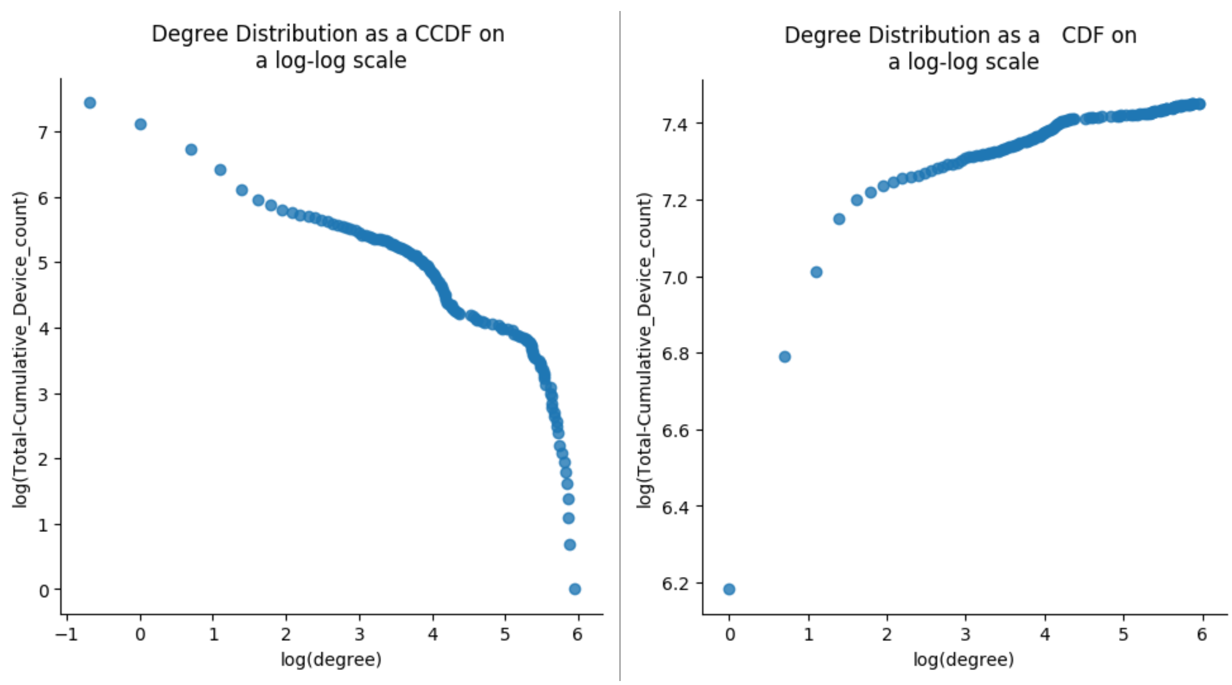
Assignment 3

The devices were divided into three classes – top S% of devices in terms of degree as super-nodes which are the most mobile and come in contact with most devices, bottom L% of devices which are the least mobile and will require special treatment to be reached, and the remaining devices which are the middle class ordinary nodes. Upon a sighting, instead of sending the chunk to the first K neighbours a device sees, a device sends the chunks to super-nodes it encounters with a different probability than to the ordinary nodes, and always with probability 1 sends the chunk to the least mobile devices as a special treatment.

Part A

Degree Distribution

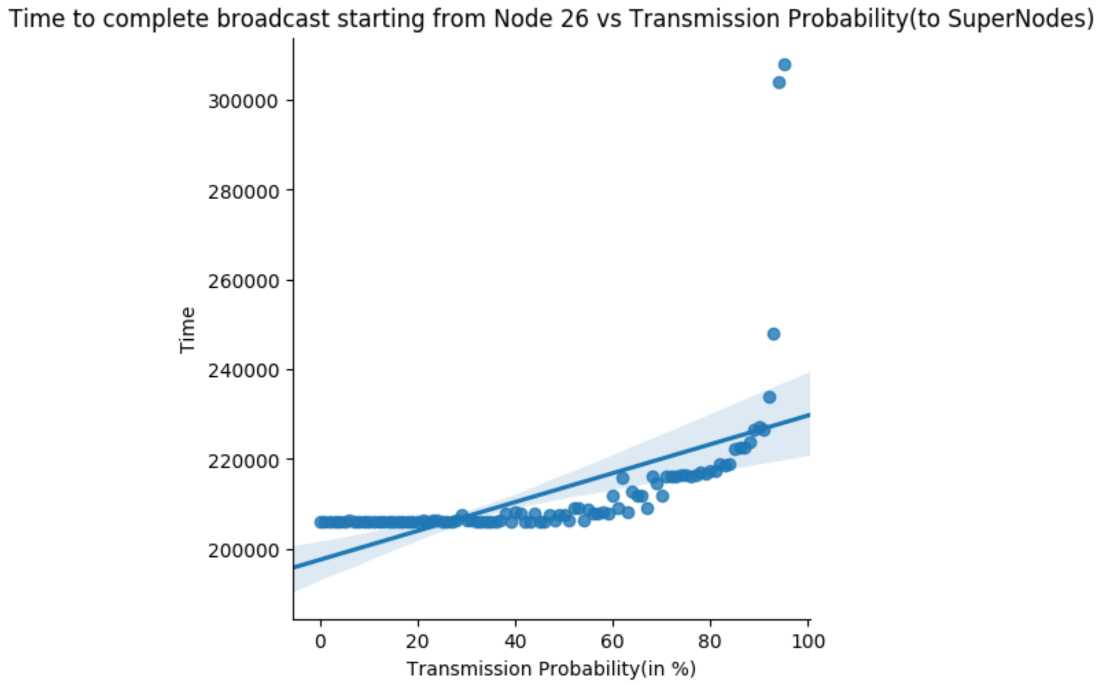
As part of a pre-classification exercise, we ran through the entire data and found the degree of the devices – the degree is the count of unique neighbours the device sees in the trace. The degree distribution was plotted as a CCDF on a log-log scale.



Graphs

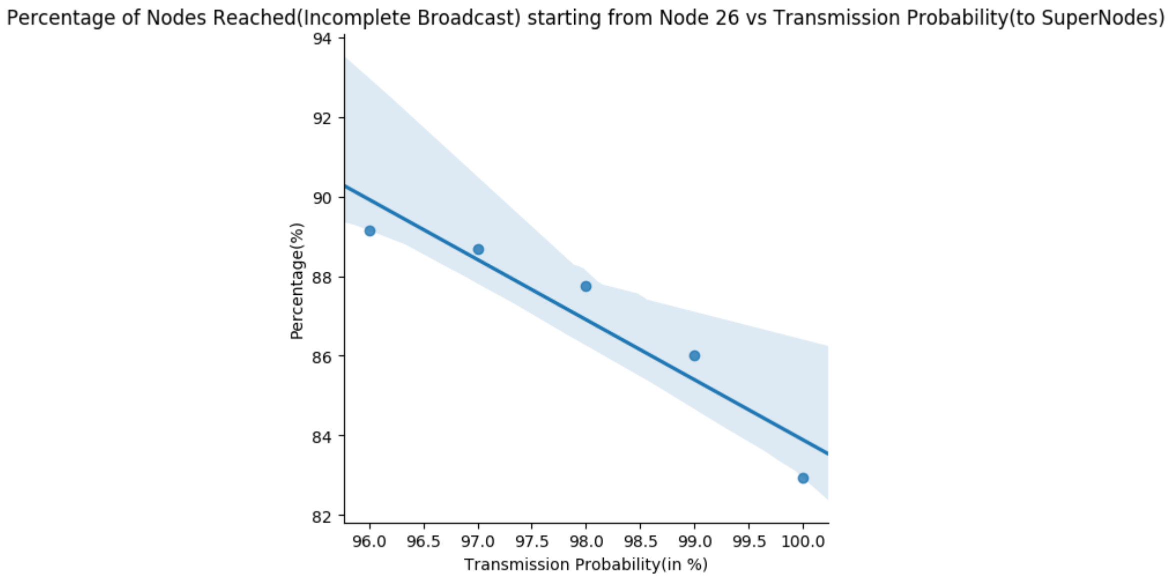
The following graphs were obtained by varying the transmission probability to super nodes(X%) for this algorithm, with the starting node as 26. The values of S and L were 0.5% and 70% respectively. The sample space was (X%,Y%), where Y = 100-X.

Broadcast Time



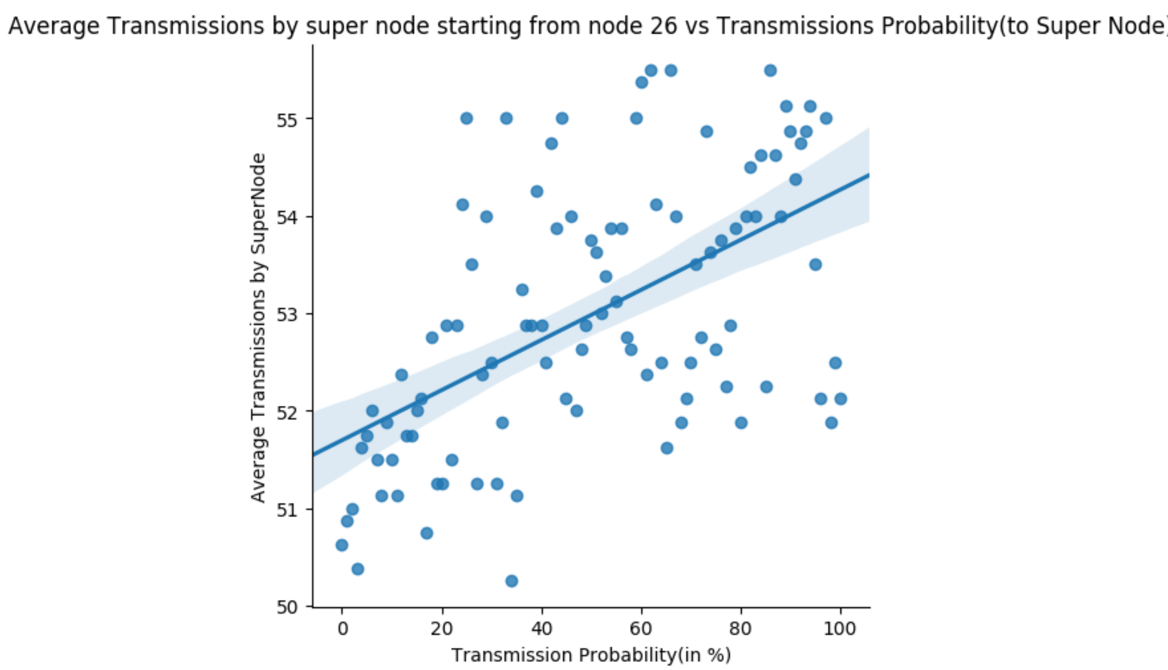
As the transmission probability to super nodes increases, we see there is an overall increasing trend in the time taken to reach 90% of the nodes. This is confirmed by the linear regression obtained from the graph. The possible reason for this is that as the Probability of Transmission to Super Nodes increases, the probability of transmission to ordinary nodes decreases. Since the number of ordinary nodes in the graph (~30%) is much larger than the number of super nodes (~0.5%), there is an increase in the overall time reach 90% of the nodes. However, we see that time remains constant in the beginning as X increases. This might be because the increase is balanced by a counter factor, i.e. as the transmission probability to super nodes increases, more of the super nodes will have the chunk, and more will be their ability to forward or distribute the chunk in the network.

Percentage of Nodes Reached (<90%)

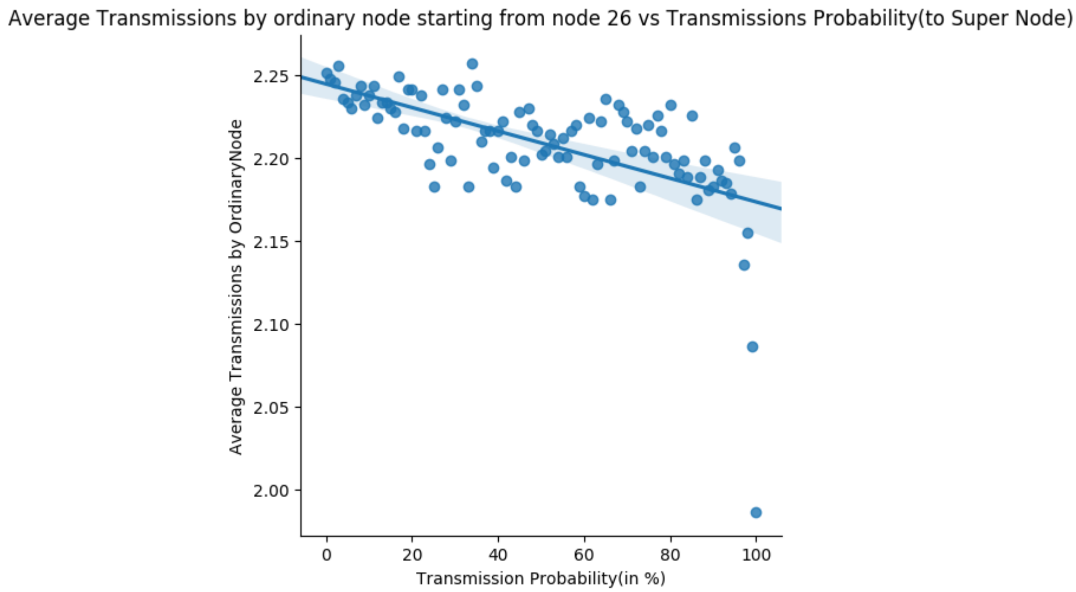


For the cases the chunk was able to reach 90% of the nodes, however for very large values of X, the chunk wasn't able to reach 90% of the nodes. For such large values of X, the corresponding values of Y, i.e. the probability of transmission to Ordinary Nodes is very less (<4%) and since they comprise of almost 30% of the devices, the chunk isn't able to reach a lot of these devices. Hence as the value of Y decreases, the percentage of nodes reached further decreases. The same is obtained on applying a linear regression to the data.

