# Power Grid Infrastructure Security

**David Frankel**

Washington University in St. Louis, 63130

February 5, 2020

**Abstract**

Electrical power is a crucial resource within the United States' infrastructure system. However, this network of generators, transformers, and substations is vulnerable to attack. This paper explores the extent of this vulnerability by simulating attacks and their outcomes.

***Keywords*— resilience, weakest link, centrality**

## 1 Introduction

As power becomes an increasingly critical resource, its delivery has impact on countless lives. Most recently, crippled power infrastructure in California led to a hospital forced to decide between losing vaccines or electronic health records[1]. While this hospital anticipated the blackout, cyber-attacks on power grids do not offer the same privilege. In 2015, Ukraine faced an attack on its power grid that led to power outages for more than 230,000 residents. Control systems in the United States are even less protected than those in Ukraine[2]. This September, the U.S. faced its first large-scale power grid intrusion in the western United States – and it was caused by an automated script scanning for vulnerabilities (not even directly aimed at harming the grid)[3].

While a cybersecurity defense focus can help prevent attacks from succeeding, it is also important to increase the resilience of power grids when generators, transformers, or substations go offline. By analyzing network data on the U.S. grid[4] using Gephi[5] and NetworkX[6], we can investigate the importance of certain nodes and the general dependencies of the network in order to understand how structure may influence the grid's ability to function while under attack.

This project will explore the importance of nodes to network stability and resilience through simulated attacks against the network and their effective results. Since power cables have a limited amperage, systems are built with a limited buffer for power draw variance (ex. more electricity needed during the day for air conditioning). However, as nodes disappear and they no longer link different sections of the network, other nodes must rely on alternative routes for power. In addition, it is important to realize that attacks target multiple nodes simultaneously (as was the case in Ukraine). As such, we will analyze the risk and damage created by attacks on multiple nodes. Overall, the power grid is shown to be vulnerable to attacks with limited scope when targeted nodes are chosen for their impact on shortest paths.

## 2 Data

We will be using a dataset curated by the Koblenz Network Collection[4] and originally cited in an article by Watts and Strogatz[7]. This undirected network contains unlabeled nodes that represent three structures in the power grid: generators, transformers, or substations. An edge represents a power supply line. With regards to these nodes, a generator (otherwise known as a power plant) produces high voltage three-phase power. A transmission substation then converts the voltage to extremely high levels for efficient long-distance

transmission. A power substation then generally steps down transmission voltages and may split the power in multiple directions. Transformers continue to lower the voltage for home use. Since the nodes are not labeled, we face limitations in the conclusions we can draw from network measures and edges. We are also limited by the fact that the edges do not have weights relative to their ampere capacity (ampacity) rating. In addition, this dataset only covers the power grid of the Western states of the U.S. and is dated to the year 1998.

The basic properties of the graph are as follows: the number of nodes $n = 4941$, the number of edges $m = 6594$, the average clustering coefficient is approximately 0.10, the average shortest path length $\bar{l} = 18.989$, and the network is one giant connected component.

# 3 Results

First, we examine the degree distribution of the network. From Figure 1, we notice that a large number of nodes have a degree of one or two. This speaks to the large volume of node chains, which we can also see an example of highlighted in Figure 2b. From Figure 1, we also learn that the plot with a log-log scale shows a somewhat linear trend, implying that this is scale-free network following a power law distribution. The exponent can be calculated to be a value of $\alpha = 3.84451$. This is in line with previous analysis[7].

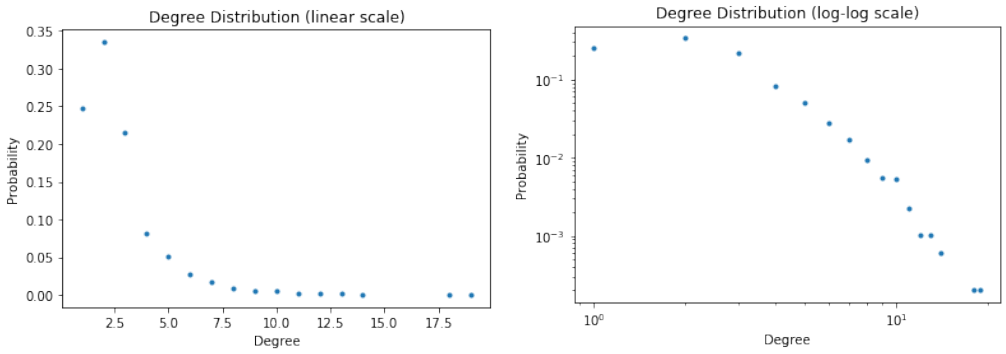



Figure 1

In order to draw Figure 2, we use a combination of different tuned Gephi algorithms in order to accentuate possibly important structures. After spreading the graph out using ForceAtlas2[8] with increased gravity to prevent over-spreading, we then use Yifan Hu[9] to provide more proportional expansion energy to small clusters. These algorithms have the benefit of providing efficient expansion that does not settle on a specific local minimum and makes further changes to the network apparent.

