Neighborhood Overlap in Power Grid Network

By Rajveer Singh and Ramyya Hari

Abstract

The US power grid's functioning is critical to the everyday working of society. This network is complex, containing many types of infrastructure and connections between them. Generator, transmission and distribution substations span the expanse of the land, connected by high voltage transmission lines, providing the US with the electrical power we so heavily rely on.

Here we examine key attributes of the system, including edge betweenness, degree distribution and other important attributes of a network. We bring in the concept of neighborhood over lap and see how the strength ties are distributed over the entire network and how it varies over we start removing the various nodes of highest and lowest edge ties. Finally, we examine the effects on the system of removal of key nodes over degree, betweenness and modular clusters. We conclude by comparing our network with a random network and see how the various parameters vary over them and that the system is difficult to model through simple, theoretical representations, and that it is not robust in the face of targeted node failure.

Introduction

The United States power grid has an interesting infrastructure. It is a combination of generator, transmission, and distribution substations span the country, connected by high voltage transmission lines, providing the US with the electrical power we so heavily rely on. As in the case of most of the high scale infrastructure, many people do not recognize its importance until they realize the consequences of its failure. The biggest issue with such kind of an infrastructure is in case of a fault, that is what if something goes wrong, what if one substation or the transmission line fail. There will be spread of blackout over the entire network if this is not constrained. We cannot solve this problem completely but we can increase the security measures both cyber and physical over the most important nodes of this system. This would reduce the chance of the blackout spreading over a large scale.

For a better understanding of the power network, we have analyzed the US power grid network. 4,941 nodes represent the various substations, and 6,594 edges represent the high voltage transmission lines that connect them. The data was also analyzed in a 1998 Nature paper by Watts and Strogatz, as well as in "The Power Grid as a Complex Network: a Survey", a paper by Giuliano Andrea Pagani and Marco Aiello.

In this case the nodes lack geographical coordinates and the edges are not weighted however we are able to analyze the properties of this system in considerable depth. We analyze parameters such as edge betweenness, degree distribution, betweeness, shortest path length along with modular clusters to determine the relationship between betweenness and degrees and how it varies when the nodes are removed. We adapt Granovetter's social theory on the power grid network to determine the role of the strong and weak ties and how they result in forming local bridges. Isolating and identifying these nodes will help us know what may lead to the network to fail and form clusters We define a value 'neighborhood overlap' which will help identify which edges are almost local bridges and hence analyze these values to see if they satisfy Granovetter's theoretical predictions.

Analysis Tools

We first created a visualization of the network using Gephi which is a visualization tool that gives a approximate model of our dataset. Then the nodes were positioned using Yifan Hu's Multilevel layout method. R-studio was use as a platform to calculate the network parameters plus actually calculating the neighborhood model.

