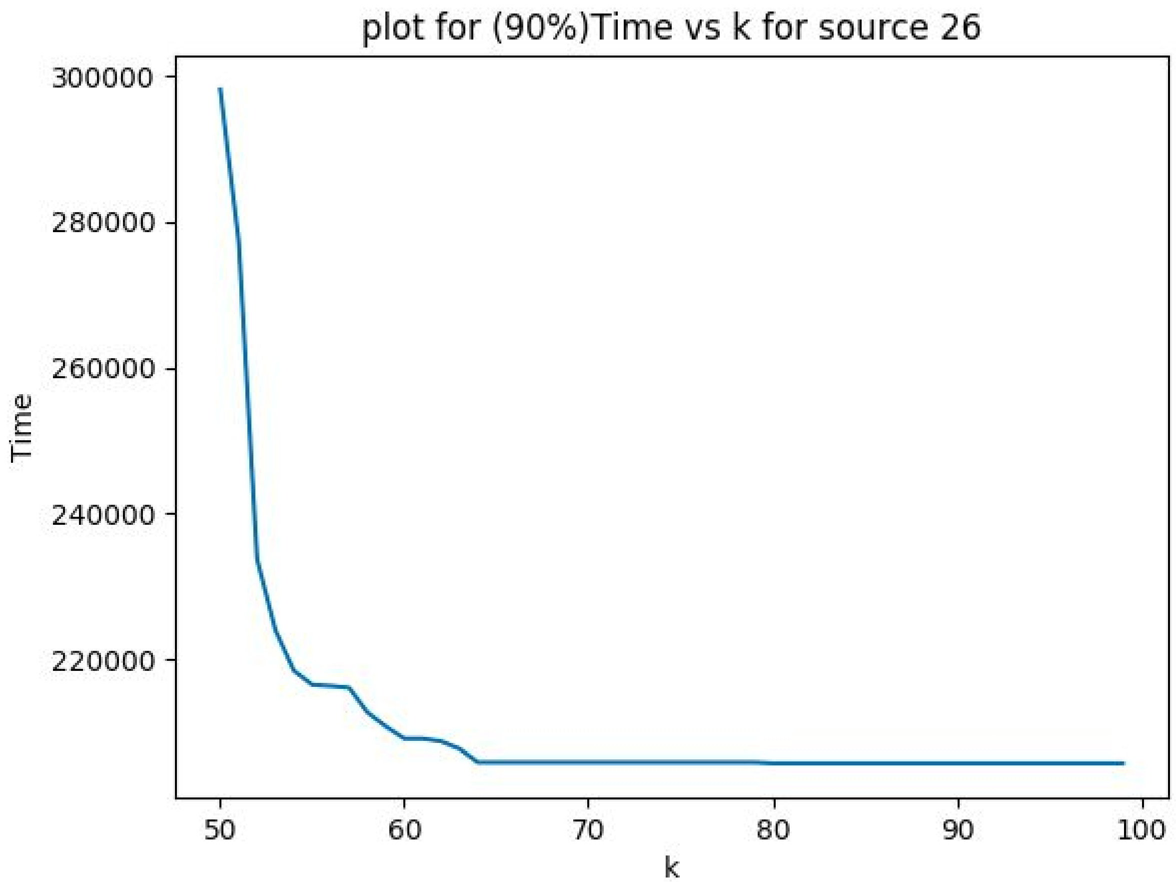
Networks Assignment 3

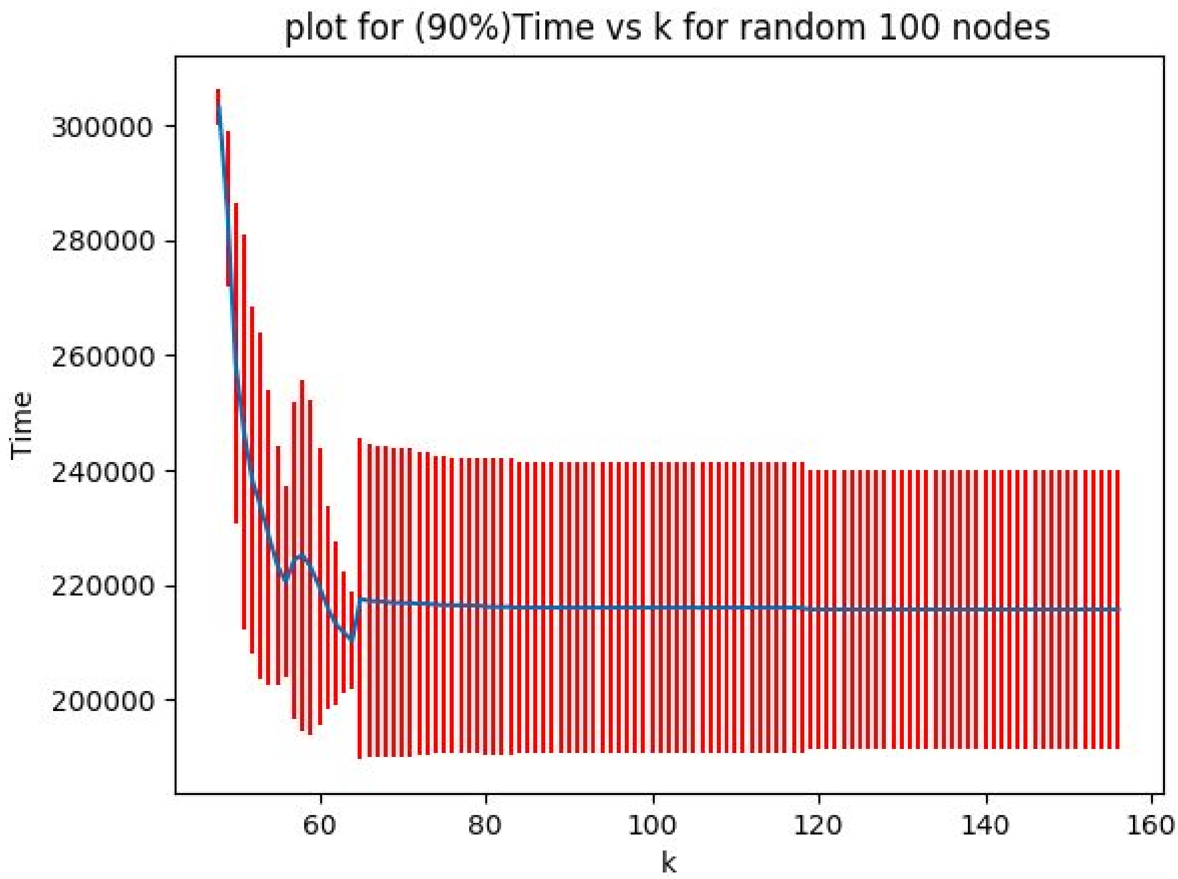
Ronak Agarwal(2015CS10252)
Saket Dingliwal(2015CS10254)

Part 1

Variation of **Completion time**(Packets reached to 90% of the total nodes) with **k**

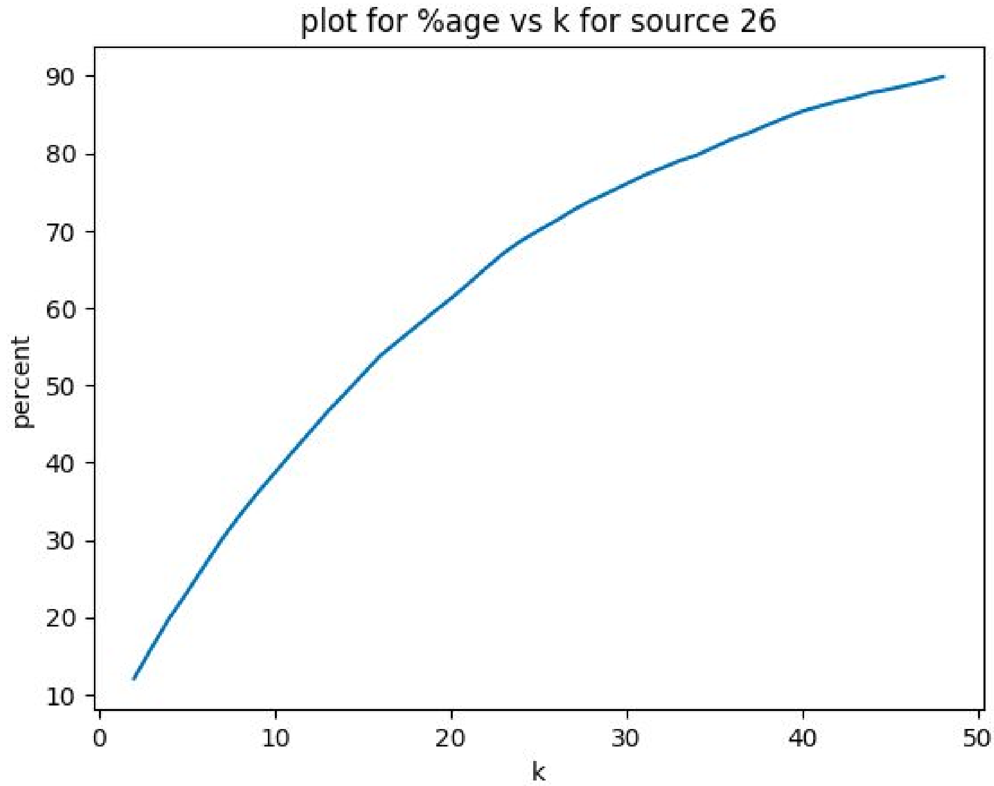
Completion Time = Time taken for packet to reach to total 90% of the total nodes.
k = Maximum number of neighbours a node will send a packet to