## 3.1 Determining Nodes Critical to Infrastructure

In order to determine which nodes to attack, we must decide on measures for selecting nodes and measures of the damage inflicted on the network. The most intuitive method for finding critical nodes is calculating their centrality. We tested five calculation methods: betweenness centrality, degree centrality, eigenvector centrality, closeness centrality, and information centrality (current flow closeness centrality)[11]. When comparing these different calculation methods, we define damage as an increase in the average shortest path length $\bar{l} = \sum_{s,t \in V} \frac{d(s,t)}{n(n-1)}$ where $V$ is the set of nodes and $d(s,t)$ is the shortest path from $s$ to $t$. As previously described, the limited ampacity of wires means that increasing load throughout the network causes systemic problems. In addition, transformers and substations have limits to the quantity of electrical current they can shift safely on demand. Another focus of the damage is causing network havoc on the global scale — we hope not to isolate a small group of nodes, but rather to increase the load across the network in such a way that service could possibly be interrupted for all those in this grid. The average shortest path length $\bar{l}$ measure fulfills all these needs.

(a) Working layout of network. All further network figures use these node positions as their starting point unless otherwise specified.

(b) Example small chain highlighted. Many more exist throughout the network, both between and within communities.

(c) Preliminary communities colored by modularity class and found with Louvain[10] method at resolution $r = 40.0$.

Figure 2

Through an approach to select nodes that were repeated between the top 10 of each centrality measure, we found a max $\bar{l}$ increase of 1.055, but average increase $\overline{\Delta\bar{l}} = 0.448$. While this is a small increase, we can choose our centrality measures more carefully. Of those measures experimented with, only betweenness, centrality, and information centrality account for geodesics, so they will perform better when using $\overline{\Delta\bar{l}}$ as a damage measure. Betweenness had the best performance, with $\overline{\Delta\bar{l}} = 0.863$ and $\max(\Delta\bar{l}) = 2.087$. While this nearly doubles the increase, attacks have the potential to be much more devastating to the network. With progressive removal of the nodes found to have the highest betweenness centrality, we can simulate an attack more similar to what nation-state actors may levy against other nations (as was the case in Ukraine). While the nodes were ordered in a specific manner here, this attack is based on a variable order, and as such nodes are chosen based on the their betweenness centrality in the original network.

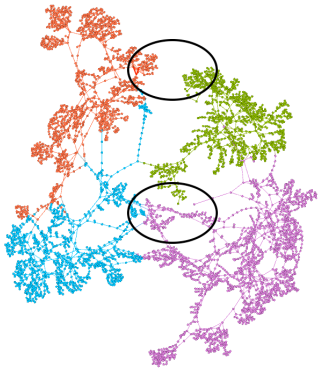


Figure 3: Communities colored with same method as Figure 2c. Easily visible missing edges circled.

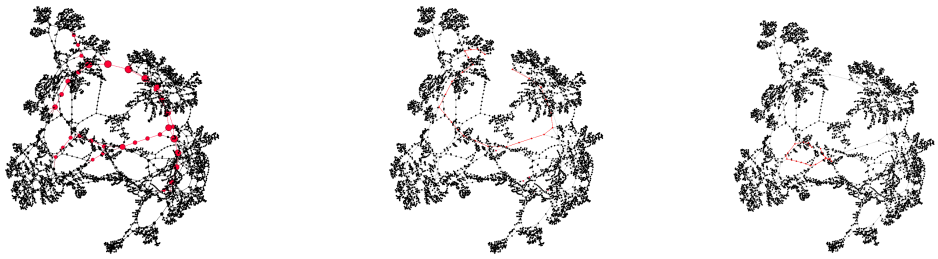
| Nodes Removed | $\bar{l}$ |
|---|---|
| 1 | 21.07636542 |
| 2 | 21.150084 |
| 3 | 21.20147787 |
| 4 | 21.39200252 |
| 5 | 23.68979521 |
| 6 | 24.22299476 |

Table 1: Multi-node Attack Results

In Figure 3, we can see several changes from the original network, with edges originally spanning between clusters now missing. There is a large $\overline{\Delta\bar{l}} = 5.233$, implying an increase in the load across the network. Without more detailed network data, we cannot examine the exact impact of this increase, but network strain is apparent. Figure 3 also demonstrates a shift in community organization due to the node removal. If the previous communities colored in Figure 2c are assumed to be the ground truth for geographic proximity and locality, then we can see that there will also be significant increases in physical distances that electricity will have to travel. This also amplifies the amount of electrical loss during transmission.

