Kevin Wang (604036846)

EE 232E Homework 2

The coding for this assignment was done in R. There is a separate .R file for each problem, named in the fashion hw2-x, where x is the problem number.

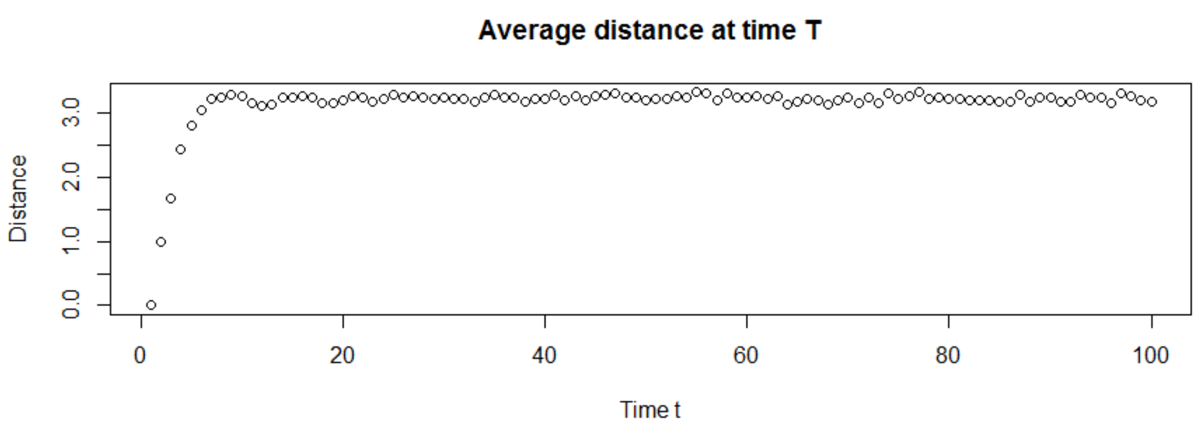
Problem 1

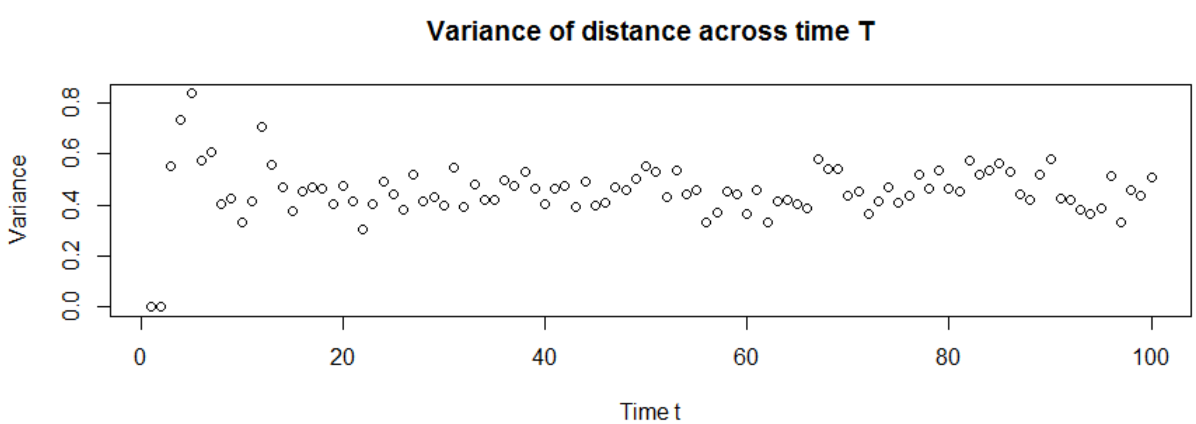
Part 1a

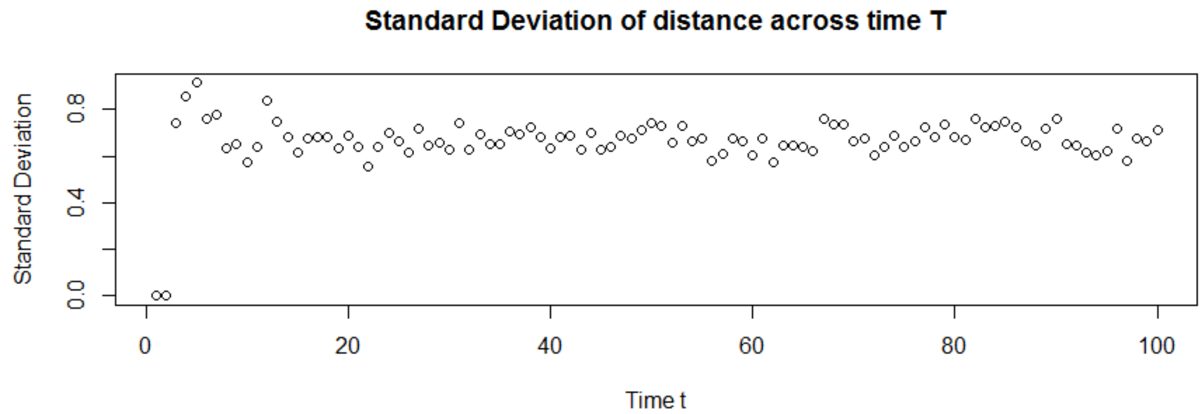
The random networks are generated using *random.graph.game*.

Part 1b

The *netrw* function is used. The default time steps T=100 and 100 walkers is used with damping = 1 (no damping).







Part 1c

The random network mean distance and standard deviation plots are not that similar to the known behavior of random walkers in d dimensional spaces. For the random networks, both the mean and SD/Variance plots seem to converge on a value for high values of time t. Perhaps the mean plot did not converge to 0 since the sign of the walk for the random network did not matter, so there were no negative distance values to counterbalance the positive distance values. The behavior may still be somewhat similar in that they both converge to a final value for (mean distance).

