Limits in Modeling Power Grid Topology

Brian Cloteaux
Applied and Computational Mathematics Division
National Institute of Standards and Technology
Gaithersburg, Maryland, USA
Email: brian.cloteaux@nist.gov

***Abstract*—Because of their importance to infrastructure, a number of studies have examined the structural properties of power grids and have proposed random topological models of them. We examine the ability to create generalized models of power grid structure by comparing real data sets to see where inherent modeling limitations occur. We then propose a possible mechanism for why these networks differ in their structure, and the implications these differences have in creating power grid models. Finally, we introduce a new model for power grids with a radial-style architecture.**

***Index Terms*—power system modeling**

I. Introduction

Observations over the past two decades have shown that there are a number of systems that contain embedded networks having a number of common features. The importance of these similar networks, commonly called complex networks, lies in the areas in which they are found. They have been shown to arise naturally in systems of both biological [1]–[3] and social [4] interactions. In addition, there also have been claims that similar complex networks appear in many engineered systems such as the Internet [5], software components [6], and an area in which we are particularly interested, power grids.

Being able to measure and model engineered systems is an essential step towards being able to maintain and understand the reliability, safety, and in some cases the security of these systems. This is especially true for electric power grid systems, where reliability questions, such as how blackouts propagate, are still open. These questions highlight the need for a formal measurement and simulation program for understanding power grids. Towards this goal, we examine a claim that is often made in the literature that power grid networks have similar properties and can be treated homogeneously. Specifically, there have been several publications implying that power grids can be treated as scale-free complex networks (such as [7]–[9]).

This idea that power grids in various parts of the world would have similar properties is a surprising one. Power grids are fundamentally different than the other engineered systems mentioned above, such as the autonomous system topology of the Internet. It is expensive and sometimes impossible to arbitrarily add or modify the substations and power lines of a power grid, i.e., the nodes and edges of the network. Because of these constraints, power grids are highly engineered and tend to have extremely static topologies.

This claim poses several questions about how we should analyze power grid networks. Should we model and simulate power grids as complex networks? Why would a highly engineered system show characteristics of complex networks? How did they acquire any shared structure that we see in them? What are the limits to creating realistic random models of power grids? We examine these questions. Since the issues associated with power grid reliability are complicated, we will only focus on the topological characteristics of networks. We start by examining where topological differences show up between the different power grids, and what implications this has for the modeling of these networks.

II. Examination of Power Grids

The idea that power grids, and particularly their topology, are complex networks can be traced back to a paper by Watts and Strogatz [1]. In their seminal paper, they showed that the power transmission grid covering much of the western states in the United States shows characteristics that are far from the expected behavior of random graphs with an identical degree distribution. Specifically, they showed that the western power grid has a much larger average shortest path and clustering coefficient than expected from random models. This fits their definition of a small-world network, and so their conclusion was that this power grid is an example of this type of network. At about the same time, several research groups showed that that degree distributions of various power grids are predictable [4], [10], [11].

To determine what static characteristics hold between power grids, we examined three data sets: Strogatz and Watts' original data set of the western United States power grid, the power grid of Poland [12], and a test power grid produced by the IEEE [13]. The basic characteristics of these sets are shown in Table I.

In determining if power grids are small-world networks, in general, we see that the average path length L of one of these power grids all appears to be proportional to the logarithm of number of nodes N in the network, in other words

$$L \sim \log N \tag{1}$$

From the data the diameter does not disqualify any of these power grids from being small-world, although we would note though that graphs with relatively small diameters are very

Data Set	# Nodes	# Edges	Avg. Deg.	Diam.	Avg. Path Len.	Avg. Clustering Coeff.
Western States	4941	6594	2.67	46	18.99	0.080
Poland	2383	2886	2.42	30	12.76	0.009
IEEE 300	300	409	2.73	24	9.94	0.086

TABLE I

WE EXAMINED THREE DATA SETS. THESE DATA SETS INCLUDE THE POWER GRID TOPOLOGIES OF THE WESTERN UNITED STATES [1], [14] AND OF POLAND [12]. IN ADDITION, WE INCLUDED THE IEEE 300 TEST DATA SET [13]. BASIC INFORMATION ABOUT THESE SETS IS RECORDED IN THE ABOVE TABLE, INCLUDING THE DIAMETER, AVERAGE SHORTEST PATH LENGTH, AND CLUSTERING COEFFICIENT.

common. One striking feature of these data sets is that the Polish power grid has a much smaller average clustering coefficient than the other data sets. In fact, the Polish set does not show an average clustering coefficient that is significantly above the expected value for a random network of its size. This calls into question whether Watts and Strogatz would have considered the Polish power grid as a small-world network at all. Thus it seems that we cannot make the generalization that all power grids show small-world characteristics.