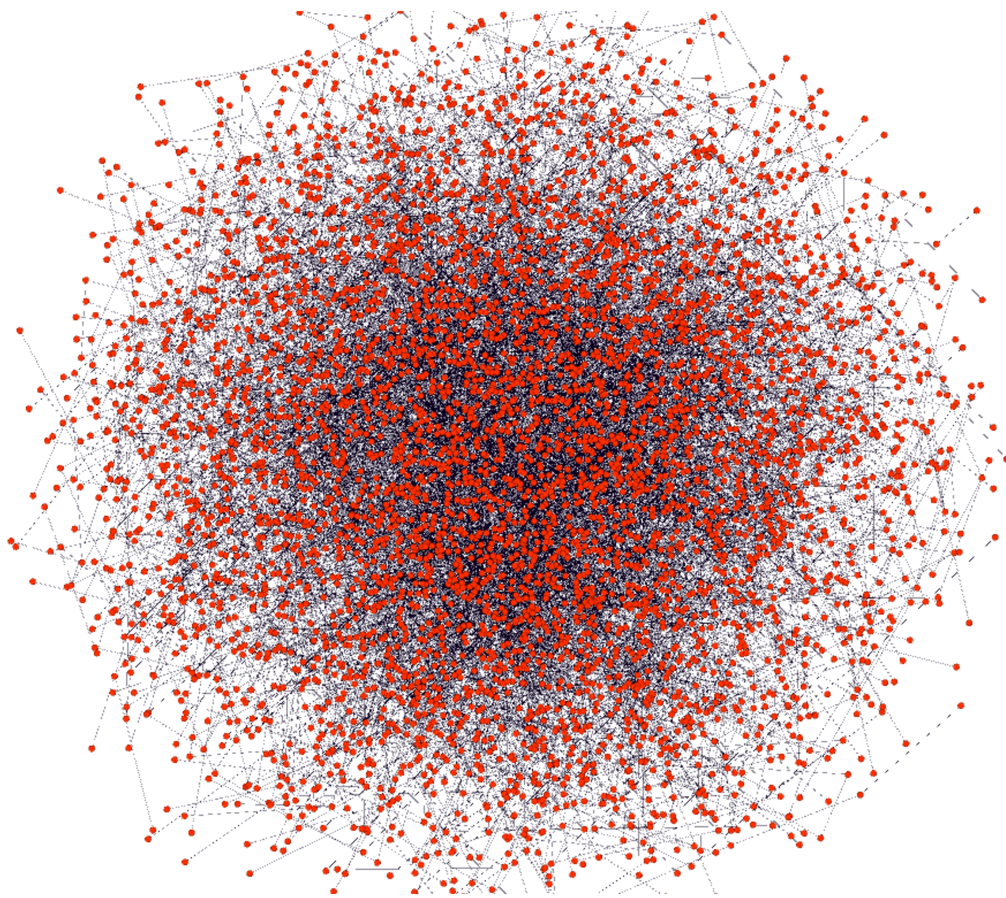


Figure 1: Visualization of the network made using Gephi. Nodes are represented in red while the edges are in black.

We can see the model of our graph in Figure 1. It is highly clustered and highly connected. Analyzing this model led us to interesting results.

If we look at the degree distribution of the network as show in Figure 2. We can see that in the center of the graph the degree distribution is more accumulated. The nodes with high degrees have a darker color (green) while the nodes which have faded color are the ones with low degree. From our basic understanding of the power system, generators are points which are highly connected to different nodes because they distribute energy over to the different substations over the transmission lines.

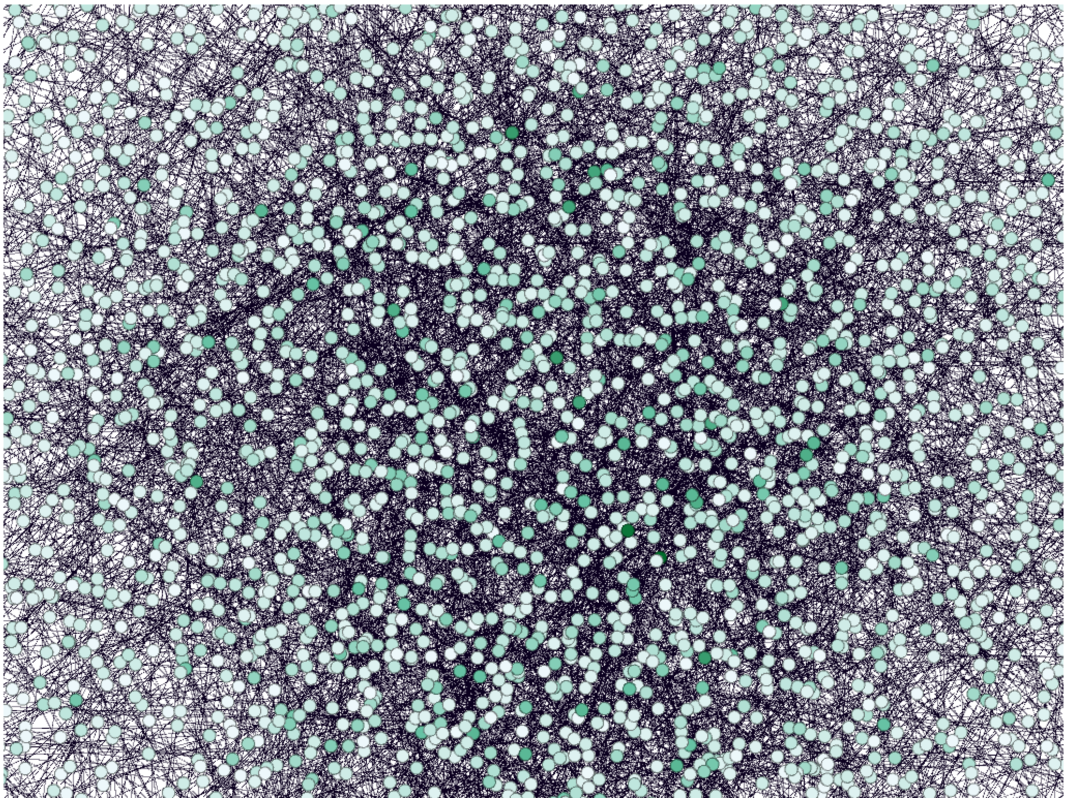


Figure 2: The degree distribution of the network

Strong and Weak Ties

One of the powerful roles that networks play is to bridge the local and the global clusters. This offers a clear explanation as to how a simple process at an individual node level can have a ripple effect over the entire network. Strong and weak ties are both relevant and important in real world networks. Let's understand this with a social network example. Are all of your relationships "strong ties?" Do you count all of your connections as good friends? Or are they colleagues who you occasionally interact with? Are they important to you at all? Should they be? A **strong tie** can be a node whom we know well. Someone who you know well. You've probably got their number on your phone. You interact with them on social networking sites. There is good 2-way conversation, and even if you don't know everything about them, you know them pretty well and information flows freely.