Part 1d

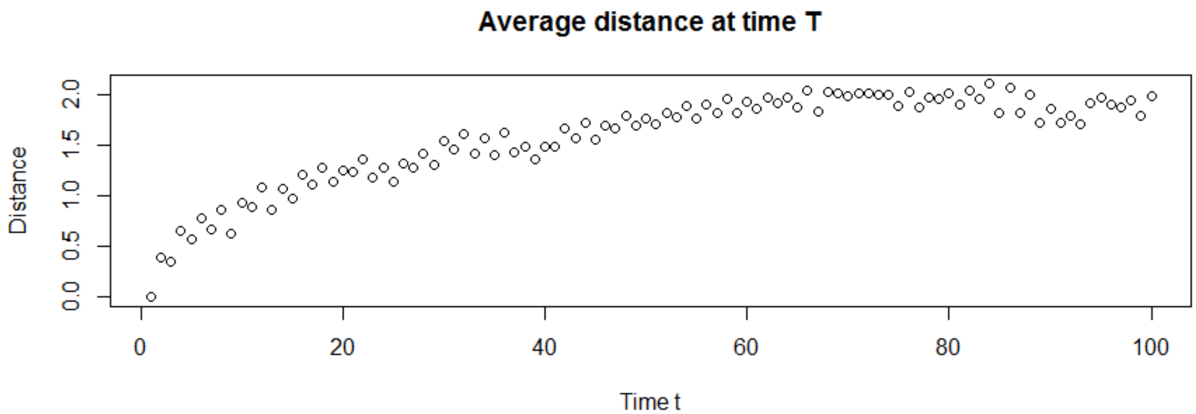
Nodes	100	1,000	10,000
Diameter	14	6	3

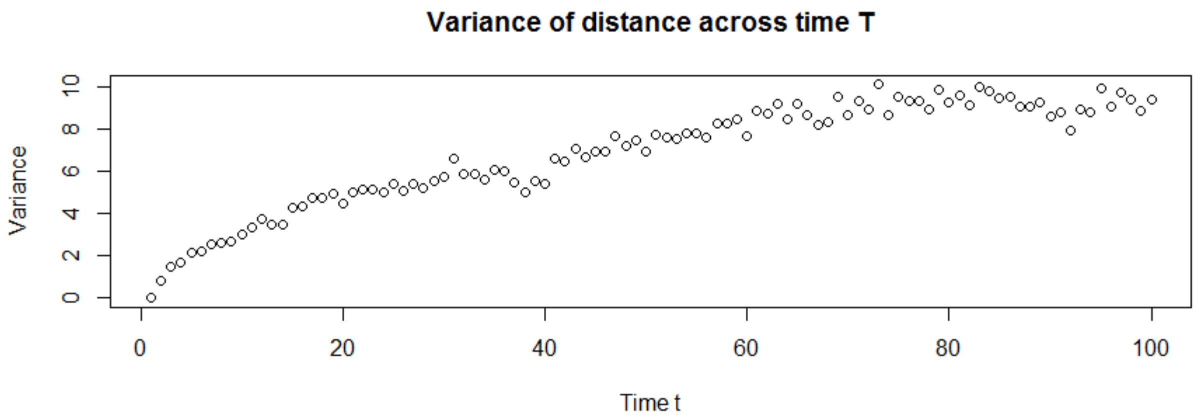
It is noted that the 100 node network was not connected.

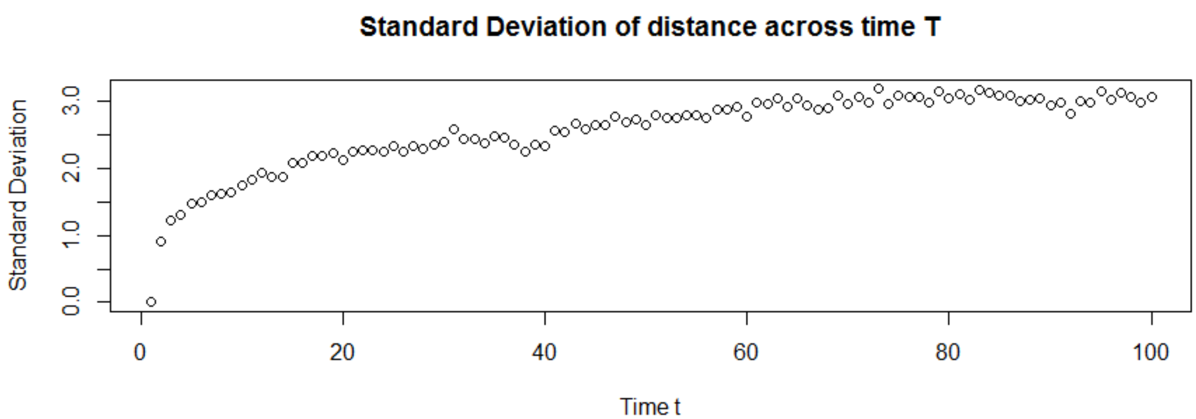
From the plots on the following pages, it is apparent that even if the number of nodes in the network is changed, then the plots for average and standard deviation still appear to converge to a value as t grows. The plots all start from zero and then rise until approaching the convergence value. The only difference is that for fewer nodes, it took longer to converge, especially for the average plots. For standard deviation and variance, all of the plots were still somewhat scattered but there is still visual evidence that the plots approach some sort of final value.

The diameter of the plots decreased as the number of nodes went up. The average distance when looking at the 1000 and 10,000 node plots went down as diameter went down/number of nodes went up (the 100 node network was not fully connected which likely caused issues with the *get.shortest.paths* distance finding method). The standard deviation seemed to go down as diameter went down/number of nodes went up. Diameter likely somewhat influences the distance results since it is a measure of the maximum distance possible between the end node and the start node of a random walk.

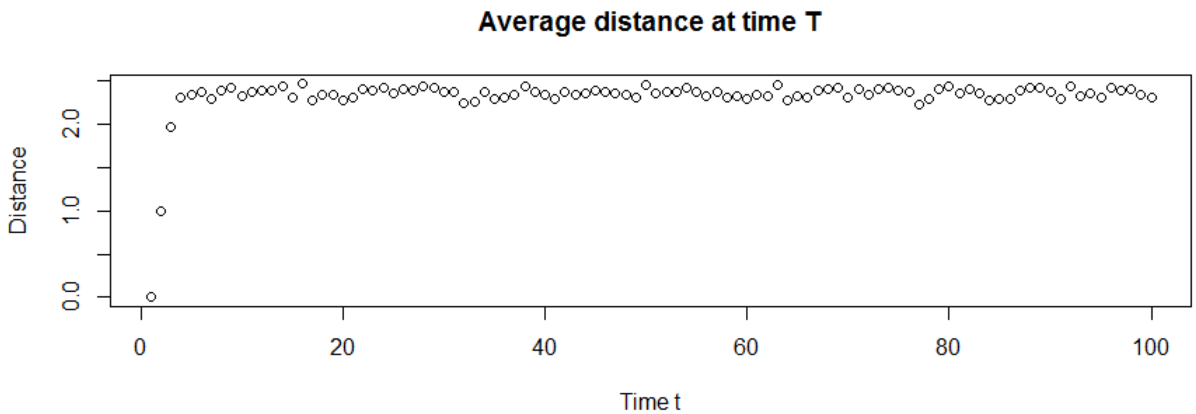
For 100 nodes:

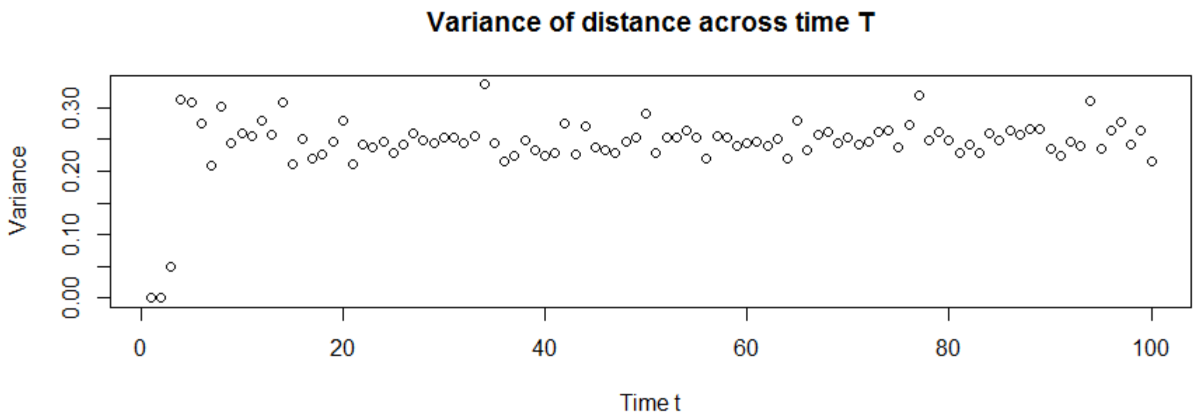


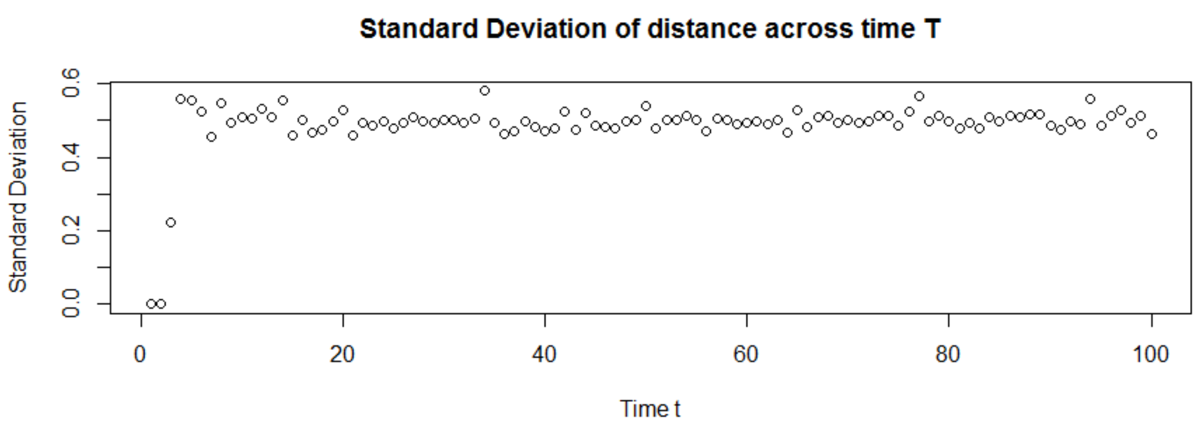




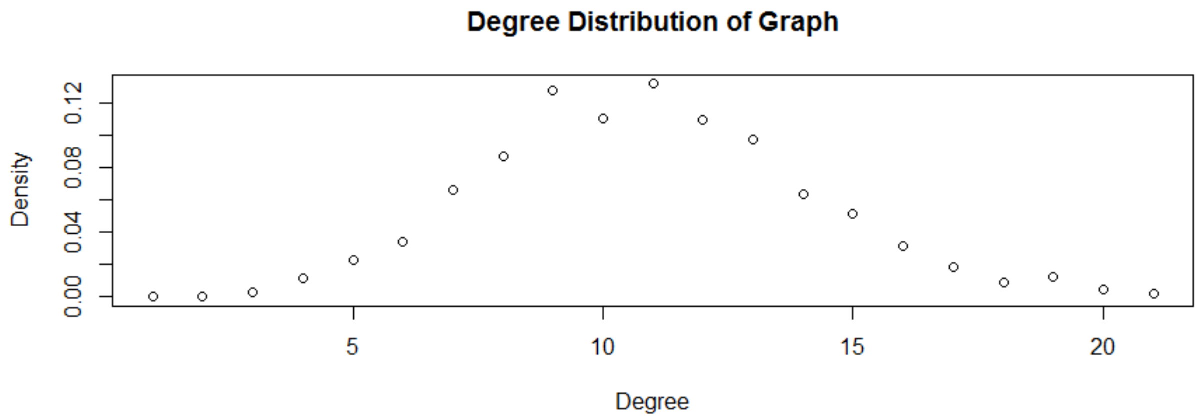
For 10,000 nodes:

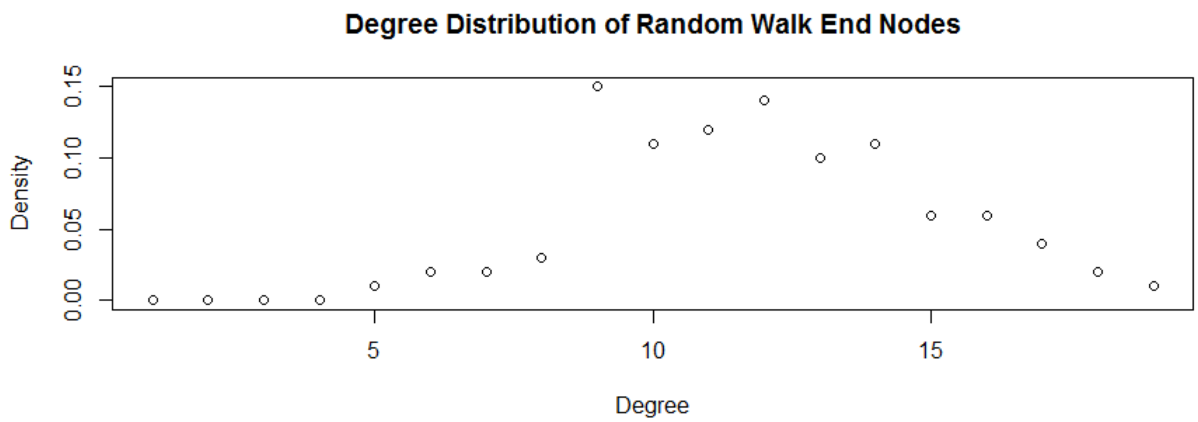






Part 1e





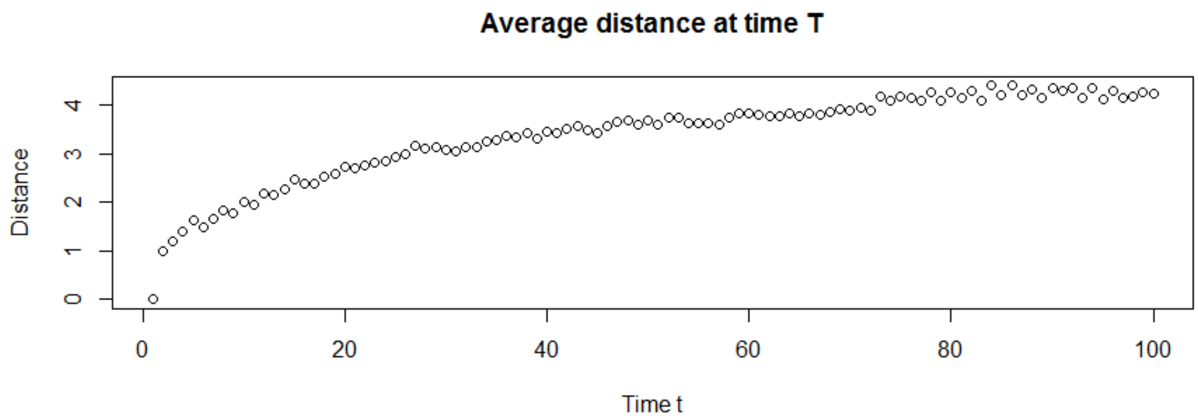
The plot for the degrees of the end nodes is less smooth than the plot for the degree distribution of the overall network, likely since there were fewer nodes in the end node sample. However, the shapes of the plots are still similar, which makes sense since the end node of random walks is like randomly sampling the original network.