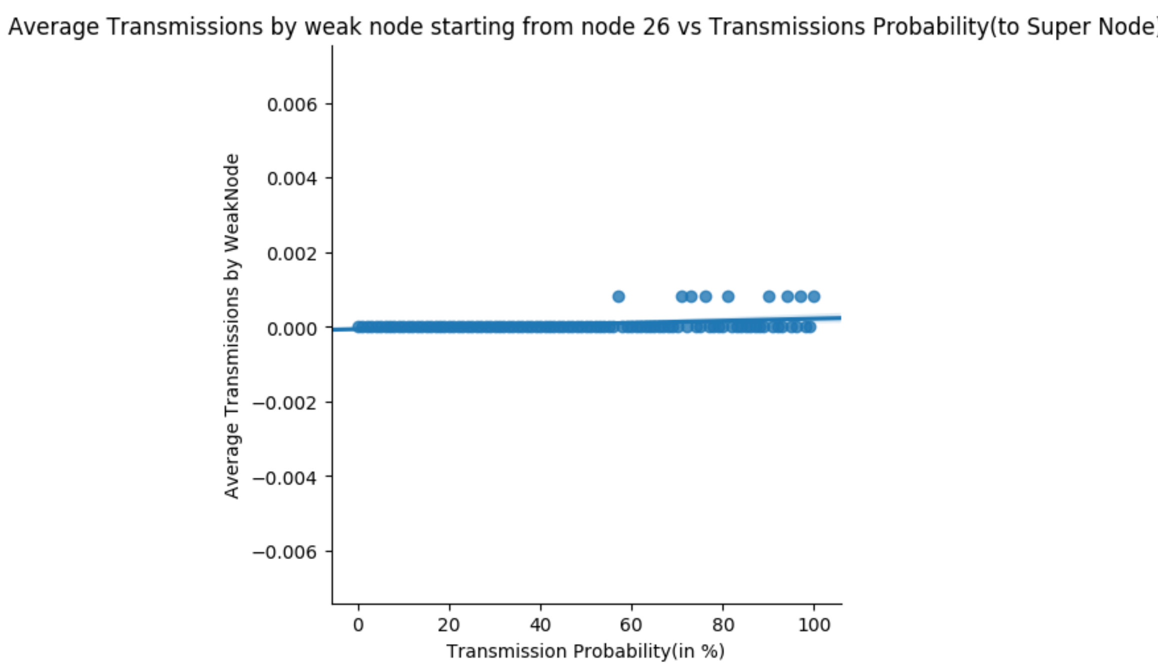
Average No. of Transmission by Super, Ordinary and Weak Nodes



As we see in the above graph, there is an overall increasing trend in the Average No. of transmissions made by Super Nodes as X increases. This is because as X increases, the percentage of super nodes with the chunk increases which thus increases the average number of transmissions made by them.



As the Transmission Probability to Super Nodes increases, the Transmission Probability to Ordinary Nodes (Y) decreases and hence lesser number of ordinary nodes get a copy of the chunk due to which the average number of transmissions made by ordinary nodes also decreases.

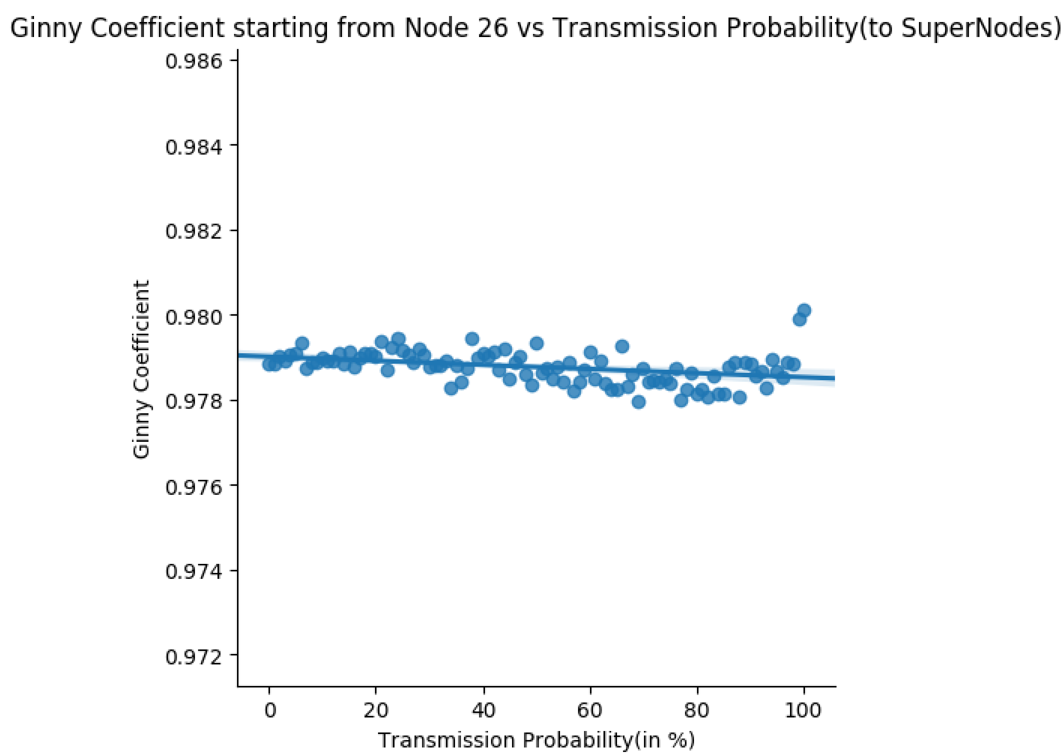


No significant change is seen in the average number of transmission to weak nodes as for any value of X, the probability of transmission to weak nodes remains the same (=1). Since these nodes are weakly connected, these values are very close to 0 for most of the cases.

Stress in the Network

The stress in the network is calculated using the Gini Coefficient. The Gini coefficient measures the inequality among values of a frequency distribution (for example, levels of income). A Gini coefficient of zero expresses perfect equality, where all values are the same (for example, where everyone has the same income). A Gini coefficient of 1 (or 100%) expresses maximal inequality among values (e.g., for a large number of people, where only one person has all the income or consumption, and all others have none, the Gini coefficient will be very nearly one). The following formula was used to calculate the value:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2\sum_{i=1}^{n}\sum_{j=1}^{n}x_j} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n\sum_{i=1}^{n}x_i}$$



As seen from the graph, there is a decreasing trend in the overall stress of the network with increase in Transmission probability to Super Nodes. The same is obtained by applying a linear regression to the data. This falls in line because increasing X reduces the network's potential to transmit (as explained in previous parts). So, in a way, it's actually decreasing the stress at the cost of reducing the actual work that is done! As X is increased, Y decreases and hence, super nodes are even more stressed out and the ordinary nodes are much lesser stressed, thereby increasing the disparity. This is what we see in the graphs for average number of transmissions made by super nodes and ordinary nodes. The perpetual high gini coefficient value is because some nodes having large network coverage (strong) as compared to most with only 1 neighbour.

Best Combination

As seen from the data, for S = 0.5% and L = 70% the best combination would be having about 40-60% Probability of Transmission to Super Nodes (X). In this range, the broadcast time isn't very large and the stress in the network is also not very much. For higher values of X, the stress in the network decreases but the time for broadcast increases considerably. For lower values of X, the time for broadcast slightly decreases but the stress in the network increases. Hence there is a tradeoff between these two goals.

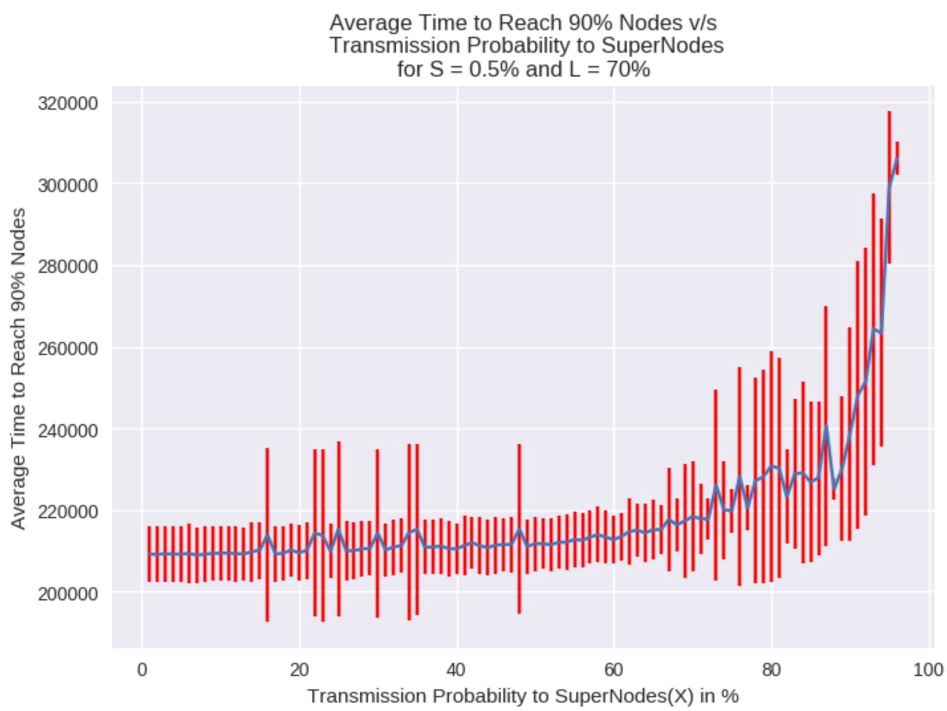
Part B

The same experiment was repeated 100 times taking any random node as the starting point to see if the trends persist statistically.

Graphs

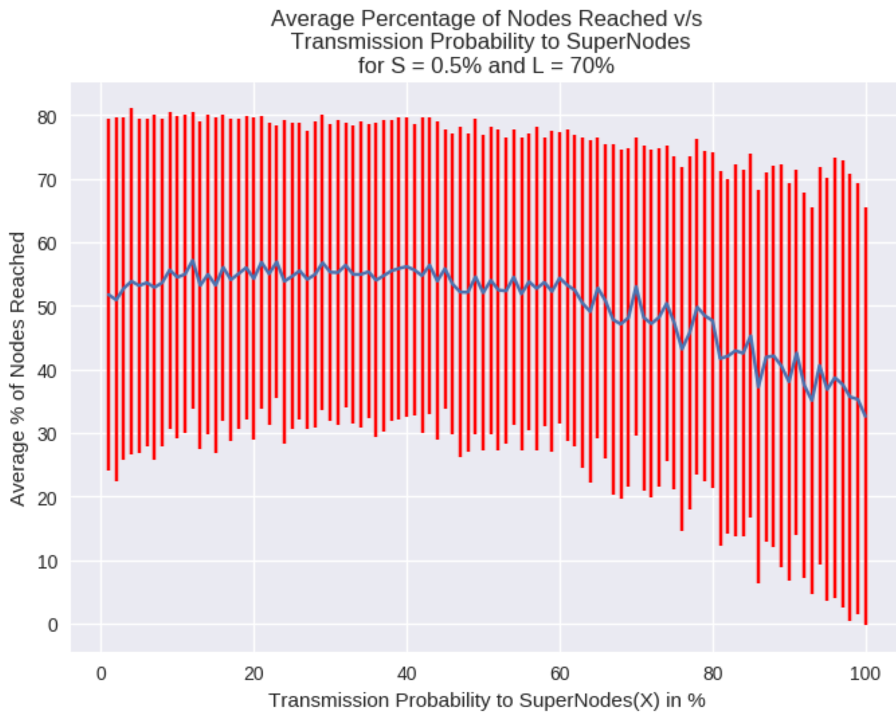
The following graphs were obtained for the average values (the red lines represent the standard deviations).

Broadcast Time



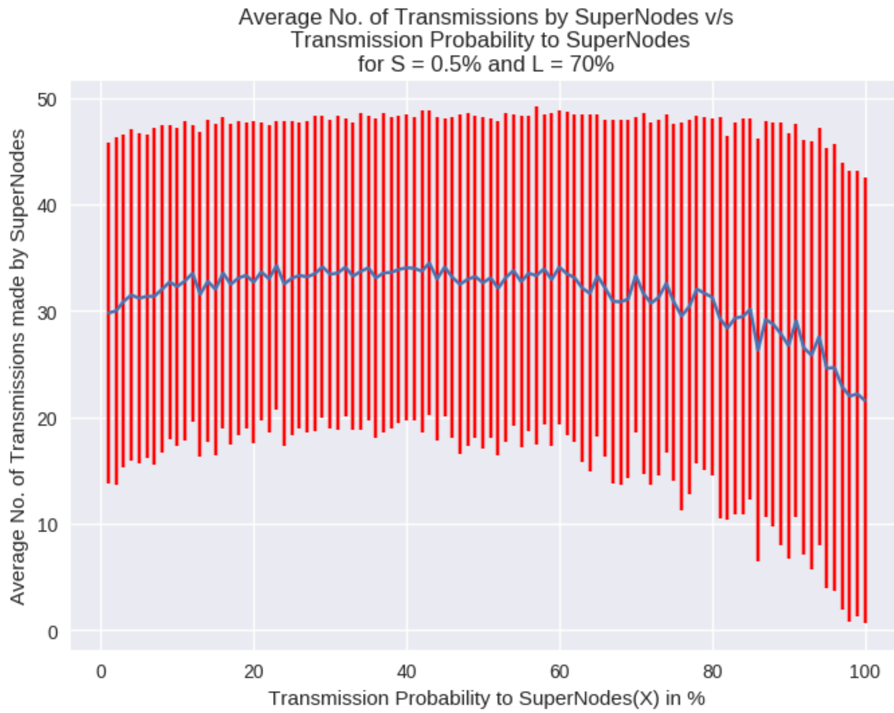
The same trend is seen for the case where the starting node was 26. The time to reach 90% of the nodes first slightly increases with increase in Transmission Probability to Super Nodes (X) and then sharply increases. The reason is same as explained before.