We next look at the degree distributions of these networks. The specific degree distribution of power grid networks has been a source of conflicting conclusions. A claim that is often made in the literature when examining specific power grids is that they are scale-free networks (such as [7]–[9]). The source of this statement stems from Barabási and Albert [10] who stated that the distribution of power-grid networks can be approximated by a power-law distribution with an exponent of 4.

Although the larger degree nodes in many power grids can be approximated with a power-law distribution, this distribution does not fit the lower degree nodes well. As first noted by Amaral, Scala, Barthélémy, and Stanley [4], a closer examination of these networks shows that their degrees are better fitted by an exponential distribution. Later, Albert, Albert and Nakarado [11] stated that the cumulative degree probability distribution is approximately

$$P(k \leq K) \sim 1 - e^{-0.5K} \tag{2}$$

An exponential distribution can be shown graphically as a linear distribution on a linear-logarithm scale. We can see the results of fitting to a semi-log distribution in Figure 1. As shown in the figure, the various power grids do appear to fit exponential distributions[1].

This raises the question on why we see this type of distribution in such a highly engineered system at all. It has been suggested that cost constraints can distort an underlying power-law distribution into an exponential distribution [4], but that raises the question of how can a power-law distribution be associated with a power grid in the first place?

As an engineered system, the power grid does not organically evolve, but instead is a reflection of design choices. Thus for the power grid to be influenced towards some power-law-like distribution requires that there must exist an underlying mechanism that influences the designers towards these structures. Accounting for a precise mechanism remains an open question, but we can speculate as to the source behind this observed structure. From an engineer's perspective, the placement of substations and power lines, is principally a function of servicing population areas while adhering to constraints imposed by the capacity and cost for construction.

Using this viewpoint, we postulate a guess as to why the larger degrees in the network seem to follow a power-law distribution. Recent research has shown that the populations of various communities tend to follow a consistent power-law distribution. In particular, it was shown that the size of natural population centers in the United States follow a power-law distribution with exponents of 1.74 in general to 1.91 for the largest communities [15]. If power substations are laid out such that their number of connections are roughly proportional to the population of the surrounding community they service, then their degree distribution should, to some extent, reflect the surrounding population distribution.

For an exponential distribution for power grids to arise from the surrounding population distributions, the importance of redundant paths into a substation should be dependent on the number of people served by that substation. In other words, the number of transmission lines into a substation takes into consideration the cost of adding the additional transmission lines versus the number of people that would benefit from the increased reliability. If the number of power lines going into the power substations are roughly proportional to the logarithm of population of the surrounding community they service, then the resulting degree distribution will be exponential. To see this, consider a population whose probability density function is the Zipf distribution

$$P_{pop}(k) = c \cdot k^{-(\rho+1)} \tag{3}$$

where the constant $c = \frac{1}{\gamma(\rho+1)}$ and $\gamma(z)$ is the Riemann zeta function. If we take the logarithm of the value produced by such a distribution to create a new distribution P_{grid} then the resulting cumulative distribution is

$$P_{grid}(k \leq K) = P_{pop}(k \leq e^K) = c \cdot \int_0^{e^K} k^{-(\rho+1)}\, dk \tag{4}$$

$$= \frac{c}{\rho}\left(1 - e^{-\rho K}\right) \tag{5}$$

Since this is proportional to the cumulative density function of an exponential distribution, it implies that the resulting power grid would have an exponential distribution.

[1]It should be noted that Holmgren [9] points out that the distributions of the Western States and the Nordic power grids do not pass the chi-square and Kolmogorov-Smirnov tests for an exponential distribution.