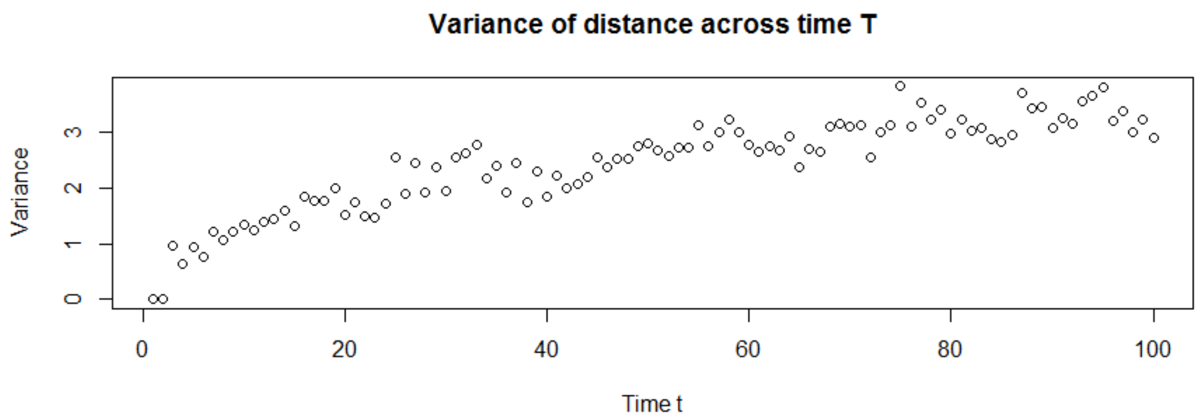
Problem 2

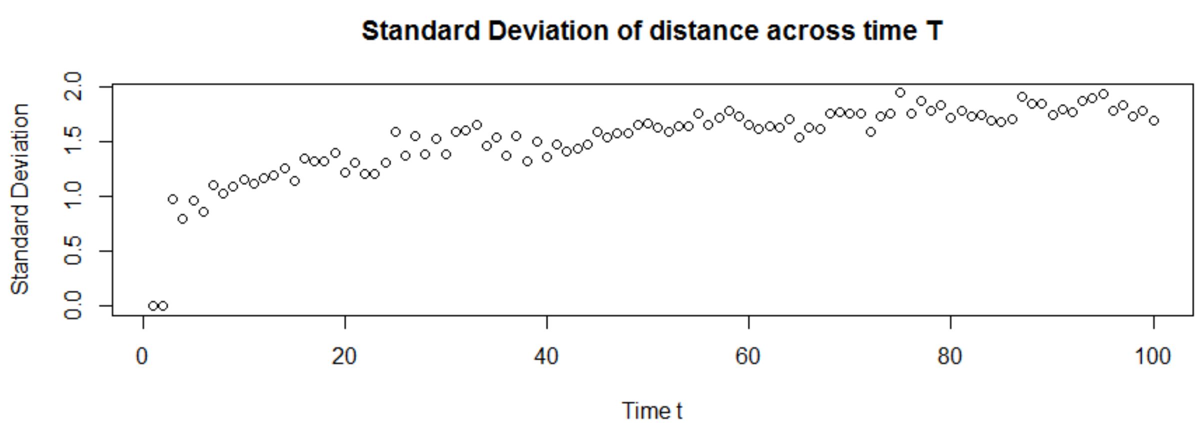
Part 2a

barabasi.game was used to generate undirected fat-tailed degree distribution networks.

Part 2b







Part 2c

Once again, the mean did not result in 0 since there were no negative distances accounted for. However, when examining the standard deviation/variance plots, there is enough scatter that it is indeed possible that the plots are $\propto \sqrt{t}$, but there is still enough scatter that it is possible that the plots do converge to some value instead like before. This might be due to the preferential attachment nature of the network generating algorithm, resulting in more small-degree nodes which might have resulted in long chains of nodes for high distances.