Percentage of Nodes Reached (<90%)

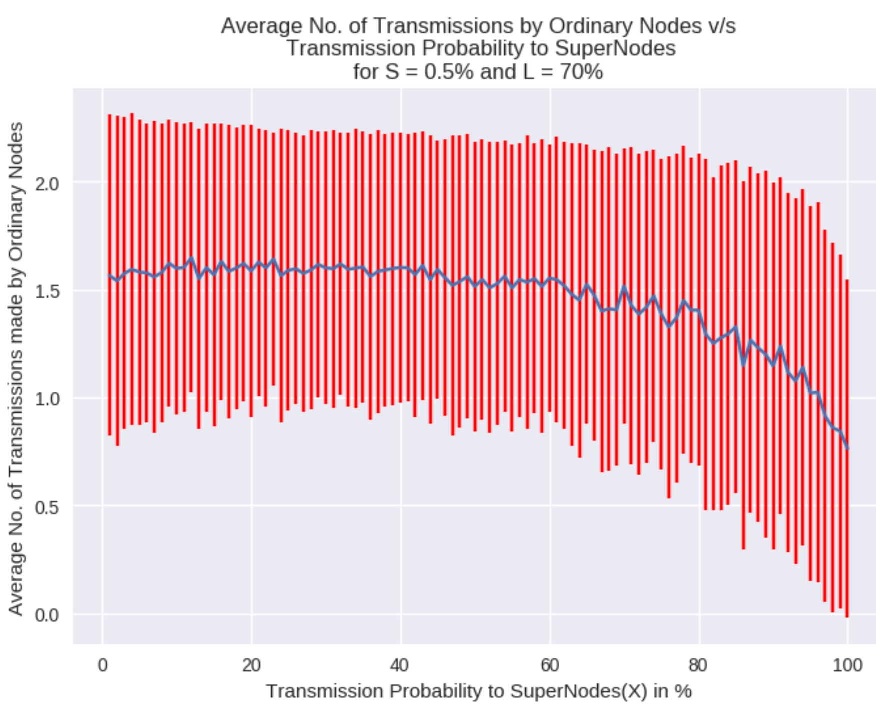


As we saw before, the average percentage of nodes reached decreases (for cases when the chunk couldn't reach 90% of the nodes). As value of X increases, the corresponding values of Y, i.e. the probability of transmission to Ordinary Nodes decreases and since they comprise of almost 30% of the devices, the chunk isn't able to reach a lot of these devices. Hence as the value of Y decreases, the percentage of nodes reached further decreases. Also, in the beginning, there is a small dip in the percentage of nodes reached as the value of X is very small. Since the probability of transmission to super nodes i.e. the most connected nodes is very small, the chunk isn't able to spread over the network very easily.

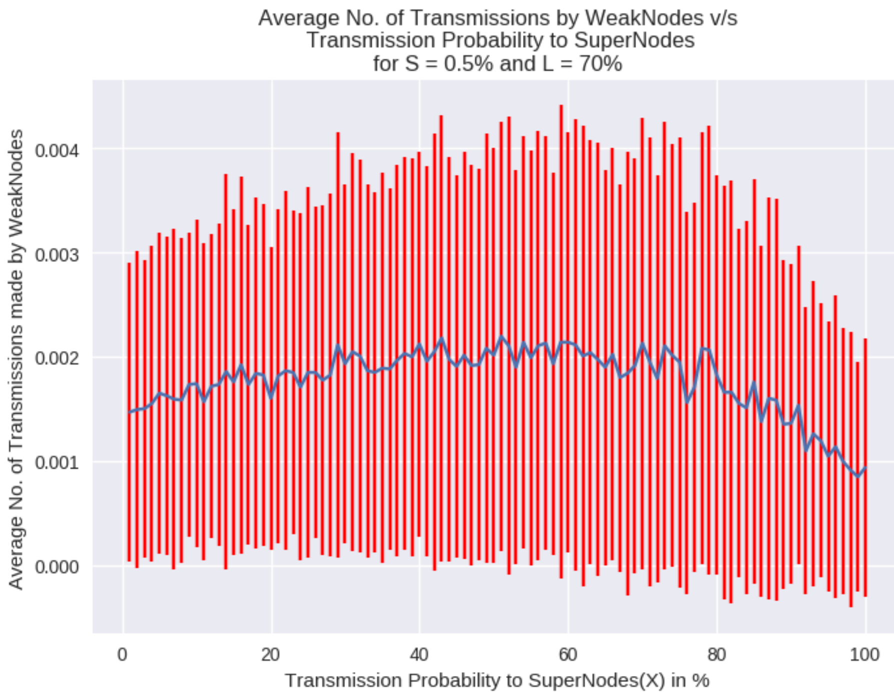
Average No. of Transmission by Super, Ordinary and Weak Nodes



Here, we see that the trend is different from what we saw for the 26 as starting node case. Averaging over a 100 nodes stabilized the graph. For large values of X, Y is very small. Since about 30% of the nodes are ordinary nodes and transmission probability to them is very low, the average number of transmissions by super nodes decreases.

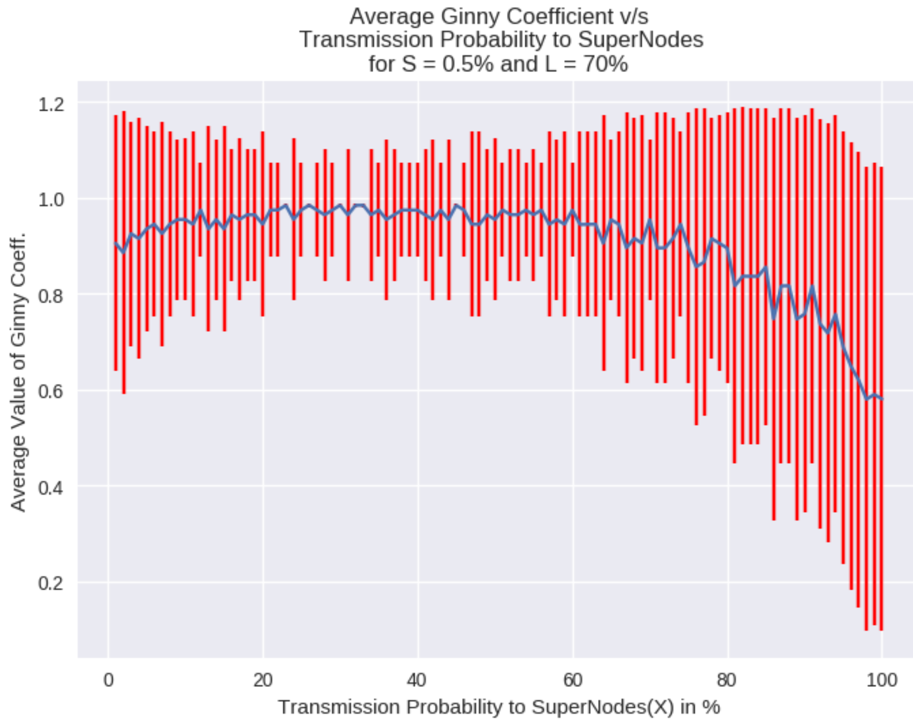


The average number of transmissions by ordinary nodes shows a decreasing trend similar to what was seen in the case with starting node as 26.



Similar to the case where starting node was 26, the average number of transmissions by weak nodes is almost zero and there is very little variation in the values.

Stress in the network



We observe that the graph obtained for Gini Coefficient corresponds to the graph for Average Number of Transmission by the super nodes. This is because Gini Coefficient is a measure of disparity in work load. Since most of the work is done by super nodes, they highly affect the Gini Coefficient. The ordinary and weak nodes always have average transmissions much lower than Super nodes. Hence, when the average number of transmissions by super nodes is lower, there is less disparity in the network and hence the Gini coefficient decreases. For small values of X, since probability of transmission to super nodes is lesser, lesser number of super nodes get the chunk and hence the work done by them is lower. For large values of X, since probability of transmission to ordinary nodes (~30% of the nodes) is very less, the number of transmissions made by the super nodes is less. Since the ordinary and weak nodes anyway do much lesser work, the disparity is lesser in this case and hence the Gini Coefficient reduces.

In general, we see a somewhat decreasing trend in the overall stress of the network with increase in Transmission probability to Super Nodes. This falls in line because increasing X reduces the network's potential to transmit (as explained in previous parts). So, in a way, it's actually decreasing the stress at the cost of reducing the actual work that is done! As X is increased, Y decreases and hence, super nodes are even more stressed out and the ordinary nodes are much lesser stressed, thereby increasing the disparity. This is what we see in the graphs for average number of transmissions made by super nodes and ordinary nodes. The perpetual high gini coefficient value is because some nodes having large network coverage (strong) as compared to most with only 1 neighbour.

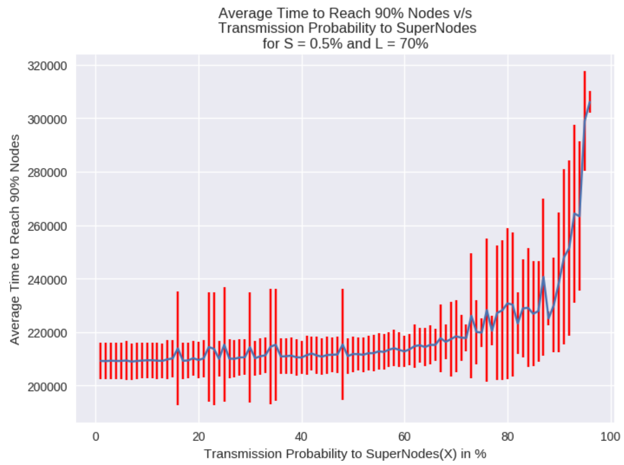
PART C

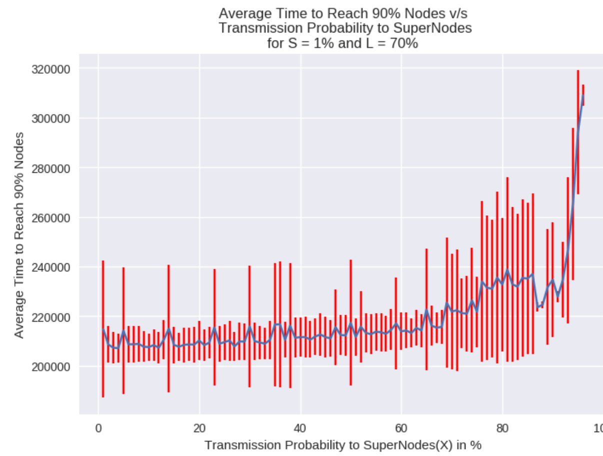
We study the variations on changing the value of S (percentage of Super Nodes)

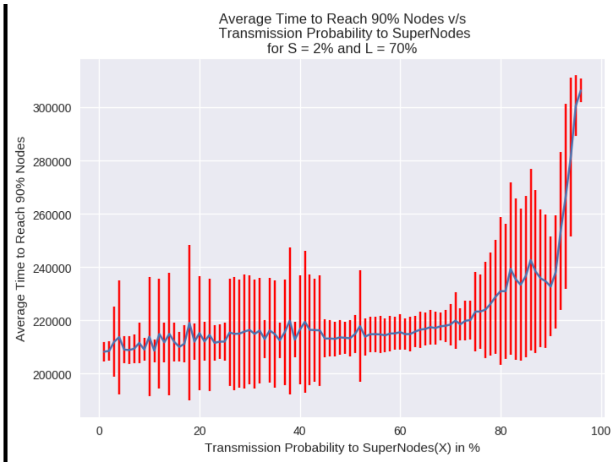
Graphs

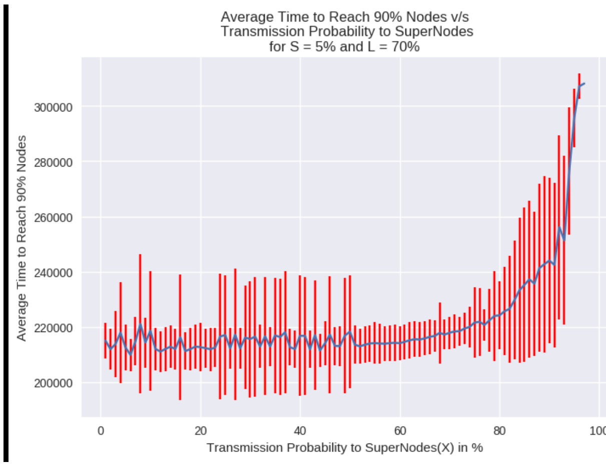
The following graphs were obtained by changing the values of S and keeping L = 70%.