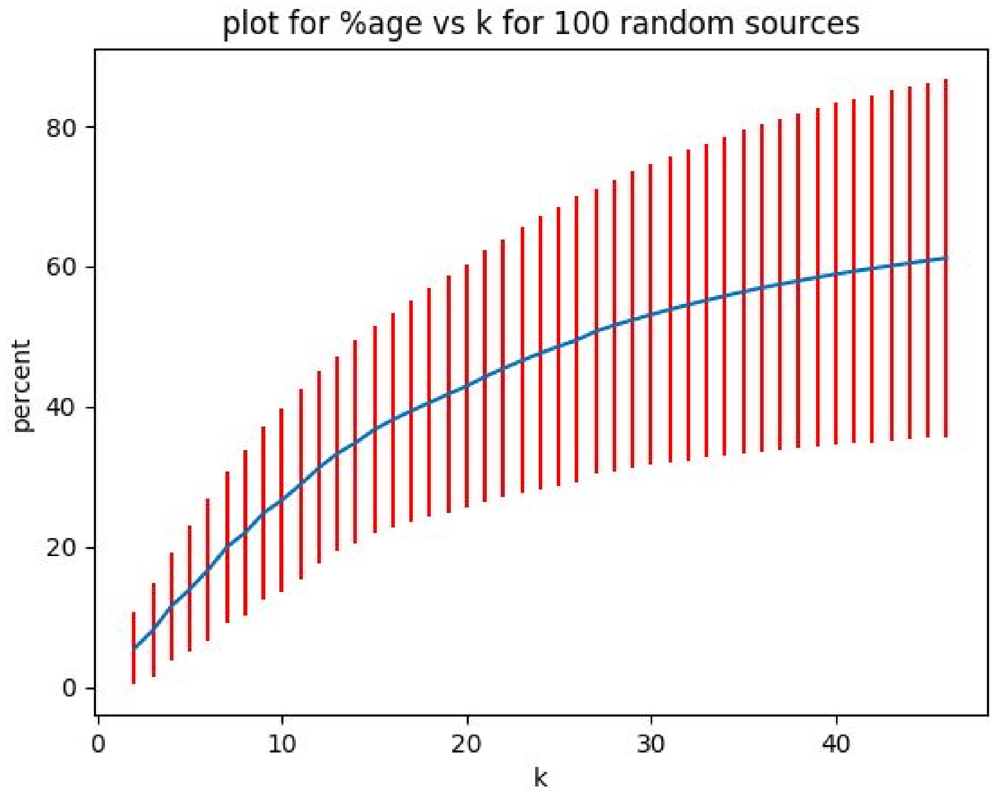


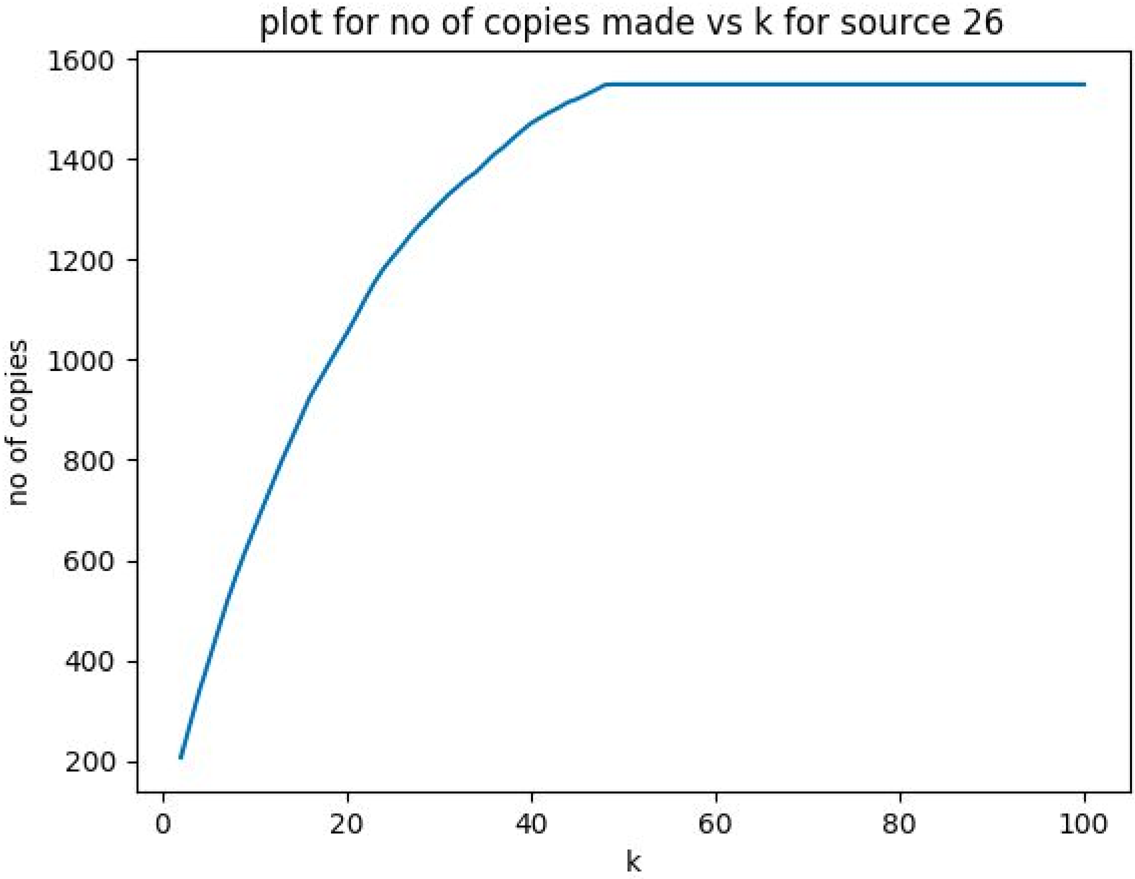


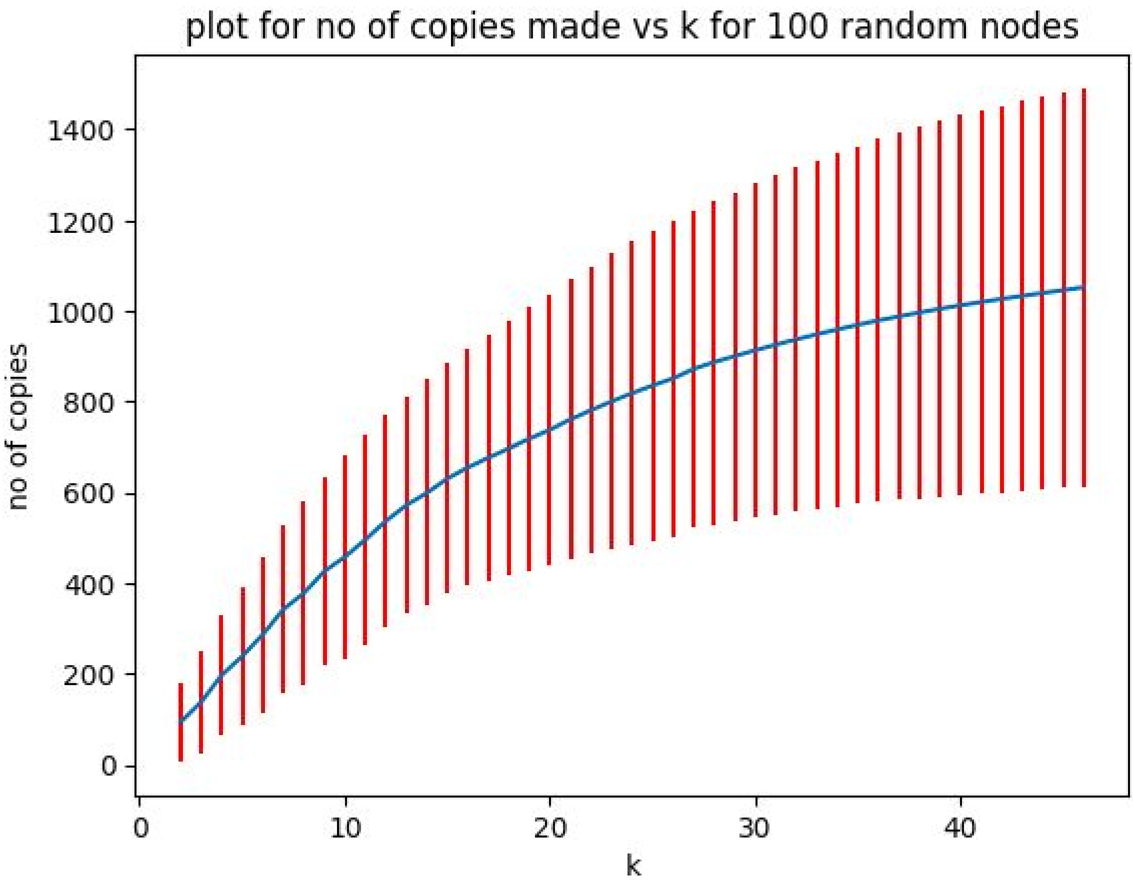
As k increases the number of transmissions allowed to be made by a single node increases. Hence as the limit of transmission increases, the node transmits data as soon as it comes in contact with some other node. Before k = 47, it was not possible for chunk to reach 90% of the nodes. After that there is decreasing trend of time upto k= 70 approximately. However after that the time remains constant as the transmissions made are not restricted by the limit k but the number of connections it makes in the CSV. A small kink is observed in data for 100 random nodes because for some sources the degree is small and chunk may not reach 90% of nodes and hence disturbing the average. This can be easily verified using plot for only node 26.

