CEN 501 : Computer Systems II: Algorithm Design Project

Submitted by Navneet Jha(1207987452), Deepthy George(1208621176), Lakshmi Srinivas(1208635554)

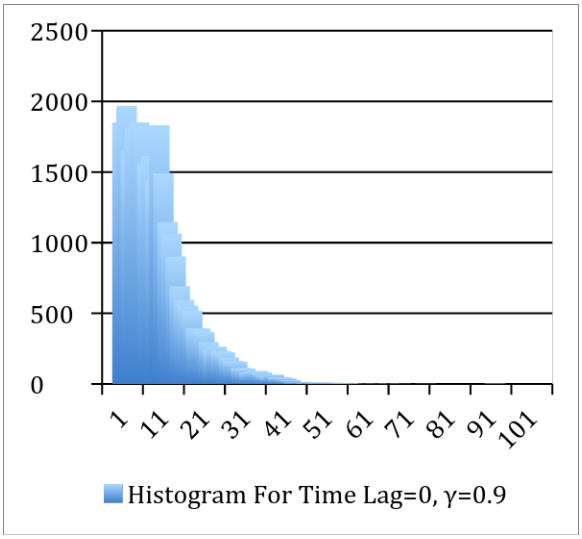
Methodology

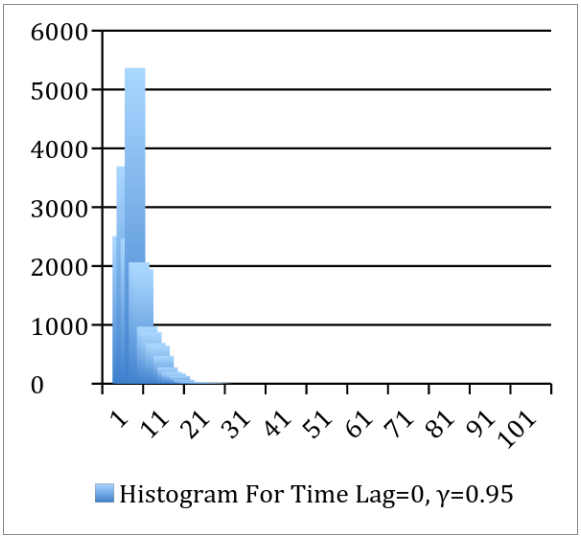
To create the graph, an adjacency list based approach was followed, which is much more efficient than the adjacency matrix based method.
Further, -O3 compiler optimization was used, which by loop unrolling and other optimizations, reduces the length of code, thus making the program run faster
The correlation coefficient is calculated right after reading from the file, and stored as a property of the node (by using structures) so that every time the particular node in involved, the same calculation doesn't have to be repeated. This was done to reduce time taken for graph creation
The characteristic path length, shortest distance between any two given nodes, is found out by using the breadth first search algorithm.

Task 1 : Analysis for 27-year period

a. Plotting histogram

The histogram is plotted for the degree of distribution for each of the nodes.





For the correlation of 0.9, The Average degree for the graph is 10.39. From the above image, we can clearly see that certain vertices have significantly higher degree than the average. These are the supernodes. In this case, the supernodes are (443, 111), (444, 117) and (445, 120). For the correlation of 0.95, the average degree for the graph is 4.726. The supernodes, for this histogram, are (204, 7), (203, 10) and (441, 102).

b. Computing Clustering Coefficient and The Characteristic path length

Clustering Coefficient

The clustering coefficient for every vertex is computed and the mean of all those values is used to compute the clustering coefficient of the graph.
The clustering coefficient for the graph with correlation = 0.9 is 0.4678
The clustering coefficient for the graph with correlation = 0.95 is .53018.

Characteristic Path Length

The characteristic path length for the graph is the mean over all the pairs of vertices.
The characteristic path length for the graph with correlation = 0.9 Is 37.462. The characteristic path length for the graph with correlation = 0.95 is 42.173.

c. Metrics for Random Graphs and Comparison to Values Computed

The clustering coefficient for the random graph is $7.62*10^{-5}$.
The characteristic path length for the random graph is 5.050.