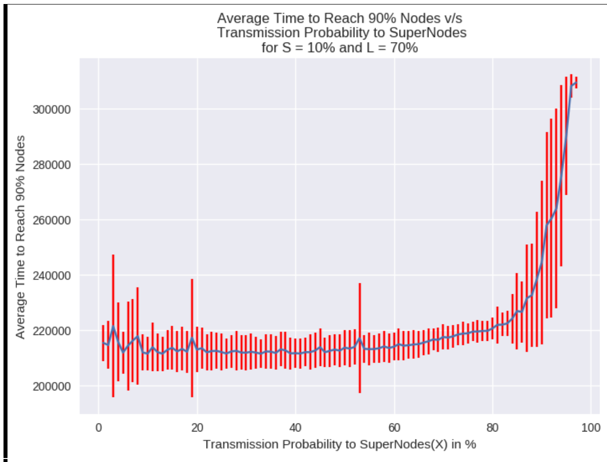
Broadcast Time

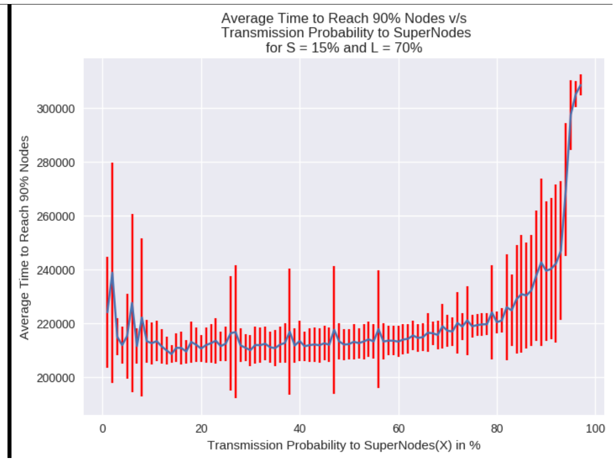


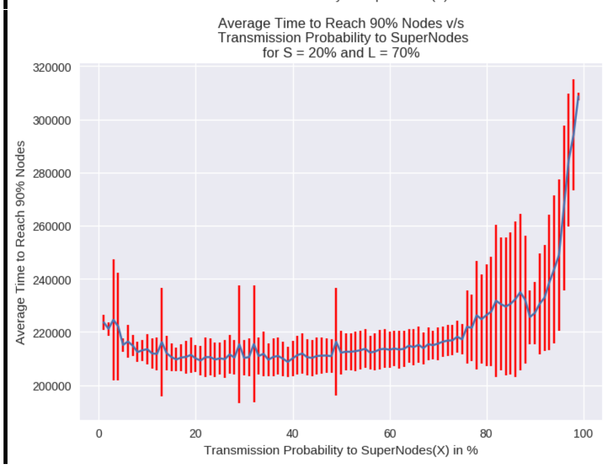






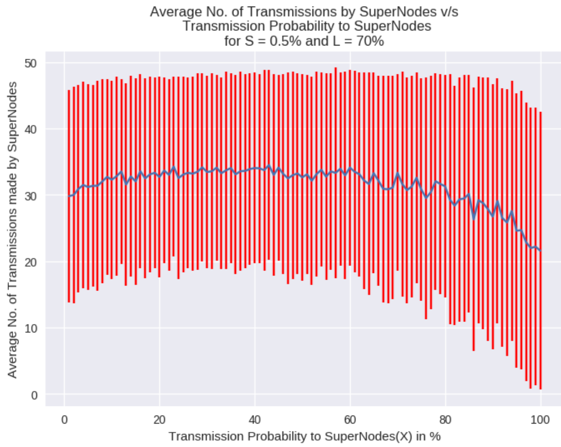


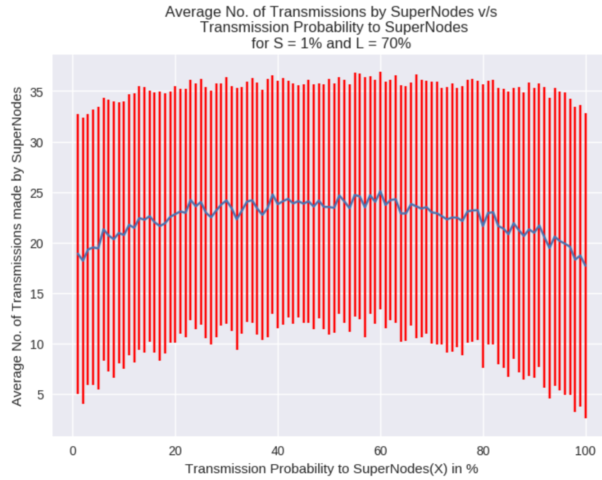


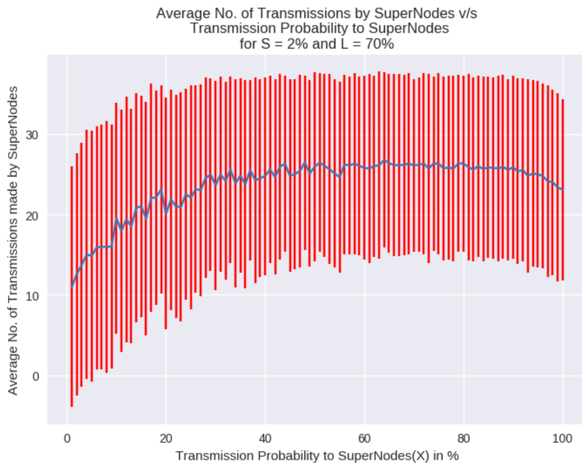


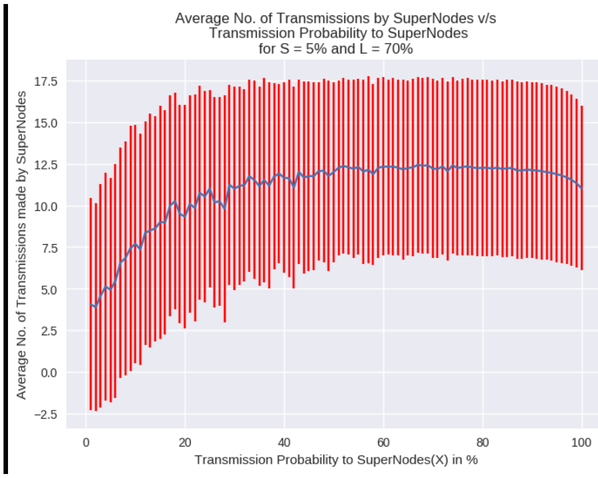
No significant change could be seen in the broadcast time with change in the percentage of super nodes. For very low values of X, the broadcast time increases with increase in S. Since the number of super nodes increases with S and the probability of transmission to super nodes(X) is very less, lesser number of nodes get the chunk in the network and hence broadcast time increases.

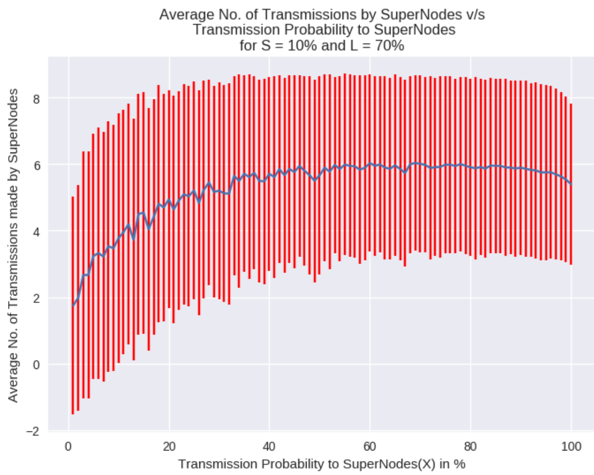
Average No. of Transmissions by Super Nodes

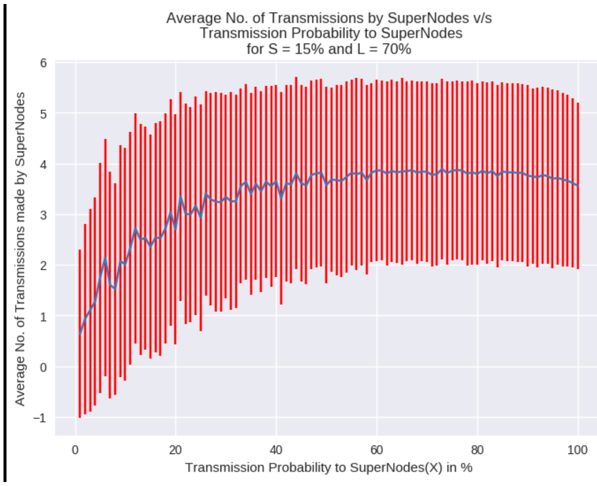


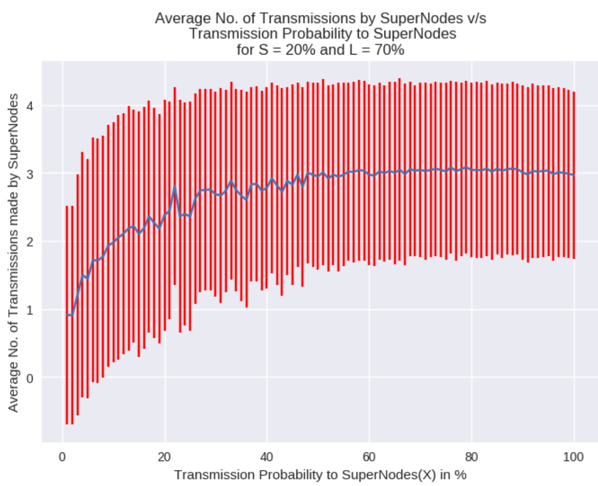






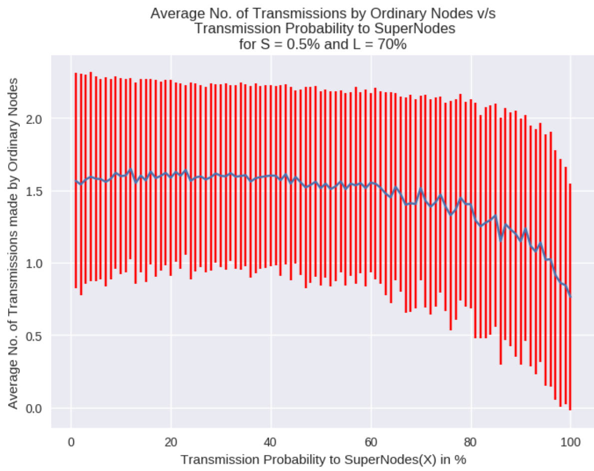


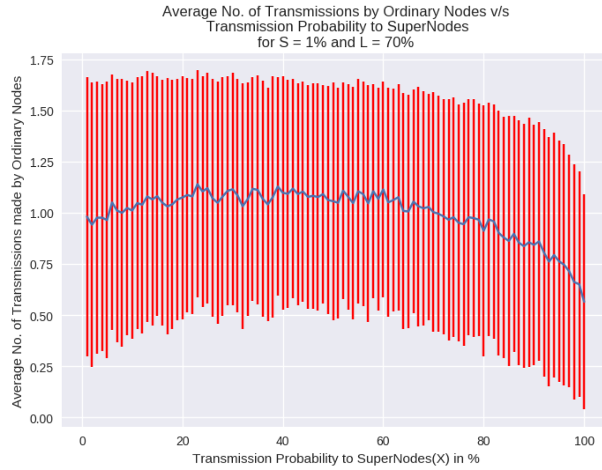


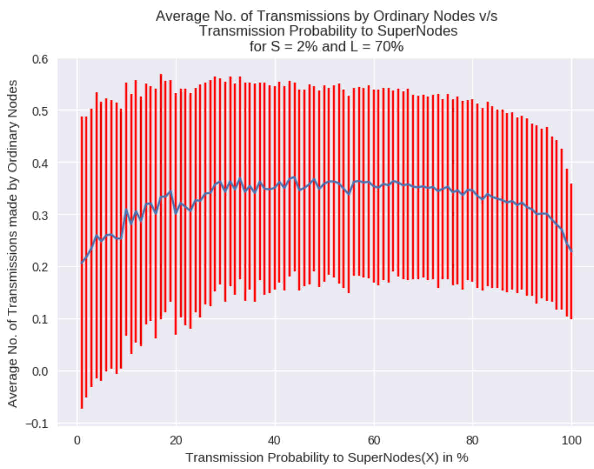


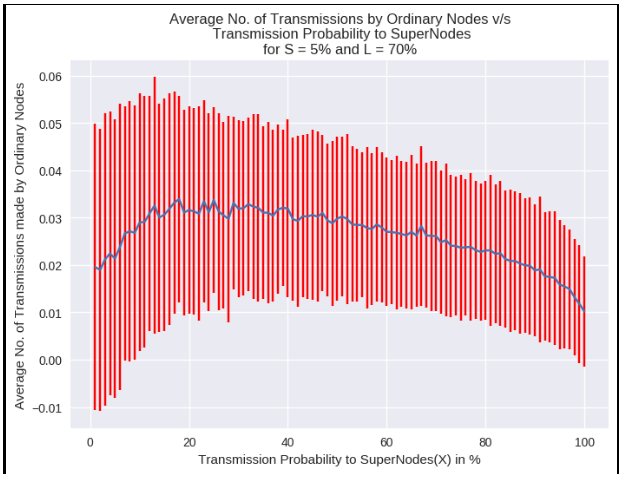
As the value of S increases, there is a significant decrease in the average number of transmissions made by super nodes, i.e. the entire graph shifts downward. Also, for small values of X, there is a greater decrease in average no. of transmissions with increasing S as the number of super nodes increases but the probability of transmission to super nodes is small and hence few super nodes get a copy of chunk. Since most super nodes are connected to other super nodes but transmission probability to super nodes is very small hence average number of transmissions by super nodes is very small.

