EE 232E Graphs and Network Flows
Homework 2

Professor Roychowdhury

Prepared by:

Arunav Singh (304 760 844)
Eric Goldfien (603 887 003)
Steven Leung (304 777 142)

NOTE: Questions 1 and 2 are programmed in R v3.3.3 with iGraph v1.0.1 while Questions 3 and 4 are programmed in R v2.15.2 with iGraph v0.7.0 and netrw v0.2.6. This is due to the use of the random_walk function for Questions 1 and 2 which requires iGraph v1.0.1 which requires R v3.3.3 to run.

<u>Question 1</u>

Part a:

An undirected random network with 1000 nodes and probability p for drawing an edge between any pair of nodes was created in R using the *erdos.renyi.game(n,p)* function.

Part b:

The plots below in *Figure 1.1* show the average distance and average standard deviation of different walkers throughout their path (steps). It is seen that both the average distance and average standard deviation plots both converge after a certain number of steps. Both plots are averaged over 1,000 walkers (each walker starts at a different node).

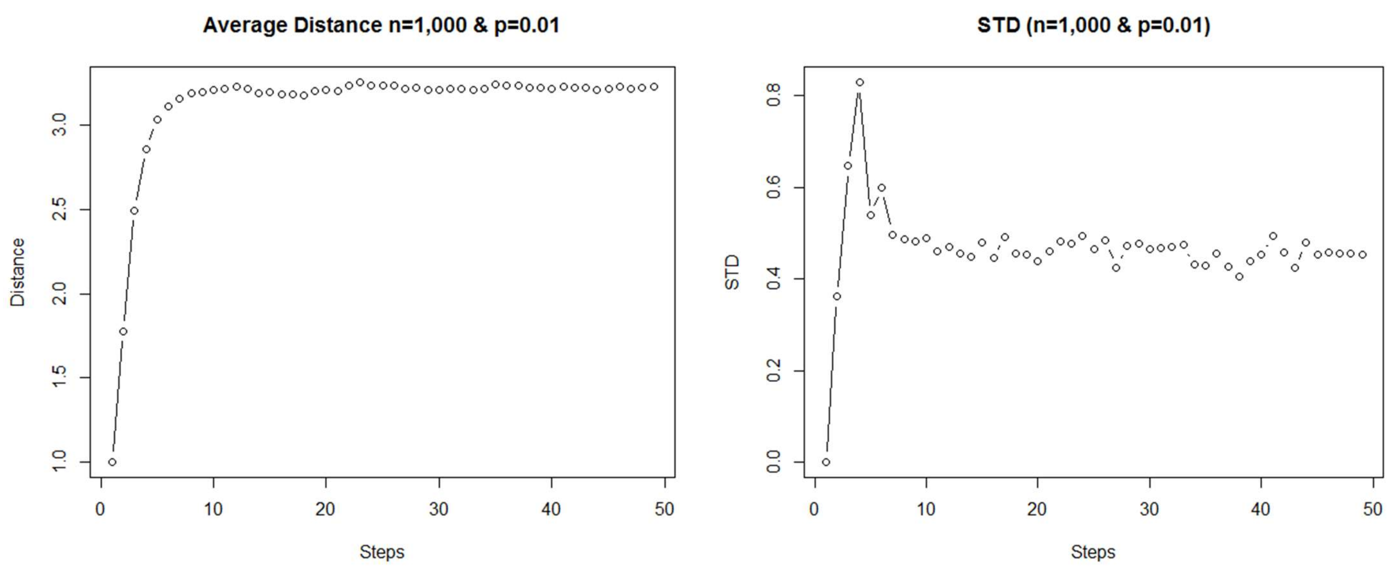


Figure 1.1 Average Distance and STD (n=1,000 & p=0.01)

Part c:

A random walker in *d* dimensional has average (signed) distance of 0 and a standard deviation (σ) proportional to $\sqrt{t}$. This is not true in a random network because all distances in a random graph are positive. The reason why a random walker in *d* dimensional has average distance zero is because the positive and negative distances cancel each other out but since in a random network there are no negative distances, this cancellation does not occur leading the averages to converge some number not equal to zero. The standard deviation is not proportional to $\sqrt{t}$ as seen in Figure 1.1. This is due to the converging nature of a random network.

Part d:

Figure 1.2 below shows the average distance and standard deviation plots for a network with 100 nodes and a probability of drawing edges of 0.01 while *Figure 1.3* shows the average distance and standard deviation plots for a network with 10,000 nodes and a probability of drawing edges at 0.01. The number of random walkers in the corresponding network is equal to the number of nodes in the network.

It is seen from *Figures 1.1-1.3* that the number of steps it takes for the average distance and standard deviation to converge is dependent on the number of nodes in the network. The larger the number of nodes in a network, the less amount of steps it will take for these measurements to converge. In addition, it is seen that even upon convergence there is a much larger standard deviation in the smaller network (n=100, STD ~3.5) compared to that of the larger networks (n=10,000, STD ~0.25). This might be because that there is a higher probability that a network with only 100 nodes is disconnected causing paths to be disrupted while networks of 1,000 or 10,000 nodes are typically always connected.