Variation of percentage transmission reached for different k





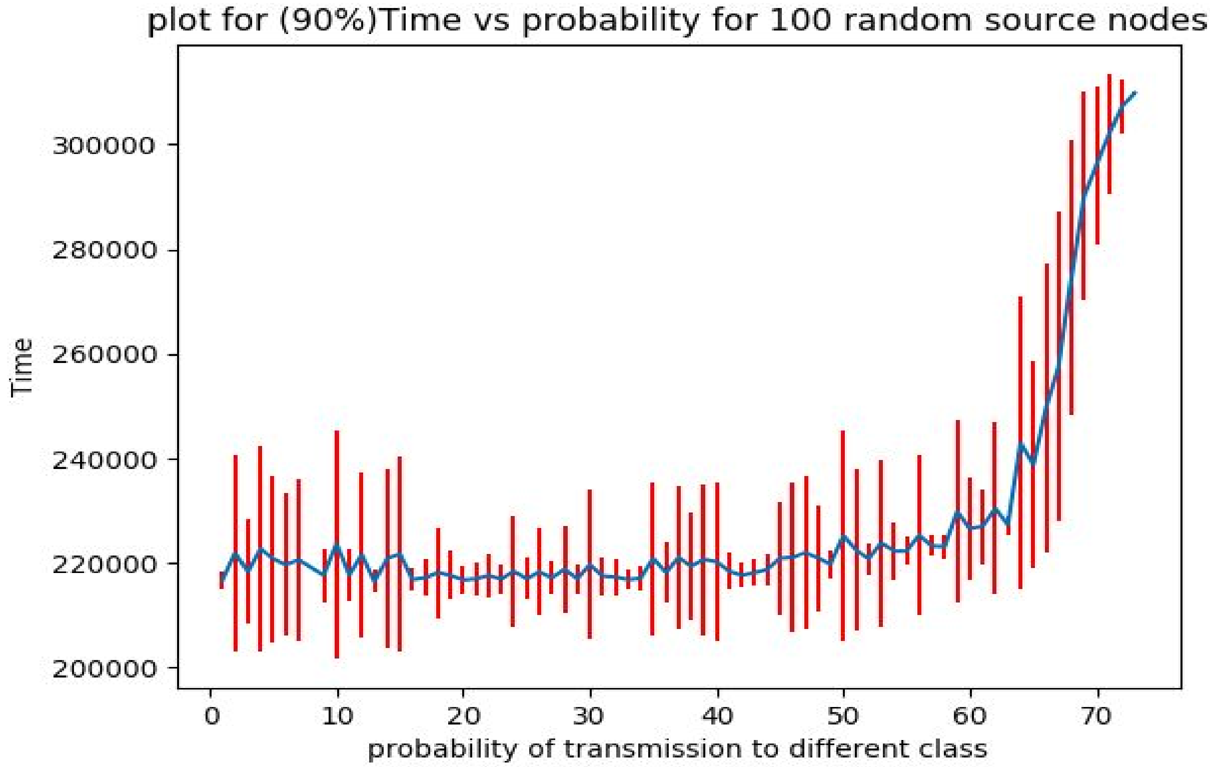

plot for no of copies made vs k for source 26
no of copies
k
1600
1400
1200
1000
800
600
400
200
0
20
40
60
80
100

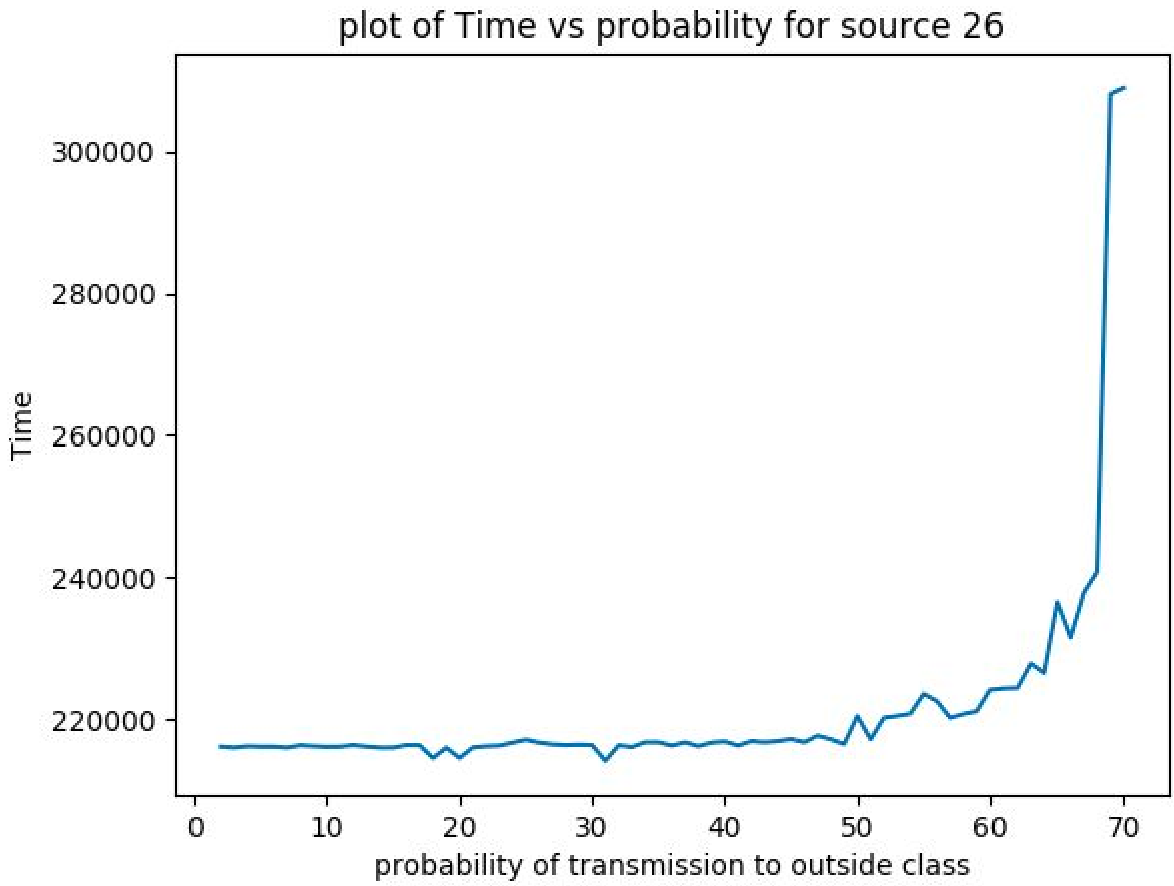

plot for no of copies made vs k for 100 random nodes
no of copies
k
1400
1200
1000
800
600
400
200
0
0
10
20
30
40

Firstly I would like to state the fact that the number of copies of chunk during the run of the broadcast is actually equal to the number of nodes the chunk has reached to, since a copy of a chunk will only be made when some node who already has a chunk encounters a node who hasn't, implying that for each copy of a chunk made a new node will be getting the chunk, the analysis for the number of copies made and the number of nodes reached would be same.

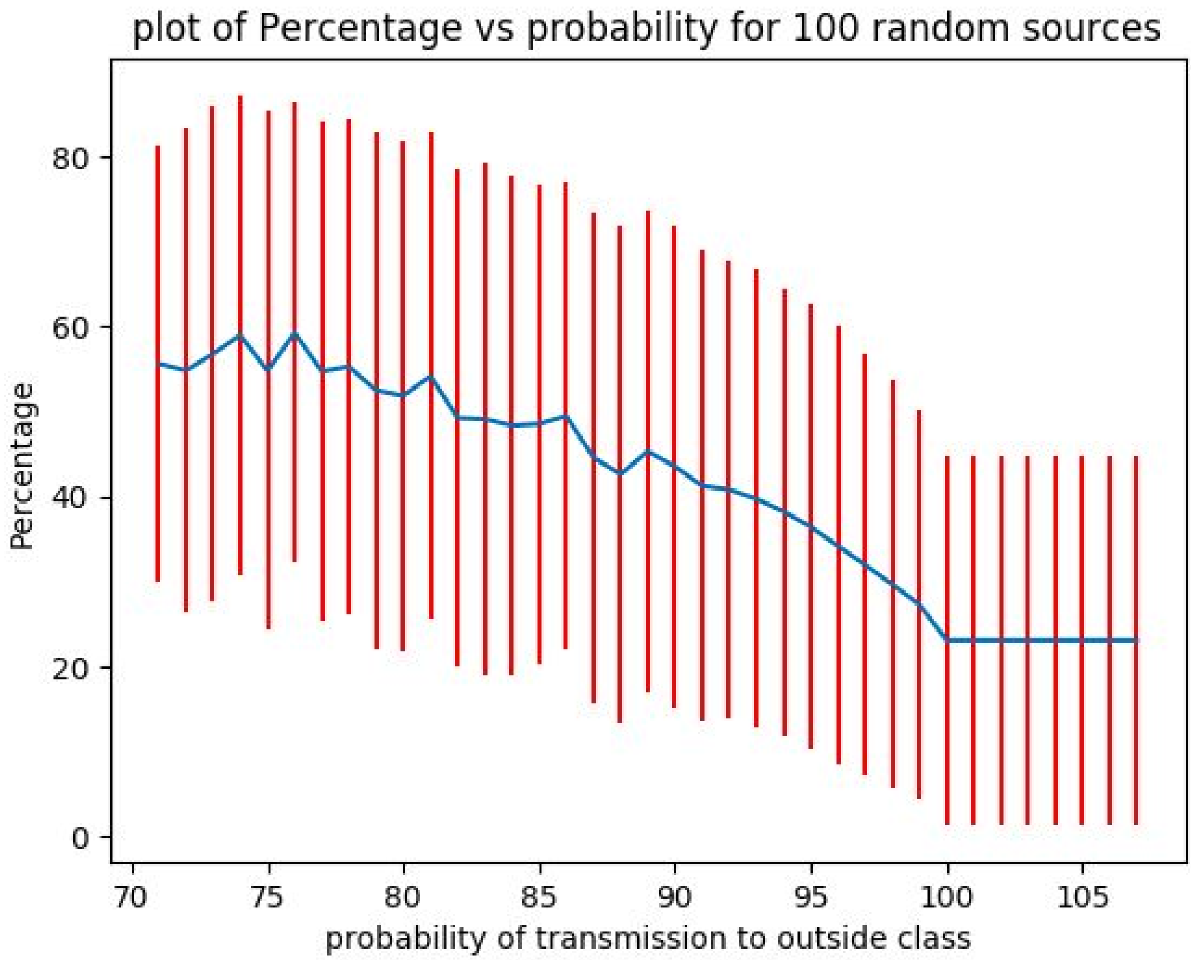
Now as we can see from the graphs(the second graph actually established the trend more strongly), clearly it can be seen that the percentage of nodes reached increases with increase in k, since at low k, a node which encounters many nodes may potentially give it to many of them but due to low k there's a restriction due to which many potential transfer are blocked and reachability is reduced.