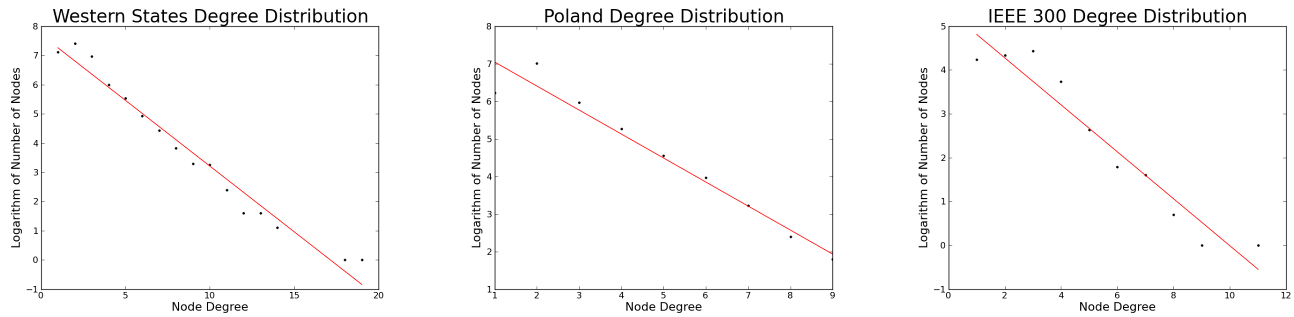


Fig. 1. A least squares fitting of the three data sets on a semi-log domain. This fitting gives support to the idea that power grid degree distribution are actually exponential distributions and not power-law.

III. Power Grid Topological Models in the dK Hierarchy

The results from the previous section imply that there are certain characteristics of power grid networks, such as degree distribution, that are independent of any underlying engineering choices. At the same time, the observed clustering coefficient values show that there are demonstrable differences in some topological characteristics between the various power grids. To understand how engineering choices affect the modeling of power grids, we examine where differences show up in the dK-hierarchy of random models.

The dK-series of models was introduced by Mahadevan, Krioukov, Fall, and Vahdat [16] as a hierarchy for random models of networks. In creating a random model for a network G, they defined three properties that they were trying to capture with their hierarchy. The first is *constructibility*, the ability to construct models having the properties in the dK-series. The second is *inclusion* or the idea that for any property $\mathcal{P}_d$ defined in the dth level of hierarchy, this property will subsume all the properties $\mathcal{P}_i$ for $0 \leq i \leq d-1$. The final property is *convergence*, which states that as d increases, the set of graphs having the property $\mathcal{P}_d$ converges to G.

For the dK-series, the properties $\mathcal{P}_d$ are all local connectivity properties of the nodes. Thus a $0K$-model is a model with the same average number of links per node. In other words it has the same number of edges as the original network. A $1K$-model has the same node degree distribution as the original graph. As mentioned in the last section, for electric power grids we can construct $1K$-models since there is empirical evidence that these grids share a common degree distribution. To create a random model with n vertices, we would simply select a n length integer sequence from an exponential distribution with $\rho = 0.5$, and any realization of this sequence would be a $1K$-model of the power grid.

The $2K$ set of models are defined by the graph's joint degree distribution (JDD). The joint degree distribution gives the degree correlations for pairs of connected nodes. In practice, this information is typically stored as a symmetric integer matrix D where the entry $D_{i,j}$ contains the number of nodes of degree i connected to a node of degree j.

In Figure 2, the joint degree distributions are shown for the three data sets. One glaring difference between the sets is the number of degree two nodes that connect to other degree two nodes, i.e., the values of $D_{2,2}$. From the IEEE test set, there are almost no connections of this type, while for the Polish set these connections predominate. The value of Western States data set falls in between the other two sets. This is a prime example that there are significant differences between power grids that cannot be captured by any generic $2K$-model.

While the precise mechanism behind why we see such radical differences is not known, we would speculate that these differences in the joint degree distribution appear to be the direct result of engineering choices. More specifically, this seems to be a reflection of the differences between the radial architectures, used more in the United States, versus the ring-main architectures favored more in Europe ([17], pg. 48).