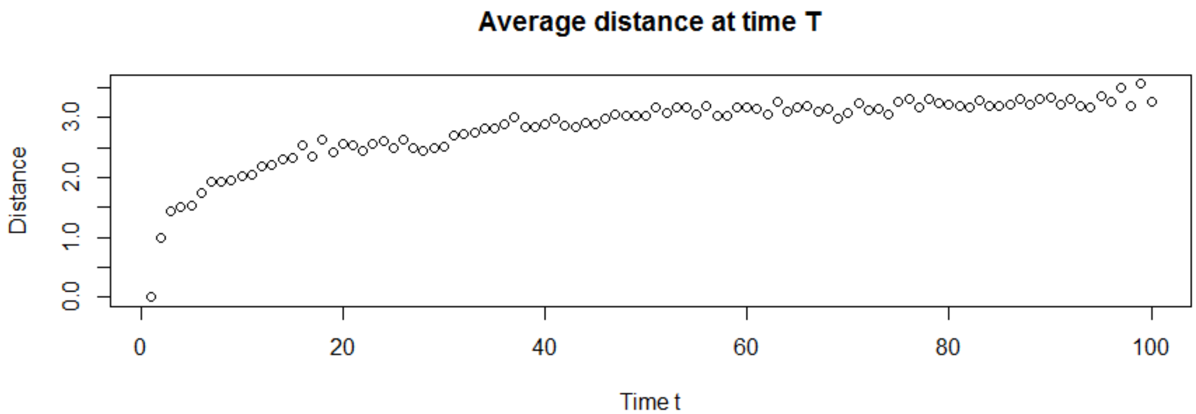
Part 2d

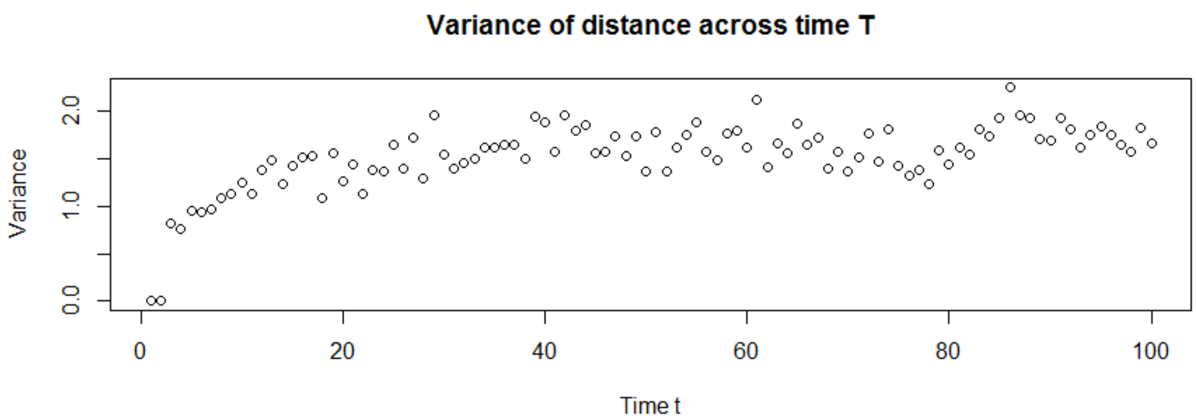
Nodes	100	1,000	10,000
Diameter	9	16	24

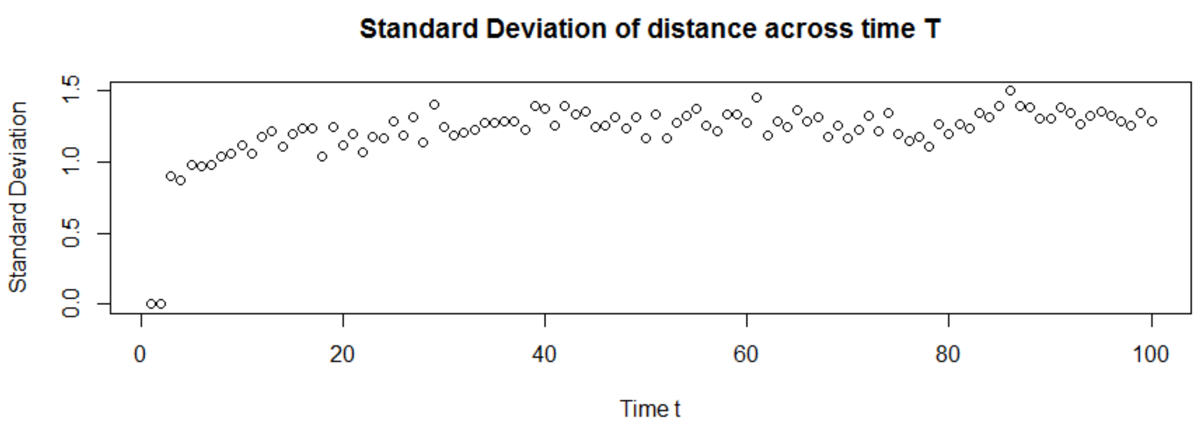
From the plots on the following pages, it is apparent that the distributions remain similar to the plots from the networks in part 1. Even if the number of nodes in the network is changed, then the plots for average and standard deviation still appear to converge to a value as t grows. The plots all seem to increase at approximately the same rate and approach a final value. Variance is still somewhat scattered but it is more apparent that the distribution does not seem to follow a $\sqrt{t}$ pattern.

For these games, the diameter increased with the number of nodes instead. The average distance went up for larger diameter/number of nodes, however the difference between the 1,000 and 10,000 networks is not very distinct. Perhaps the average distance stops increasing past a certain node size of the network. The variance plots were all scattered, but appeared to be somewhat lower for lower diameter/number of nodes.

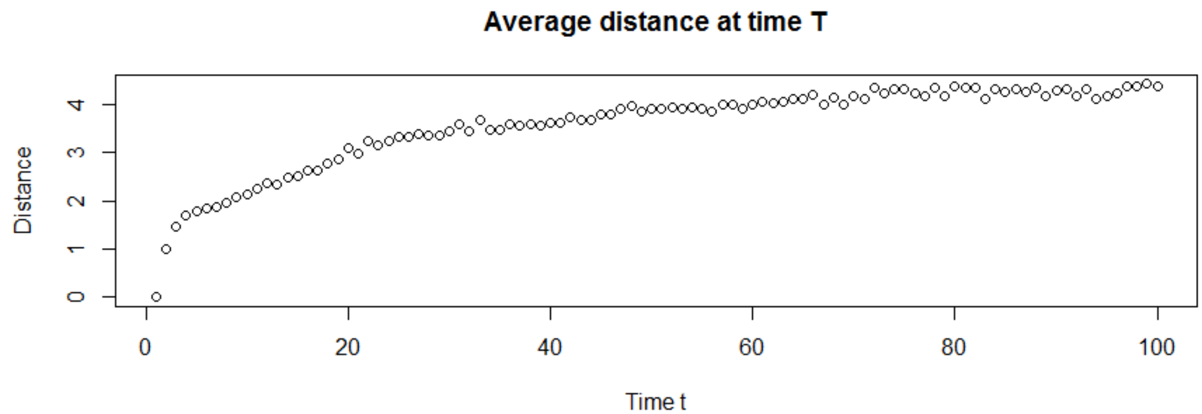
For 100 nodes:

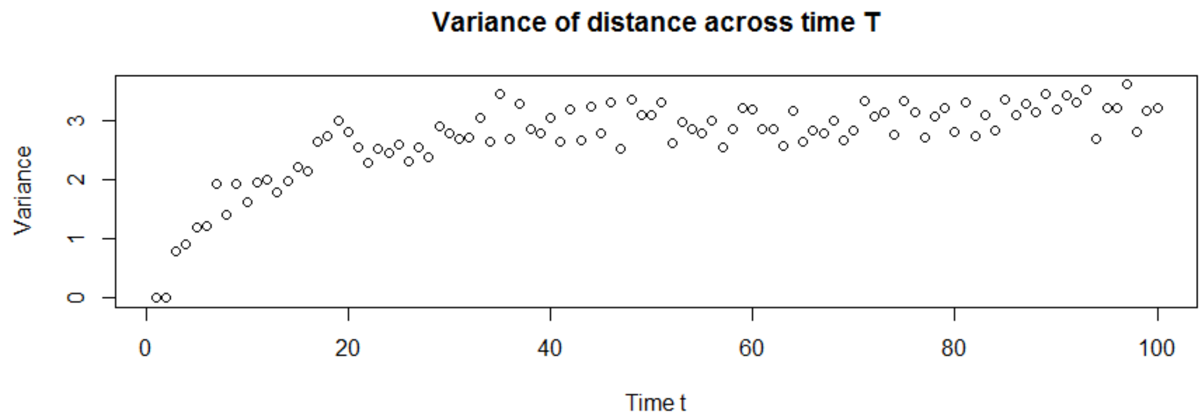


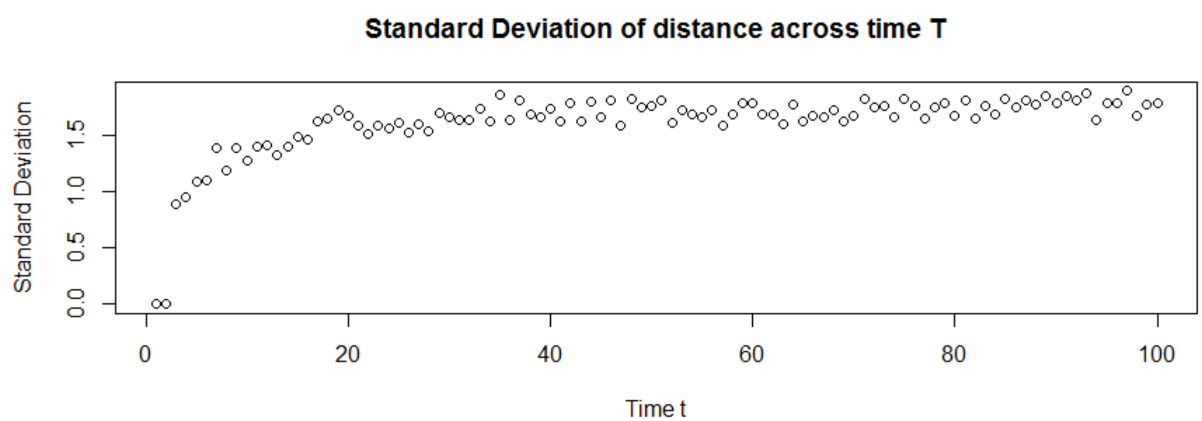




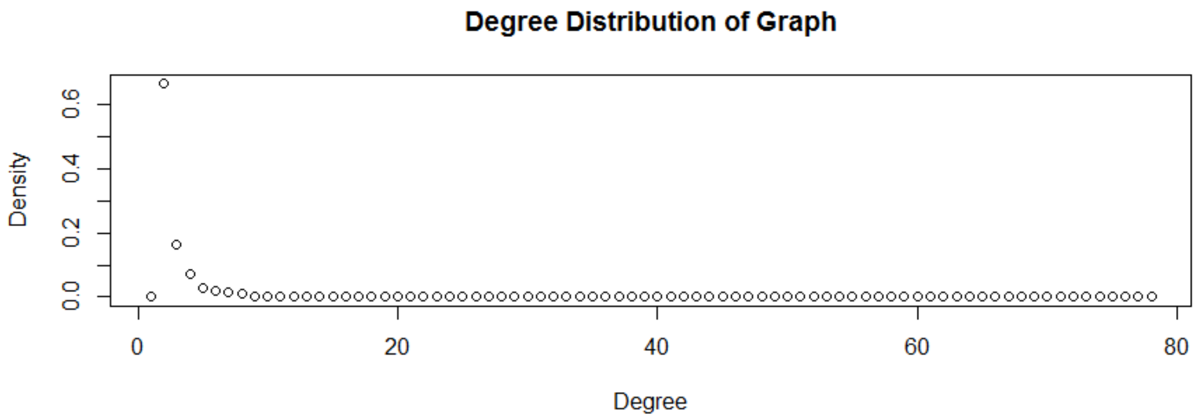
For 10,000 nodes:

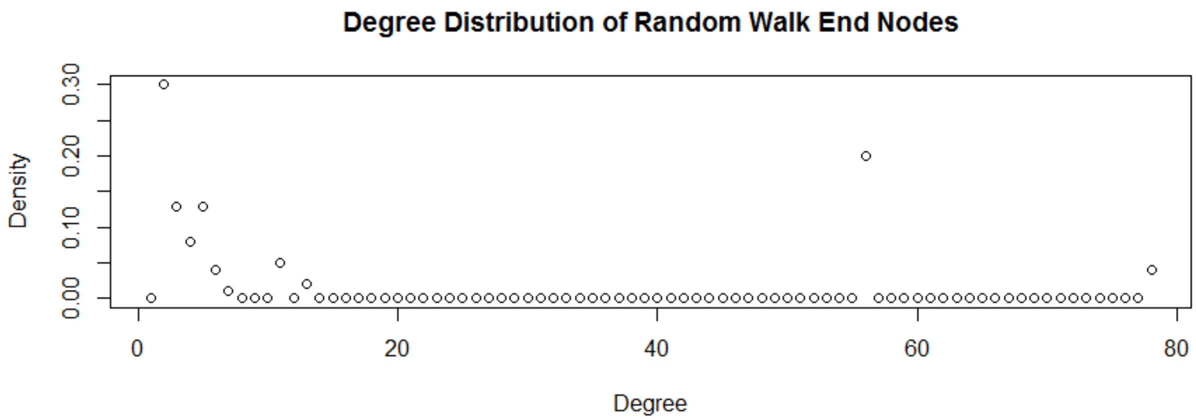






Part 2e

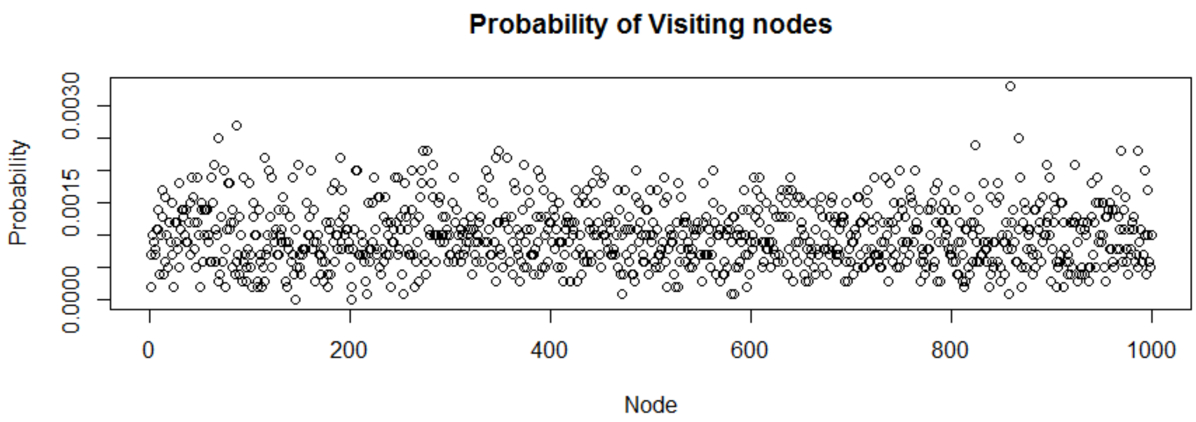


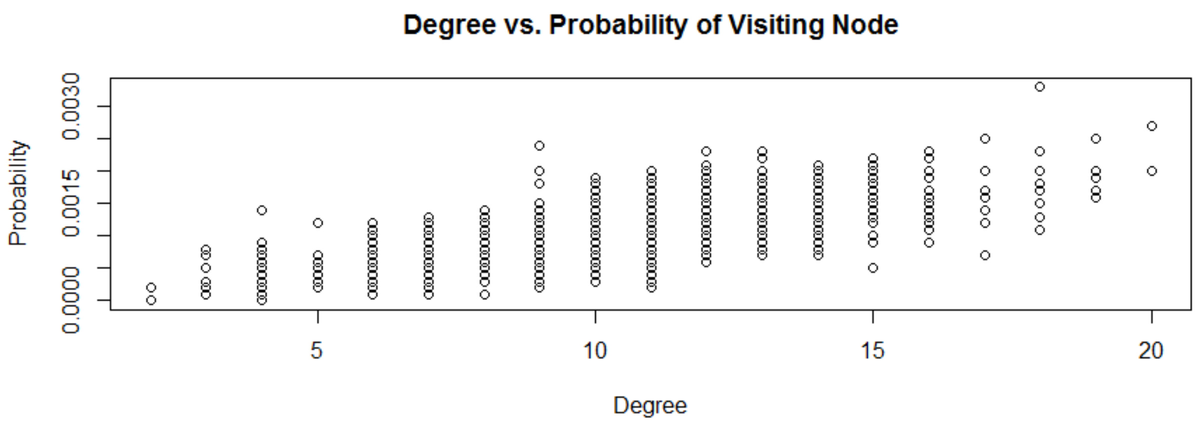


The distribution of the degrees of end nodes was similar to the degree distribution for the overall network. However, the end node plot had some outliers. This is likely due to the fact that if a high-degree node is traversed during a walk, it is possible to keep returning to that node from its large amount of neighbors since the graph is undirected and the other nodes have low degrees.