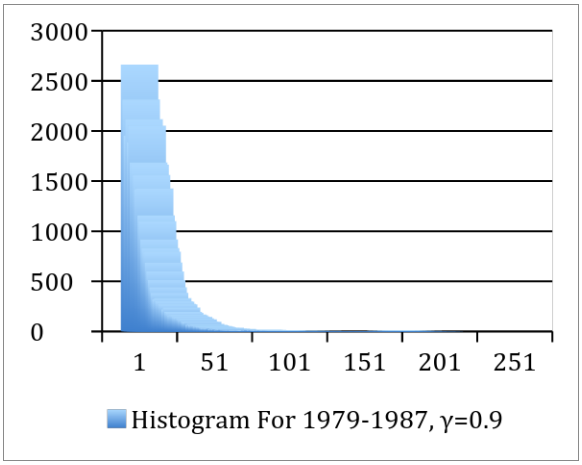
Task 2 : Analysis for 9-year periods

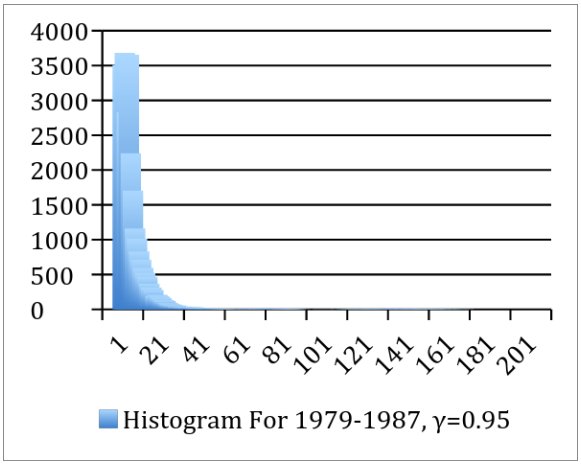
The data provided is divided into 3 sets of 9 years each. The above task is repeated for each of these periods and the data is as provided below.

a. Plotting histogram

1. For the Time frame 1979 - 1987

The histogram is plotted for the degree of distribution for each of the nodes.