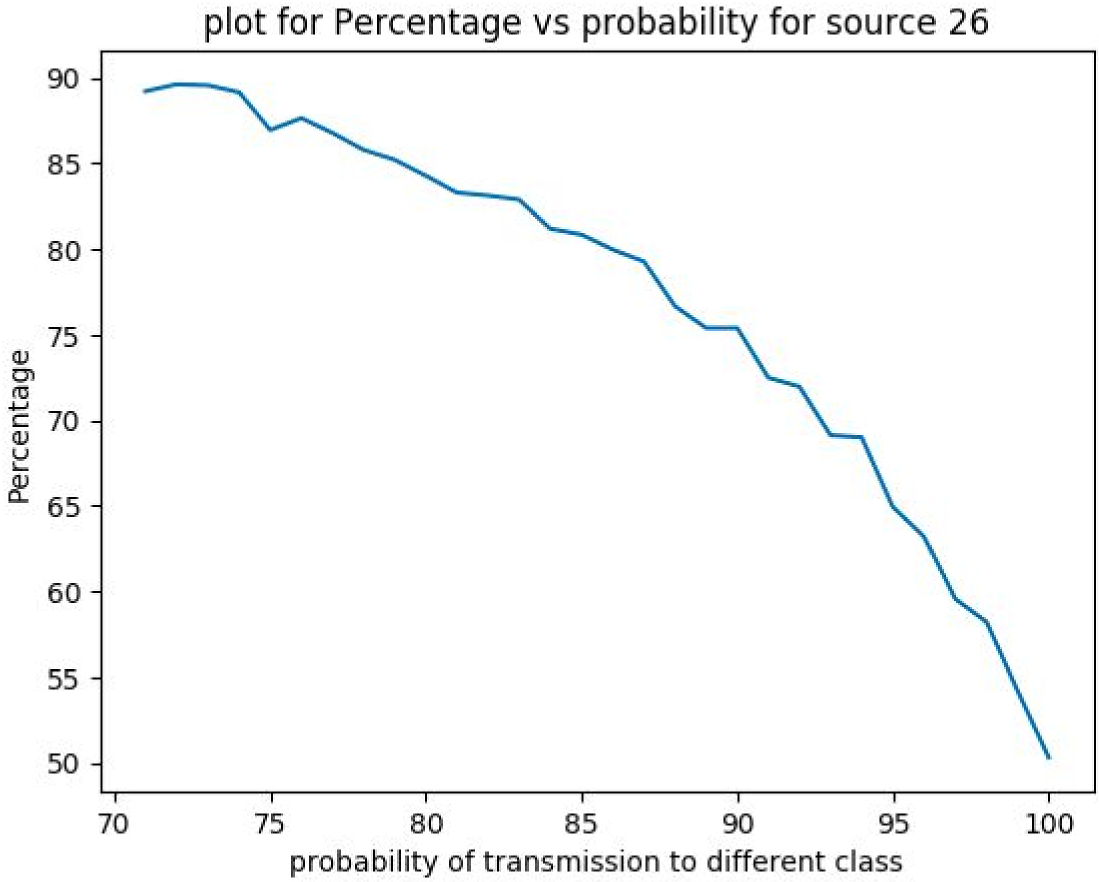
Part 3

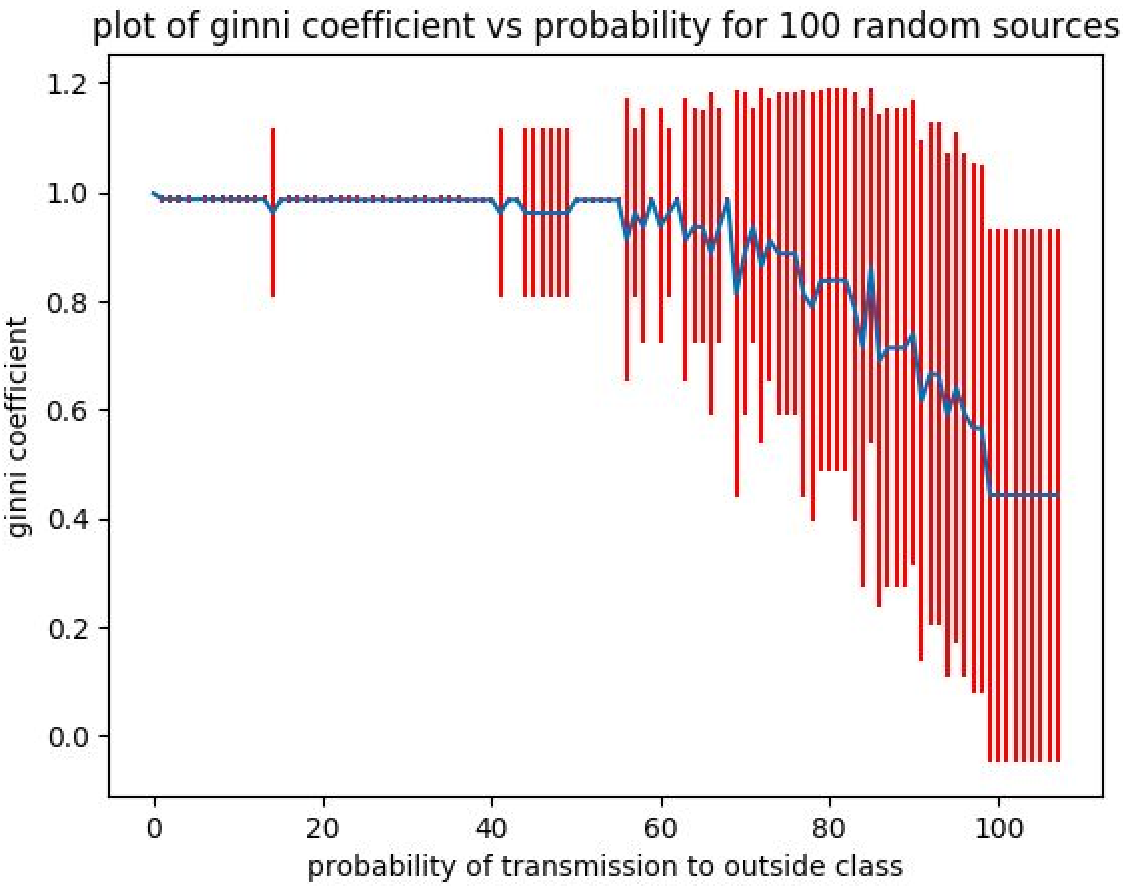

plot for (90%)Time vs probability for 100 random source nodes
Time
300000
280000
260000
240000
220000
200000
0
10
20
30
40
50
60
70
probability of transmission to different class


plot of Time vs probability for source 26
Time
300000
280000
260000
240000
220000
0
10
20
30
40
50
60
70
probability of transmission to outside class

The time taken to transmit data increases as the probability of sending the chunk outside the close class increases. This is because the probability of transmission within the same class decreases and most of the connections are within the same class therefore with such a low probability of transmission with each connection the time is expected to increase. It increases so much that after p = 0.7 the chunk is not able to reach 90% of the nodes and the percentage of the nodes reached is plotted. The percentage values show a decreasing trend as visible from their graphs.



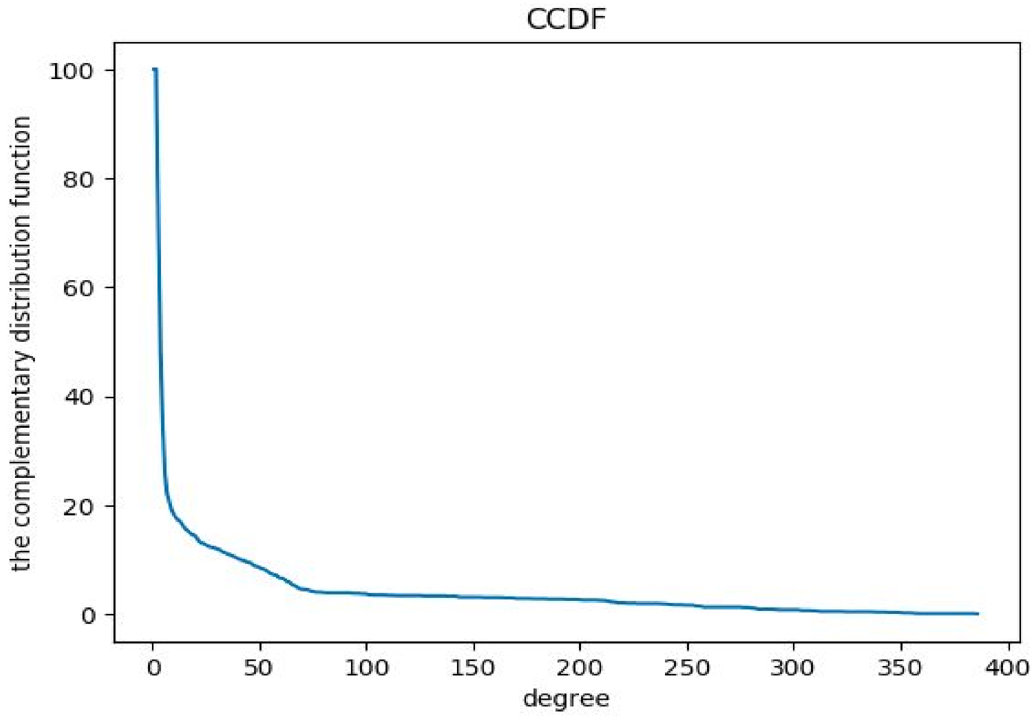

plot for Percentage vs probability for source 26
Percentage
90
85
80
75
70
65
60
55
50
70
75
80
85
90
95
100
probability of transmission to different class


plot of ginni coefficient vs probability for 100 random sources
ginni coefficient
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
probability of transmission to outside class