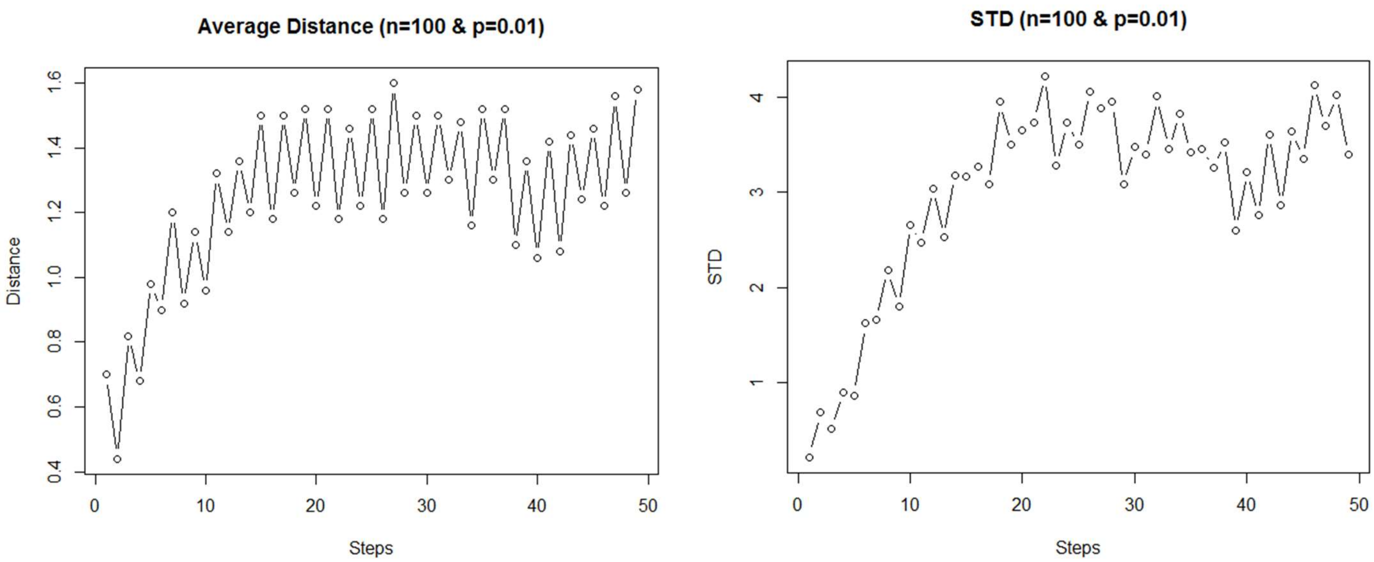


Figure 1.2 Average Distance and STD (n=100 & p=0.01)

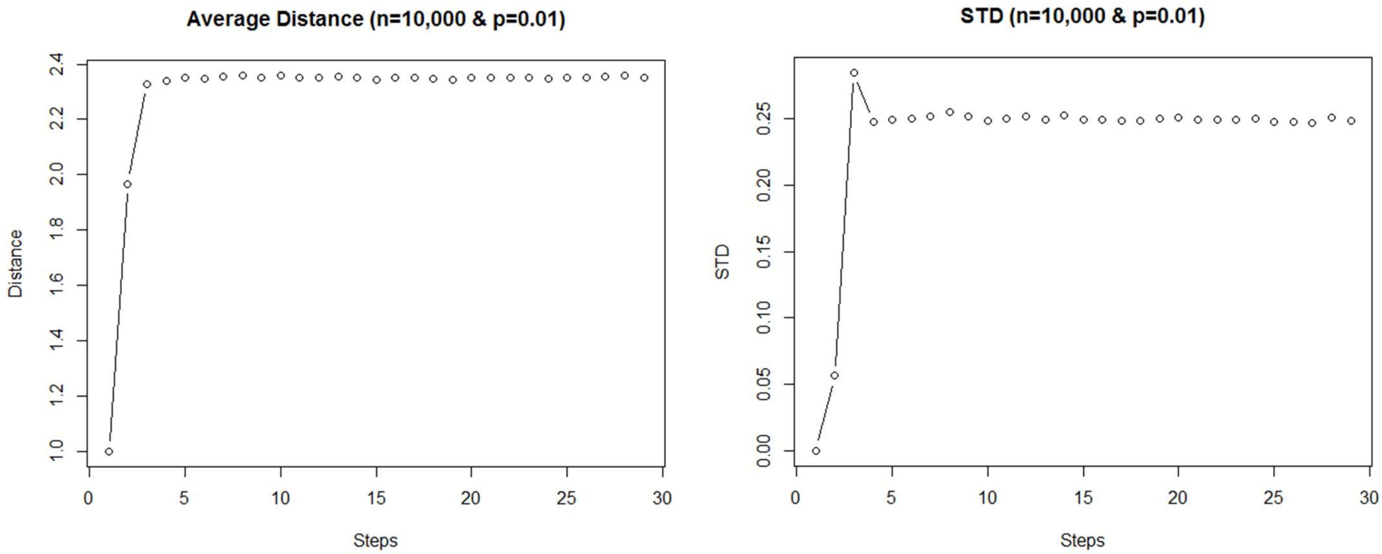


Figure 1.2 Average Distance and STD (n=10,000 & p=0.01)

Below in *Table 1.1* is a summary of the diameter of each of the networks in *Figures 1.1-1.3*. The role that the diameter of the network plays is that it serves as an upper bound for the distance that the average walker converges to. This is because diameter is defined as the longest shortest path and therefore it would be impossible for a random walker to exceed this distance.