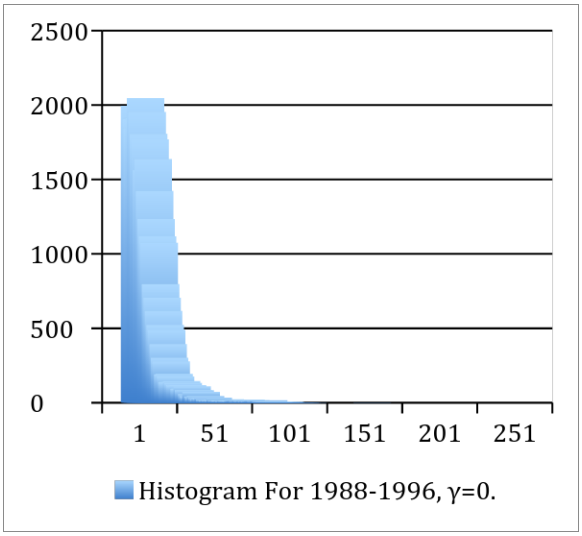


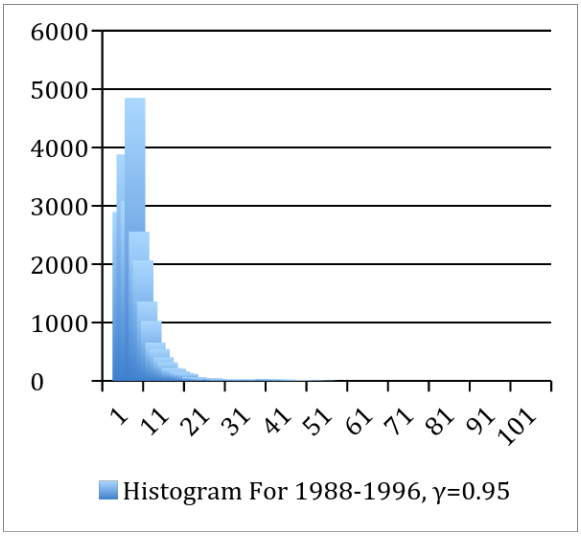


The Average degree for the graph (PCC = 0.9) 14.04. The supernodes are (74, 28), (77, 63) and (76, 65). The Average degree for the graph (PCC=0.95) is 7.512. The supernodes, for this histogram, are (77, 63), (74, 28) and (76, 65).

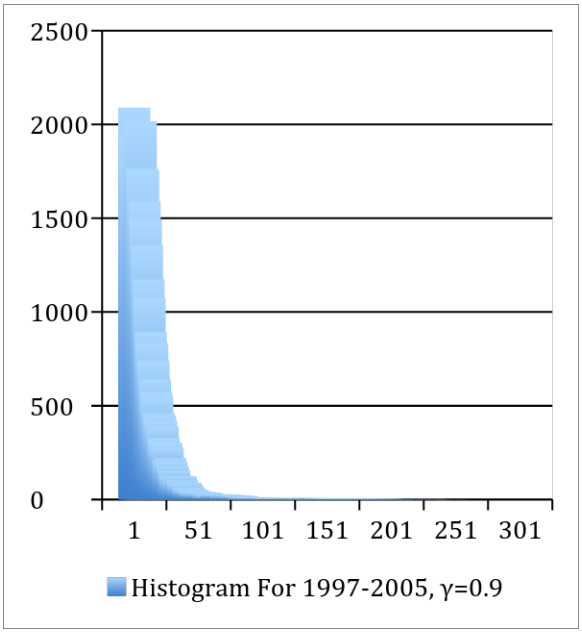
2. For the Time frame 1988 – 1996

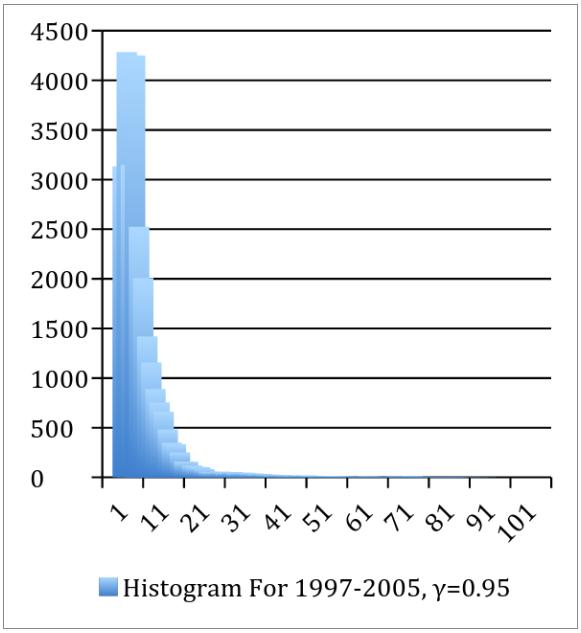
The Average degree for the graph (PCC = 0.9) is 11.70. The supernodes are (296, 279), (297,279) and 295,279). The Average degree for the graph (PCC = 0.95) is 5.506. The supernodes, for this histogram, are (203, 10), (440, 102), (441, 102) and (441, 103).





3. For the Time frame 1997 – 2005





The Average degree for the graph (PCC=0.9) is 14.71. From the histogram, we can see that certain vertexes have significantly higher degree than the average. These are the supernodes. In this case, the supernodes are (412, 140), (413, 139) and (414, 139). The Average degree for the graph (PCC=0.95) is 6.51. The supernodes, for this histogram, are (366, 133), (365, 133) and (365, 132).

b. Computing Clustering Coefficient and The Characteristic path length

1. For the Time frame 1979 – 1987

Clustering Coefficient

The clustering coefficient for the graph with correlation = 0.9 is 0.521.
The clustering coefficient for the graph with correlation = 0.95 is 0.586.

