US Power Grid Network Analysis

Madeleine Bairey & Shanté Stowell

May 14, 2014

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. In this paper, we create a network model in which nodes represent substations and links represent high voltage transmission lines. We examine a variety of potential models to simply represent or recreate the network. We also examine key attributes of the system, including betweenness and degree distribution. Finally, we examine the effects on the system of removal of key nodes with either high degree or high betweenness. We conclude 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 Power Grid

The United States power grid is an impressive piece of infrastructure. 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 with many other large scale infrastructures, the power grid serves users who may only notice its presence when the system fails in some way. One of the main issues with the system is that failures or blackouts can cause very large scale problems across the country. However, this is not always the case, as some failures can be comparatively localized.

To better understand the effects of substation failure, 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.

Although the nodes lack geographical coordinates and the edges are not weighted, we are able to analyze the properties of the network in a fair amount of detail. We analyze some common parameters, such as average degree, shortest path length, and clustering coefficient. In addition, we analyze the betweenness and the relationship between betweenness and degree. Using this information, we determine the effects of removing nodes of high degree or high betweenness, simulating substation failure. Finally, we comment on possible further research regarding this network.

Methods and Results

Visualization

We began by creating a visualization of the network using Gephi, a network visualization tool. The nodes were colored by modularity class (which creates groups of nodes based on connectedness) and sized by degree. Then the nodes were positioned using Yifan Hu's Multilevel layout method. The results are shown in figure 1.

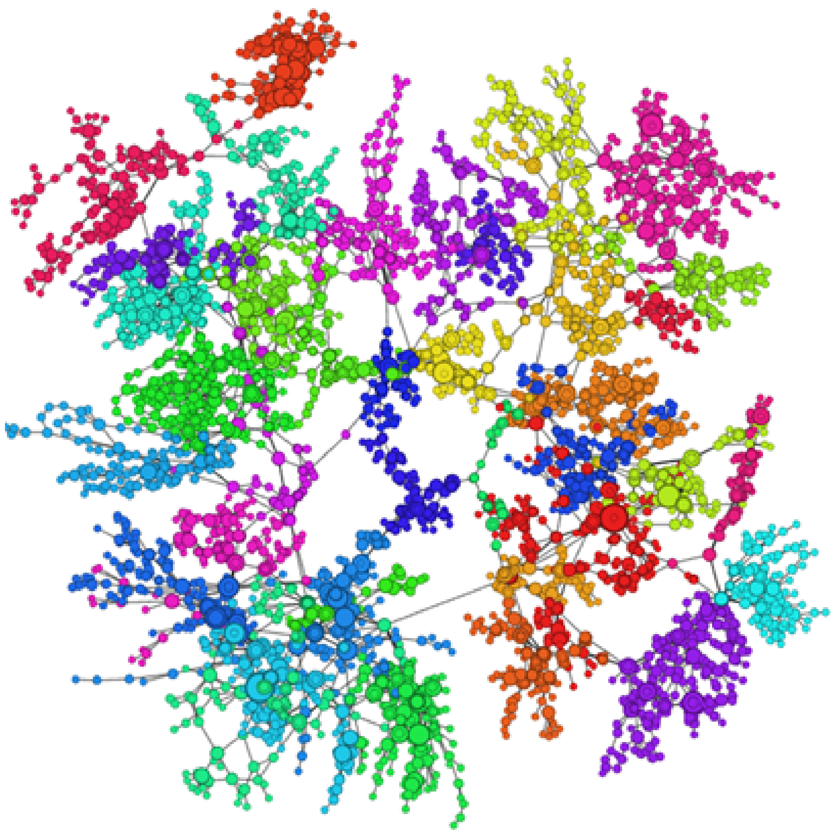

Figure 1: Visualization of the network made using Gephi. Node degree is represented by size and modularity class is represented by color.

Inspecting this graph shows an interesting pattern of node connections. As seen in figure 2, many nodes are connected in strings. These strings of nodes consist of nodes which are only connected to about two other nodes, and are organized linearly. Some of these strings form loops, while others terminate at one node, and several strings may be interconnected.