Number of Nodes (n)	Diameter
100	13
1,000	5
10,000	3

Table 1.1 Summary of Network Diameters for Erdos Renyi Networks

Part e:

Figure 1.4 shows the degree distribution for a network with n=1,000 (left) and degree distribution at the end of the random walk for the same network (right). It is seen that these two distributions are similar but there is a bit more emphasis on the degrees around the mean (Degree = 8 to 12) for the end of random walk distribution versus the normal distribution. This makes sense because nodes with more edges connected to it (higher degree) will have a higher probability that a random walker will end up at that particular node. Thus leading to the end of a (converging) random walk to end up at one of those nodes.

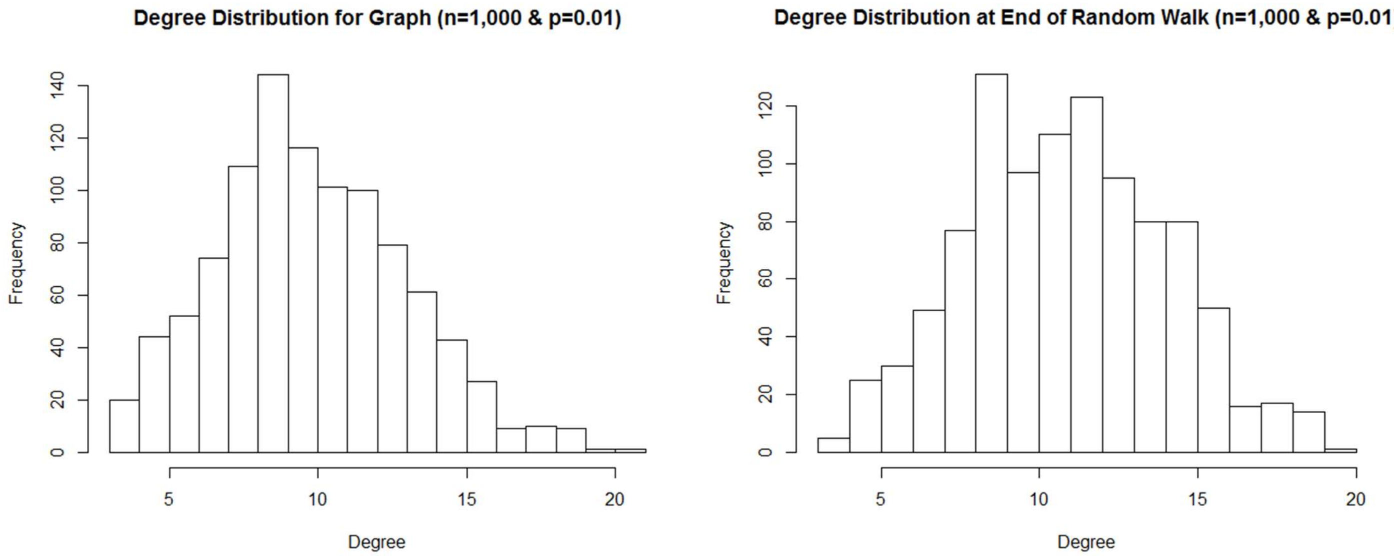


Figure 1.4 Degree Distribution of Network (Left) vs Distribution at End of Random Walk (Right)

Question 2:

Part a:

The function barabasi.game() was used to generates with a degree distribution proportional to x^{-3}in this question.

Part b:

The plots below in *Figure 2.1* show the average distance and average standard deviation of different walkers throughout their path (steps). It is seen that both the average distance and average standard deviation plots both converge after a certain number of steps. Both plots are averaged over 1,000 walkers (each walker starts at a different node).