Characteristic Path Length

The characteristic path length for the graph is the mean over all the pairs of vertices.
The characteristic path length for the graph with correlation = 0.9 is 28.672
The characteristic path length for the graph with correlation = 0.95 is 20.105

2. For the Time frame 1988 – 1996

Clustering Coefficient

The clustering coefficient for the graph with correlation = 0.9 is 0.488
The clustering coefficient for the graph with correlation = 0.95 is 0.549.

Characteristic Path Length

The characteristic path length for the graph with correlation = 0.9 is 66.498.
The characteristic path length for the graph with correlation = 0.95 is 40.593.

3. For the Time frame 1997 – 2005

Clustering Coefficient

The clustering coefficient for the graph with correlation = 0.9 is 0.494.
The clustering coefficient for the graph with correlation = 0.95 is .549.

Characteristic Path Length

The characteristic path length for the graph with correlation = 0.9 is 46.293.
The characteristic path length for the graph with correlation = 0.95 is 43.879.

c. Metrics for Random Graphs and Comparison to Values Computed

1. For the Time frame 1979 – 1987

The clustering coefficient for the random graph (PCC = .9) is $1.030 * 10^{-4}$and for (PCC = .95) is $5.515*10^{-5}$.
The Characteristic Path Length for a random graph (PCC = .95) is 4.474 and for (PCC=.95) is 5.862.

2. For the Time frame 1988 – 1996

The clustering coefficient for the random graph (PCC=.9) is $8.59*10^{-5}$ and for (PCC=0.95) is $4.042*10^{-5}$.
The Characteristic Path Length for a random graph (PCC = .95) is 4.806 and for (PCC=.95) is 6.93.

3. For the Time frame 1997 – 2005

The clustering coefficient for the random graph (PCC=.9) is $1.08*10^{-4}$ and for (PCC=0.95) is $4.78*10^{-5}$.
The Characteristic Path Length for a random graph (PCC = .95) is 4.397 and for (PCC=.95) is 6.310.

