Grid Structural Characteristics as Validation Criteria for Synthetic Networks

Adam B. Birchfield, *Student Member, IEEE*, Ti Xu, *Student Member, IEEE*,
Kathleen M. Gegner, *Student Member, IEEE*, Komal S. Shetye, *Member, IEEE*, and Thomas J. Overbye, *Fellow, IEEE*

Abstract—This paper presents a methodology and set of validation criteria for the systematic creation of synthetic power system test cases. The synthesized grids do not correspond to any real grid and are, thus, free from confidentiality requirements. The cases are built to match statistical characteristics found in actual power grids. First, substations are geographically placed on a selected territory, synthesized from public information about the underlying population and generation plants. A clustering technique is employed, which ensures the synthetic substations meet realistic proportions of load and generation, among other constraints. Next, a network of transmission lines is added. This paper describes several structural statistics to be used in characterizing real power system networks, including connectivity, Delaunay triangulation overlap, dc power flow analysis, and line intersection rate. The paper presents a methodology to generate synthetic line topologies with realistic parameters that satisfy these criteria. Then, the test cases can be augmented with additional complexities to build large, realistic cases. The methodology is illustrated in building a 2000 bus public test case that meets the criteria specified.

Index Terms—Clustering, delaunay triangulation, power grid topology, synthetic networks.

I. INTRODUCTION

SYNTHETIC networks, that is, entirely fictitious but realistic power grids, have many potential benefits for aiding power systems research and development. Because of confidentiality requirements on critical infrastructure information, data on actual power systems often cannot be freely shared, published, and compared among researchers. It is important that results of research be reproducible by peers, and synthetic networks can serve as test cases for new algorithms which can then be applied to the actual grid models. Because synthetic networks are entirely fictitious, they are free for public use and can spur innovation for future grid challenges.

An important requirement in building synthetic power system models is that they match the size, complexity, and characteristics of actual grids. Public test cases do exist, such as the IEEE test cases, but most of them have 300 buses or fewer, and they lack complexity and additional specifications that actual grids have and which various kinds of research need. Actual grid models often include many phase-shifting transformers, transformer impedance correction tables, remote tap-changing bus voltage regulation, switched shunt reactive power compensation, and other complexities, which are not always included in current, simple test systems.

Pioneering work on this topic includes an effort described by [1], which used public information from utilities and regulatory agencies to develop an approximate dc power flow model of the continental European grid. This approach differs from the concept of synthetic networks, which do not attempt to approximate the actual grid, but only to create fictitious models which match characteristics of actual grids. The network of [1] is large with 1500 buses and is designed for optimal power flow and locational marginal price analysis. However, it does not contain specifications sufficient for an ac power flow solution, nor does it include other complexities which would be present in a real power system.

The work of [2]–[5] outlines an approach to generating truly synthetic transmission line topologies, based on the small world approach first described in [6]. Other graph-theory and statistical analyses of power systems, including [7]–[9], highlight bus nodal degree, average clustering coefficient, and average shortest path length as properties that characterize real power systems. This paper builds on these observations and applies them to creating synthetic networks of full power systems, not only the branch topologies.

The authors' previous work in [10] introduces the principles of an approach which builds a set of synthetic substations located on a real geographic footprint in North America, utilizing public information about energy generation and consumption in a region. The authors' work in [11] adds a basic statistical analysis of power grids' transmission line topology and presents a pilot application to a 150-bus public test case suitable for geomagnetic disturbance studies.

This paper presents a substation placement method from real energy and population data based on [10] and [11]. This paper adds to the substation placement process with a detailed clustering methodology which matches substation characteristics of actual grids more realistically. In particular, the distribution of

load and generation proportions are matched across a system, allowing load and generation to be clustered together in the same substation subject to constraints.

The next step in the network synthesis approach presented in this paper is to connect the substation buses with transmission line topologies at multiple voltage levels. This paper describes some structural characteristics of actual power systems, which can serve as validation metrics to determine how realistic the resulting cases are. The metrics considered in this paper are based on 1) topological connectivity, 2) overlap with the Delaunay triangulation and minimum spanning tree of the substations, 3) ac power flow convergence, and 4) geographic intersections between same-voltage-level transmission lines. Data are given for each of the metrics based on real power systems, and an algorithm is proposed to generate transmission line topologies which match the criteria described. This method builds on the simplified process introduced in [11] to match the proportion of the minimum spanning tree, Delaunay segments, Delaunay neighbors, and line intersections found in actual grids' structure.