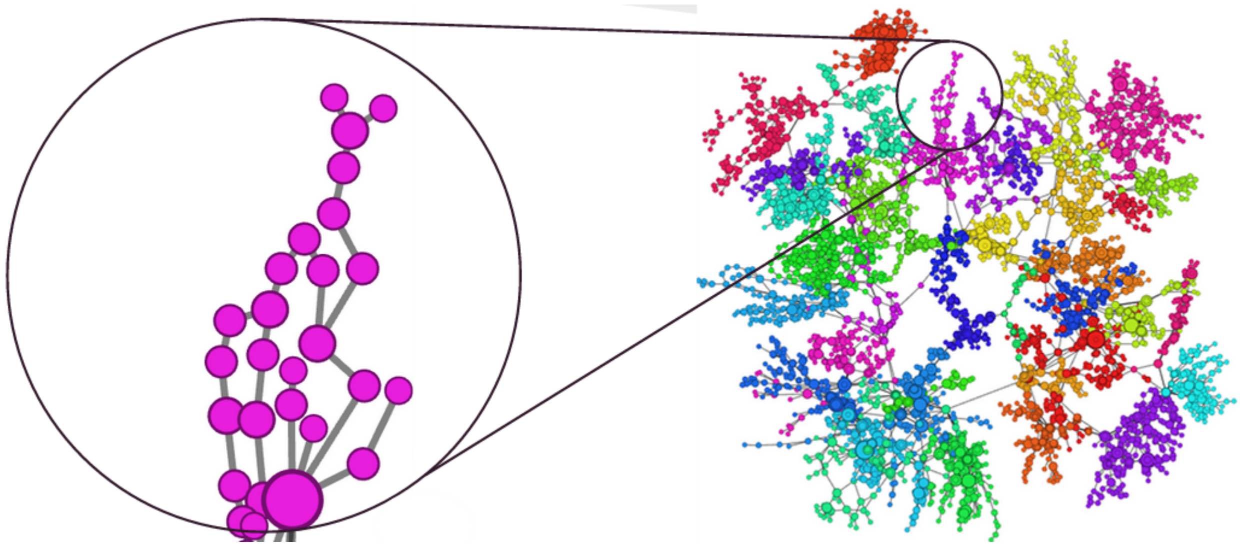

Figure 2: Modular segment of Gephi model. Note that many nodes are connected in "strings"

The structure found above can be simply explained through the distinction of three types of nodes: generator, transmission, and distribution substations. Generators provide the source of electricity and are connected to a large number of transmission substations, therefore they commonly have a high degree. Transmission substations most commonly connecting to one to three other substations, so they have a lower degree. This type of substation is most likely responsible for the near linear portions of the Gephi model. Distribution substations have a single high voltage transmission line providing power, which is then distributed to communities through low voltage lines that are not considered in this network. These substations are the nodes scattered around the outskirts of the network with degree one.

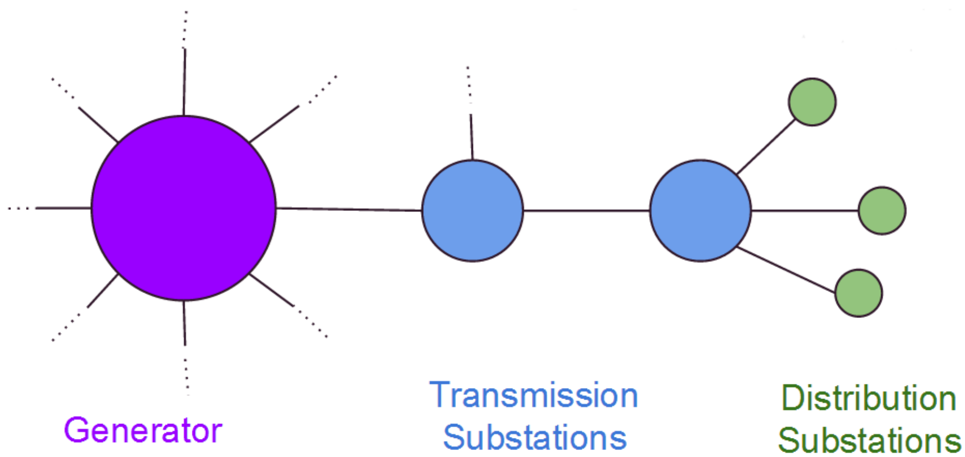

Figure 3: Generator, transmission, and distribution substation distinction. Generators are likely to be connected to many other nodes, and therefore high degree nodes may be generator nodes. Transmission substations connect generators to distribution stations, and are therefore unlikely to have a degree greater than 2 or 3. At distribution substations, power is diverted to low voltage lines, so these nodes may have a degree of only one.