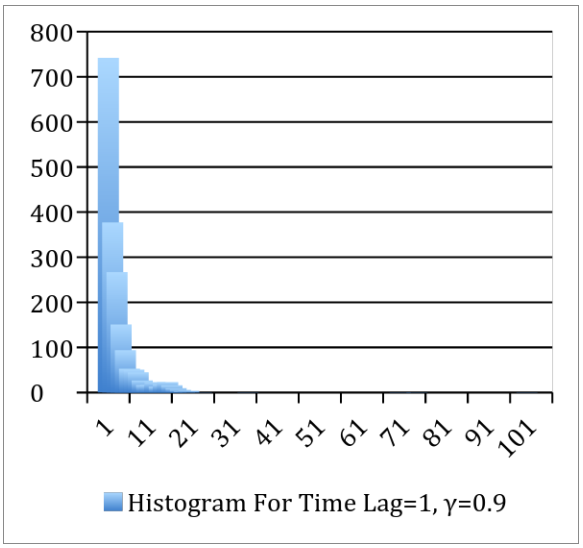
Task 3

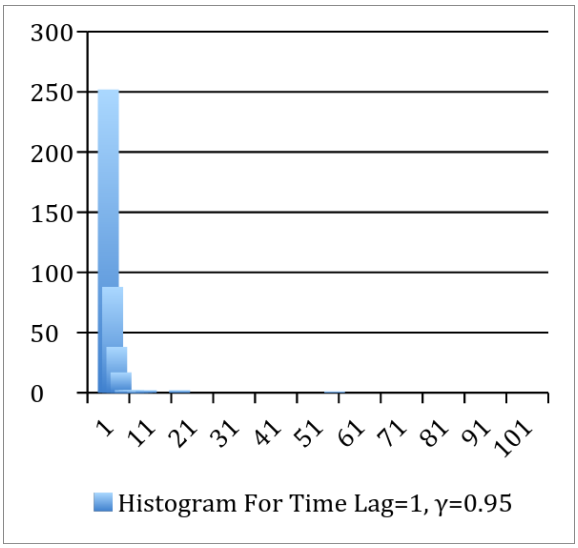
Task 1 is repeated to include Time Lags. The time lag is calculated using the following formula. Time lags of s= {1,2,3,4} weeks are considered.

a. Plotting histogram

1. For the Time Lag s=1

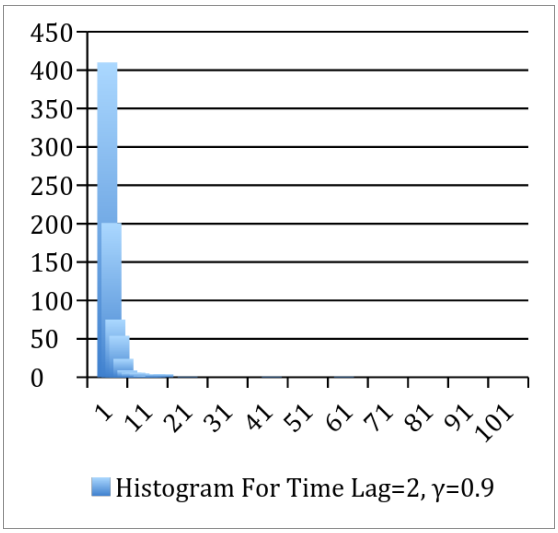
For the graph with correlation 0.9, the average degree for the graph is 7.00. From the histogram, we can see that certain vertexes have significantly higher degree than the average. These are the supernodes. In this case, the supernodes are (421, 152), (78, 13) and (84, 8). The Average degree for the graph(PCC=0.95) is 1.88. The supernodes, for this histogram, are (330, 225), (333, 223) and (84, 8).
The histogram is plotted for the degree of distribution for each of the nodes.

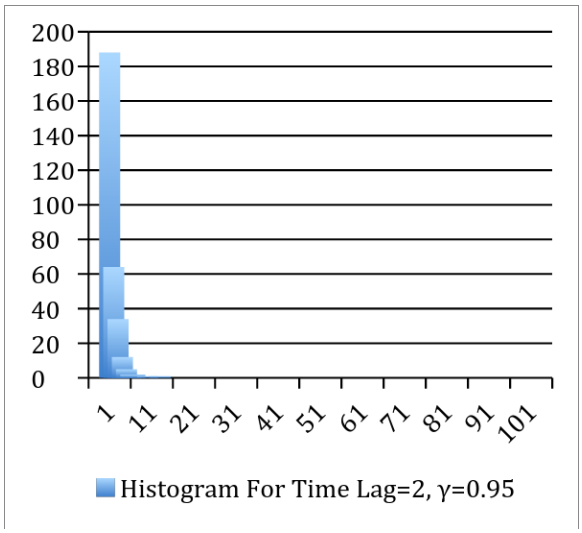




2. For the Time Lag s=2

The histogram is plotted for the degree of distribution for each of the nodes.