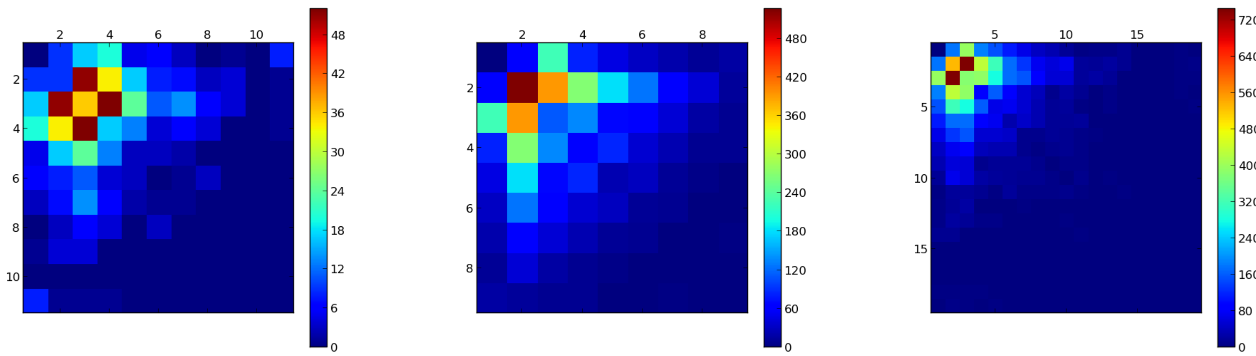
The fact that we cannot create a generic $2K$-model for power grids is further reinforced when we examine the $3K$-characteristics between the sets. For a graph to be a $3K$-model, the local clustering coefficient for each node must also be matched. The local clustering coefficient for a node n is the ratio of number of edges between its neighboring nodes to the number needed to form a complete subgraph [1]. In other words, for the node n its local clustering coefficient C_n is

$$C_n = \frac{|\{e_{ij} : v_i, v_j \in N_n, e_{ij} \in E\}|}{\binom{|N_n|}{2}} \tag{6}$$

where N_n is the set of nodes adjacent to n.

From the definition of the dK-series of models, if power grids had a general $3K$-model, that would imply that it would also have a $2K$-model. Obviously, that does not hold in this case. In addition, we have already seen large differences in the average clustering coefficient between the Polish and Western States data sets, but if they had similar local clustering coefficients then we would see almost no difference in the averages.

This section shows an important limit of our ability to perform analysis using random power grid models. To capture most topological characteristics, we require at least a $2K$-model and often a $3K$-model. For a generic power grid model, this is impossible; at best, we can only guarantee $1K$-models.



(a) IEEE 300 (b) Polish Power Grid (c) Western States Power Grid

Fig. 2. This figure shows the joint degree distributions of the three test data sets. These figures show the large differences between the joint degree distributions of the three data sets. In particular, the entries $D_{2,2}$ show radically different values between all three sets.

IV. A Topology Model for Radial-Style Grids

Our contention is that non-trivial topological characteristics of power grids can be captured by random models only if we take into account the engineering principles behind the creation of the grids. Extending this idea means that there is no one model that accurately reflects all power grids. Thus any realistic model would have to reflect the design principles behind the power grid that they were simulating.

We applied some simple assumptions to how radial style power grids are formed to see if we could create topologically better fitting models. Based on these assumptions, this leads to a straightforward algorithm to creating models. We simulate a population over a geographic area using random uniform distribution to distribute a set of points on the unit square. Each point is weighted with a random value taken from some power-law distribution. If we were trying to accurately replicate the population distribution for the United States, our distribution would use one of the exponents given by Jiang and Jia [15], but for our purposes of creating a model that tries to match known results [11], we used a Zipf distribution with the parameter $\rho = 0.5$.

For our models of power grids, we required two topological characteristics of our random model. The first is that the resulting network is connected. If we were to have separate components in a model, then we would have to consider each of these components as a separate power grid. We ensure connectivity by first computing a minimal spanning tree connecting all of the points in the network. The weight of an edge is computed by some function that accounts for construction cost of the transmission line.

The second characteristic we require is that they must be planar. We assume that the layout of the transmission lines is such that they do not cross each other, but rather converge at substations. In our algorithm, we guarantee this by only adding edges that do not cross existing edges in the network.

In order to construct our models, we made some additional simplifying assumptions. In real power grids, cost constraints for connecting communities come not only from the distance between substations, but also come from having to navigate geographical barriers such mountain ranges or bodies of water. For our purposes, we assume that distance is the only cost constraint for connecting substations.