A solvable ac power flow case with geographic substation definitions, areas, line and transformer parameters can then be augmented to include additional complexity such as reactive power control devices, transformer tap-changing control, and parameters necessary for optimal power flow, transient stability, and geomagnetic disturbance analysis. These topics are introduced in this paper; future work will add more detail.

The final section of this paper presents a case study of the approach outlined in the paper. It uses a geographic footprint located within the U.S. state of Texas to build a case with 1500 substations, about 2000 buses, and about 50 000 MW of peak load. The case is publicly available online [12]. It does not correspond to the actual grid, but it meets the validation criteria of this paper and shares realistic characteristics with actual grids, making it useful for research and education.

II. SUBSTATIONS

The creation of a synthetic power grid starts with the assignment of geographic coordinates, loads, and generators to substations. In a recent Eastern Interconnect planning case, about 80% of substations contain purely load, and 5% substations contain only generation, while 5.5% contain both [10]. In addition, only 2.5% of buses are directly connected to both generators and loads. In the synthetic networks created here, we assume that each individual substation contains loads or generators or both and may include multiple buses, each of which may be directly connected to a load, a generator, or neither [13]. Suppose that there are $N_s = N_l + N_g + N_b$ total substations to be placed in the synthetic network; denote N_l as the number of substations (Type l) containing only loads and N_g as the number (Type g) containing only generators, along with N_b as the number (Type b) including both.

A. Load

Since actual load information is not fully public and electricity demand is highly correlated with population [10], we use the geographic coordinates, described in terms of latitude and longitude, and population of each postal code obtained from the

TABLE I
POSTAL CODE CLUSTERING ALGORITHM

|	
1.A	Collect geographic and population data for N postal codes;
1.B	Compute the Haversine distance between two postal codes;
1.C	Set $K = N_l + N_0$ and $k = N$; Initialize each postal code as one cluster;
2.A	Combine two clusters c_1, c_2 with the minimum distance given by (1) as a new cluster c $= c_1 \cup c_2$;
2.B	Update the coordinates of cluster c as the average of the coordinates of clusters c_1, c_2, weighted by the sum of the population of its members; update the population in cluster c as the sum of the population of clusters c_1, c_2;
2.C	Remove clusters c_1, c_2, and update the distances between each of the remaining clusters and the new cluster c; Set $k = k - 1$;
2.D	If $k > K$, go to step 2.A; otherwise, stop.
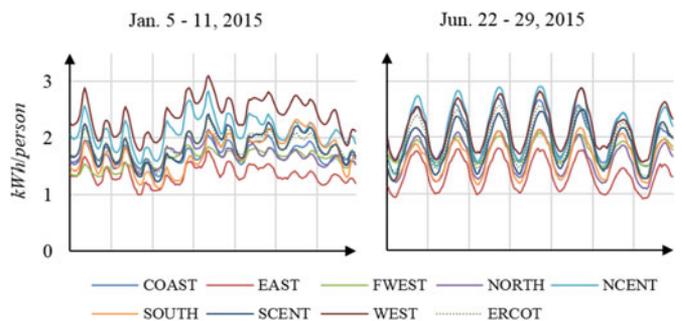
Fig. 1. Hourly electricity consumption per capita for eight ERCOT weather zones in one winter week and one summer week [16].

public U.S. census database [14] to determine the load substation (Type l and Type b) locations and their corresponding electricity consumption. Given a system footprint with N postal codes, we modify and apply the hierarchical clustering algorithm [15] to group those N postal codes into $N_l + N_b$ non-overlapping clusters, each of which represents a load substation and includes several postal codes close to each other. Equation (1) gives the distance between two clusters as the weighted average distances of their postal code members. The weight $w(m_1, m_2)$ is the sum of populations of members m_1, m_2. Substations are not allowed to cluster beyond a maximum population limit, so that urban areas will not conglomerate into a few load substations with unreasonably high electricity consumption. Table I summarizes the postal code clustering process.

$$\text{dist}_c (c_1, c_2) = \frac{\sum_{m_1 \in c_1} \sum_{m_2 \in c_2} w(m_1, m_2) \, \text{dist}(m_1, m_2)}{\sum_{m_1} \sum_{m_2} w(m_1, m_2)}$$
(1)

Based on a study of state-level load and population regression, previous work assigned a peak summer load of 2.00 kW per capita and 0.57 kvar per capita to each load [11]. However, the load behaviors vary over different locations; Fig. 1 shows load variations of areas within Texas as an example. For this paper, the per-capita MW consumption is determined from public aggregate consumption records in the geographic area. A fixed power factor is assumed for each load.