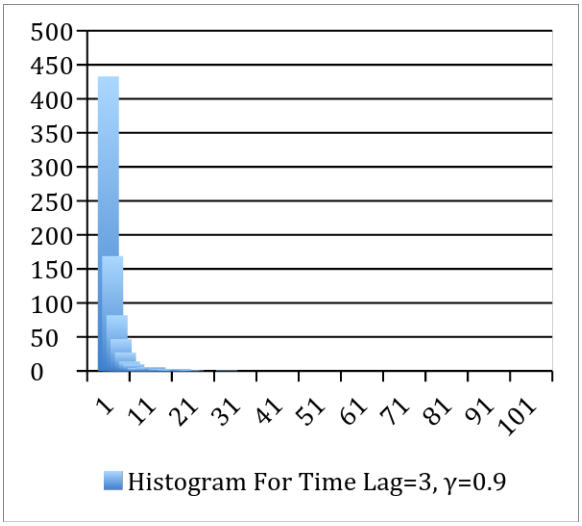


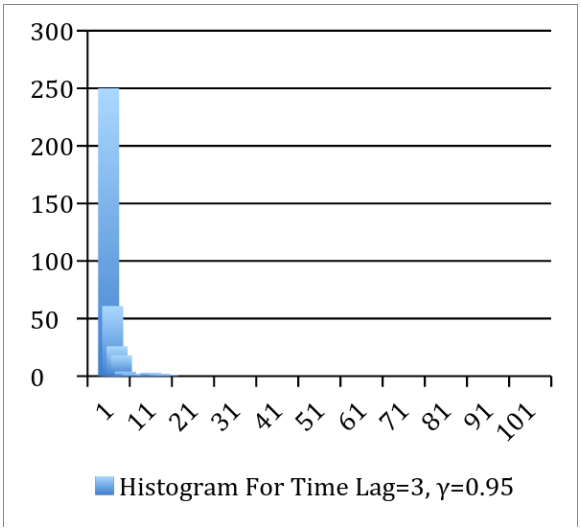
The Average degree for the graph (PCC = 0.9) is 4.56. Supernodes are (442, 113), (425, 152) and (412, 151). The Average degree for the graph (PCC=0.95) is 1.74. The supernodes, for this histogram, are (51, 76) and (442, 113).

3. For the Time Lag s=3

The Average degree for the graph (PCC=0.9) is 2. 382.The supernodes are (440, 96) and (104, 46). The Average degree for the graph (PCC=0.95) is 3.46. The supernodes, for this histogram, are (104, 70), (440,96) and (104, 46).

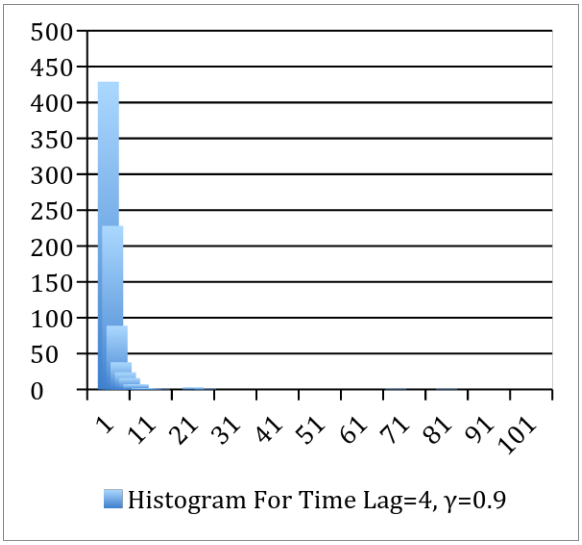
The histogram is plotted for the degree of distribution for each of the nodes.

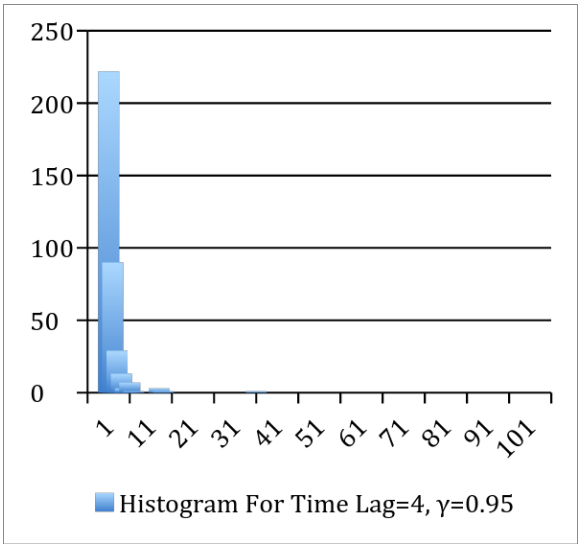




4. For the Time Lag s=4

The histogram is plotted for the degree of distribution for each of the nodes.