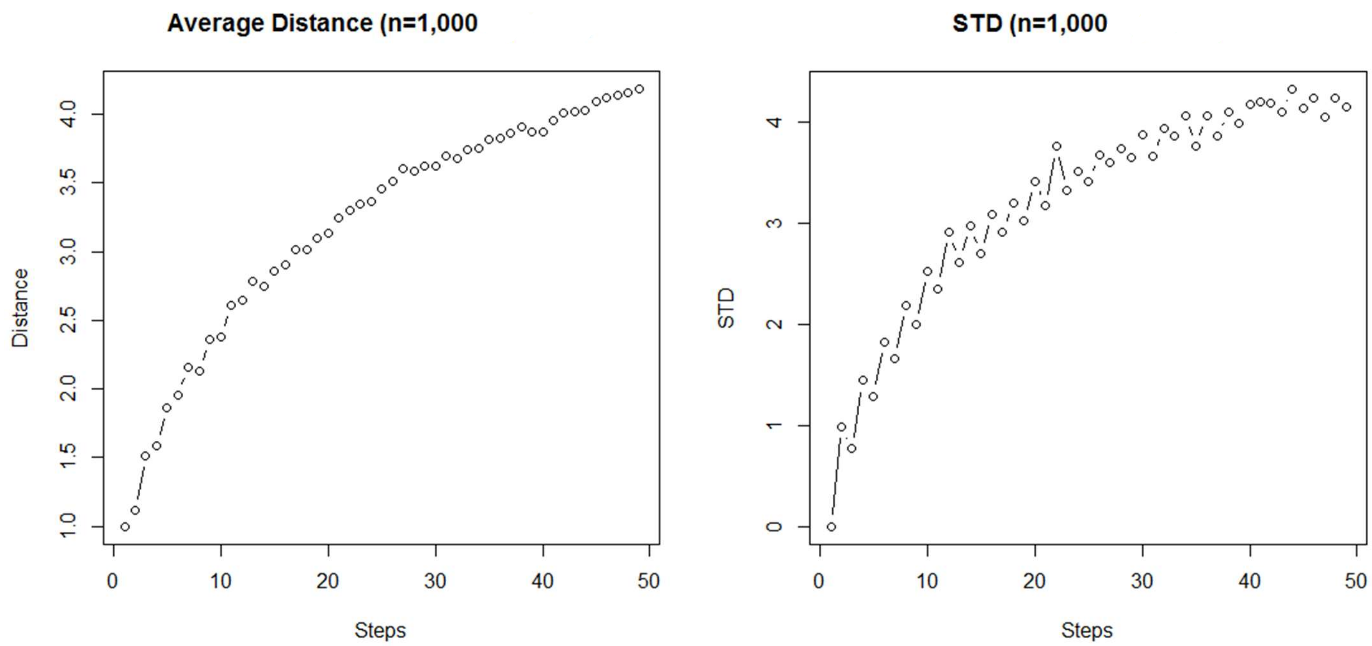


Figure 2.1 Average Distance and STD (n=1,000)

Part c:

A random walker in *d* dimensional has average (signed) distance of 0 and a standard deviation (σ) proportional to $\sqrt{t}$. This is not true in a random network because all distances in a random graph are positive. The reason why a random walker in *d* dimensional has average distance zero is because the positive and negative distances cancel each other out but since in a random network there are no negative distances, this cancellation does not occur leading the averages to converge some number not equal to zero. The standard deviation is not proportional to $\sqrt{t}$ as seen in Figure 2.1. This is due to the converging nature of a random network.

Part d:

Figure 2.3 below shows the average distance and standard deviation plots for a network with 100 nodes and a degree distribution proportional to x^{-3} while *Figure 2.3* shows the average distance and standard deviation plots for a network with 10,000 nodes and a degree distribution proportional to x^{-3}.The number of random walkers in the corresponding network is equal to the number of nodes in the network.

It is seen that the Barabasi networks in *Figures 2.1-2.3* have a similar curve and values for average distance and standard deviation compared to that of the Erdos Renyi networks in question 1. This could be because regardless of the number of nodes in a Barabasi network, most of nodes have a low degree versus an Erdos Renyi network where the degree distribution follows a bell curve. In addition, Barabasi networks are always connected.

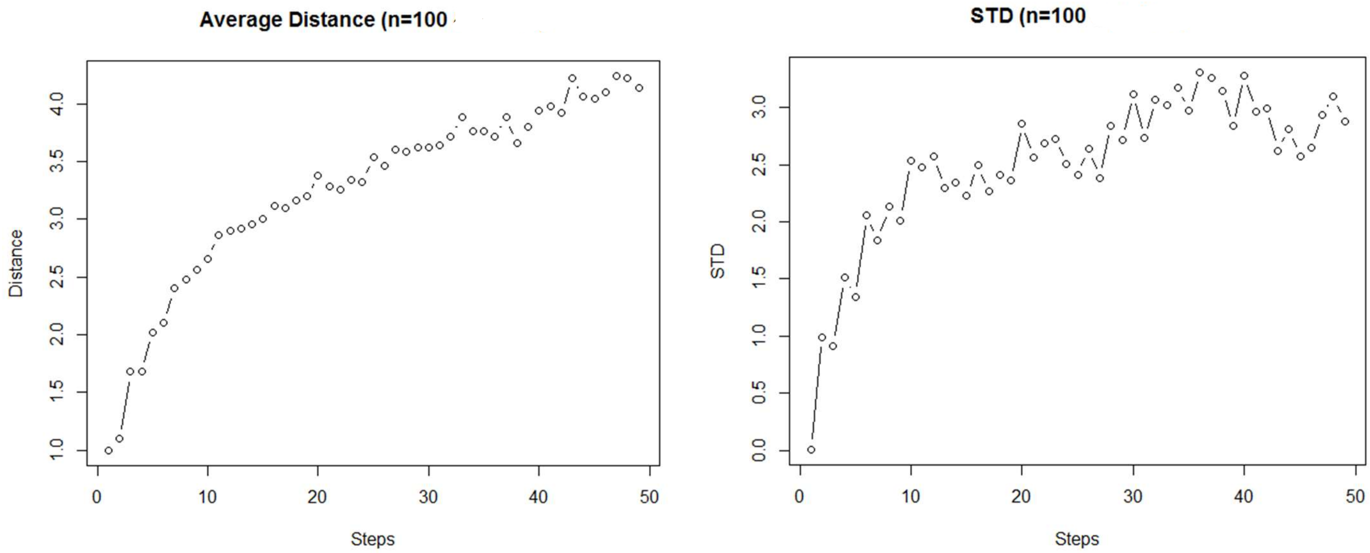


Figure 2.2 Average Distance and STD (n=1,000)