Problem 3

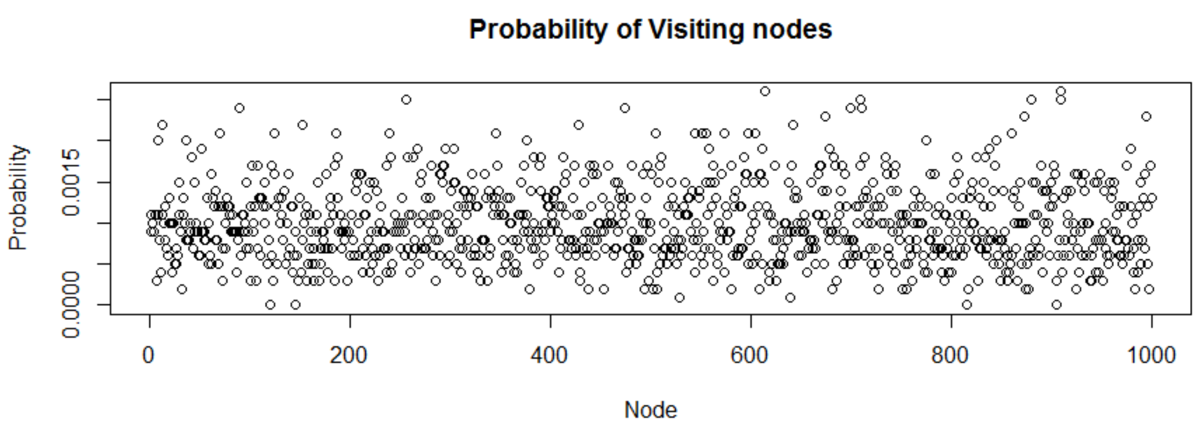
Part 3a

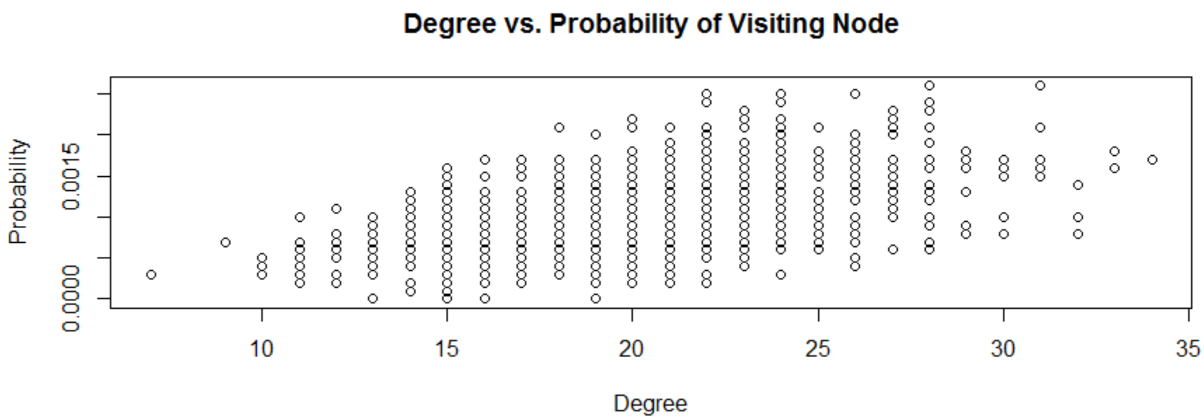




For the undirected graph, it appears that as the degree of the node increased, the probability of visiting the node trended upwards. This might be because larger degrees mean there is a chance of a walker going back and forth between the large degree node and its neighbors.

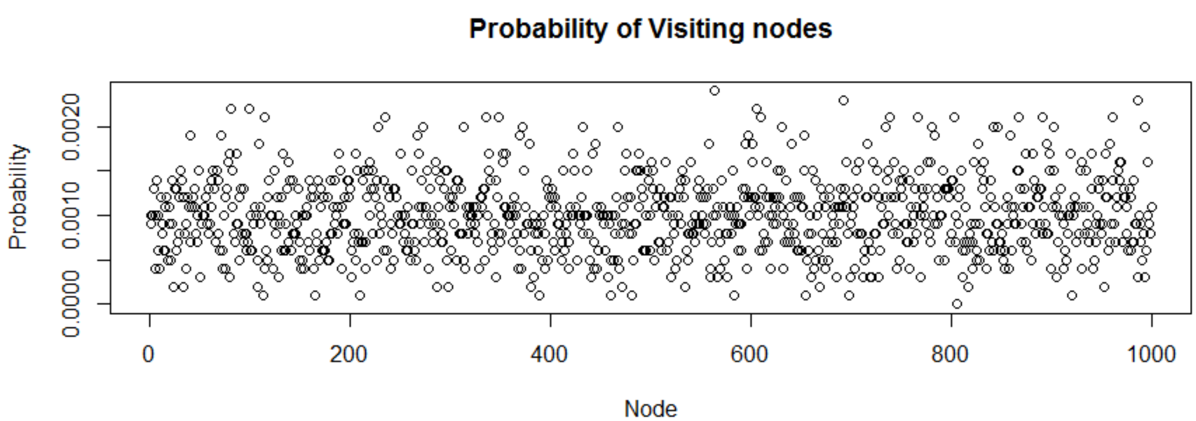
Part 3b

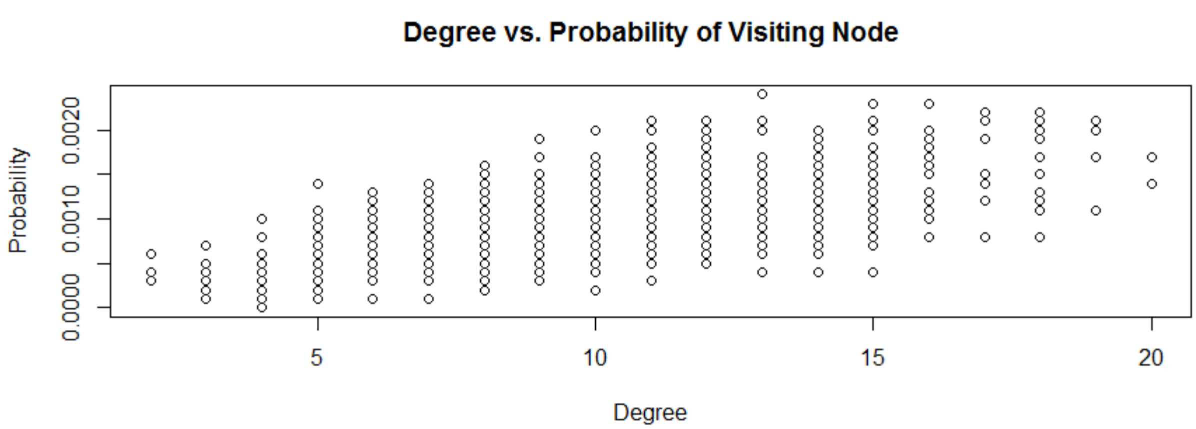




With the directed network, once again it appears that for higher degrees, there is increased probability for visiting the node.

<u>Part 3c</u>

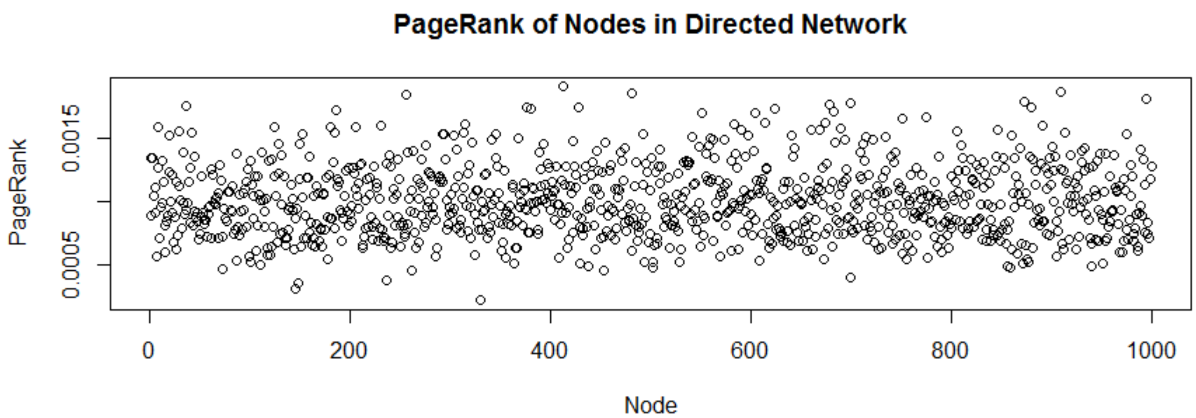




Once again, for the undirected network, even with damping=0.85, it appears that higher degrees correspond to higher probability of visiting the node during a random walk.

