# A STATISTICAL METHOD FOR SYNTHETIC POWER GRID GENERATION BASED ON THE U.S. WESTERN INTERCONNECTION

*Saleh Soltan, Gil Zussman*
*Electrical Engineering, Columbia University, New York, NY*
*{saleh,gil}@ee.columbia.edu*

SIAM Workshop on Network Science 2015
May 16-17 · Snowbird, Utah

## Summary

In order to develop algorithms that identify power grid vulnerabilities, there is a need to evaluate their performance with real grid topologies. However, due to security reasons, such topologies (and particularly, the locations of the nodes and edges) may not be available. Therefore, we focus on a method for generating test networks with similar characteristics to the real grid. Small-world and scale-free networks [1, 7] are two of the models that were suggested for representing power grids. However, [4] showed that none of these models can properly represent the U.S. Western Interconnection (WI) power grid transmission network (see Fig. 1). An alternative model was proposed in [6] but does not consider the nodes' *spatial distribution*. While there are models for generating spatial networks [2], most of them were not designed to generate networks with properties similar to the power grid. Hence, we present a procedure to generate synthetic networks with similar structural properties and spatial distribution to the WI. It is based on a Gaussian Mixture Model (GMM) that generates the positions of the nodes and a Quadratic Discriminant Analysis (QDA) that is used to connect them. We show that obtained networks have properties similar to the ones of the real grid.

## Positions of the Nodes

The positions of the nodes, denoted by $x_i \in \mathbb{R}^2$ ($i = 1, \ldots, 13992$ in the WI), in the power transmission network are correlated with populations and geographical properties (see for example Fig. 1). Thus, the nodes can be clustered into groups based on their geographical proximity. There are several clustering techniques that can be used. However, we are interested in generating similarly distributed set of points on the plane. Hence, we use the GMM for clustering the points in our data set which represents the WI.[1] To apply GMM to our data set, we used the `mclust` library in `R` to divided the WI into 27 clusters and obtained the mean and variances $(\mu_k, \Sigma_k)$ of the nodes in the cluster $k = 1, \ldots, 27$, along with the categorical probability of the clusters $\pi = (\pi_1, \ldots, \pi_{27})$.

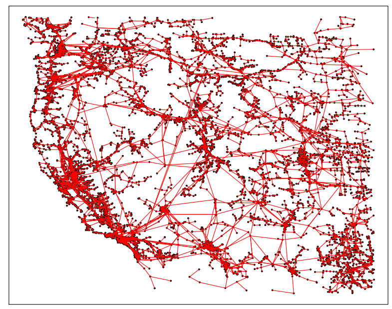

Figure 1: The U.S. Western Interconnection (WI) power grid with 13992 buses (nodes) and 18681 lines (edges) [3].

## Connections between the Nodes

Fig. 2 illustrates the degree distribution of the nodes in the WI. It can be seen that the distribution is heavy-tailed, and therefore, we used linear regression with the nodes whose degree is greater than 3 (in log-log scale) to obtain the exponent ($= -3.4$). We also observed that overall the distribution of the degrees is very similar to the response function of the second order system in control theory in the form of $\frac{a}{\sqrt{((a-d^2)^2+(pd)^2)}}$. Hence, we fit $\mathbb{P}(d) \propto \frac{a}{\sqrt{((a-d^{3.4})^2+(pd^{1.7})^2)}}$ with parameters $p = 4$ and $a = 30.54$ to the degree distribution of the nodes. As can be seen in Fig. 2, the blue dashed line representing this function fits the degree distribution in the WI very well.

We observed the distribution of log-length of the lines ($\log |x_i - x_j|$) in the WI and realized that a Gaussian distribution is a good fit for the log-length of the lines. This suggests that QDA could be used to decide whether

This work was supported in part by CIAN NSF ERC under grant EEC-0812072, DTRA grant HDTRA1-13-1-0021, and the People Programme (Marie Curie Actions) of the European Unions Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no [PIIF-GA-2013-629740].11.

[1] The data is from the Platts Geographic Information System (GIS) [5] (for more details, see [3, Section VI]).

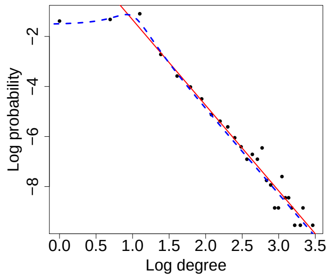


Figure 2: The degree distributions of the nodes in WI (in log-log scale). A (red) solid line with slope −3.4 is fitted to the distribution of nodes with degree greater that 3. A (blue) dashed line shows the fitted probability distribution function $\mathbb{P}(d) \propto \frac{a}{\sqrt{((a-d^{3.4})^2+(pd^{1.7})^2)}}$.