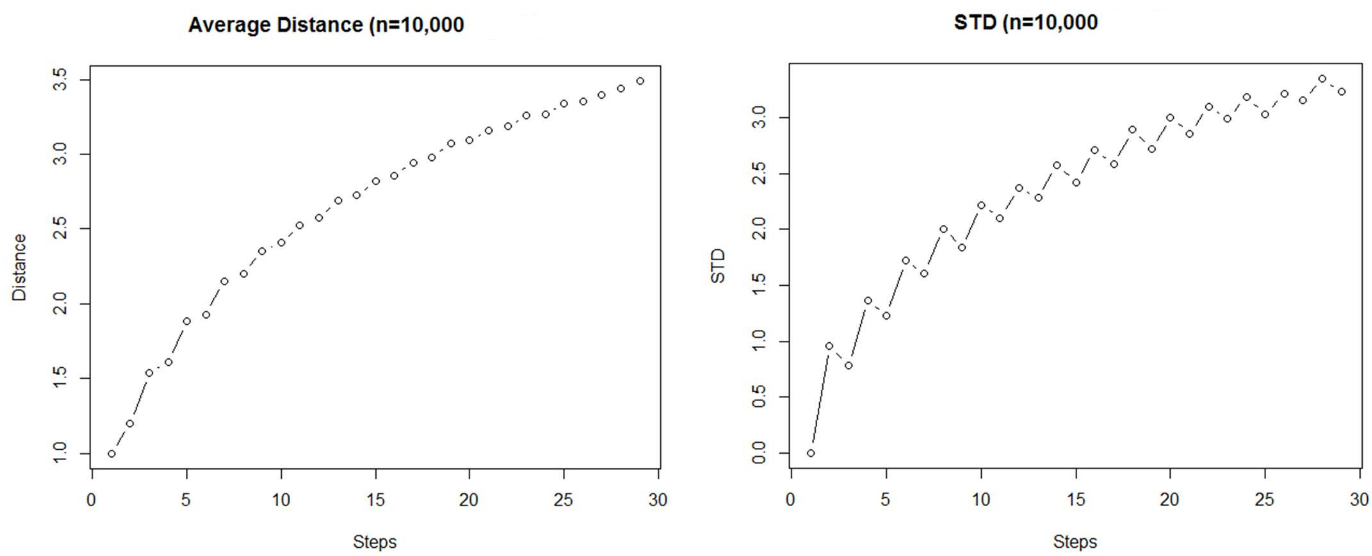


Figure 2.3 Average Distance and STD (n=1,000)

Below in *Table 1.1* is a summary of the diameter of each of the networks in *Figures 1.1-1.3*. The role that the diameter of the network plays is that it serves as an upper bound for the distance that the average walker converges to. This is because diameter is defined as the longest shortest path and therefore it would be impossible for a random walker to exceed this distance. It is noted that as the number of nodes in a Barabasi network increases, the diameter also increases while for a Erdos Renyi network, as the number of nodes increases, the diameter decreases. This may be another reason why the average distance and standard deviation curves for a barabasi network remain the same regardless of the number of nodes while that is not true for an Erdos Renyi networks.

Number of Nodes (n)	Diameter
100	12
1,000	21
10,000	29

Table 2.1 Summary of Network Diameters for Barabasi Networks

Part e:

Figure 1.4 shows the degree distribution for a network with n=1,000 (left) and degree distribution at the end of the random walk for the same network (right). It is seen that the distributions of both plots are similar. The little variations in the plot are due to randomness of the walks.

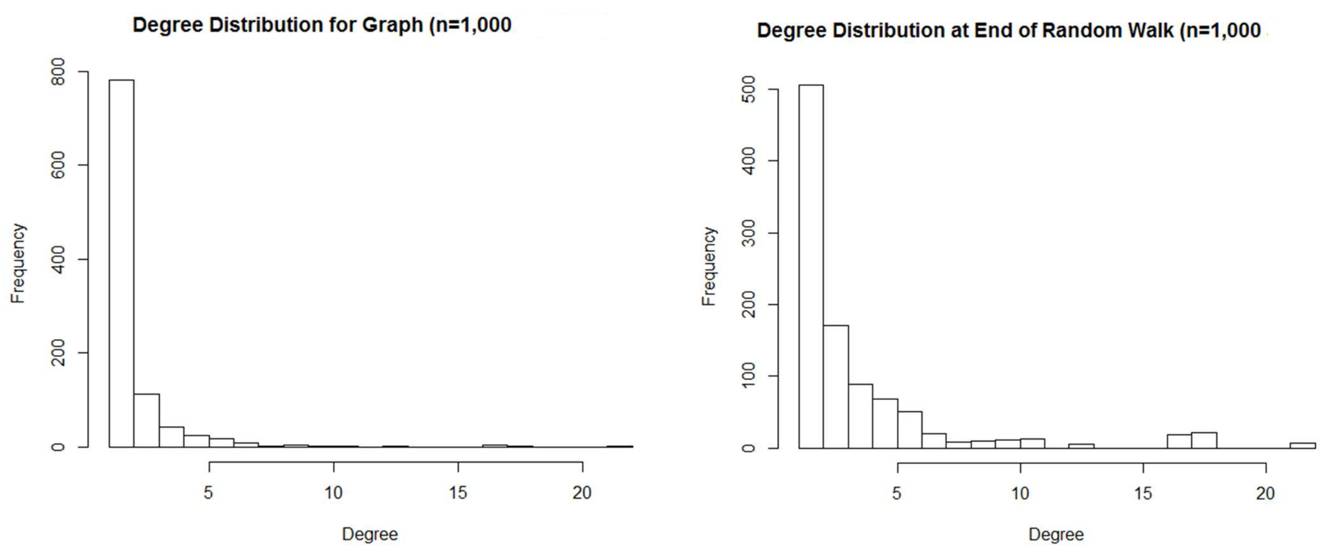


Figure 2.4 Degree Distribution of Network (Left) vs Distribution at End of Random Walk (Right)

<u>Question 3:</u>

Part a:

As *Figure 3.1* shows below, there is a linear relationship between the degree of a node and its visit probability. As the degree of a node increases, so does its visit probability. Visiting probability was generated using the *netrw()* function.