A second simplification is that for a real power substation, there is a maximum capacity on the amount of power it can handle. Exceeding that capacity requires building additional substations to service the neighboring areas. As a simplification, we assumed that each natural population center is serviced by one substation. We can justify this simplification by thinking of all the substations needed to service a community as a type of supernode.

Since we are trying to capture a radial style development, our method seeks to connect to nodes with the lowest cost. We added a second restriction in choosing node connections in that we only looked at the nearest articulation points in the network. An articulation point is a node that if removed would cause the network to become disconnected. By making the arbitrary choice to connect only to articulation points in the network, we are trying to reduce the overall number of articulation points in the model and thus make the resulting grid more robust to node outages.

Similar ideas of connecting to nearest nodes to simulate a radial style of development has been previously proposed [18], [19]. Our improvements to these methods have been to ensure that the grid is planar, has some level of node robustness, and approximately matches measured degree distributions.

The complete method is outlined below.

1) On a unit square, randomly (using a uniform distribution) place n nodes
2) Assign each node n a weight ω_n drawn from a power-law distribution (in our case, a Zipf distribution with parameter $\rho = 0.5$)
3) Using the Euclidean distance as the edge weight between

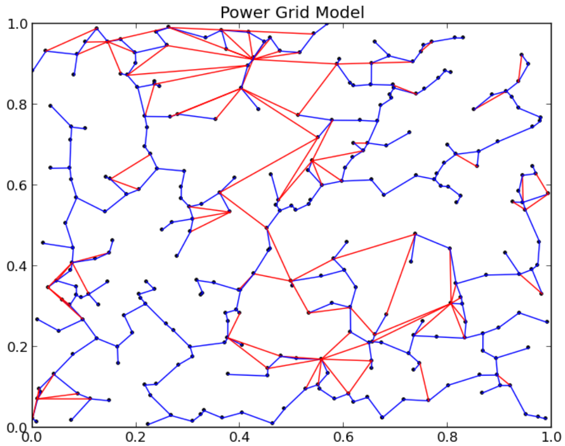


Fig. 3. This figure shows a typical model created by the algorithm. The blue edges represent those selected for the spanning while the red edges are the ones added afterwards. The size of the model was chosen to match size of the IEEE 300 data set. This model had 300 nodes, 395 edges, an average degree of 2.63, a maximum degree of 13, an average clustering coefficient of 0.16, an average path length of 13.58, and a diameter of 41.

two nodes, compute the minimal spanning tree for the edges between the nodes

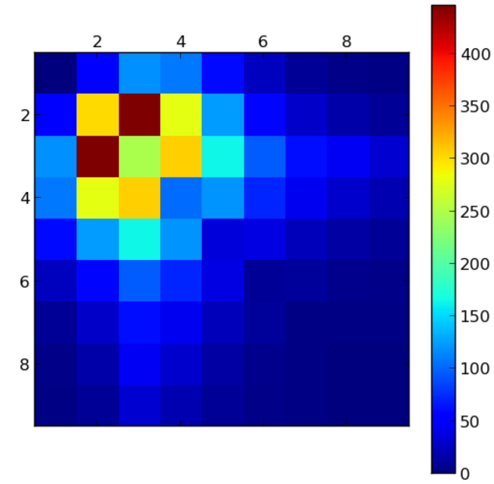
4) For each node n, while $|N_n| < \lfloor \log \omega_n \rfloor$
    a) Find the set of all nodes P that can be connected to n by a straight line that does not cross any other edge
    b) From the set P, connect to the non-neighboring articulation point with the smallest edge weight (distance)

Figure 3 shows a typical model that is produced by this algorithm. In this figure, the edges added while creating the minimal spanning tree are shown as blue, while the remaining edges are shown in red. For this example, we created a model with the same number of nodes as the IEEE 300 data set.