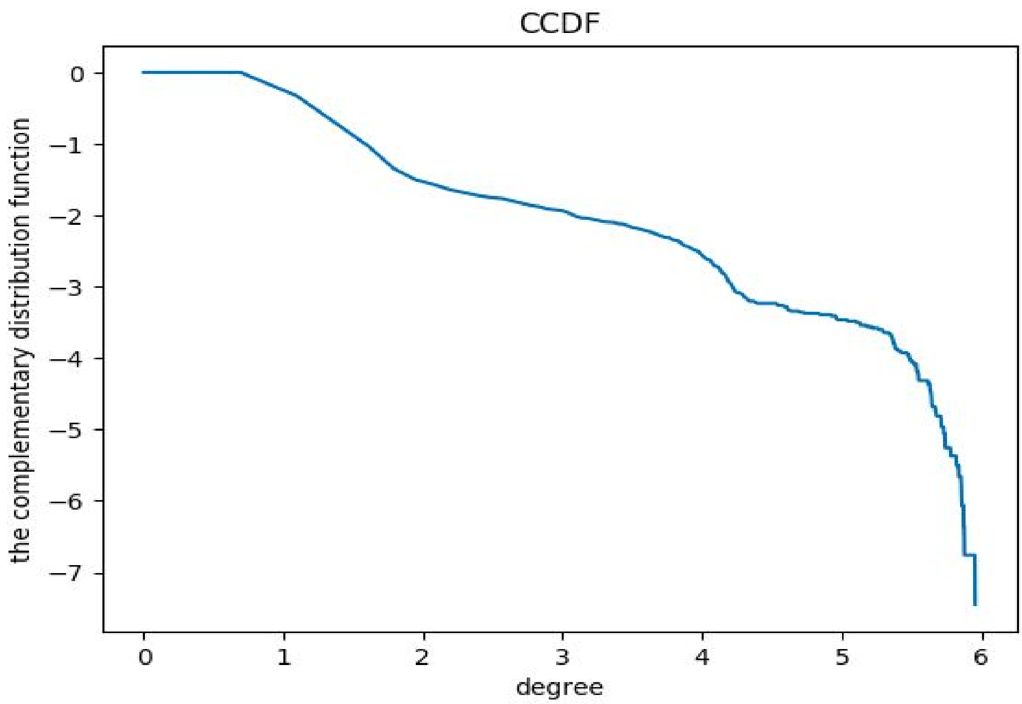
As we can see from the analysis of the Gini Coefficient which measures the variation in number of transmissions by different nodes, the stress is high for lower probability of transmission to outside class. This means the data is transmitted within the same class with high probability. And there are larger number of connections within the same class, stressing out the nodes to make multiple copies. Therefore to handle the trade off between time and stress on the nodes, choosing the probability value to be around 65 - 70 is fine as the time for computation increases significantly later. The gini also reduces to 0.9 around this range and hence lesser stress on the nodes.

Part 2

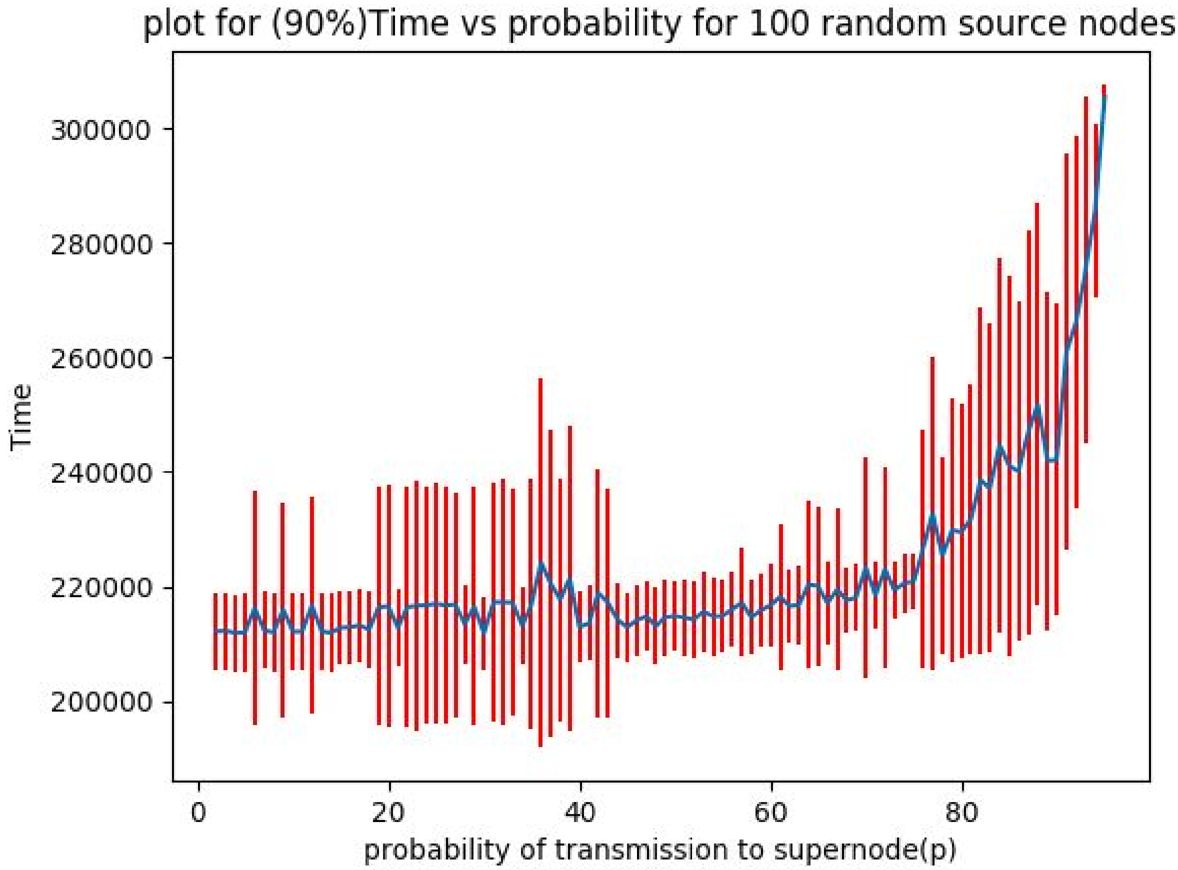
CCDF plot for degrees of various nodes in the graph



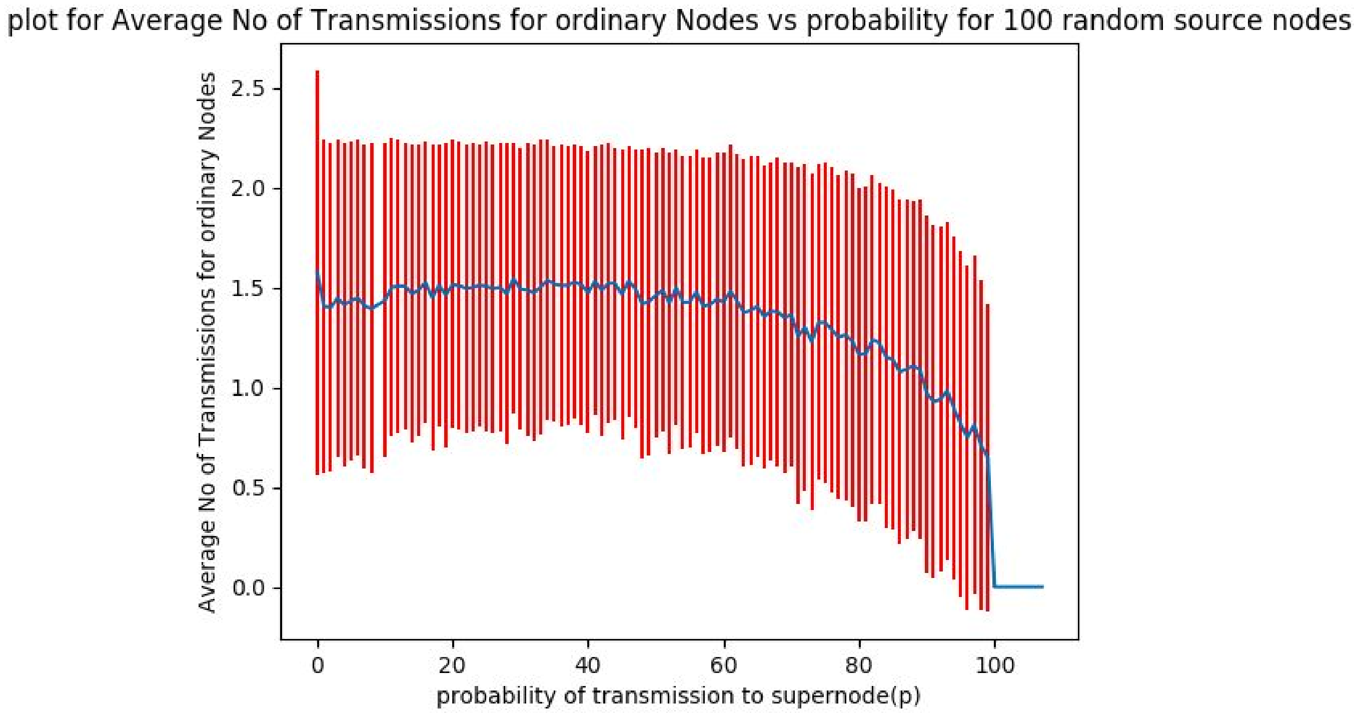
When drawn on a log-log scale the variation of degree becomes very clear how the variation of degrees is there in the csv.