A **weak tie** on the other hand is a more tenuous relationship. Once a year, you may send them a Christmas message promising to be in touch more often. If you look up their number, they are surprised to hear from you. You have different interests and don't interact much. You might have kept their business card in case it comes in handy one day.

However, these weak ties are crucial in binding groups of strong ties together. They bring circles of networks into contact with each other, strengthening relationships and forming new bonds between existing relationship circles. Edge between A and B is a **bridge** if, when deleted, it would make A and B lie in 2 different components (Figure 4) and an edge is a **local bridge** if its endpoints have no friends in common that is deleting the edge would increase the distance of the endpoints to have a value more than 2 as shown in Figure 5.

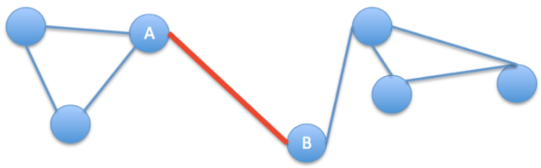


Figure 4: A Bridge

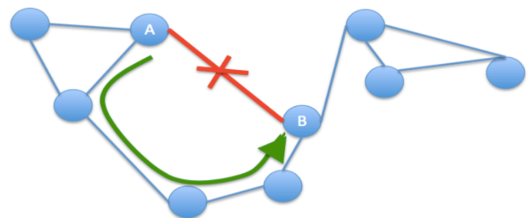


Figure 5: A Local Bridge

Local Bridges and Weak Ties

Now that we have a clear idea about the different sort of ties and links between the edges lets try establishing a connection between them based on triadic closure which states that if two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future.

According to Granovetter if a node A in a network satisfies the Strong Triadic Closure Property and is involved in at least two strong ties, then any local bridge it is involved in must be a weak tie which can be shown in the following Figure 6.

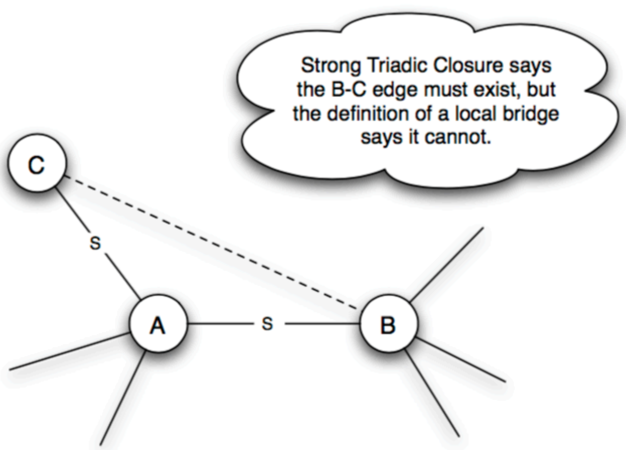


Figure 6: The figure illustrates the reason why: if the A-B edge is a strong tie, then there must also be an edge between B and C, meaning that the A-B edge cannot be a local bridge.

Granovetter's theory

According to Granovetter Networks consist of local and global components that are bridged together.

Individual nodes and links can create a ripple effect through the entire network. Different nodes act differently in diffusing information. We see the concept of weak ties and local bridges and how it takes part in diffusion. When two nodes A and B have a strong tie with a node C then it is most likely to have a weak tie between A and B. We could define the strength of the ties based on few parameters such as information shared, frequency of interaction, emotional affect, and trust.
Since the theorem is based on social interaction and our adaptation is to that of a power grid the frequency of interaction and information shared would have to be what we take into account.

Since a very small fraction of the edges in the network form local bridges, we take into consideration the nodes which are on the verge of becoming the local bridges or "almost" local bridges. We determine this by defining a term called as neighborhood overlap of an edge connecting A and B which is shown as below.

$$\text{Neighborhood overlap} = \frac{\textbf{number of nodes who are neighbors of both A and B}}{\textbf{number of nodes who are neighbors of at least one of A or B}}$$

As from the definition, in the numerator we consider all the neighbors of A and B excluding them.

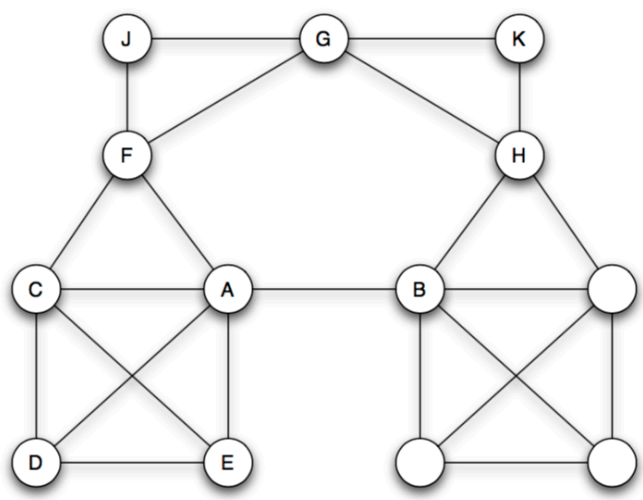


Figure 7: Show Neighborhood overlap

Let's consider this example to show how neighborhood overlap works from Figure 7. The denominator of the neighborhood overlap for AF is determined by the nodes B, C, D, E, G, and J, since these are the ones that are a neighbor of at least one of A or F. For the numerator, we can see that only C is the common neighbor of both A and F so the neighborhood overlap will give a value of 1/6.