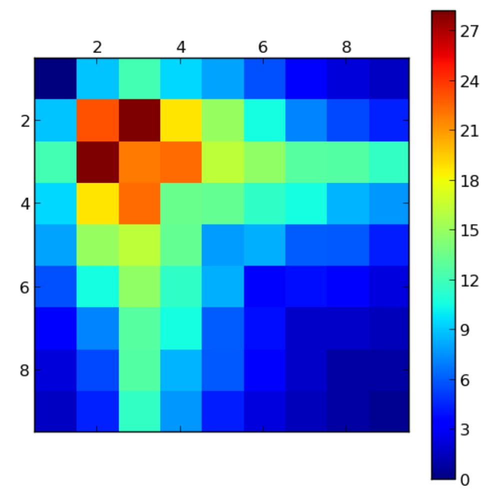
To examine the structure of these models, we examined the average joint degree distribution produced by the algorithm. For Figure 4, we created 200 models with the same number of nodes as the Polish power grid data set. We then computed the average and standard deviation of each entry of the joint degree distribution of these models. The structure of the matrix is representative of the behavior of the algorithm across all size data sets, and gives us an idea of the topological characteristics of the resultant models. We notice that the models tend to produce a high value in the $D_{2,3}$ entires, while producing a much lower number of $D_{2,2}$ connections.

To understand how closely the algorithm approximates a $2K$-model, we computed the average of each entry of the joint degree distribution for 200 model instances for each size of data set. We then subtracted these average values from the original data set and normalized by the magnitude of the largest value in the data set. The results of this comparison are shown in Figure 5.

As we would expect, we see large differences in $D_{2,2}$ entries



(a) Average of Joint Degree Distribution

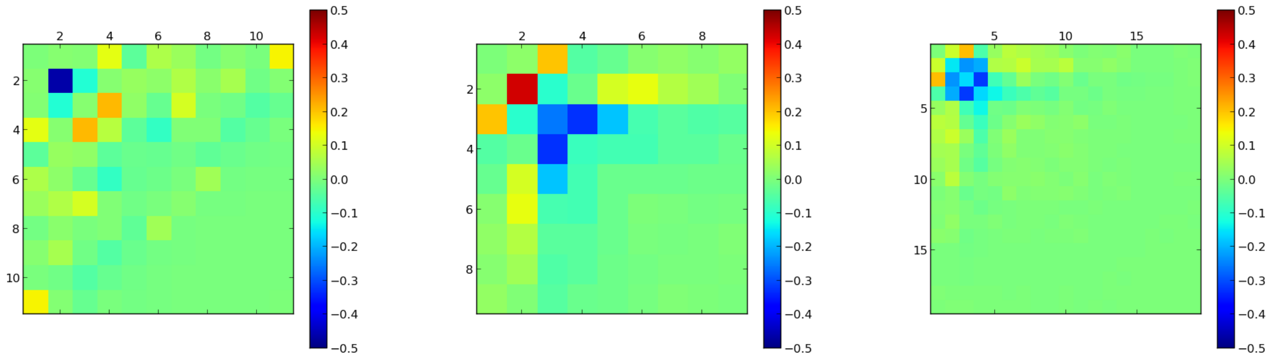


(b) Standard Deviation of Joint Degree Distribution

Fig. 4. The average joint degree distribution of 200 models of the Polish power grid topology using Algorithm 1. This distribution is typical of the joint degree distribution from the model created by the algorithm.

for the IEEE 300 and Polish Power Grid data sets. The model severely overestimates this value for the IEEE 300 data set and underestimates it for the Polish power grid. In both cases, such a model would have limited usefulness in capturing important topological characteristics. In the case of the Western State data set, however, the model average shows a lower relative error across all its entries, and is a much closer approximation of this data set.

In addition, we see from the table in Figure 5 that the average clustering coefficient from the model is about twice



(a) IEEE 300 (b) Polish Power Grid (c) Western States Power Grid

	Avg. Deg.		Avg. Clustering Coeff.	
Data Set	Mean	Std. Dev.	Mean	Std. Dev.
IEEE 300	2.794	0.106	0.195	0.025
Poland	2.795	0.040	0.170	0.008
Western States	2.797	0.027	0.167	0.006

Fig. 5. This figure shows the relative difference between the joint degree distribution of the three test data sets and the average joint degree distribution of 200 model created with our modeling algorithm. The difference between the two data sets is normalized by the magnitude of largest value in the original data set. In addition, we show a table containing the average degree and clustering coefficient in order to establish where these models fall in the dK-hierarchy.

as large as the Western states network. This highlights that while the algorithm approximates a $2K$-model of radial architectures, there are further refinements needed for it to become a $3K$-model.