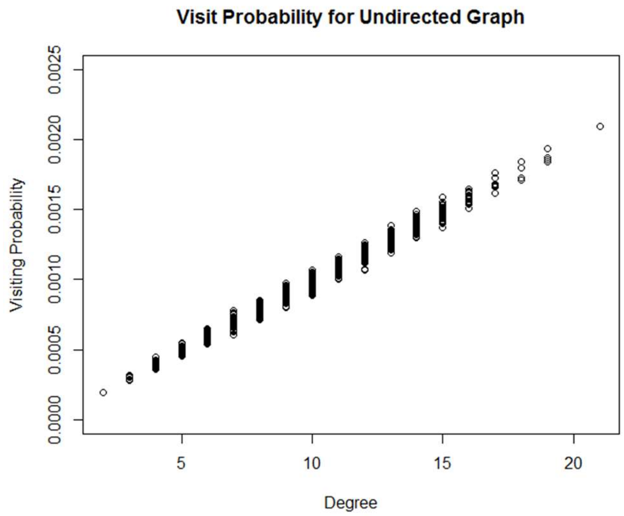


Figure 3.1: Visit Probability vs Degree for Undirected Graph

Part b:

We created a similar network to the one in part a) however this time the graph that is generated was a directed graph. Therefore when we observe the visit probability for each node in this network compared to its degree, we must observe the in-degree and the out-degree. In *Figure 3.1* we can see that there is a much weaker linear relationship between visit probability and in-degree for nodes in a directed graph compared to that of a undirected graph.

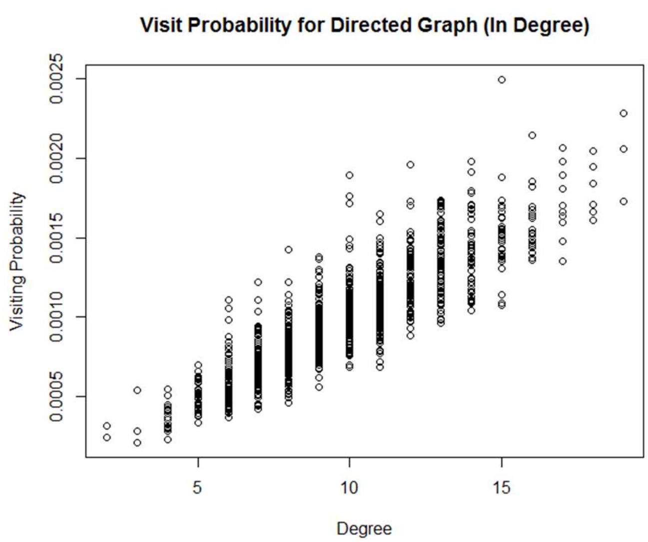


Figure 3.2: Visit Probability vs In-Degree for Directed Graph

In *Figure 3.2* however we can see that there is relationship between the visit probability and out-degree of a graph. This is sound with our intuition as a node is more likely to be visited if it has a large in-degree where as having a large out-degree has no effect on the number of times a node will be visited.

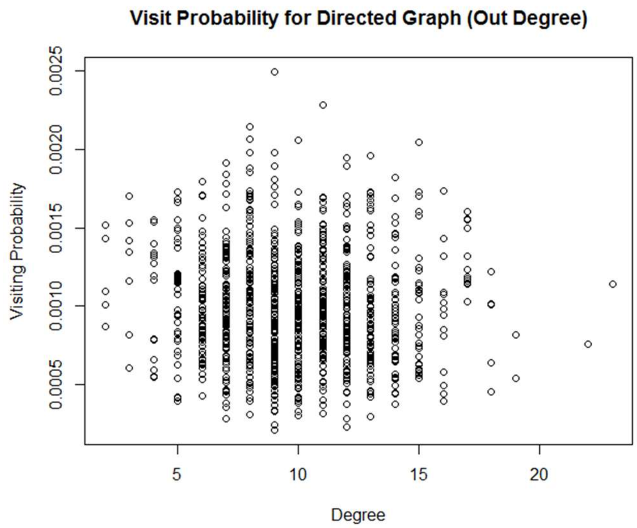


Figure 3.2: Visit Probability vs Out-Degree for Directed Graph

If we combine average the results we get in *Figure 3.1* with the results in *Figure 3.2* we achieve the graph shown in *Figure 3.3.* A linear relationship between the visit probability and the degree in a directed graph is not visible anymore because this graph takes into account the in-degree as well as the out-degree.