Comparison to models

Network analysis starts through computation of common network parameters, such as average degree, average shortest path length, and clustering coefficient. Such parameters are also calculated for commonly used models in an attempt to determine which model best represents our network. However, as shown in Table 1, the parameters are not well matched. In addition to the parameters, we also compare degree distributions to find a match.

	Power Grid	**Random**	**Small World**	**Scale Free**
Nodes	4941	4941	4941	4941
Links	6594	6594	9882	4941
Model parameters	-	p=0.00054	k= 4, p=.402	m=1, n=4941
<k>	2.67	2.67	4	2
<l>	18.99	8.66	7.3	8.86
<CC>	0.107	0.00054	0.107	0

Table 1: Parameters for the power network and various network models

Degree Distribution

We first compared the degree distribution of our network to that of several simple network models, as seen in figure 4. However, none of these models are a good fit for the data.

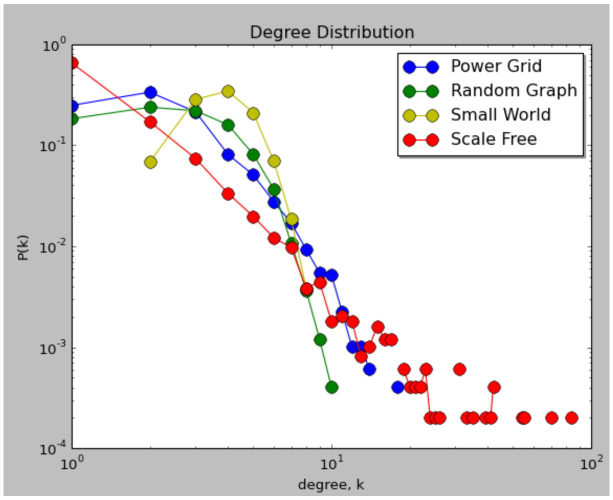

Figure 4: Degree distributions of the power free network and several model networks. None of the model networks are good fits for the data.

We then went on to compare the degree distribution with several analytic solutions, including exponential, power law, and power law with cutoff distributions. By using Python's optimize.curvefit function, we fit each analytic solution to the data. However, none of these fits worked well for the data. Ultimately, we were able to obtain an exponential fit for the data for nodes with degree greater than 3. The data follows an exponential fit, where $P(k)=0.65e^{-k/1.9}$. k is the degree and P(k) is the probability that a node will have that degree. This fit is shown in figure 5.

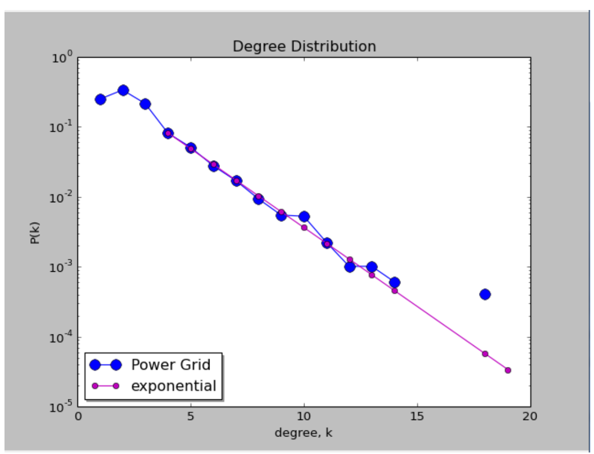

Figure 5: Degree Distribution of the power grid network, and an exponential fit for this data. The fit was obtained using Python's optimize.curvefit function and only holds for degrees greater than 3. The data is modeled by $P(k)=0.65e^{-k/1.9}$

We believe that to truly model the degree distribution of this data, a combination of two models would be required. This may be due to the nature of the network, which contains several well-connected portions of the graph which are connected together by strings of nodes. It may be that fitting one degree distribution model to the well connected portions and combining this with a degree distribution model for the node strings would result in a model that well represents the distribution of both high and low degrees.