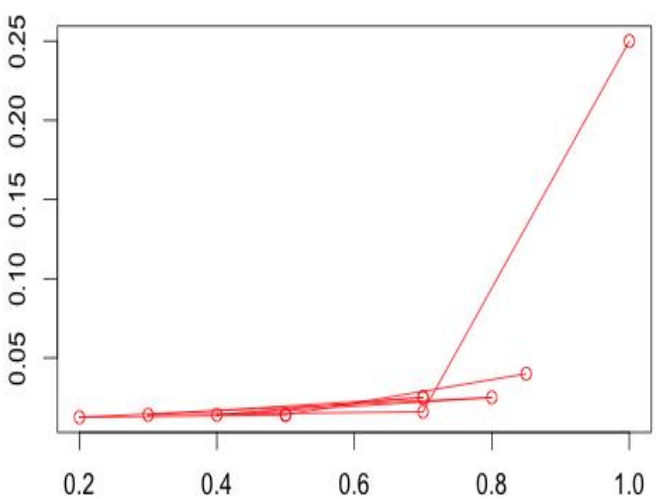
From the definition we can see that the ratio in question is 0 when the numerator value is 0 and hence we can say that the edge is a local bridge. So the notion of a local bridge is contained within this definition — local bridges are the edges of neighborhood overlap 0. Because of this we can say that the edges with very small neighborhood overlap value can be considered to be "almost" local bridges.

We apply the same to our power network model to determine the local bridges and “almost” local bridges. Since the network is huge, we cannot calculate the neighborhood overlap for all the nodes so for our convenience we sort all the nodes on the basis of edge betweenness and then calculate neighborhood overlap for 10 edges distributed evenly over the entire network. Following table shows the edge Id for which we calculate the neighborhood overlap for our model.

Edge Id	Neighborhood Overlap
1	¼
658	0/6
1974	0/7
2632	0/4
3290	0/4
3948	0/7
4606	0/4
5264	0/8
5922	0/7
6584	0/9

Since in our case we can see that for the neighborhood overlap the numerator values are 0 for most cases, we consider them to be 0.0001 so as to get our results.

Now we plot the the neighborhood overlap of edges as a function of their percentile in the sorted order of all edges by tie strength which in our case is edge betweenness which is shown in Figure 5. We can see that the fact that overlap increases with increasing tie strength is consistent with the theoretical Granovetter's predictions which is shown in Figure 6.



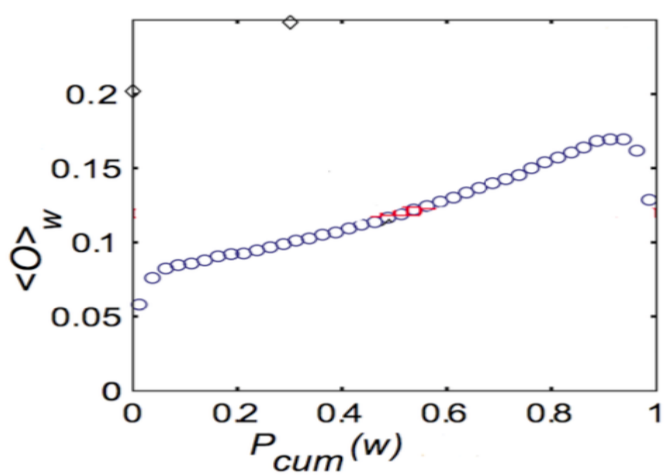


Figure 8: Neighborhood overlap vs tie strength *Figure 9 : Neighborhood overlap vs tie strength(Granovetter's prediction)*

Similar to Granovetter's predication we can see there is a linear rise in the graph as we move on the X-axis. The the nodes with high tie strength (edge betweeness in our case) have a higher neighborhood overlap value.

This picture presented is more based on a local node level, let's try exploring the same at a global level and see how the weak ties play a role in binding the communities together. We can do this by deleting the links/edges along with the nodes they are connecting to form clusters. First we eliminate the edges with the maximum tie strengths along with the nodes they are connecting to, and working down towards the edges with weak ties and they node they are connecting. For our implementation, since the graph was huge we have to eliminate around 20 strong ties to see the formation of clusters.