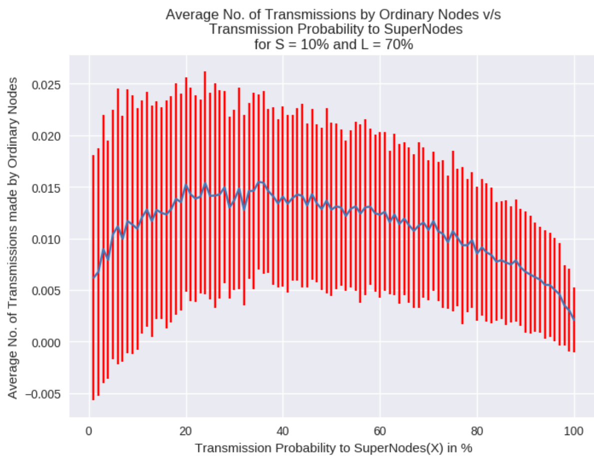
Average No. of Transmissions by Ordinary Nodes

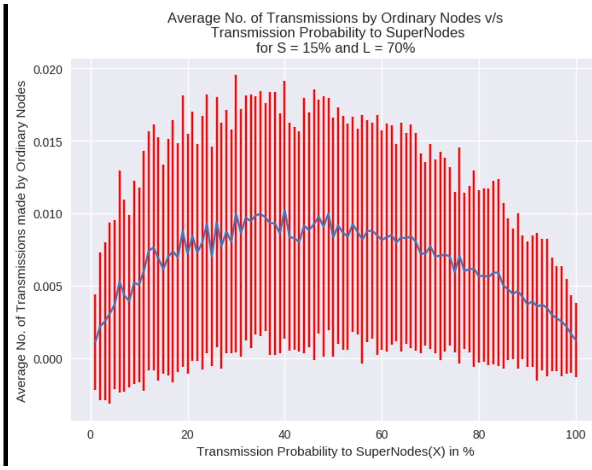


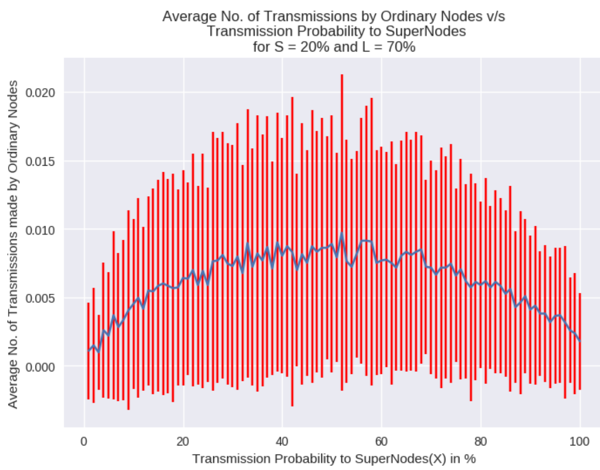






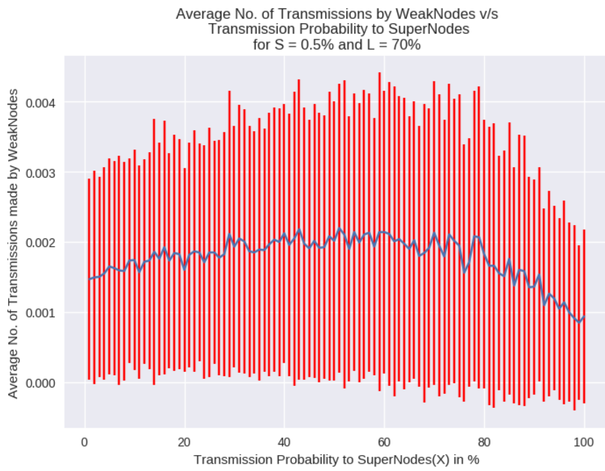


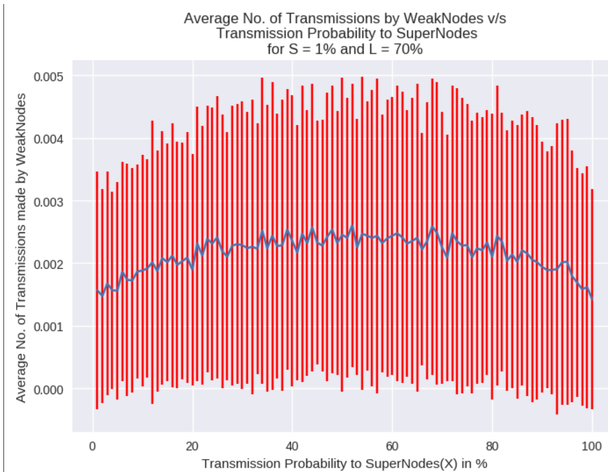


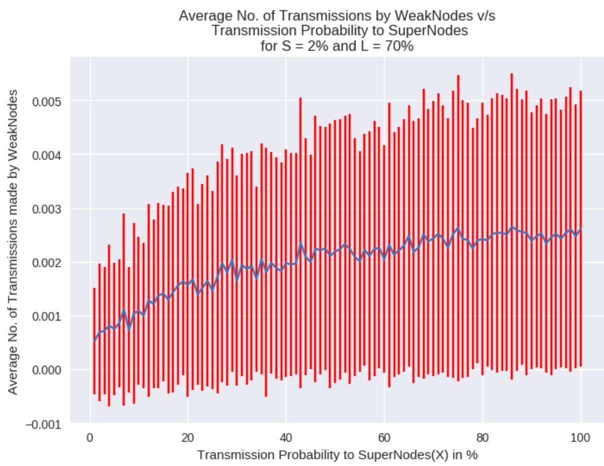


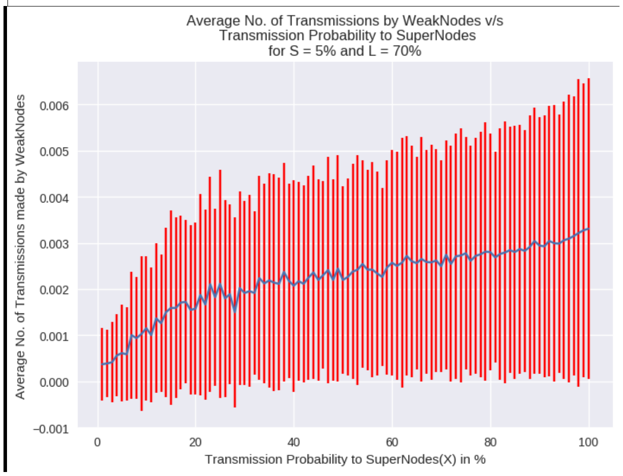
Here also, there is an overall decrease in the average no. of transmissions made by Ordinary Nodes with increase in value of S, i.e. the graph shifts downward. Also, for small values of X, there is a greater decrease in average no. of transmissions with increasing S as the number of super nodes increases but the probability of transmission to super nodes is very small and hence the chunk isn't distributed easily over the network. Since most ordinary nodes would be connected to some super node and transmission probability to super nodes is very small hence average number of transmissions by ordinary nodes is very small.

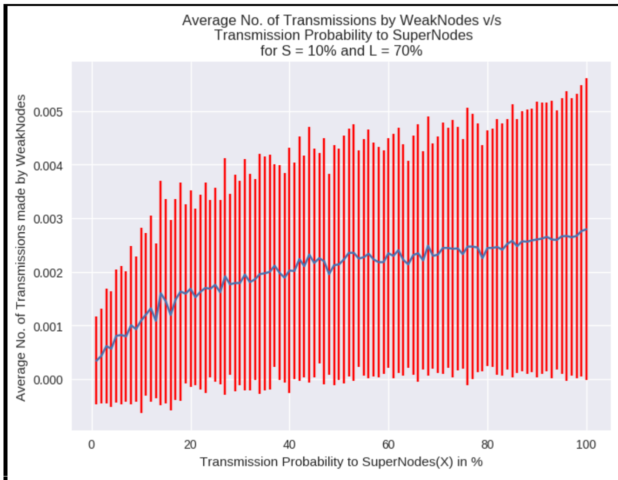
Average No. of Transmissions by Weak Nodes

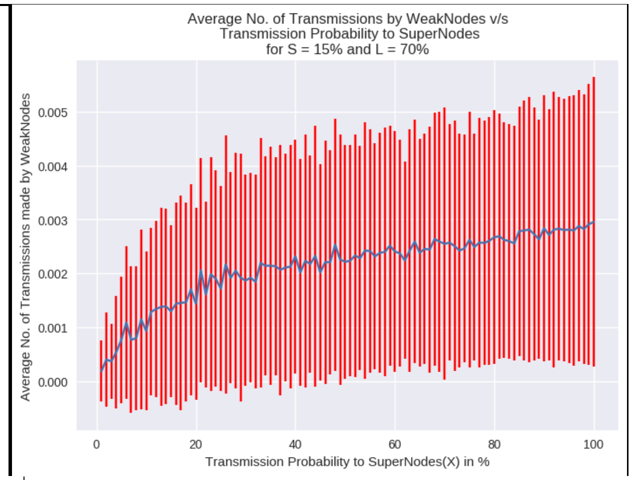


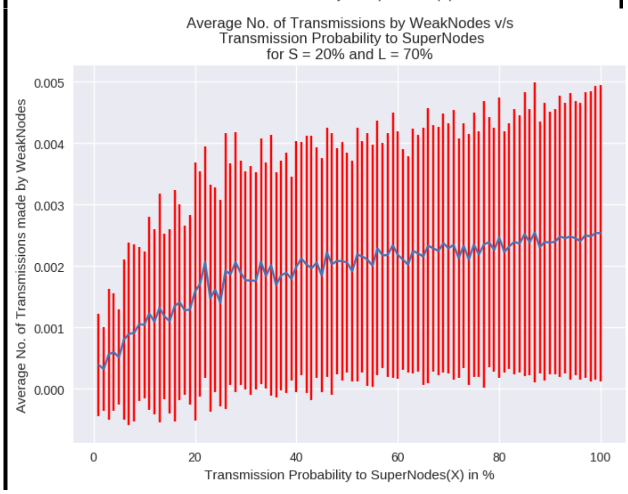






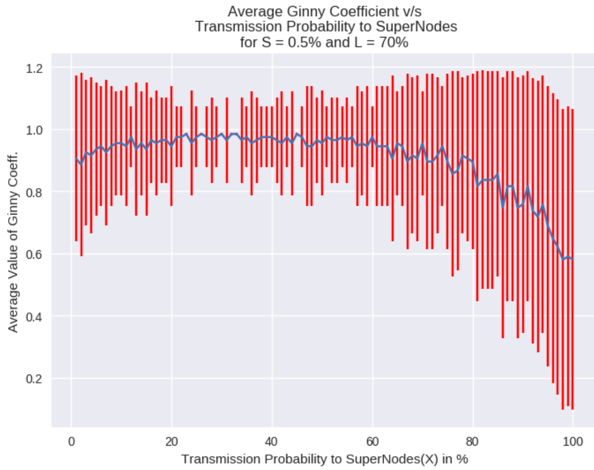


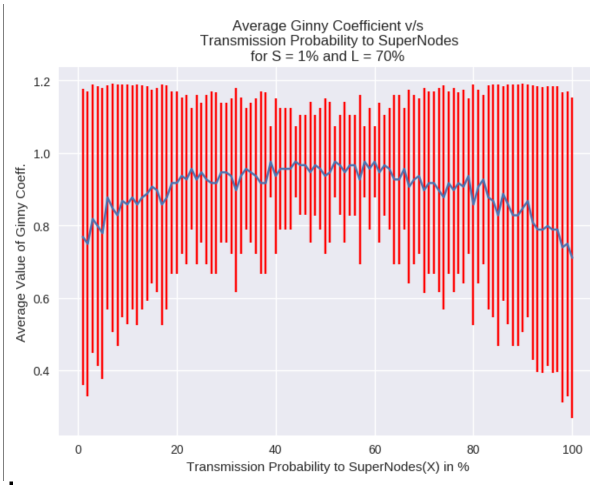


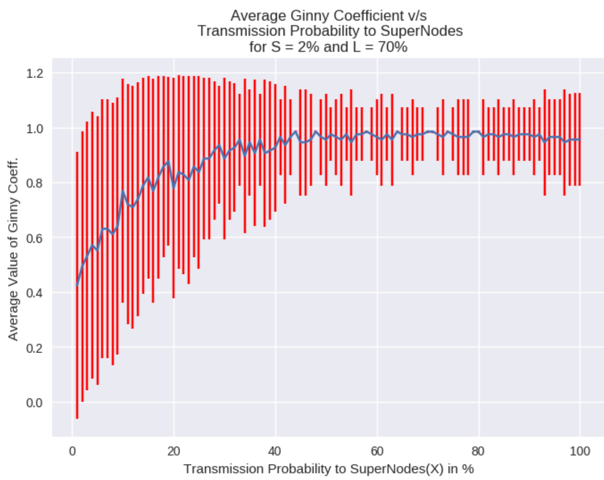


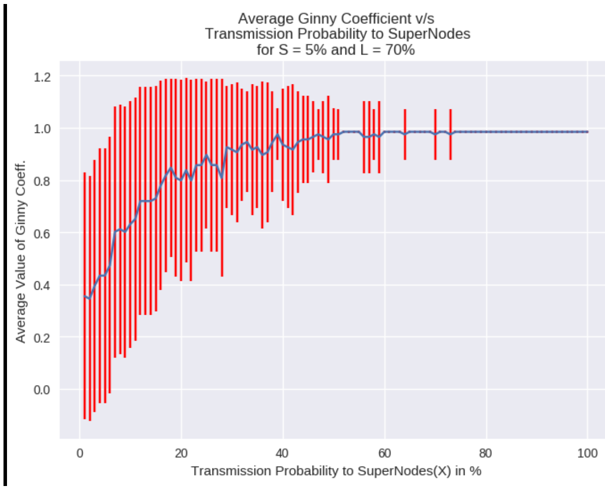
As such, we don't see any trend in the above graphs because the value of average number of transmissions by weak nodes is very close to zero i.e. almost negligible.