The Average degree for the graph (PCC=0.9) is 2.254. Supernodes, in this case are (312, 235), (298, 220) and (444, 130). The Average degree for the graph (PCC=0.95) is 2.238. The supernodes, for this histogram, are (441, 97), (441, 98), (441, 99), (298, 220) and (444, 130).

b. Computing Clustering Coefficient and The Characteristic path length

1. For the Time Lag s=1

Clustering Coefficient

The clustering coefficient for the graph with correlation = 0.9 is 0.68.
The clustering coefficient for the graph with correlation = 0.95 is 0.831.

Characteristic Path Length

The characteristic path length for the graph with correlation = 0.9 is 3.272
The characteristic path length for the graph with correlation = 0.95 is 2.182

2. For the Time Lag s=2

Clustering Coefficient

The clustering coefficient for the graph with correlation = 0.9 is 0.769
The clustering coefficient for the graph with correlation = 0.95 is 0.826

Characteristic Path Length

The characteristic path length for the graph with correlation = 0.9 is 1.669.
The characteristic path length for the graph with correlation = 0.95 is 1.479.

3. For the Time Lag s=3

Clustering Coefficient

The clustering coefficient for the graph with correlation = 0.9 is 0.766
The clustering coefficient for the graph with correlation = 0.95 is 0.853

Characteristic Path Length

The characteristic path length for the graph with correlation = 0.9 is 1.773.
The characteristic path length for the graph with correlation = 0.95 is 1.539.

4. For the Time Lag s=4

Clustering Coefficient

The clustering coefficient for the graph with correlation = 0.9 is 0.741
The clustering coefficient for the graph with correlation = 0.95 is 0.853

Characteristic Path Length

The characteristic path length for the graph with correlation = 0.9 is 1.773
The characteristic path length for the graph with correlation = 0.95 is 1.563

c. Metrics for Random Graphs and Comparison to Values Computed

1. For the Time Lag s=1

The clustering coefficient for the random graph (PCC = .9) is $5.139 * 10^{-5}$ and for (PCC = .95) is $1.380*10^{-5}$.
The Characteristic Path Length for a random graph (PCC = .95) is 6.075 and for (PCC=.95) is 18.726.

2. For the Time Lag s=2

The clustering coefficient for the random graph (PCC = .9) is $3.348 * 10^{-5}$ and for (PCC = .95) is $1.277*10^{-5}$.
The Characteristic Path Length for a random graph (PCC = .95) is 7.7912 and for (PCC=.95) is 21.343.

3. For the Time Lag s=3

The clustering coefficient for the random graph (PCC = .9) is $1.749 * 10^{-5}$ and for (PCC = .95) is $2.540*10^{-5}$.
The Characteristic Path Length for a random graph (PCC = .95) is 13.62 and for (PCC=.95) is 9.5239.

4. For the Time Lag s=4

The clustering coefficient for the random graph (PCC = .9) is $1.655 * 10^{-5}$ and for (PCC = .95) is $1.643*10^{-5}$.
The Characteristic Path Length for a random graph (PCC = .95) is 14.546 and for (PCC=.95) is 14.675.

Summary of Observations:

In the project, we created a graph corresponding to the Sea Ice Concentration data provided to us. We made a few important observations in our results. Firstly, we observe that often enough, the supernodes are located close together. Secondly, the number of edges are very low for the graphs created with time lags. Thirdly, the clustering coefficient of the created graphs are much higher than that of random graphs. And Lastly, the characteristic path length of the graphs created are much greater as compared to a random graph, if we don't include time lags. When we include time lags, the random metrics are higher.