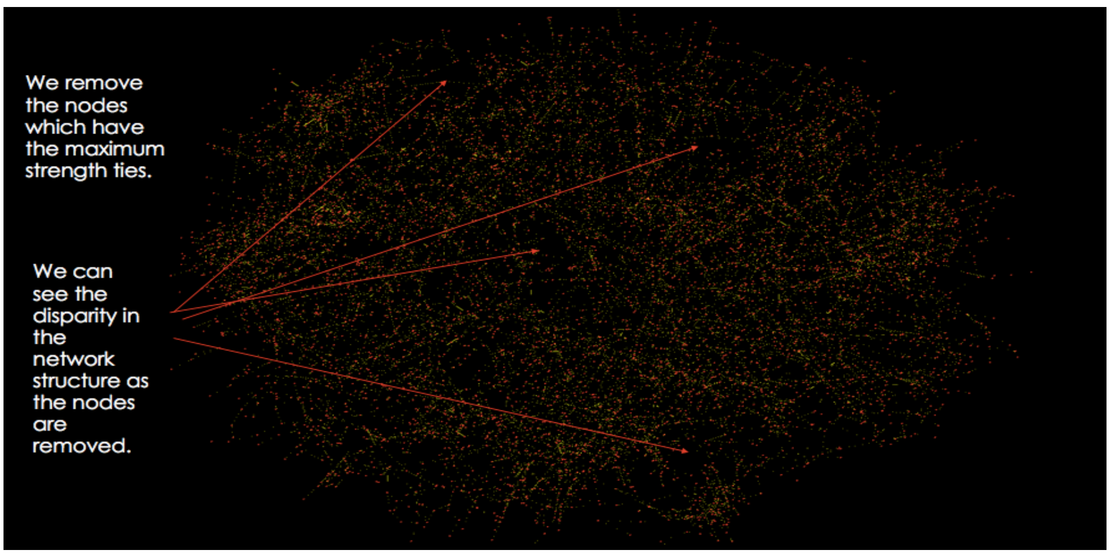


Figure 10: Our power model with all the strong ties with the connecting nodes removed.

This can be shown from the Figure 10 above. The graph showed the obvious result, it formed clusters because of the elimination of the edges and the nodes. Now let's see what we find if we start removing the weakest ties (nodes with minimum edge betweenness) and the nodes they are connected to. In this case we find that the network become clustered again once we eliminated a critical number of weak ties and the respective nodes. Another important thing that we observed was that the clustering of the graph was more rapid and it is more dispersed when compared to when the strong ties and the edges were removed. This can be seen in Figure 11 below.

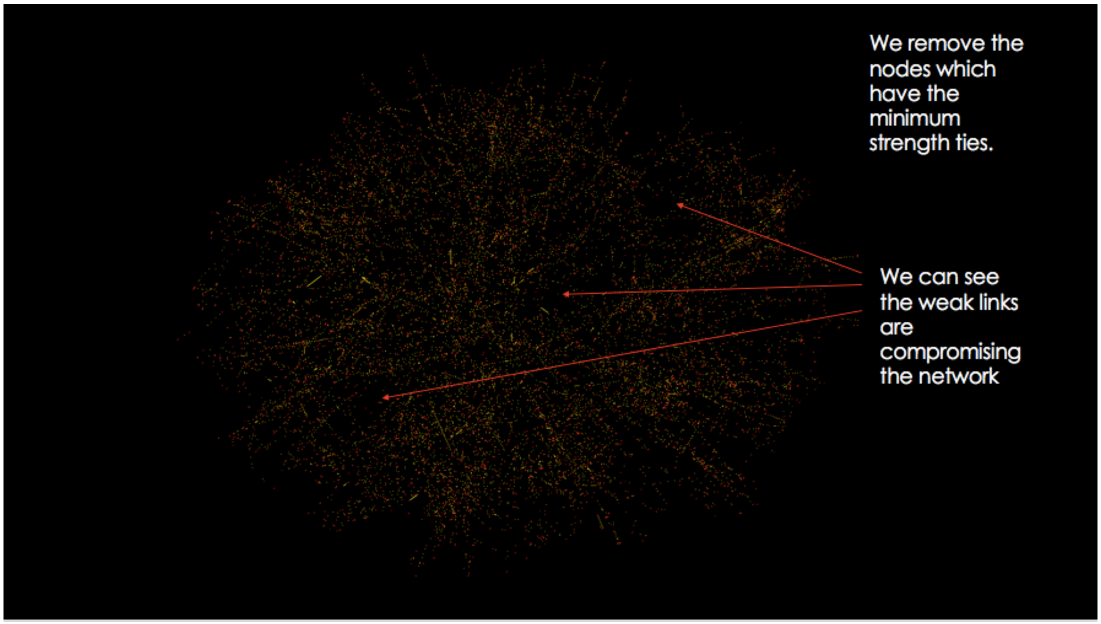


Figure 11: Our power model with all the weak ties with the connecting nodes removed

With the above observation we can say that it is consistent with a Granovetter's theory in which he remarks that the weak ties provide the more crucial connective structure for holding the communities together, and for keeping the global structure of the giant component intact.