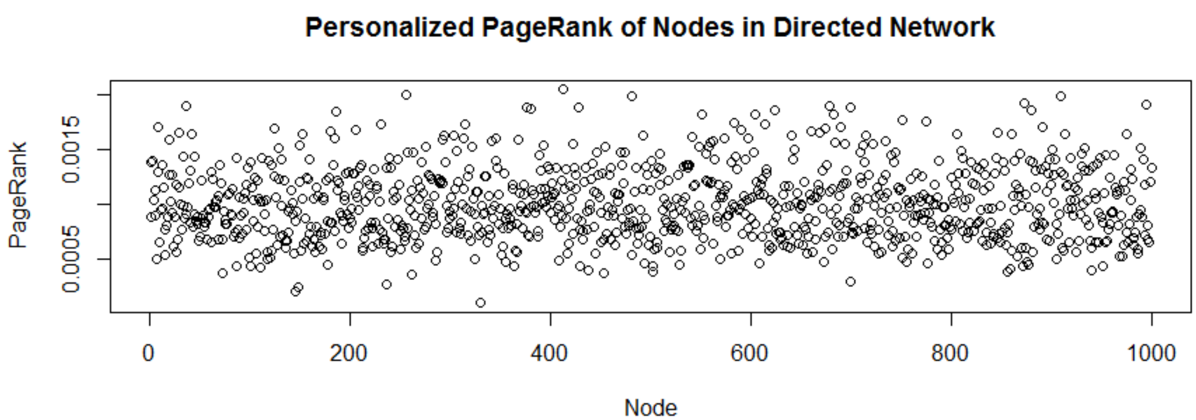
Problem 4

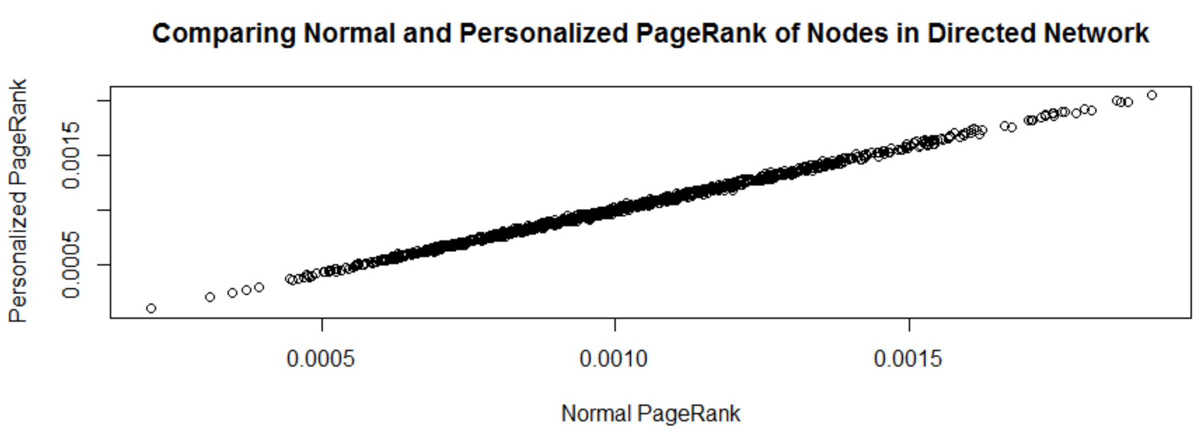
Part 4a

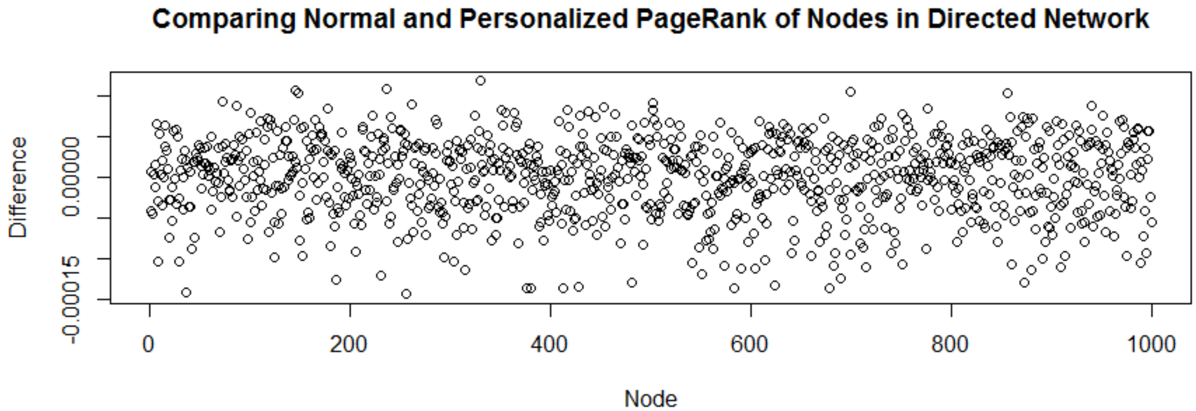


The PageRank is calculated using the *page.rank* function from igraph. An alternative is using *personalized.pagerank* from netrw.

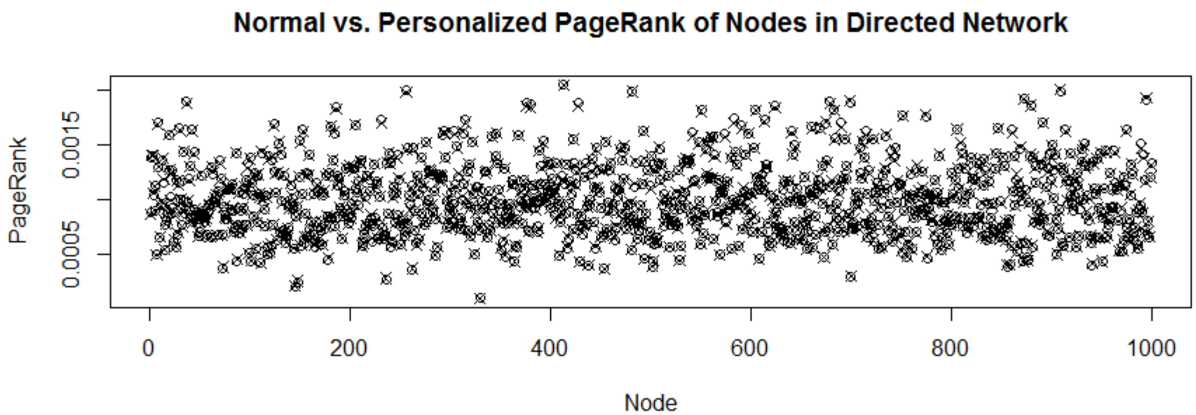
Part 4b







The mean difference between the two PageRanks is 5.08655e-15.



This plot plots the normal PageRank as X's and personalized PageRank as O's. It is apparent that the two PageRank methods produce very similar results, whether the teleportation probabilities to each node is the same or weighted according to PageRank.

<u>Part 4c</u>

The likelihood of teleporting to each node can be accounted for by specifying them in the PageRank calculating functions. However, when the personalized PageRank method described in this assignment was implemented (weighting the probabilities according to PageRank), no obvious difference was detected. Thus, it is possible that no adjustments are needed to account for the self-enforcement.