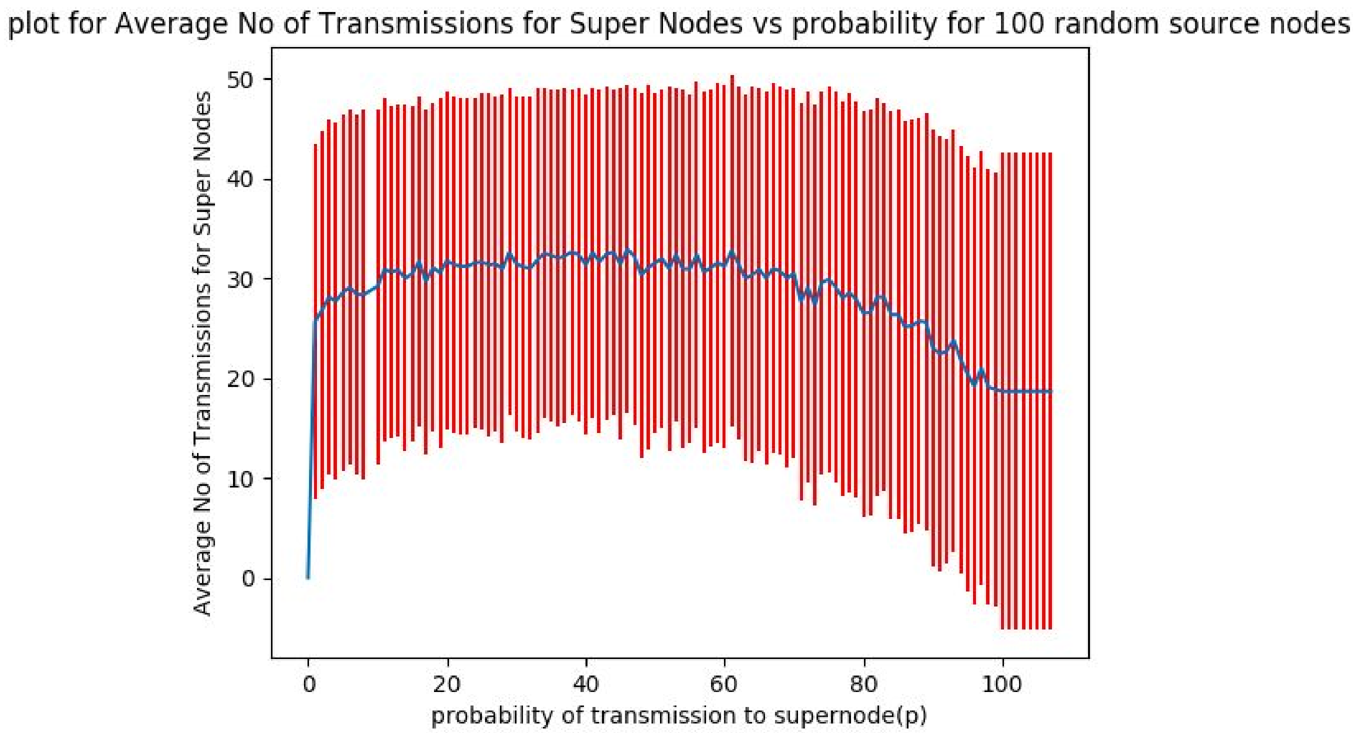
Various Aggregate trends for varying probability



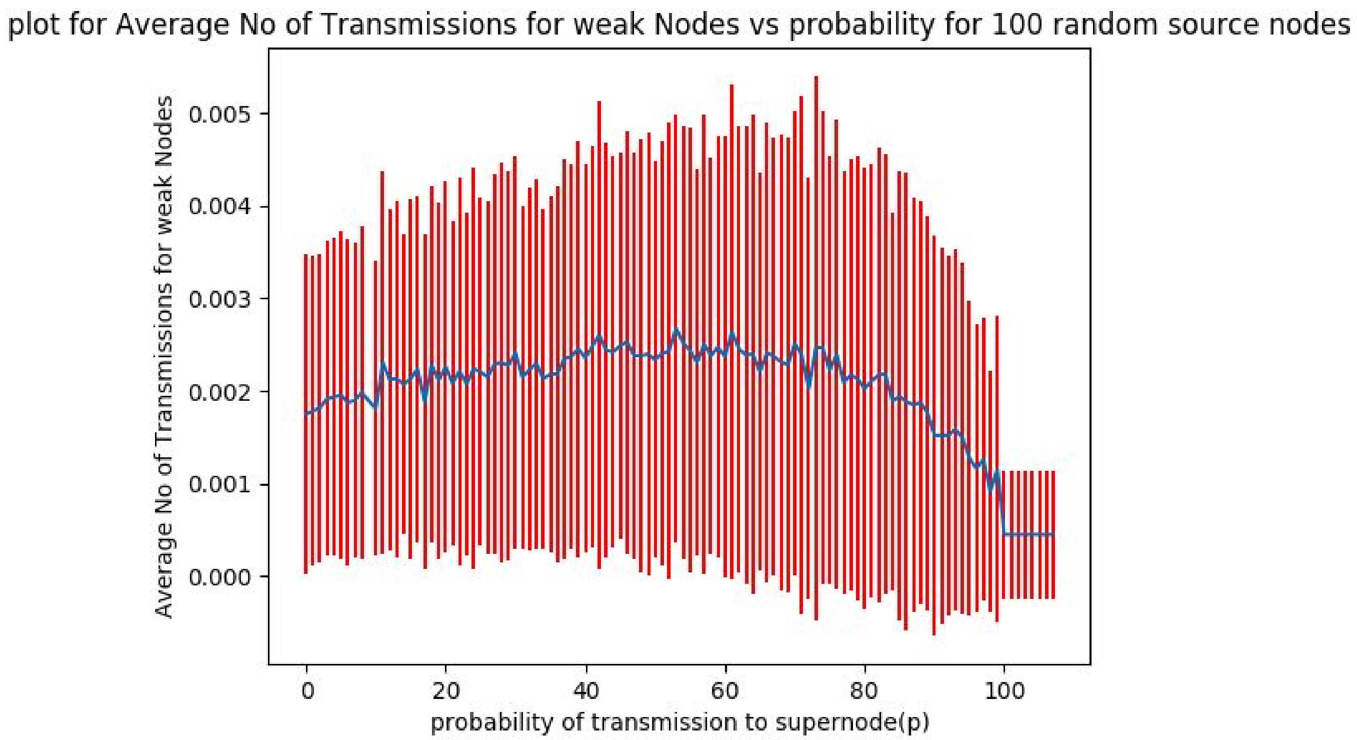
As the probability of transmission to supernode is increased, the probability of transmission to ordinary nodes is decreased. Since there are many ordinary nodes in the network, the transmission to such nodes is reduced and hence more time is taken to reach 90% of the nodes. For very high probability even at the end of the file 90% of the nodes are not reached.



As the probability of transmission to ordinary nodes decreases the average number of transmission made by ordinary nodes also decreases. Most of the ordinary nodes receive the chunk after large number of connections and hence there are fewer number of their connections left for further them to transmit it further.

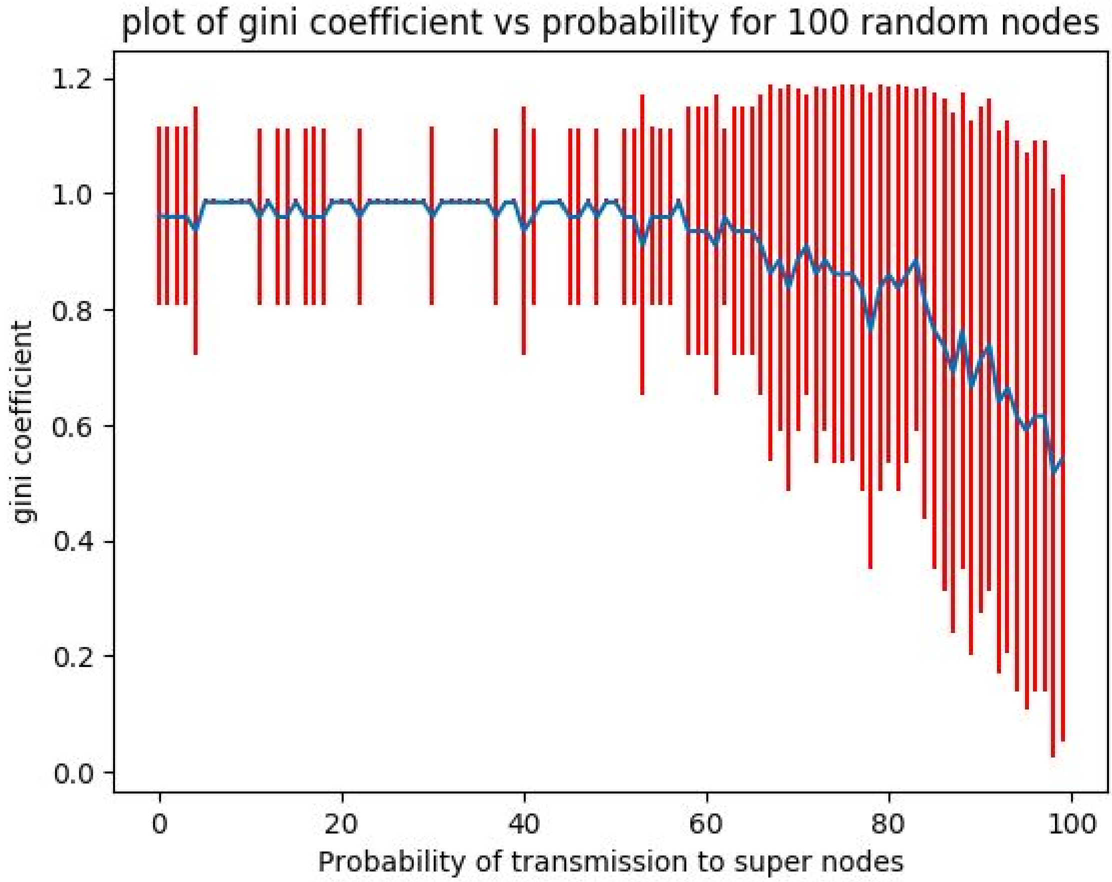


Initially as the probability for transmission to super nodes increases, the nodes receive their chunk early ie in its first few connections and hence are able to transmit the data more and more. Thus an increasing trend in the average number of transmission is observed till probability reaches 60. After that it decreases because at high probability of super nodes the probability of transmission to ordinary nodes is low. There are only a few supernodes and most of the nodes it comes in contact with are ordinary nodes which do not receive the chunk due to low probability of transmission.