Analysis of different parameters before and after the removal of nodes

Let's look the basic parameter like degree distribution, eccentricity distribution and size distribution of the power network. From the following graphs (Figure 12) we can see that degree distribution of the entire network is more concentrated around the values of 2 i.e. most of the nodes have a degree of 2 while there is only one node with the maximum degree as 19. Another interesting feature we notice is the gradual fall in degrees from 4-14 suggesting that there is a linear decrease in the number of transmission lines between the substations. We can say that nodes with the maximum degrees are mostly generators producing and distributing energy.

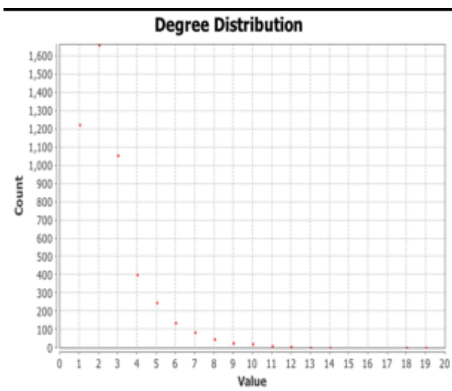


Figure 12: Actual Network

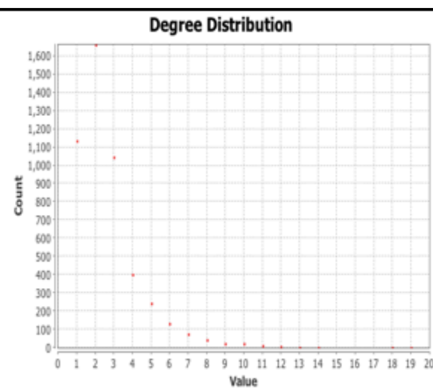


Figure 13: Network with minimum strength and connecting nodes removed

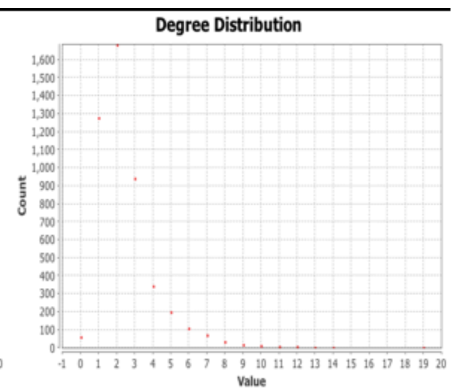


Figure 14: Network with maximum strength ties and connecting nodes removed

Even after removal of the maximum strength ties and the minimum strength ties we do not see any spiking change in the three graphs. However, we can see that the degree distribution of 18 in Figure 13 have been removed however the node with the maximum degree 19 is still present suggesting that the node with highest degree did not have the maximum edge betweenness.

Now look at the eccentricity distribution of the graph. Eccentricity distribution is nothing but the length of the longest path starting at a node. From the Figure 17, we can see that once the nodes with maximum edge betweenness are removed, the graph gets clustered towards the right hand side suggesting the increase the in the length of the longest path starting at that node. We can see that the length is above 80 for most of the nodes in the graph. However, there is not much of a change in Figure 16 where the nodes with minimum edge betweeness are removed.

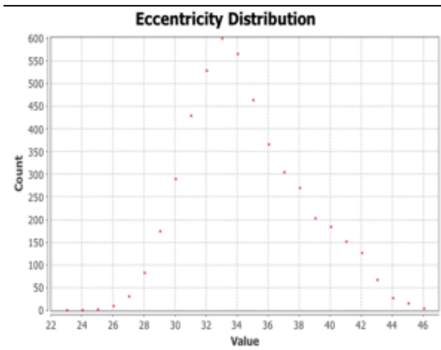


Figure 15: Actual Network

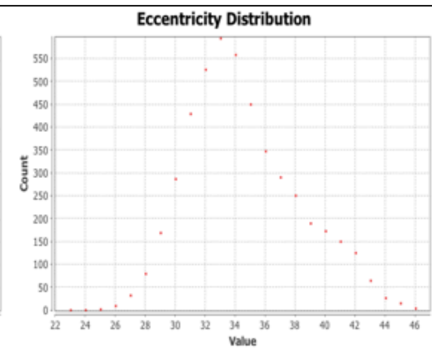


Figure 16: Network with minimum strength and connecting nodes removed

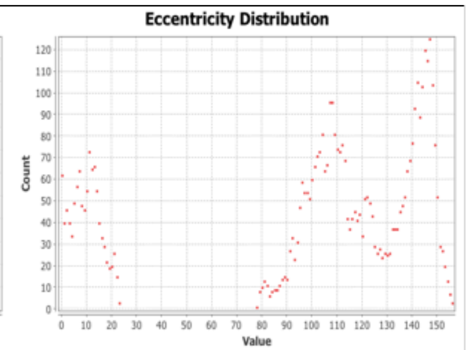


Figure 17: Network with maximum strength ties and connecting nodes removed

The size distribution of a graph is the number of nodes in different clusters. The modularity class here refers to the number of clusters in the network. We can see from Figure 19 that once the nodes with maximum edge betweenness are removed the number of clusters increase and the size of the clusters is more or less close to 0 suggesting that the individual nodes form individual clusters. However, if we look at Figure 17, we don’t see much of a change from the initial Figure 18 which is the actual network.