V. Concluding Thoughts

Although it is common in the literature to treat power grids as homogeneous networks, beyond the basic degree distributions, the various power grids show little commonality in their structure. They seem to be poor candidates to treat, both by modeling and analysis, as general complex networks. Whatever common structure they have that is reminiscent of complex networks, seems likely to be a reflection of the underlying population distributions. We could go so far as to speculate that if populations were distributed according to some non-fat-tailed distribution (such as a normal or Poisson distribution), we would probably see the degree distributions in power grids more closely resembling meshes and sharing none of characteristics we typically associate with complex networks.

Further, we specifically showed that there are large differences between power grids in their joint degree distributions. This implies that there does not exist a general purpose model of the topology of power grids that is better than being a $1K$-model. It seems that an accurate understanding of a power grid's topology must take into account the engineering principles used in the design of each specific grid.

We then extended this idea to propose a new algorithm for generating radial-style power grid topologies. Further research will involve making the algorithm more closely match the reference data set, and investigate how other grid architectures can be successfully modeled. In addition, we are interested in seeing whether application of these ideas can lead to new and useful analysis techniques and metrics in dealing with power grid networks.

Acknowledgments

The author would like to thank the referees, Ron Boisvert, and Craig Weinschenk for their helpful comments.

References

[1] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.

[2] H. Jeong, B. Tombor, R. Albert, Z. Oltvai, and A. L. Barabási, "The large-scale organization of metabolic networks," *Nature*, vol. 407, no. 6804, pp. 651–654, October 2000.

[3] R. J. Williams and N. D. Martinez, "Simple rules yield complex food webs," *Nature*, vol. 404, no. 6774, pp. 180–183, March 2000.

[4] L. A. N. Amaral, A. Scala, M. Barthlmy, and H. E. Stanley, "Classes of small-world networks," *Proceedings of the National Academy of Sciences*, vol. 97, no. 21, pp. 11 149–11 152, 2000.

[5] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," in *Proceedings of the 1999 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (SIGCOMM)*. New York, NY, USA: ACM, 1999, pp. 251–262.

[6] A. Potanin, J. Noble, M. Frean, and R. Biddle, "Scale-free geometry in OO programs," *Communications of the ACM*, vol. 48, no. 5, pp. 99–103, 2005.

[7] D. P. Chassin and C. Posse, "Evaluating North American electric grid reliability using the Barabási-Albert network model," *Physica A: Statistical Mechanics and its Applications*, vol. 355, no. 2-4, pp. 667–677, 2005.

[8] P. Crucitti, V. Latora, and M. Marchiori, "A topological analysis of the Italian electric power grid," *Physica A: Statistical Mechanics and its Applications*, vol. 338, no. 1-2, pp. 92–97, 2004.

[9] A. J. Holmgren, "Using graph models to analyze the vulnerability of electric power networks," *Risk Analysis*, vol. 26, no. 4, pp. 955–969, 2006.

[10] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[11] R. Albert, I. Albert, and G. L. Nakarado, “Structural vulnerability of the North American power grid,” *Physical Review E*, vol. 69, p. 025103, Feb 2004.

[12] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, “Polish power grid data set,” http://www.pserc.cornell.edu/matpower/, retrieved on August 8, 2012.

[13] R. D. Christie, “IEEE 300 test power grid data set,” http://www.ee.washington.edu/research/pstca/, retrieved on August 19, 2008.

[14] D. J. Watts and S. H. Strogatz, “Western states power grid data set,” http://www-personal.umich.edu/~mejn/netdata/power.zip, retrieved on August 26, 2008.

[15] B. Jiang and T. Jia, “Zipf’s law for all the natural cities in the United States: a geospatial perspective,” *International Journal of Geographical Information Science*, vol. 25, no. 8, pp. 1269–1281, 2011.

[16] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 4, pp. 135–146, Aug. 2006. [Online]. Available: http://doi.acm.org/10.1145/1151659.1159930

[17] J. Schlabbach and K. Rofalski, *Power System Engineering: Planning, Design, and Operation of Power Systems and Equipment*. Wiley, 2008.

[18] Z. Wang, R. Thomas, and A. Scaglione, “Generating random topology power grids,” in *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, Jan. 2008, p. 183.

[19] P. Hines, S. Blumsack, E. Cotilla Sanchez, and C. Barrows, “The topological and electrical structure of power grids,” in *Proceedings of the 43st Annual Hawaii International Conference on System Sciences*, Jan. 2010, pp. 1–10.