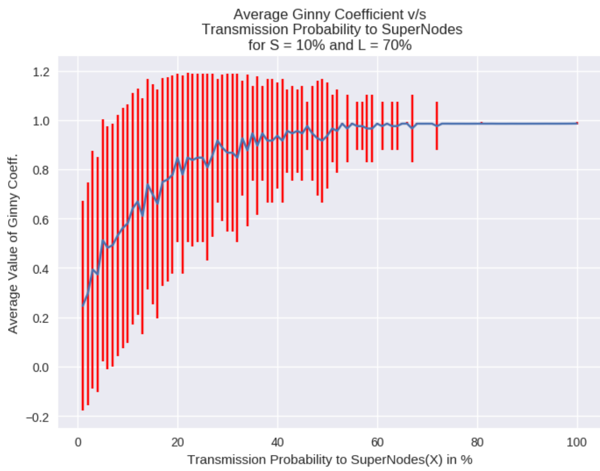
Stress in the Network

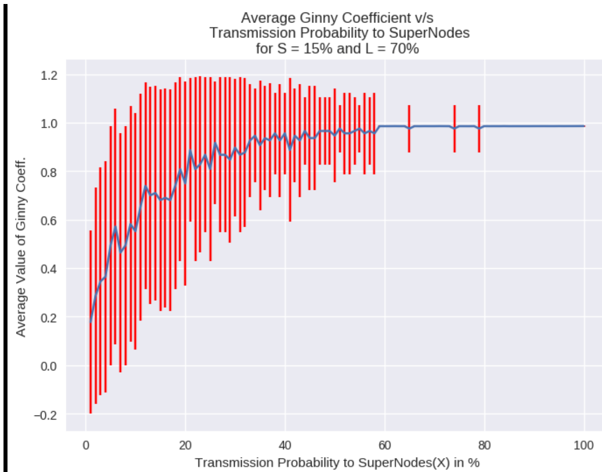


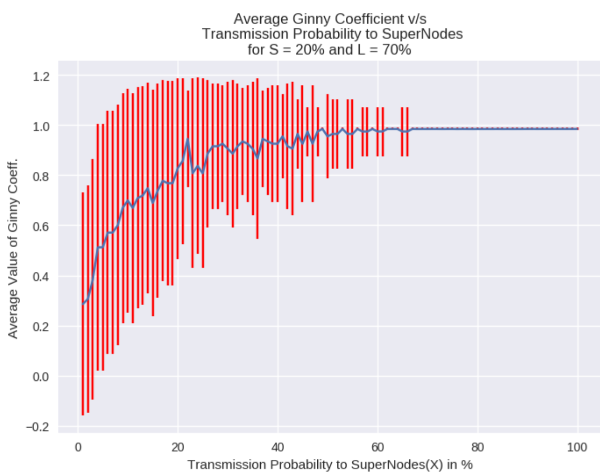












As the value of S increases, we see that Gini coefficient converges to 1 for large values of X which implies that the stress in the network is very large. For small values of X, it reduces.

We observe that the graph obtained for Gini Coefficient corresponds to the graph for Average Number of Transmission by the super nodes. This is because Gini Coefficient is a measure of disparity in work load. Since most of the work is done by super nodes, they highly affect the Gini Coefficient. The ordinary and weak nodes always have average transmissions much lower than Super nodes. Hence, when the average number of transmissions by super nodes is lower, there is less disparity in the network and hence the Gini coefficient decreases. For small values of X, since probability of transmission to super nodes is lesser, lesser number of super nodes get the chunk and hence the work done by them is lower. As the value of S increases, the number of Super Nodes in the network increases. And since all of them are doing lesser work, the disparity in the network decreases.

For large values of X, the probability of transmission to super nodes is high. Since the average number of transmissions by super nodes increases, and the number of of super nodes increases with increasing S, the disparity in the network increases and hence Gini coeff approaches 1.

For small S, for larger X the Gini coeff reduced as there was large number of ordinary nodes in the network not getting the chunk. However, for large S, there are many super nodes and very few ordinary nodes. Since super nodes do a lot of work and weak nodes hardly do any, there is a lot of disparity in the network.

Best Combination

As we see from the graphs of broadcast time and Gini coefficient, we get the best results for small values of X (<20%) for almost all values of S. As value of S increases, Ginny coefficient reduces for small X and broadcast time is also small.

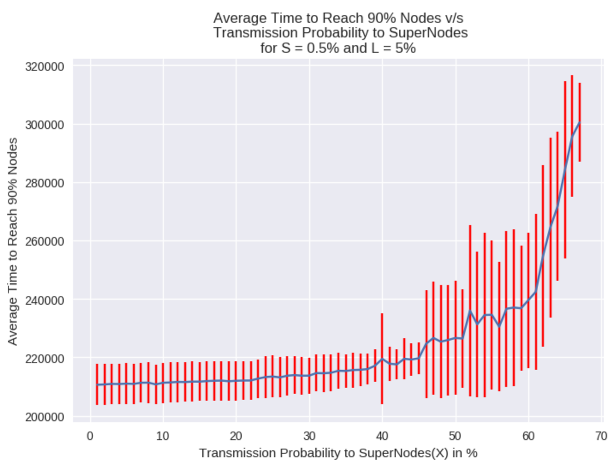
PART D

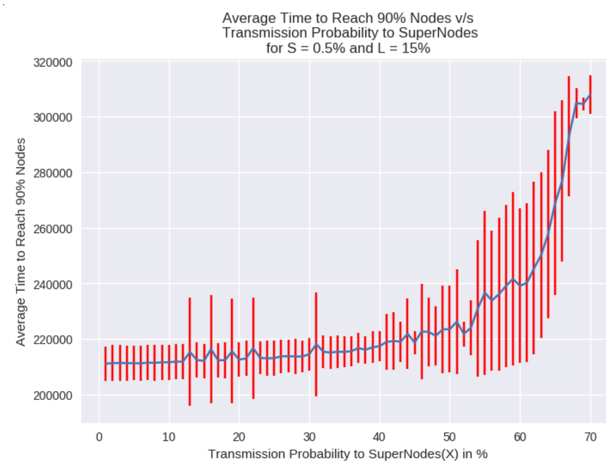
We study the variations on changing the value of L (percentage of Weak Nodes)

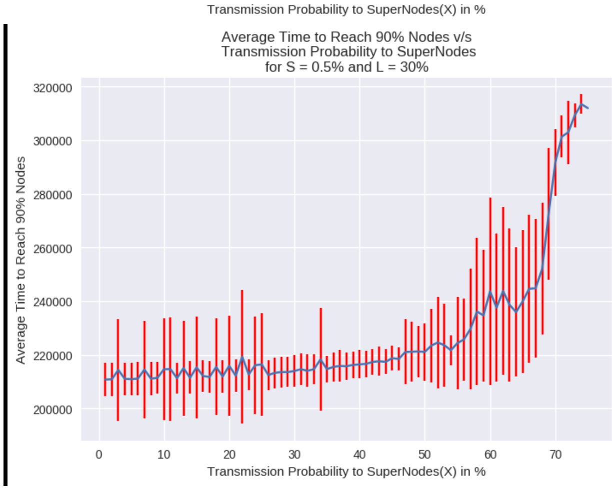
Graphs

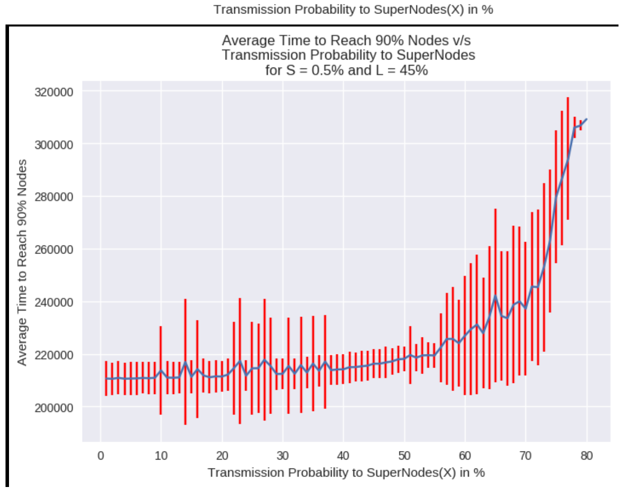
The following graphs were obtained by changing the values of L and keeping S = 0.5%.