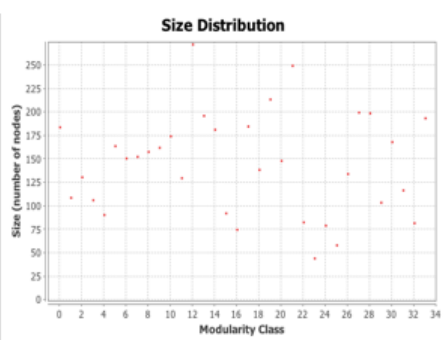


Figure 18: Actual Network

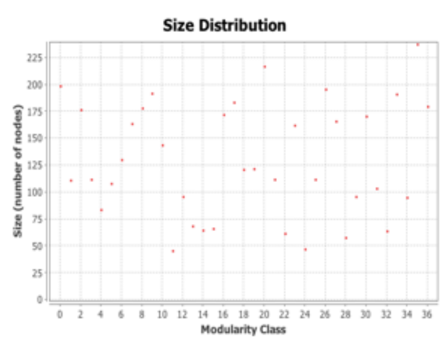


Figure 19: Network with minimum strength and connecting nodes removed

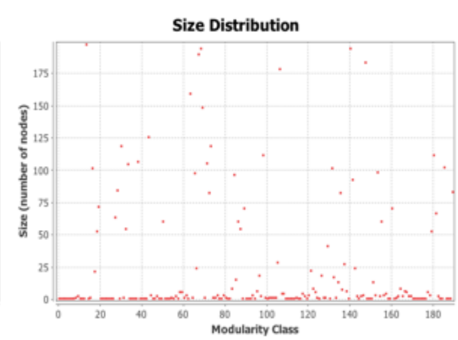


Figure 20: Network with maximum strength ties and connecting nodes removed

Comparison with the Random network and Small world network

The small-world structure of networks are based on the Watts-Strogatz model. It is characteristic of short average path lengths and high clustering. The small-world network had a mean degree k=4 and p=0.402. The random network is a Erdos-Renyi model with the probability p=0.00054. The parameters that were used to simulate all the network models are based on a research on US Power Grid Network Analysis where they have calculated the parameters such that all the models are comparable.

	Power Grid	Random	Small World
Number of nodes	4941	4941	4941
Edges	6594	6594	9882
Model Parameters		p=0.00054	k=4, p=0.402
Mean degree(k)	2.67	2.67	4
Average Path Length(l)	18.99	8.66	7.3
Clustering Coefficient	0.107	0.00054	0.107

Fig.21 Network models and their parameters

We compare how the betweenness, degree distribution, degree are in the different models. Theoretically the nature of behavior of small-world networks and power grid network should be similar we see results that are contradicting of that fact. From Fig.22 of betweenness we observe that in the power model we have an even distribution in the range of values, in small-world networks and random networks the value is close to 0 and the values are all comparable. This will imply that the shortest length paths through all nodes are almost similar.