Betweenness

In addition to the more common parameters used in network analysis, we also determine the betweenness distribution, with the belief that it will provide additional information about the network. Betweenness, or betweenness centrality, measures the centrality of a node in a network by calculating the number of shortest paths from all vertices to all others that pass through that node. The distribution shown in Figure 6 demonstrates the low frequency with which a given nodes has high betweenness, and the higher frequency with which a given node has low betweenness. The distribution complements the modular nature of the Gephi model, where most nodes connect in groups with fewer connections to other groups.

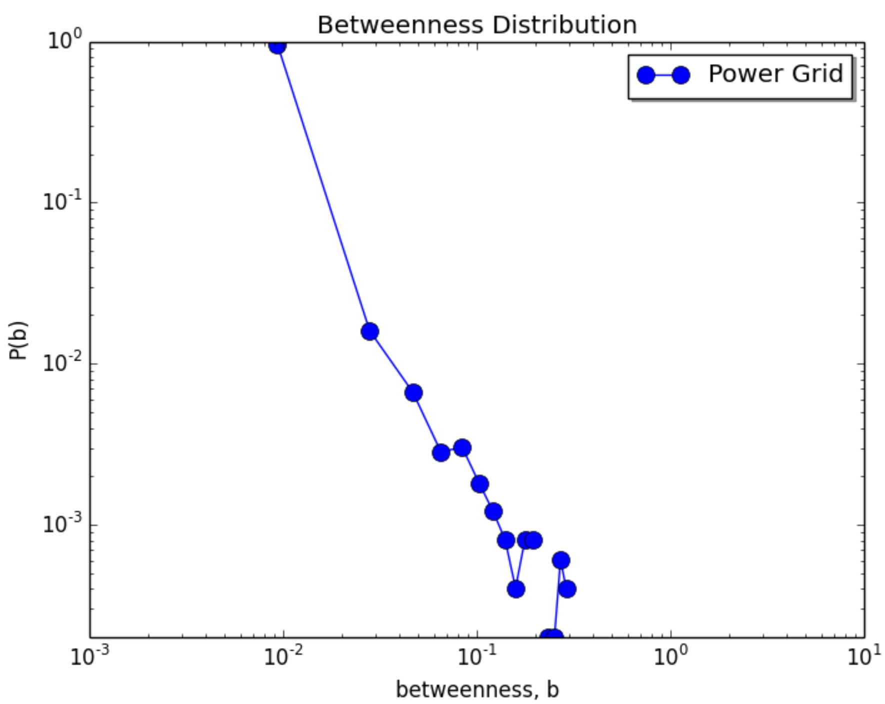

Figure 6: Betweenness distribution plotted on a log-log scale. This shows that there are a large number of nodes with low betweenness and very few nodes with high betweenness.

The correlation between degree and betweenness is also considered. The results shown in Figure 7 show no correlation between the two parameters.