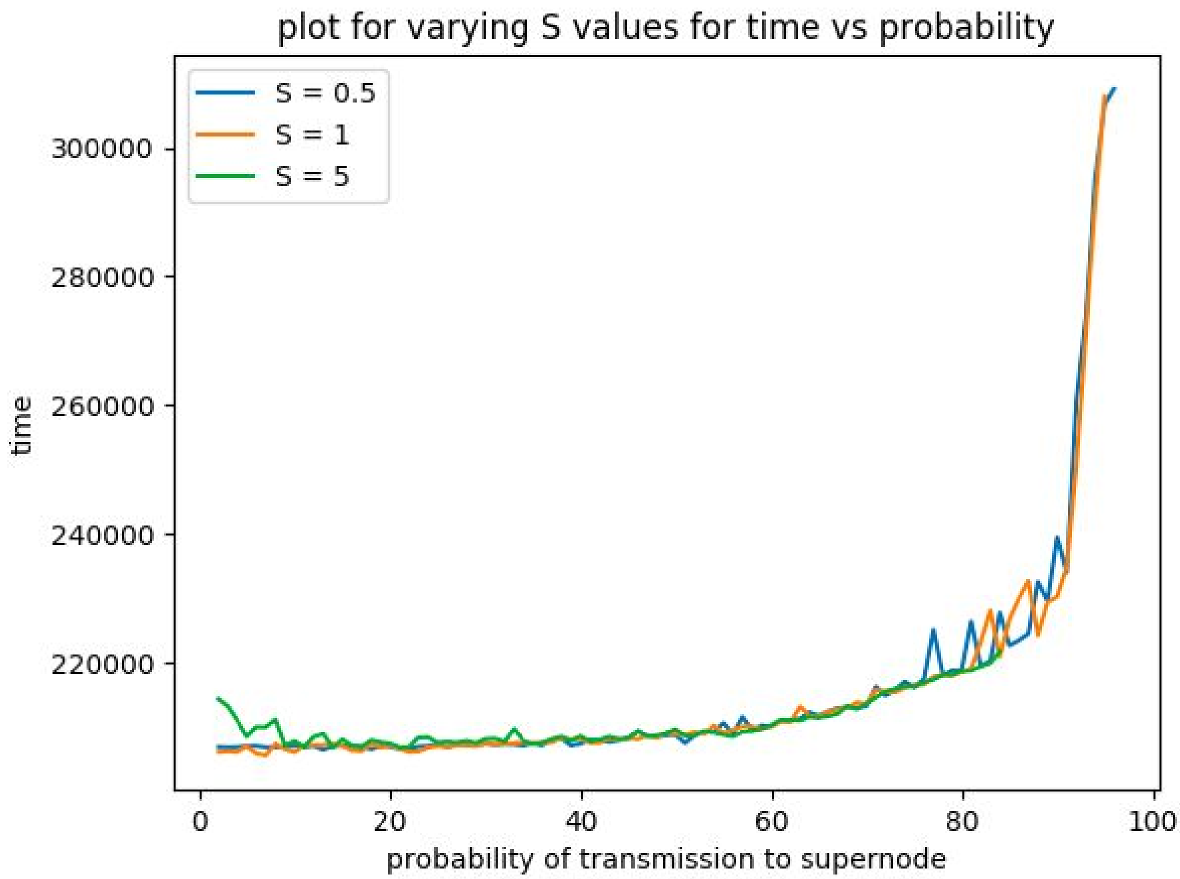


The trend for the weak nodes is largely 0 and immaterial. The connections established by weak nodes are too less to observe any significant behaviour. That is why the graph with high standard deviation and low range of values turns up.

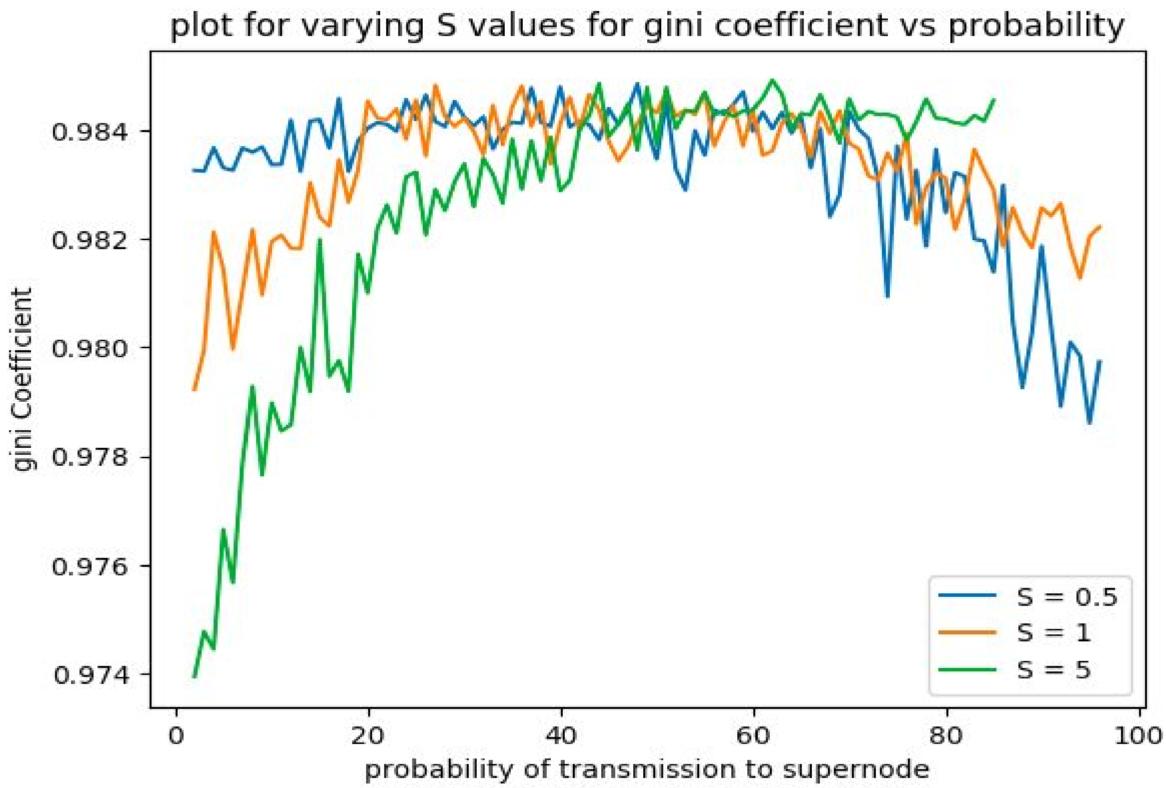
Overall the average number of transmissions made by super nodes are much higher than ordinary or weak nodes for obvious reason of their higher degree. Certainly the number of copies made by super nodes are much higher and hence they have maximum stress for broadcast. The gini Coefficient calculation verifies it as it appears to be close to 1.



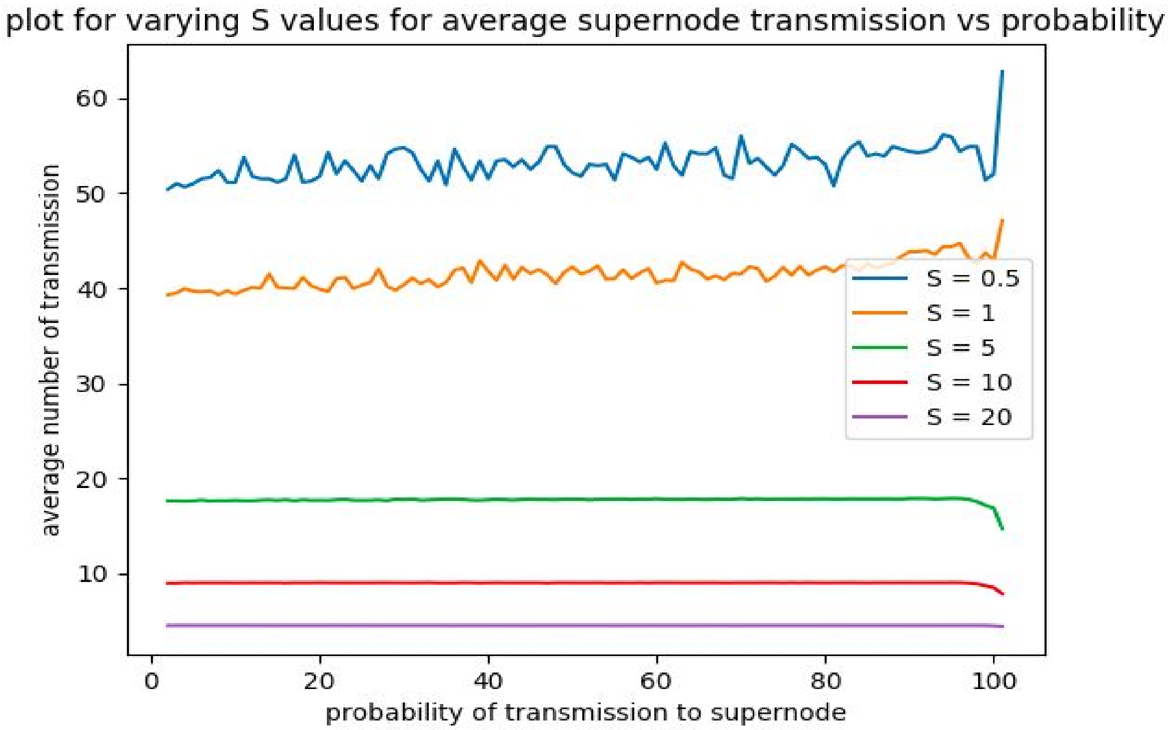
The value is close to one for probability = 0.6. Thereafter the coefficient decreases as the number of transmissions made by the super nodes decreases.
Therefore it is best to keep the probability of transmission to 0.75 to 0.8 so that the time do not increases too much and also the gini coefficient falls below 1.
This trade off is significant because it is necessary that all the broadcast work is not done by few of nodes and also the broadcast must happen as quickly as possible. Therefore it becomes essential to ensure gini coefficient is not close to 1 and time is also under a certain limit.



Not much deviation from original data is seen as the value of S is varied in time.