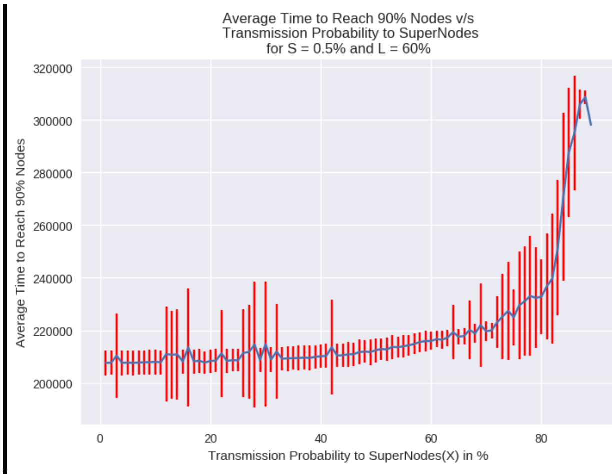
Broadcast Time

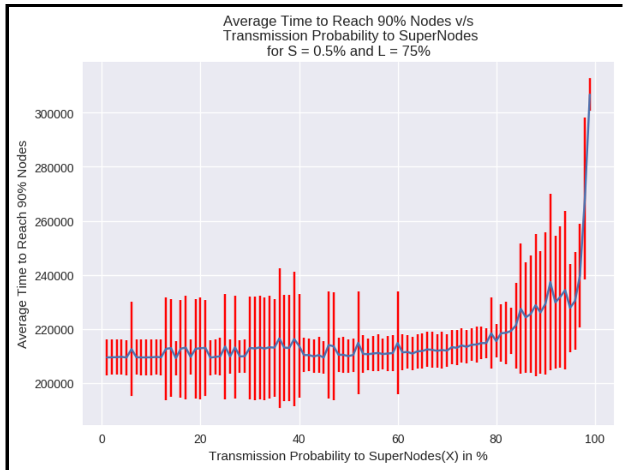


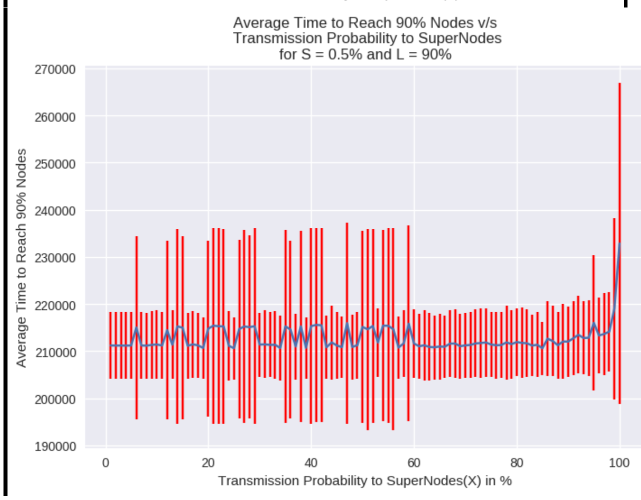






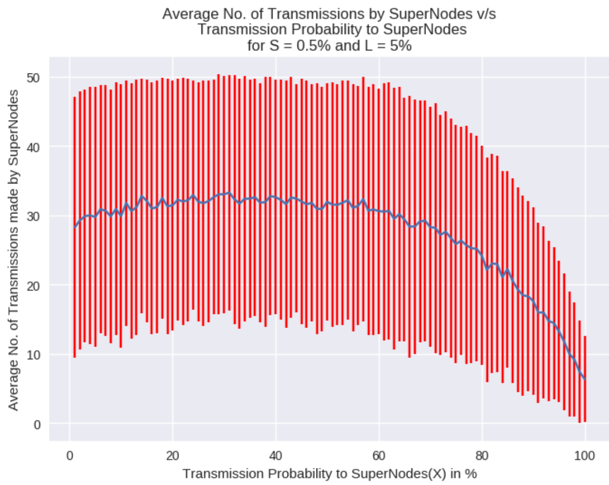


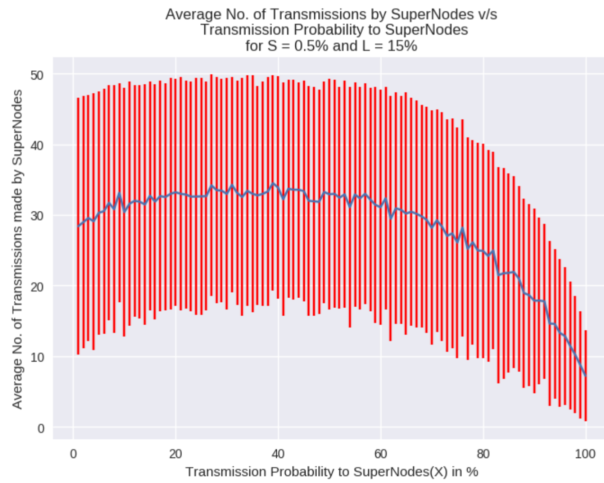


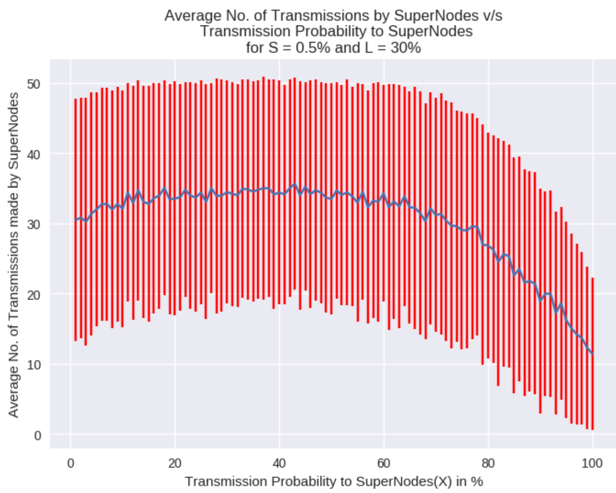


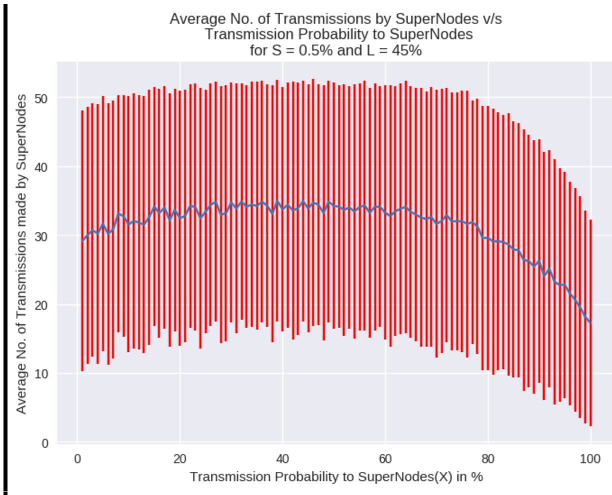
Through these graphs, we see that as the Percentage of nodes marked as Weak (L) increases, the time for broadcast significantly decreases. This is because according to our algorithm, the probability of transmission to weak nodes is 1. As more nodes are considered into this category, the rate of propagation of the chunk in the network increases and hence the broadcast time increases. For very small values of L, the chunk didn't even reach 90% of the nodes for very large values of X.

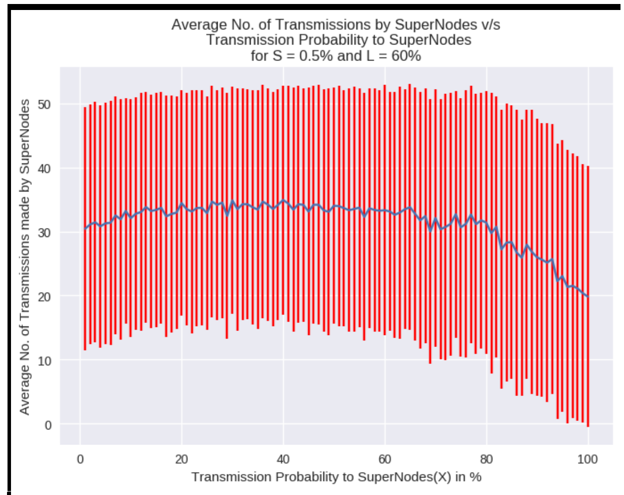
Average No. of Transmission by SuperNodes

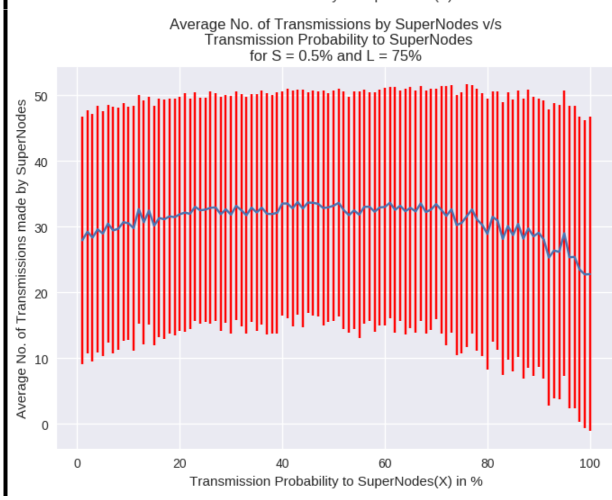


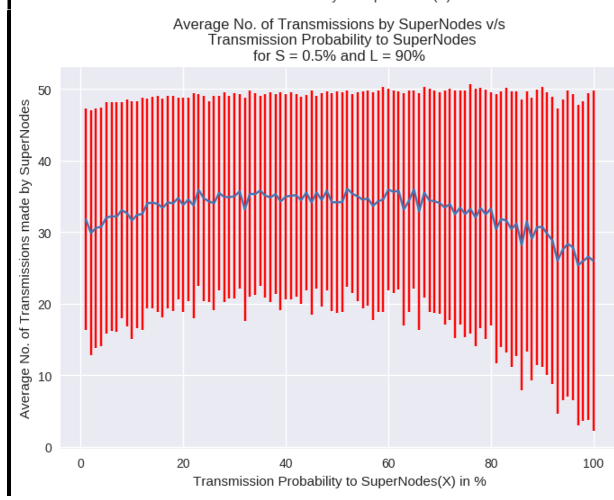






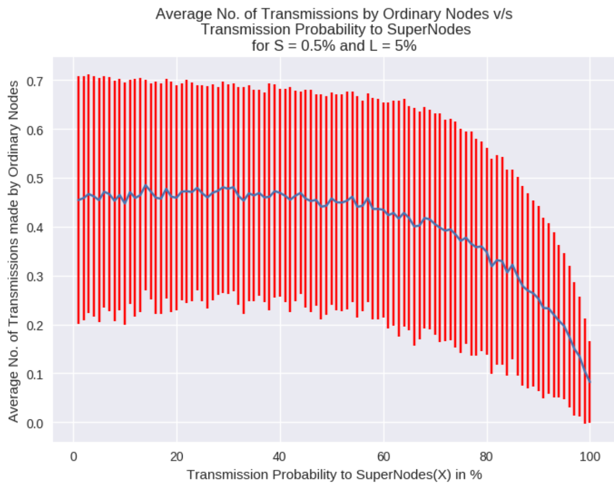


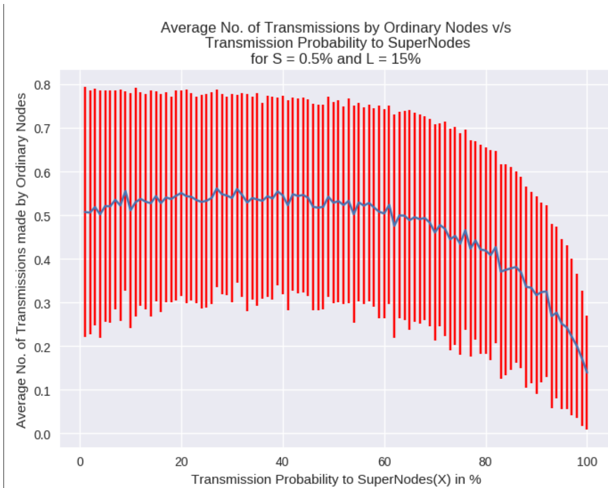


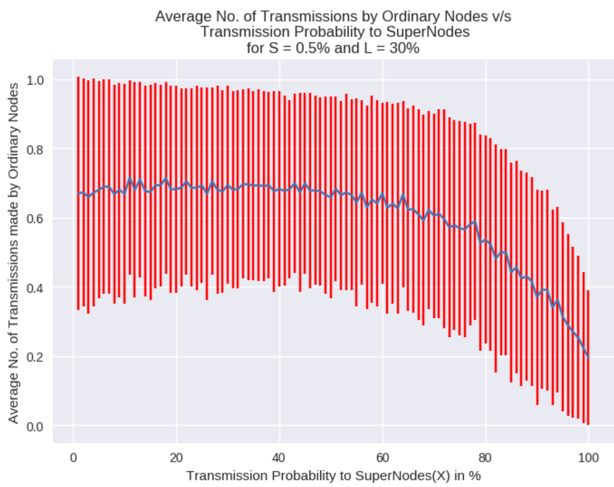


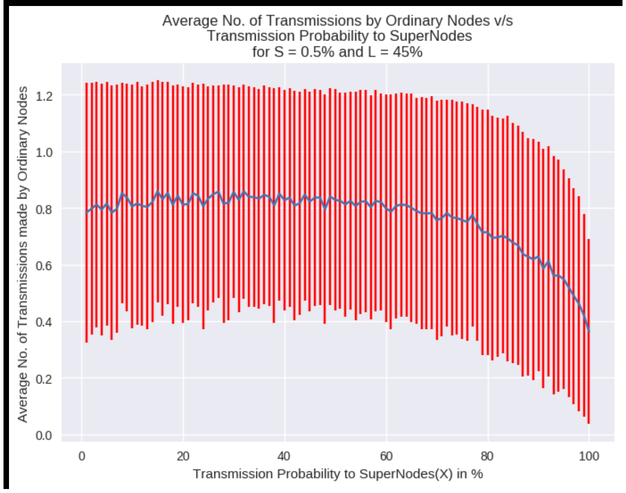
The average number transmissions made by super nodes increases for higher values of X with increase in L. This is because the probability of transmission to nodes in L is 1. Hence whenever a super node meets a node in L, it transmits for sure rather than with a probability <1. As number of nodes in L increases, so does the number of transmissions by super nodes.

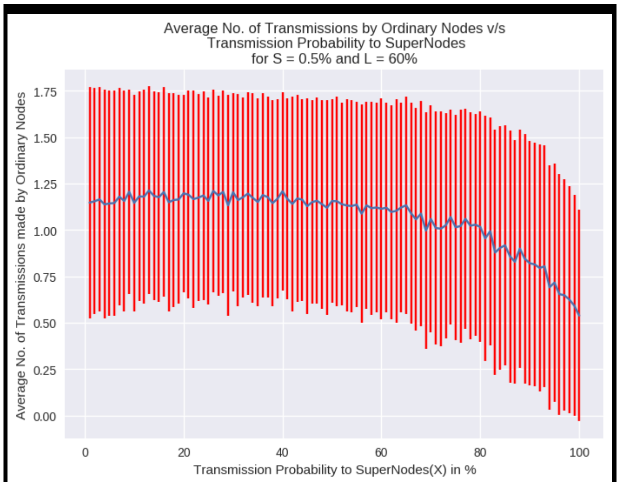
Average No. of Transmission by Ordinary Nodes