### 3.2 Modified Clustering for Node Selection

Another method to select candidates for removal depends on the importance of their bridges. This can act as another approach for selecting critical nodes. In this process, we will also be recomputing edge betweenness each time. Using an adapted form of the Girvan–Newman algorithm, we can remove the endpoints of each important edge. As this is a clustering algorithm, the process can also be used to create separations between groups of nodes. If the nodes rely on each other for power, then the $\overline{\Delta \bar{l}}$ will increase and indicate the network damage. Local bridges are critical components in the networks, since by definition, their removal indicates large increases in the shortest path lengths of their neighbors. The edges themselves cannot be directly removed as cyberattacks cannot easily target the power cables themselves. This should also help eliminate some of the problems in the previous approach. For example, using betweenness values without re-calculation can result in picking multiple nodes from the same loop (as shown in Figure 4a). In this modified approach, new loops are broken each time. In Figures 4b and 4c, the red loops were generated by highlighting the shortest path between the neighbors of the removed nodes. As shown in the figures, these are long loops that do not have paths across that could shorten path lengths (essentially, nodes in these loops follow the patterns of a Watts-Strogatz model before any rewiring has occurred). This highlights that a possible mitigation to these attacks is to find these rings and create new wiring across them.



(a) Chain of high betweenness nodes (without re-calculation) highlighted in red and sized proportional to value.

(b) First loop of nodes disrupted by modified Girvan-Newman in red.

(c) Subsequent (smaller) loop of nodes disrupted by modified Girvan-Newman in red.

Figure 4

However, the ForceAtlas2 visualization scheme also illustrates a flaw in the previous approaches. In the center of the ring of Figure 4a, there is node with a chain of three spokes. Despite, according to this visualization, being an important node to connecting different parts of the network, the chains inflate the path length counting such that reaching the center node of the three spokes is seen as expensive (and the center node is not ever marked to be attacked). This is a limitation of the under-specific data and the approach we are using to find important nodes. While this problem needs further investigation, it does also indicate that we are somewhat properly focusing on maximizing damage to the network and not a specific off-shooting chain of nodes.

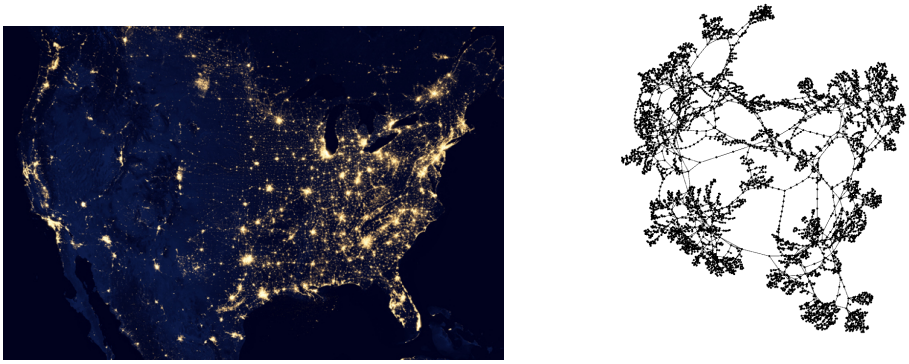
## 4 Conclusion

Throughout this project, we have seen the ways that shortest paths can provide insights into the ways we can damage networks and subject them to high strain. Considering the limited number of damaged nodes required to cause severe average path length increases, the U.S. power grid is vulnerable to attack. In addition to enhanced cybersecurity, grid-based mitigation strategies might want to be considered during future network expansion.

The GitHub repo containing the NetworkX-only code is here: `https://github.com/frankeld/power_grid`.

## 5 Future Work (Extensions)

The simplified nature of the data in this set prevents extended analysis on the specific components of the U.S. grid that may be at risk. However, it may be interesting to examine other ways that data about the power grid may be revealed. As shown in Figure 5, the initial working layout of the graph *for Western*



(a) Composite satellite view of U.S. at night[12]

(b) Initial working layout of network, rotated for comparison.

Figure 5

*states* is quite similar to the entire U.S. at night (note that the lights are possibly correlated with power infrastructure and power draw). Not only does this hint at how small-world growth patterns are already visible locally, but also at how the geography of the grid is relevant. Electric transmission efficiency is affected by the distance between endpoints, another key aspect of data missing from this set. Armed with this data, better mitigations could be designed, or we could be more vulnerable to better targeted attacks.

## References


[1] N. Wetsman, "California's blackouts reveal health care's fragile power system," *The Verge*, Oct 2019.

[2] K. Zetter, "Inside the cunning, unprecedented hack of Ukraine's power grid," *Wired*, Mar 2016.

[3] B. Sobczak, "Report reveals play-by-play of first U.S. grid cyberattack," *E&E News*, Sep 2019.

[4] "US power grid network dataset – KONECT," Sept. 2016.

[5] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," 2009.

[6] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using networkx," in *Proceedings of the 7th Python in Science Conference* (G. Varoquaux, T. Vaught, and J. Millman, eds.), (Pasadena, CA USA), pp. 11 – 15, 2008.

[7] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 1, pp. 440–442, 1998.

[8] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, "Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software," *PLoS ONE*, vol. 9, p. e98679, Jun 2014.

[9] Y. Hu, "Efficient and high quality force-directed graph drawing," *Wolfram Research Inc.*

[10] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, p. P10008, Oct 2008.

[11] K. Stephenson and M. Zelen, "Rethinking centrality: Methods and examples," *Social Networks*, vol. 11, p. 1–37, Mar 1989.

[12] NASA Administrator, "NASA-NOAA satellite reveals new views of Earth at night," Jun 2013.