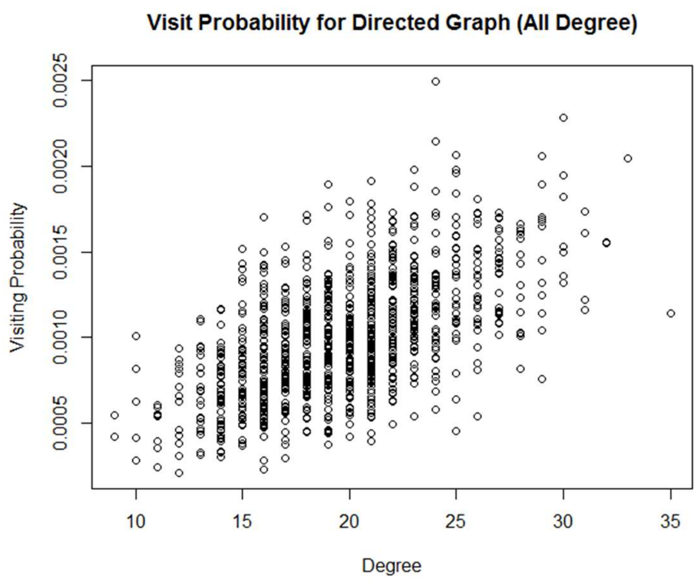


Figure 3.3: Visit Probability vs Degree for Directed Graph

Part c:

Having created a network with a damping factor < 1, we now have the possibility of a teleportation event. Given that the damping factor is relatively high (0.85), the possibility of a teleportation event is relatively low (15%). Therefore we expect the visit probability vs degree graph for an undirected graph to look similar to the one in *Figure 3.1. Figure 3.4* shows exactly that as the linear relationship seen in part one is preserved but there is slightly more variance due to the possibility of teleportation.

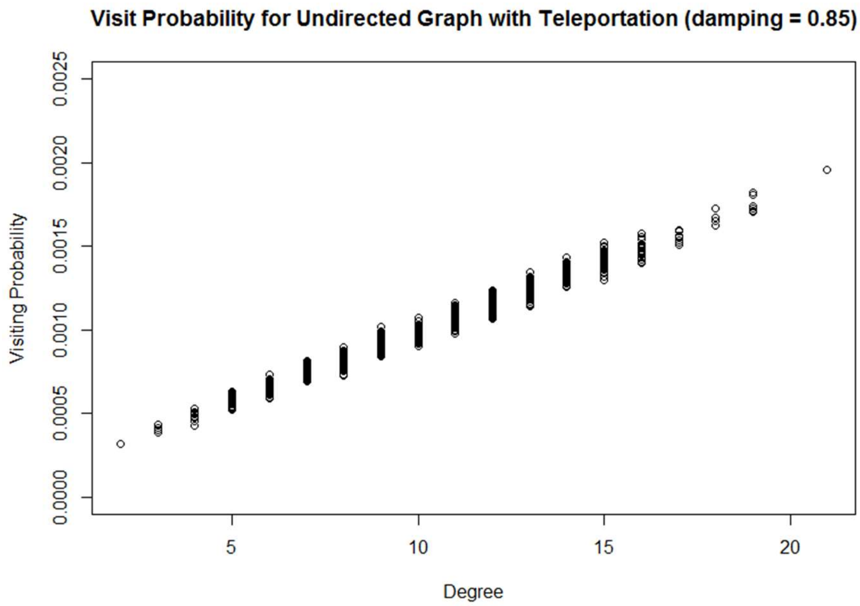


Figure 3.1: Visit Probability vs Degree for Undirected Graph with Damping = 0.85

<u>Question 4:</u>

Part a:

An undirected random network with 1000 nodes and probability p for drawing an edge between any pair of nodes was created in R using the *erdos.renyi.game(n,p)* function. Random Walkers were placed on the graph using the netrw package to generate an initial PageRank.

Part b:

The network we have created here differs from the one generated in part a) by having a non-uniform teleportation probability to another node given that a teleportation event occurs (15% chance). This teleportation probability to another node is now determined by its PageRank which we have generated from simulating the random walk from *part (a)*.

