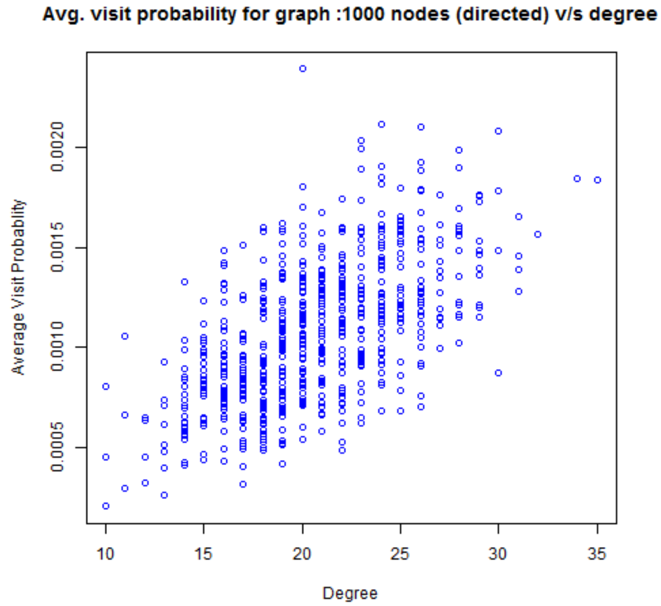

Figure 10 : Average visit probability v/s Total degree

Observation:

- The in-degrees of nodes are highly correlated with the visit probability of the nodes, as more the number of incoming links the node has, more is the visit probability of the node.
- The inverse relationship can be seen for the out-degree by the correlation value.

Part (c)

In all the previous sections, so far we have been using a damping factor = 1. Which means that there has been no teleportation. In this part, we introduce a **damping factor of 0.85** for an undirected random network. For the random walks created, we measure the probability that the walker visits each node. **The correlation between the no of degree and visit probability is: 0.964**

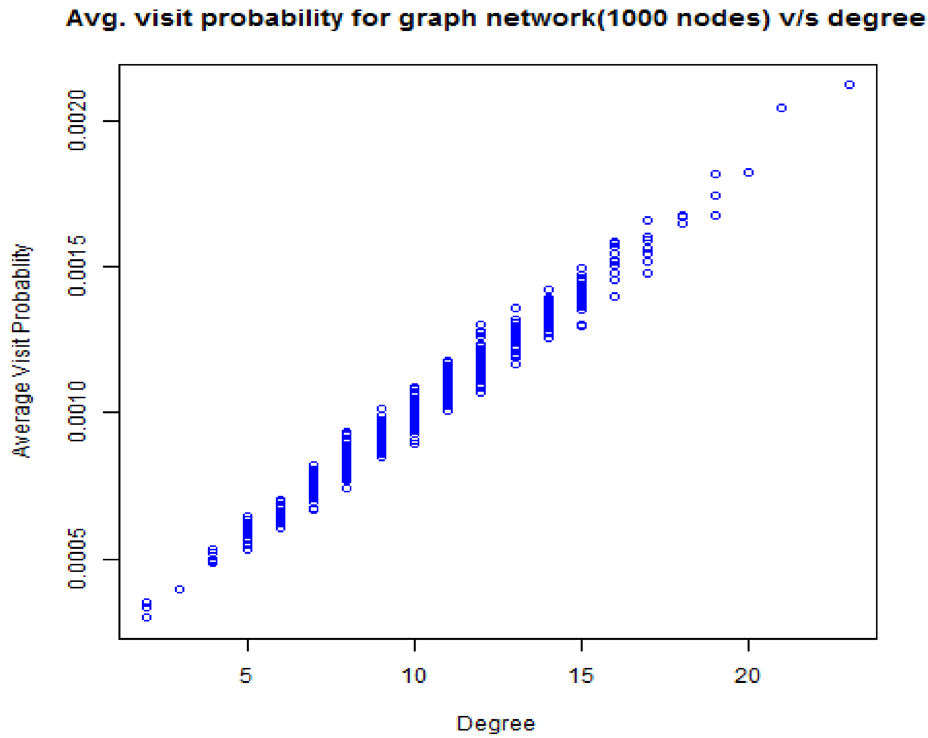

Figure 11 : Average Visit Probability v/s Total degree for 100o nodes network with d = 0.85

Observation:

The degree of nodes and their visit probability **are linearly dependent as they have high a correlation value.**

Question 4: Personalized Page Rank

Part (a)

In part(a) we create a directed random network with 1000 nodes, where the probability for drawing an edge between any pair of nodes is 0.01. Then we use random walk with **damping parameter 0.85** to simulate the PageRank of the nodes.

We, simulated the page rank of the network by plotting the visit probability of each node. The results are as follows:

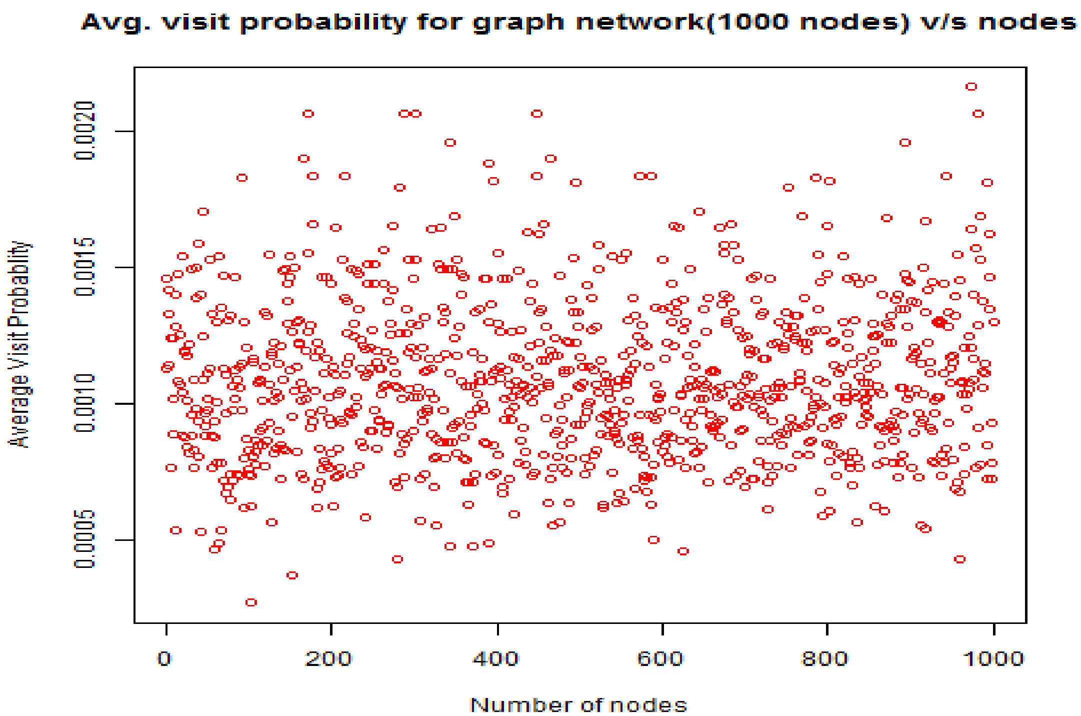

Figure 12 : Simulated PageRank - Average visit Probability v/s Number of nodes

Part (b)

- A random directed network was generated with a damping factor of 0.85, to create a Personalized PageRank.
- We used the visit probabilities of this network as teleportation probability to help create a new directed network.
- We this we intend to make the page-rank for each node to proportional to the teleportation probability instead of it being equal 1/N.

Correlation between page-rank and page-rank (with teleportation) = 0.983

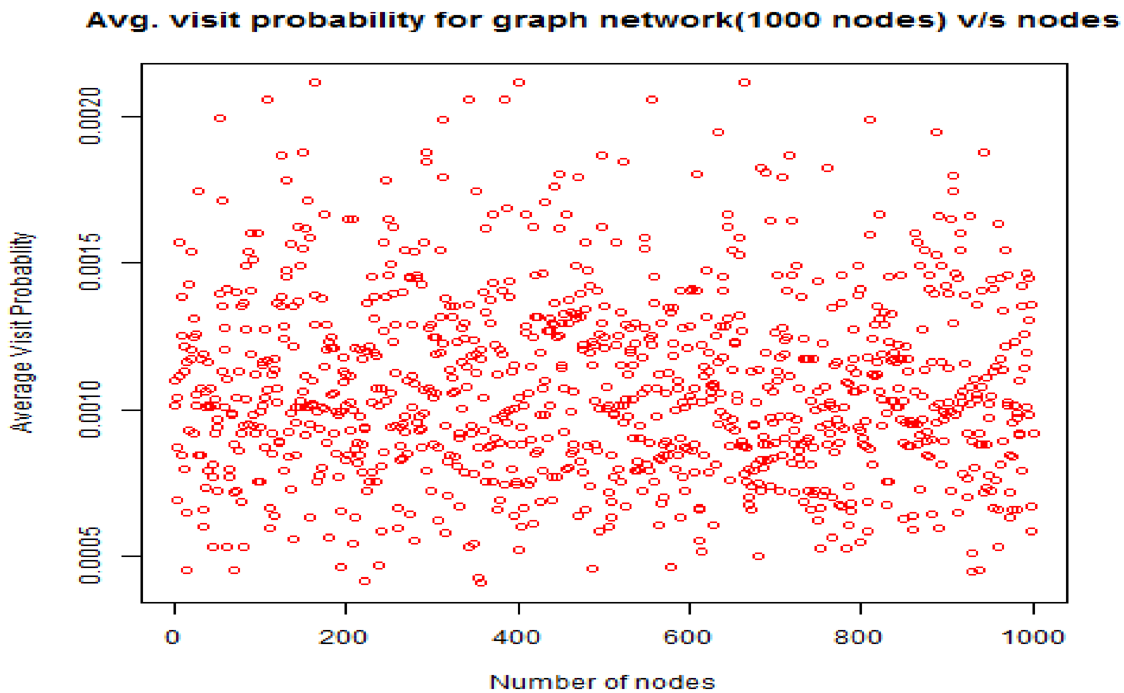

Figure 13 : Simulated Personalized PageRank with Random Walk - Average visit Probability v/s Number of nodes

Observation:

- As we can observe, the graph of visit probabilities is **more concentrated** with this approach than the graph produced in part(a).
- Nodes which have large people interest based on teleportation have **large visit probabilities and consequently larger page rank**.

Part (c)

- Following PageRank equation is considered to investigate the notion of self-importance into the PageRank of each node:

$$\mathtt{Pr}(A) = \frac{(1-d)}{N} + d\sum_{T_{in}} \frac{\mathtt{Pr}(T_{in})}{C(T_{in})}$$

We need to change the teleportation probability from 1/N to a number which is proportional to the PageRank to account the effect of self-enforcement.