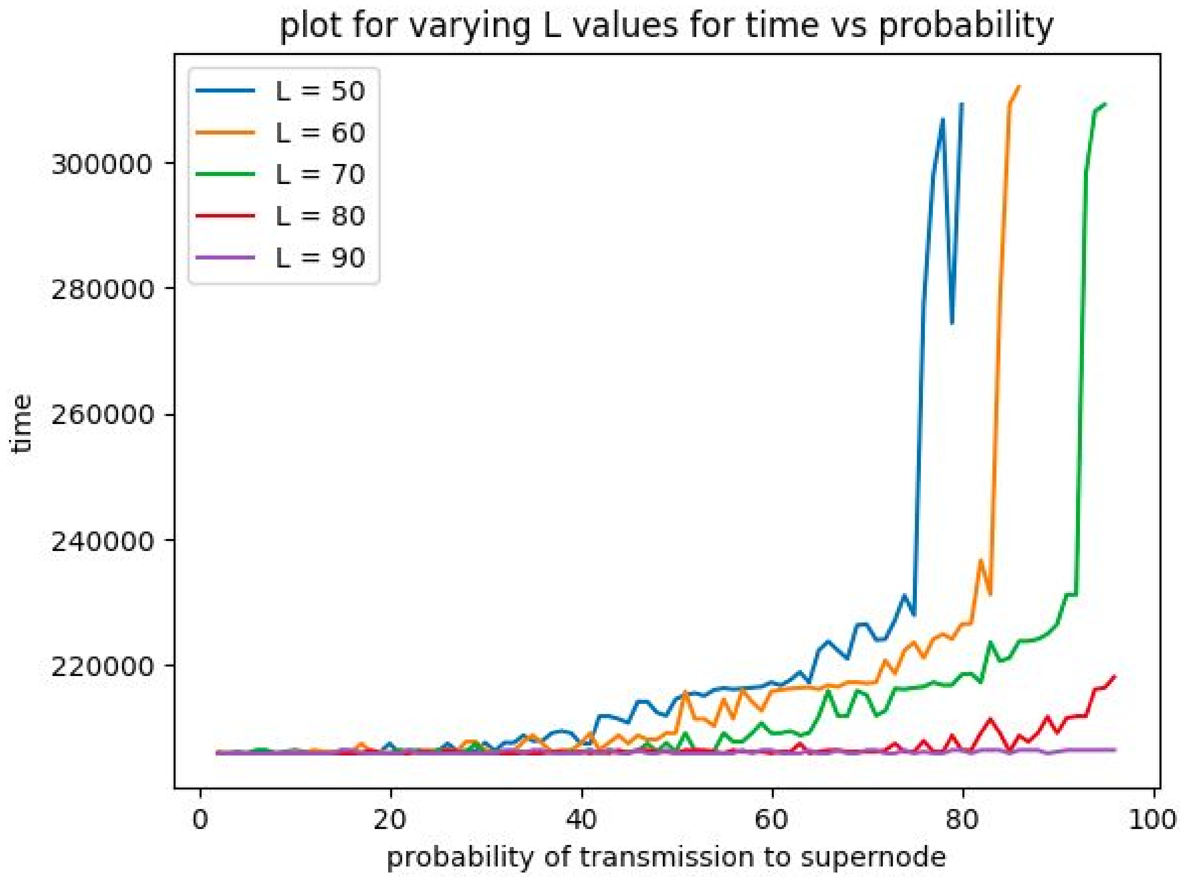
Similarly the value of gini coefficient varies only in a small range with varying S. For larger S, there is slightly less value of the coefficient for small probability of transmission between the supernodes.



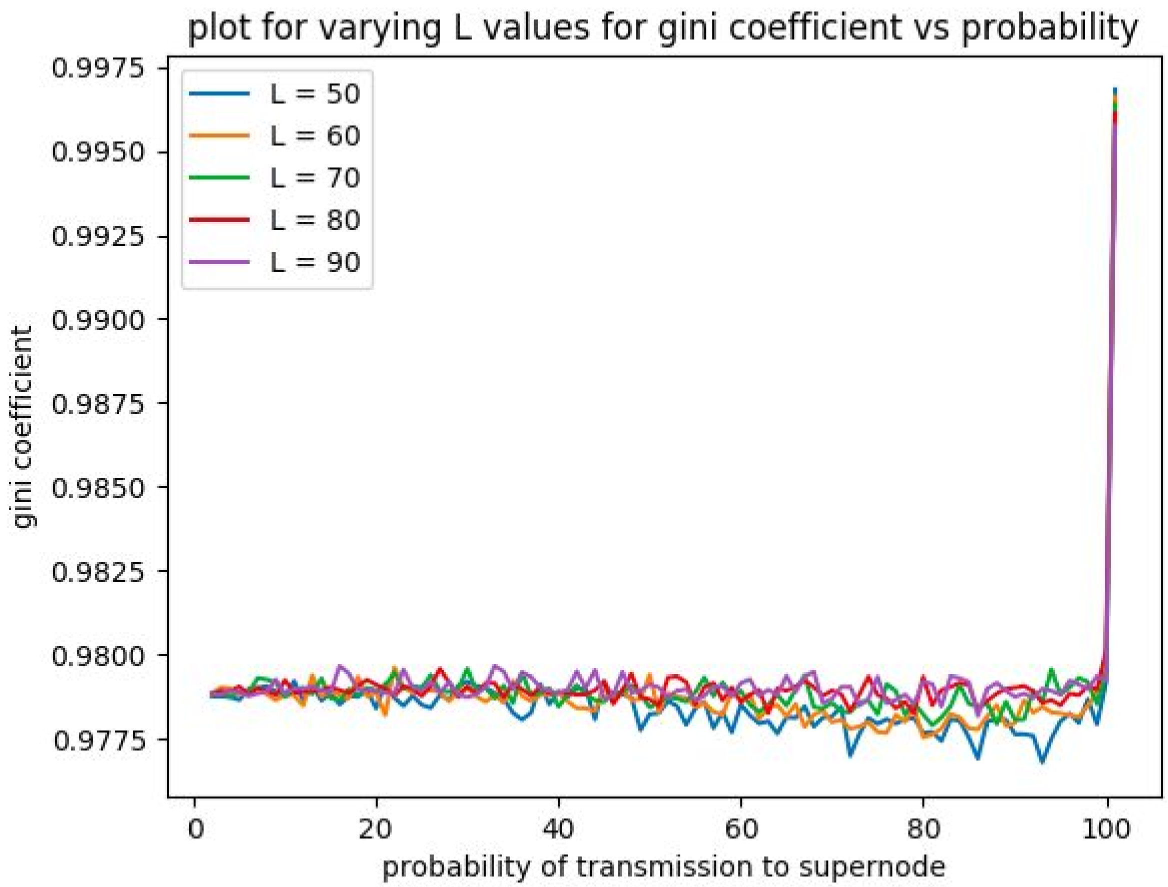
The average number of transmissions made by supernodes vary significantly with the value of S. For larger S, more number of nodes are called super nodes even though they have fewer degree. Hence the average decreases.

The average values for weak nodes have very high standard deviation and very small values and hence nothing much can be inferred from those values. Therefore, the plots are not drawn here.

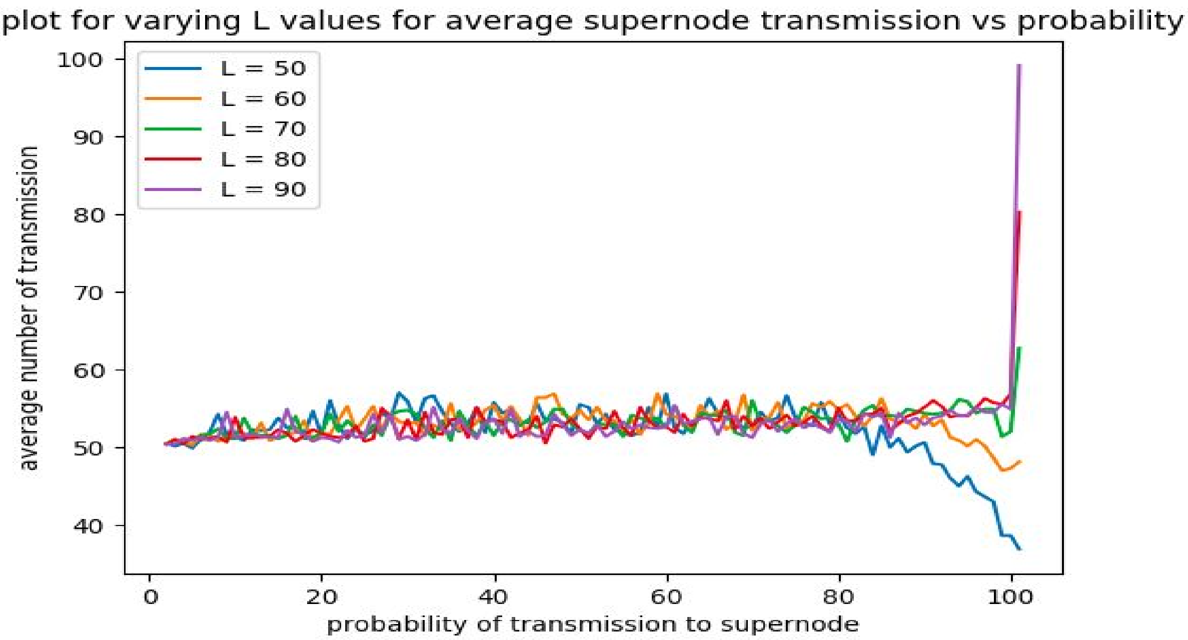
Variation of L



As L increases the time taken to transmit decreases at the higher end of probability. This is because the probability of transmission to weak node is 1 and more the number of weak nodes, the transmission to each node become certain as they come in contact with each other. For L=90 the time is significantly lower as almost all the nodes are classified as weak nodes .



The gini coefficient do not change much on changing L values as can be seen from the above plot.



Average number of transmission of supernodes remain almost the same

Comparison Of Different Algorithms

In the third part there is sharp increase in the time for transmission after probability crosses 0.6. Also the there is decrease in Gini Coefficient after that probability. So it becomes difficult to set ideal probability in this case. However there is slightly less gini for part two even when the time is same and hence a good ideal point. Therefore second method is better for getting good ideal points.