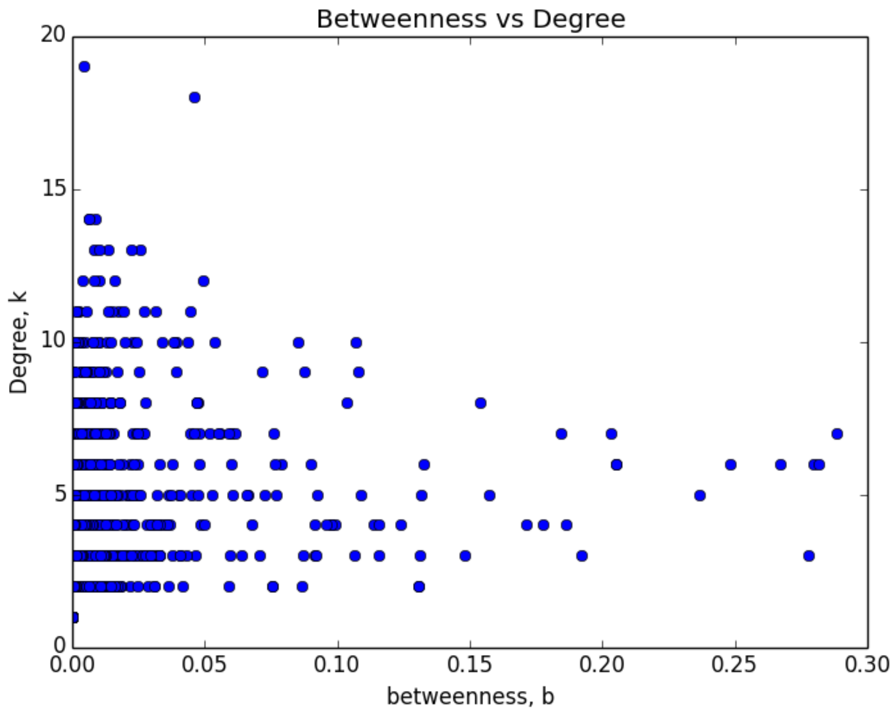

Figure 7: Plot of betweenness vs. degree. This shows that there is little to no correlation between degree and betweenness.

The lack of correlation between the two parameters fits well with what can be seen from the visualization of the model created in Gephi. Low degree nodes, such as transmission substations, may have high betweenness due to their location between the generator and distribution substations. By contrast, a high degree node such as a generator may have low betweenness, due to its location within a group and lack of connections to other such groups.

Analysis of node removal

In order to simulate the accidental or intentional failure of substations, we run two scenarios removing nodes from the network. This is done to get an idea of the effects of specific node failures across the scale of the whole network. The Pagani-Aeillo paper references numerous analyses of the power grid network, all of which confirm that random node removal usually has minimal impact on the network. By contrast, directed attacks on specific nodes may have larger impacts.

Due to the lack of correlation between degree and betweenness, the two methods of node removal involve removing 10 nodes first with highest degree and then with highest betweenness.

When the 10 nodes with the highest degree are removed, the graph is split into 33 connected components, where 51 total nodes are disconnected from the main graph. This is likely caused because some node clusters were only connected to the rest of the graph through a high degree node. This kind

of targeting removed over 1% of the nodes from the network. While this may seem trivial, a removal of 1% of users from the US power grid system could have serious detrimental effects.
Removing nodes with high degree had little effect on average shortest path length of the network.

In contrast, removing the 10 nodes with the highest betweenness split the graph into only 2 connected components, disconnecting only a single node from the main graph. This is likely due to the fact that nodes with high betweenness connect large groups of nodes to one another and removing just 10 nodes is not sufficient to entirely disconnect these large groups, so very few nodes were disconnected from the larger network.

Removing nodes with high betweenness had a large effect on the average shortest path length. Before removing the nodes, the path length was 19. Without the high betweenness nodes, the path length became 25, a 32% increase. This increase in average shortest path length has the potential to be more detrimental to the overall system than disconnection of nodes. While average shortest path length is not representative of the number of nodes power must travel through to get from a generator to a user, it can be used as an indicator of how this actual travel distance will change. When there is higher demand through a connection than the connection can supply, an automatic shutoff can occur to prevent system damage. Increasing the shortest path length means that more power will have to travel through more connections, which could result in this kind of automatic shut-off, resulting in a large number of nodes becoming disconnected from the power system. Our network does not model the capacity of the links, and therefore cannot be used to determine exactly what effect this would have on the system

Conclusions

In conclusion, our data is not easily modeled by any of the simple models used here. To better model this system, some form of compound model may be required. This kind of analysis could be performed by modeling each sub-group of nodes as a network, then modeling the connections between groups as a separate network. This kind of analysis is more complicated, but could result in better understanding of how this network might change over time or better tools for analyzing larger networks of a similar kind.

From our analysis of node removal, it is apparent that removing a small number of nodes can have a large impact on the overall system. Removing nodes with high degree results in a loss of connectivity of a portion of nodes within the network, while removing nodes with high betweenness results in a significantly larger average shortest path length. However, determining how a loss in connectivity or an increase in shortest path length would affect the overall system would require more complete knowledge of the kinds of stations each node represents.

Finally, there is still a fair amount of future analysis that can be performed to better understand the power grid network. More data, such as the capacity of transmission lines or weight of edges, or geographical locations of nodes may provide a more detailed and accurate presentation of the network. Additionally, experimenting with number and types of nodes removed may have varying levels of impact.