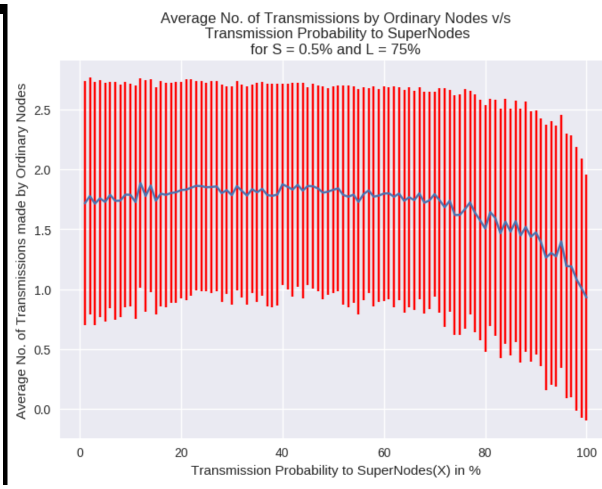


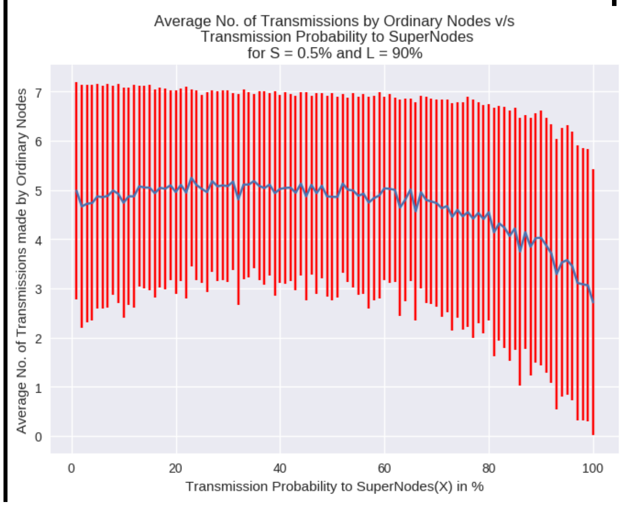






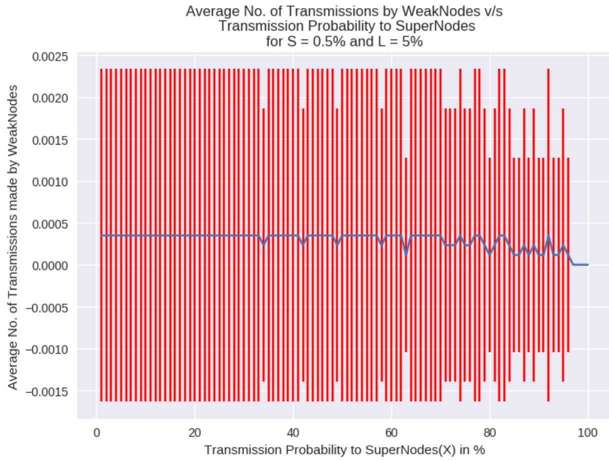


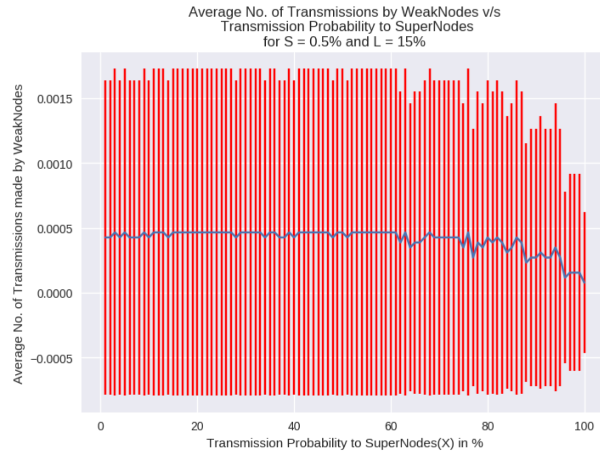


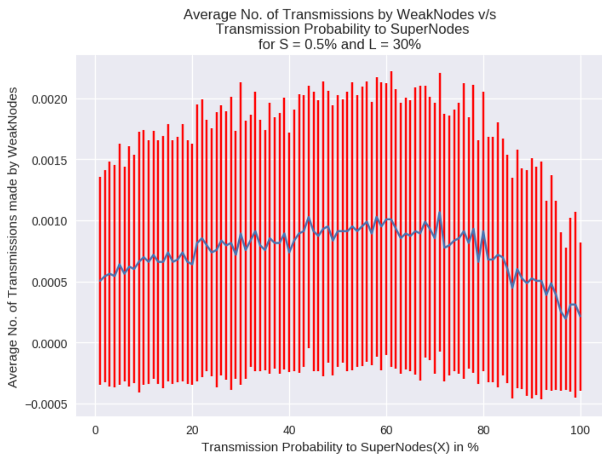


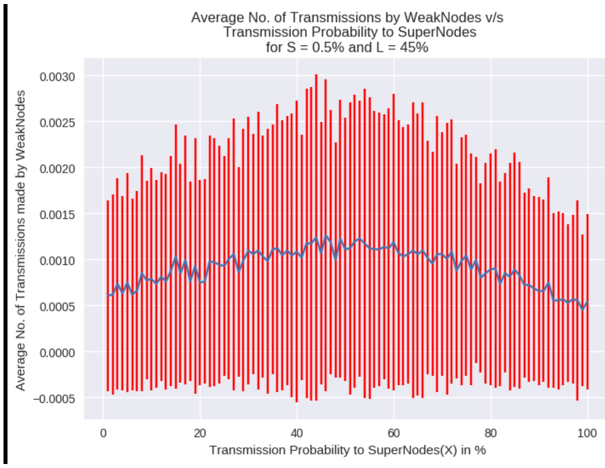
We observe that as L increases, there is a significant increase in the average no. of transmissions by the ordinary nodes i.e., the entire graph shifts upwards. The reason is similar to the one in the previous part. As probability of transmission to weak nodes is 1, as the number of weak nodes increases, the average number of transmissions significantly increases for other nodes.

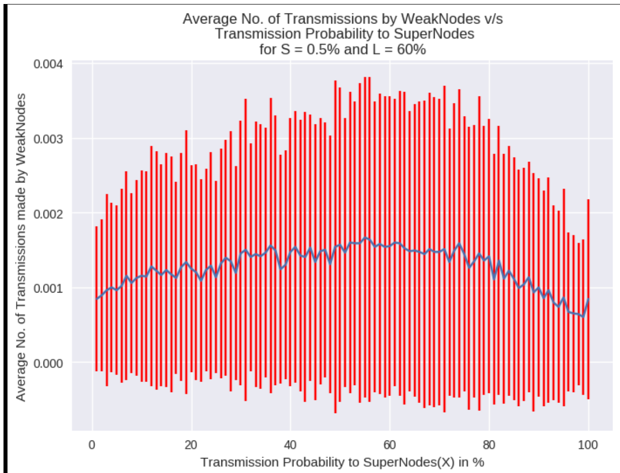
Average No. of Transmissions by Weak Nodes

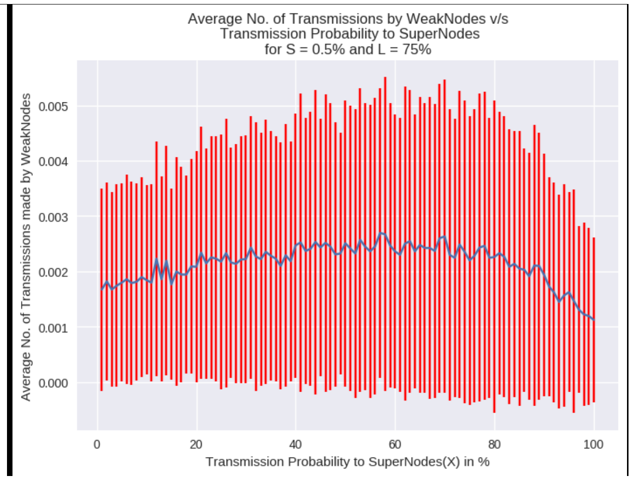


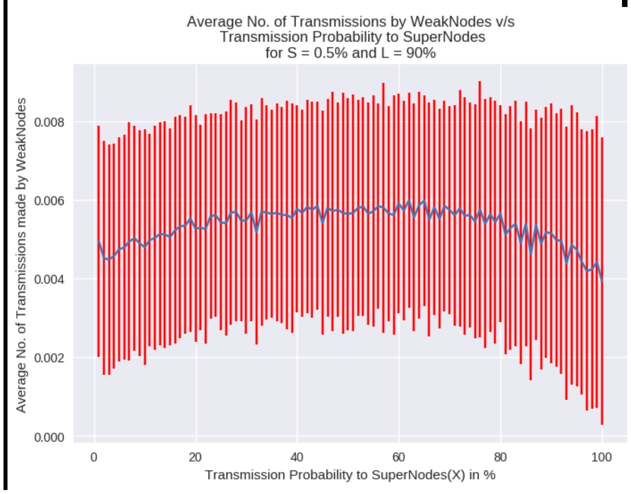






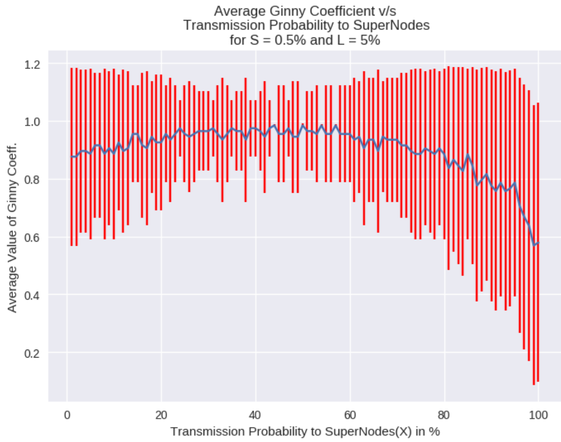


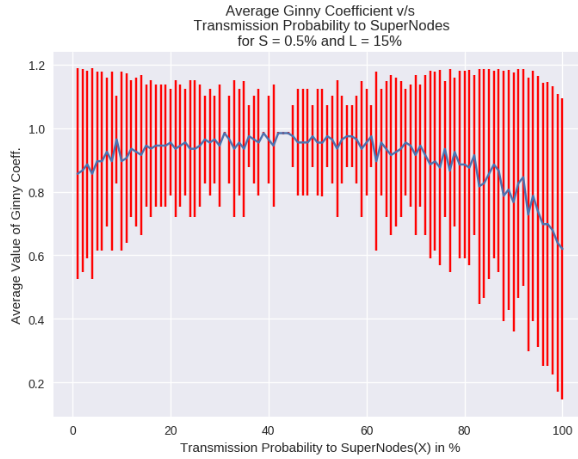


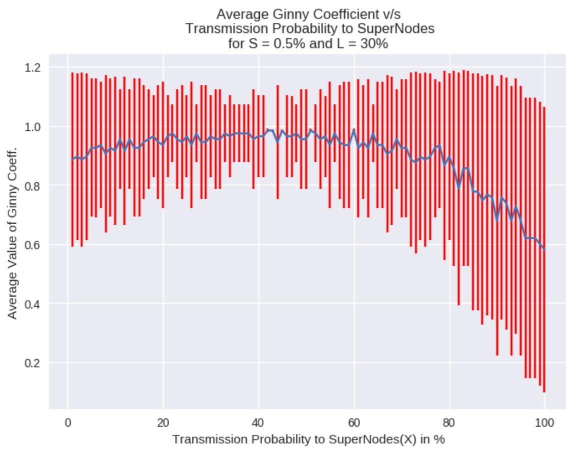


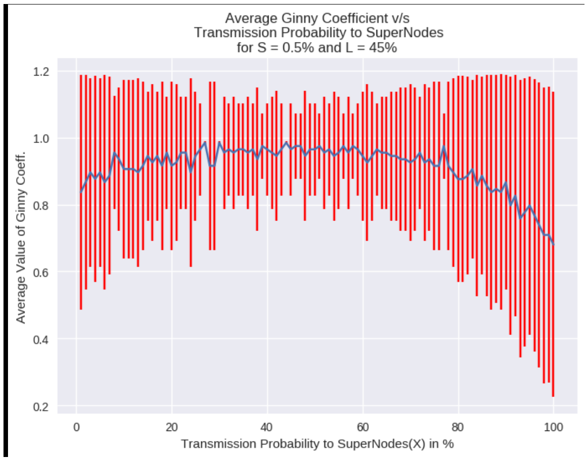
The average number of transmissions by Weak nodes increases with increase in L due to the reasons mentioned in the previous case.

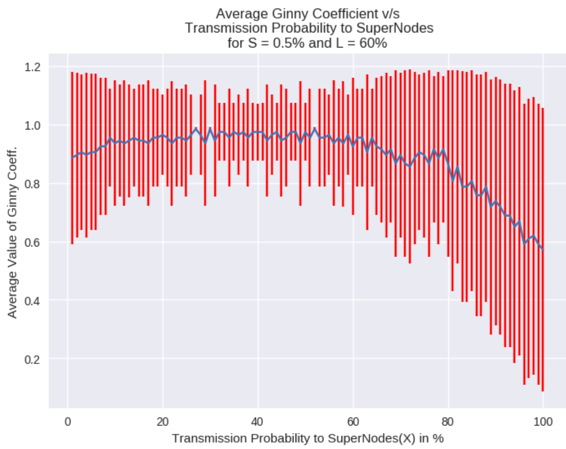
Stress in the Network

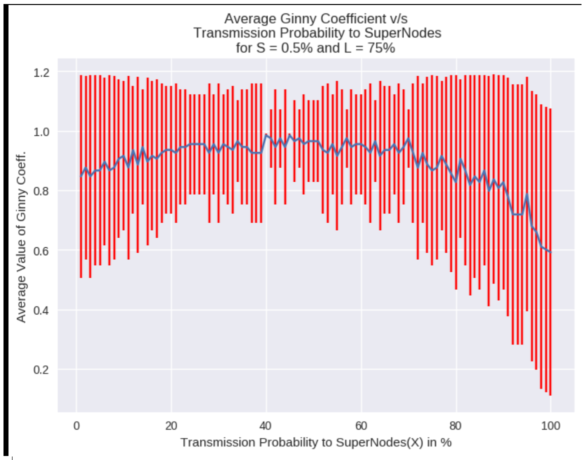


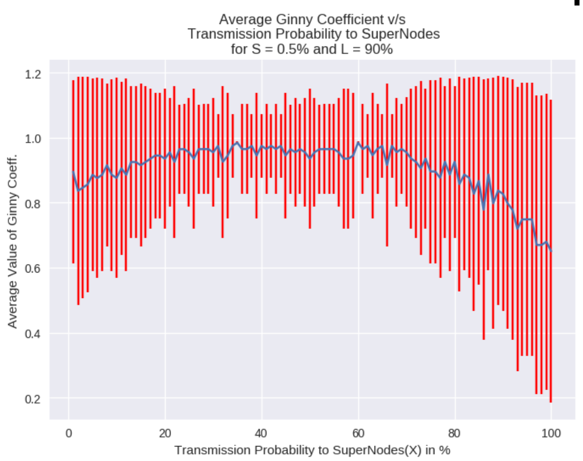












We don't see any significant change in the stress in the network with change in value of L.

Best Combination

We see that since Gini coefficient doesn't change a lot with change in L, we see at the broadcast times. For very small X (<20%), the Gini coefficient is low as well as broadcast time is small for all L. Also, for large X, Gini coefficient is smaller but broadcast times are larger so we can get a tradeoff at around X = 80-90%.