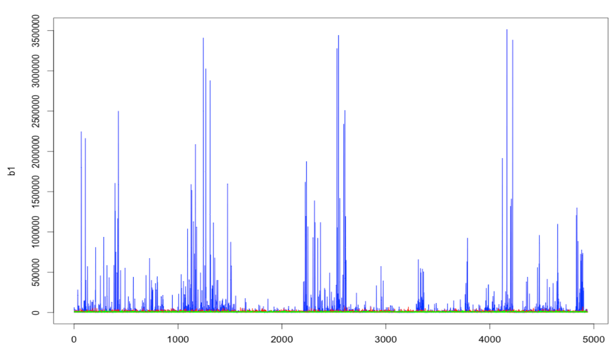


Fig.22 Betweeness Comparison

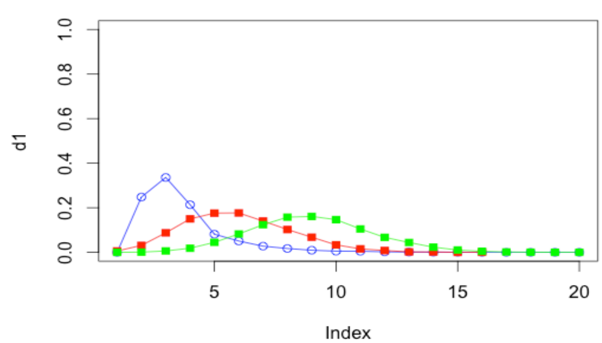


Fig.23 Degree Distribution Comparison

Degree distribution we observe that none of the models are comparable to our data. The power network, most nodes have a degree distribution of 4 where it peaks after which we observe a steep drop in the value. Random graphs and small world networks on the other hand have peak values of 6 and 9 after which there's a gradual decrease in value. What we may deduce from this is the power network is highly connected that is 3-4 out of 10 nodes will have degree of 4 versus the other two networks where the peak values are at 6-9 but the probability values are much lower meaning fewer nodes will have higher connectivity. We expect the small-world network and the power grid to be similar but recreating the exact same parameters of both are difficult which leads to contradictions from the theory. There is also the factor that every time we simulate and create the random models they are created differently which means that deducing results from just one model won't be consistent.

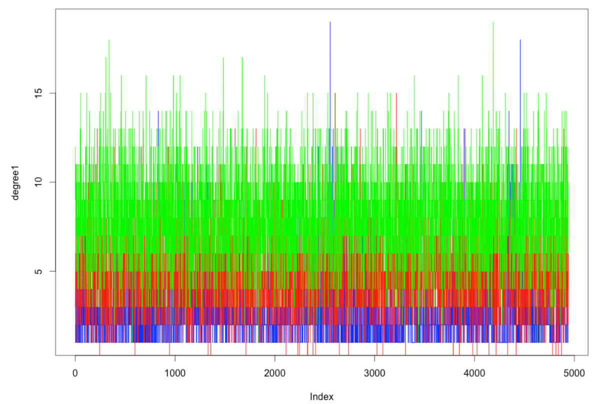


Fig.24 Degree Comparison

Degree comparison: The power network is observed to have consistently low degrees but none of them are 0. In Fig.24 we can see that the node 2553 of our network has the maximum degree of 19. The small world network has nodes which have degrees quite consistent over all the nodes and is quite high.

The random network however also has consistent degrees but the range is much less compared to the small-world network. However many nodes have degree 0 suggesting that is disconnected. What we may infer as the relation between degree and betweenness is that the low degree nodes which may be transmission substations will have high betweenness due to the fact that it will be on the route between generator and distribution substations. Similarly, generators would have a have degree but low betweenness as they would be nodes with high neighborhood overlap suggesting that they will have maximize the cascading effect on the network.

Related Work

Collective dynamics of 'small-world' networks

This is one of the earliest works in the domain where they study the effect of diffusion on small-word networks. Data was collected on many real world networks such as the neural network of the worm Caenorhabditis, the western power grid of the United States and the collaboration graph of film actors. These are all models of dynamic systems. They look at how small-world connectivity is functionally significant in dynamic systems. They modified the test case to simulate spread of infectious diseases. They have defined parameters such as critical infectiousness which we can translate to a node with high betweenness i.e. a node with high connectivity will be most likely to contract diseases to all other nodes. They talk about it in more detail on the speed and extent of spread as well, which would mean they take into account transmission speeds and the degree to which a certain node is infected and to what degree it is contractible. They note that small-world architecture has the distinct combination of high clustering with short path lengths which is very essential in simulating dynamic networks.

Networks, Crowds, and Markets: Reasoning about a Highly Connected World

A chapter in this book goes over the concept of Weak ties and Local bridges. We learn about the basic concepts that are essential in the understanding of Granovetter's theorem such as local bridges, strength of ties, triadic closure. Strength of bridges are defined based on few factors such as frequency of interaction, trust, information shared. These values will vary based on the network. We have defined more on this in our section on weak ties and local bridges.

Madeleine Bairey & Shanté Stowell , US Power Grid Network Analysis

This experiment involves using the network dataset of the Western US Power Grid. The data was visualized using Gephi. They simulate and compare to see if they can recreate a network similar to the power grid by modifying parameters. They have models of Erdos-Renyi for a random graph, Watts-Strogatz for small world and scale free model. They looked for similarities between models in degree distribution and betweenness, they observed a lot of contradictions from their theoretical predictions to the actual observations made. The betweenness is also compared to degree, the relationship between the two is also observed. The effect of removal of nodes on the network showed that if more than 1% of nodes which had high betweenness were removed, it leads to clustering of the network. To conclude, node removal has a significant effect on the network but the extent is also dependent on the node and it's characteristics.

References

D. J. Watts and S. H. Strogatz, "Collective dynamics of `small-world' networks", Nature 393, 440-442 (1998)

Dr. Cecilia Mascolo , Lecture 2, Small world and weak ties

Steve Borgatti ,Strength of Weak Ties, Structural Holes, Closure and Small Worlds ,

David Easley and Jon Kleinberg, Networks, Crowds, and Markets: Reasoning about a Highly Connected World.

Madeleine Bairey & Shanté Stowell , US Power Grid Network Analysis