**Procedure 1:** Connecting nodes based on QDA

**Input:** $n, x_i$ for $i = 1, \dots, n$.

1: For all $i$ sample $d_i$ from the probability distribution $\mathbb{P}(d) \propto \frac{a}{\sqrt{((a-d^{3.4})^2+(pd^{1.7})^2)}}$.
2: **for** each $i$ **do**
3: Connect node $i$ to $d_i$ other nodes selected with probability proportional to $\mathbb{P}_{\text{QDA}}(y_{ij} = 1 | x_i, x_j) \times \mathbb{1}\{d_j > 0\}$.
4: Update $d_i \leftarrow 0$ and $d_j \leftarrow d_j - 1$ for all nodes $j$ connected to node $i$ in the previous step.

two nodes should be connected (based on their distances). Thus, we fit a QDA model to find the probability that two nodes are connected ($y_{ij} = 1$, if nodes $i, j$ are connected, and $y_{ij} = 0$, otherwise) based on their distance: $\mathbb{P}_{\text{QDA}}(y_{ij} = 1 | x_i, x_j)$. Considering these observations, we introduce Procedure 1 to connect nodes.

**Algorithm:** Generating Synthetic Network (GSN)

**Input:** $n$.

1: For all $i = 1, \dots, n$ sample $z_i$ from the categorical probability distribution $\pi$ obtained from GMM.
2: For all $i$ sample $x_i$ from the probability distribution $\mathcal{N}(\mu_{z_i}, \Sigma_{z_i})$ obtained from GMM.
3: Connect nodes using Procedure 1.
4: Make the network connected.

## Generating a Synthetic Network

We now introduce an algorithm to generate a synthetic network similar to the WI. First, the Generating Synthetic Network (GSN) Algorithm generates the positions of the nodes using parameters obtained from GMM. Then, using Procedure 1, it connects the nodes. After this process, however, the obtained network may not be connected. Therefore, the algorithm makes the network connected by recursively connecting the pair of nodes with the minimum geographical distance that are not in the same connected component.

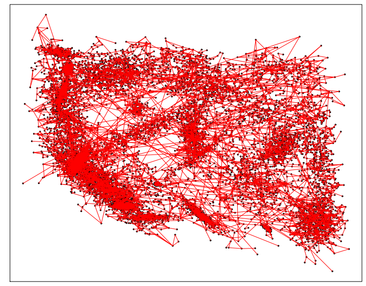

Figure 3: A network generated using the GSN Algorithm.

Table 1: Summary of networks properties. All the networks have 13992 nodes.

| Networks | Edges | C | L |
|---|---|---|---|
| Western Interconnection | 18681 | 0.053 | 18.46 |
| GSN Algorithm | 18672 | 0.043 | 22.13 |
| Random network | 19665 | 0.0001 | 9.81 |
| Scale-free network | 27981 | 0.001 | 4.42 |
| Small-world network | 27984 | 0.353 | 12.97 |

## Evaluation

Two of the most important network properties are the *average path length* (denoted by $L$) and the *clustering coefficient* (denoted by $C$). We use these along with the number of nodes and edges to evaluate the structural similarities of the generated networks and the WI network (an example of a generated network appears in Fig. 3). For comparison, we also generated a random network, a scale-free network, and a small-world network.[2] As can be seen in Table 1, the network generated by the GSN Algorithm has very similar number of edges and very similar $C$ and $L$ values to the WI network. The values of $C$ and $L$ for other networks with the same number of nodes significantly differ from the WI's values.

## References


[1] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[2] M. Barthelemy. Spatial networks. *arXiv:1010.0302v2*, 2010.

[3] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman. Power grid vulnerability to geographically correlated failures - analysis and control implications. *arxiv:1206.1099*, 2012.

[4] E. Cotilla-Sanchez, P. D. Hines, C. Barrows, and S. Blumsack. Comparing the topological and electrical structure of the North American electric power infrastructure. *IEEE Syst. J.*, 6(4):616–626, 2012.

[5] Platts. GIS Data. `http://www.platts.com/Products/gisdata`.

[6] Z. Wang, A. Scaglione, and R. J. Thomas. Generating statistically correct random topologies for testing smart grid communication and control networks. *IEEE Trans. Smart Grid*, 1(1):28–39, 2010.

[7] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.


[2] We generated these networks several times. However, since the clustering coefficient and average path length are average values, we obtain very similar values for different instances.