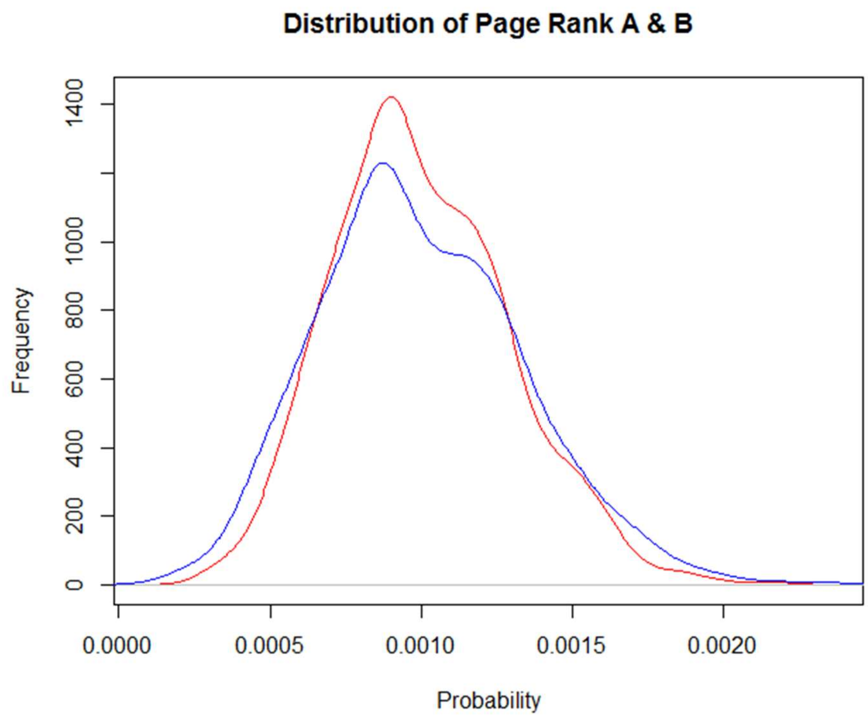


Figure 4.1: Distribution of PageRank (Red = PageRank A | Blue = PageRank B)

Figure 4.1 above shows the range of probabilities (X-axis) contained in the **resulting** PageRank for parts (a) and (b) as well as the frequency of each probability (Y-axis). We see here that the high probabilities get higher and the low probabilities get lower. Given that the sum of the probabilities must add up to 1, the frequency of middle probabilities decreases while the probabilities of both ends increase.

Part c:

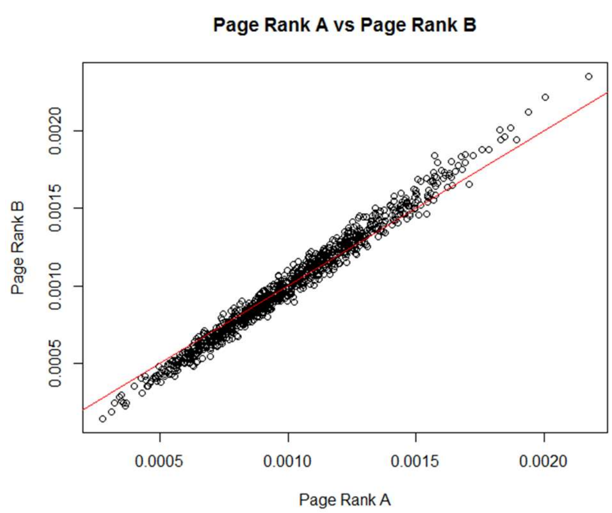
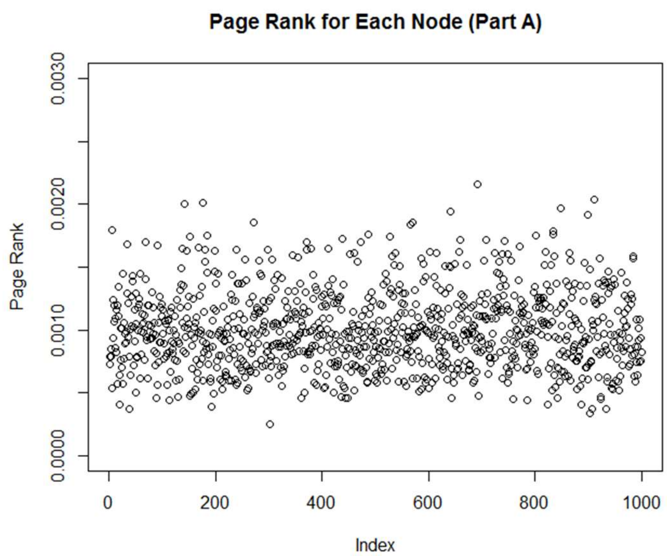


Figure 4.2: Comparing Probabilities of PageRank A and B

We can see in *Figure 4.2* the changes in the probabilities from PageRank A to PageRank B. Comparing the distribution of the points to the redline, which represents a line of slope 1, we see that the higher probabilities in PageRank A are indeed mapped to higher probabilities in PageRank B vice versa for lower probabilities.



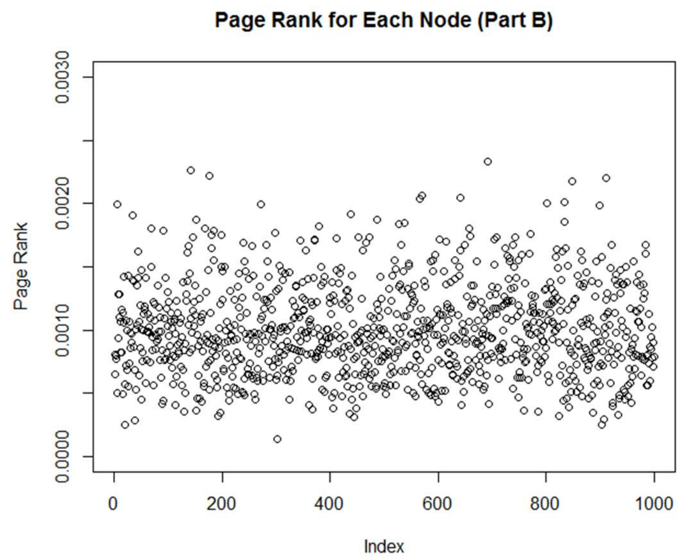


Figure 4.3: Index of Node and Corresponding PageRank for Part A (Left) and B (Right)

Figure 4.3 also shows the polarization of these probabilities as we can see the probabilities have become more spread out vertically in Part B when compared to Part A.