B. Generators

The U.S. Energy Information Administration maintains a yearly survey of the nation's generators and has already re-
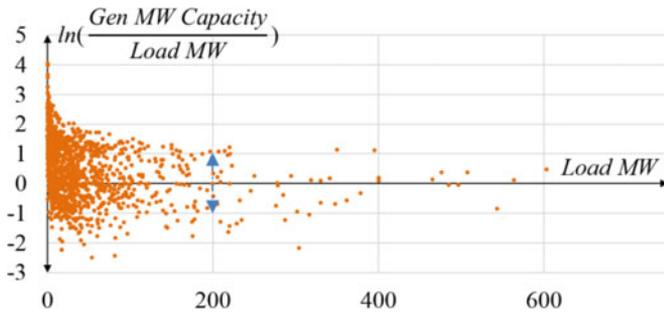
Fig. 2.   Ratio of generation capacity to load of Type b substations in Eastern Interconnect case.

TABLE II
GENERATOR CLUSTERING ALGORITHM

1.A	Collect geographic and generation unit data for all power plants;
1.B	Randomly select N_b load substations as the Type b; set $p = N_b$
2.A	Pick one of the p selected load substations, s, and, based on its load MW and the statistics in Fig. 2, randomly choose a value as its gen-load ratio;
2.B	Compute the distance of substation s to each of the available power plants;
2.C	Sort and choose power plants in the increasing order of their distances computed in 2.B as the power plant candidates for substation s until substation s achieves its minimum generation capacity;
2.D	If $p-1>1$, set $p = p-1$ and go to 2.B; otherwise, continue;
3.A	Set k as the number of the remaining, unassigned power plants and initialize each generator as one cluster;
3.B	Compute the Haversine distance between two generators;
3.C	Combine the pair of two clusters c_1, c_2 with the minimum distance as a new cluster c $= c_1 \cup c_2$;
3.D	Update the coordinates of cluster c as the average of the coordinates of clusters c_1, c_2, weighted by the total generation capacity of the clusters' members; update the total generation capacity in cluster c as the sum of the generation capacities of clusters c_1, c_2;
3.E	Update the distances between each of the remaining clusters and the new cluster c, and remove clusters c_1, c_2; Set $k = k-1$;
3.F	If $k > N_g$, go to step 3.A; otherwise, stop.

leased the 2014 survey data to the public [17]. Data is collected on the fuel type, generation capacity, geographic coordinates and unit information of all power plants.

From the clustered population data of the previous section, N_b load substations are randomly selected as the Type b substations with both generators and loads the remaining N_l load substations are Type l substations purely containing loads. Fig. 2 shows the ratio of the generation capacity to the load MW value – referred to as "gen-load ratio" - for each Type b substation in the actual Eastern Interconnect planning case. This paper uses the load MW to randomly pick a gen-load ratio value for each of the N_b selected substations using the statistics in Fig. 2. For instance, given a substation s with a load of 200 MW, a ratio value will be randomly drawn from the range of e^{-1} to e^1 and will be used to compute the minimum generation capacity (gen-load ratio × 200 MW) that the substation s must achieve. Then, all available power plants are sorted and assigned to substation s in the increasing order of their distances to substation s until substation s satisfies its minimum generation capacity requirement. In the created synthetic network, only conventional coal- and gas-fueled power plants are allowed to be located in the same substations with loads.

The remaining power plants are clustered into N_g Type g substations with only generators in a similar way, in which the postal codes are grouped into $N_l + N_b$ load substations. An additional constraint is enforced such that any hydro, nuclear or renewable energy resource cannot be grouped together with a power plant of a different fuel type. Table II summarizes the generator clustering algorithm.

III. TRANSMISSION LINES

Next, the substations are connected with synthetic transmission line topologies at multiple nominal voltage levels. For the approach of this paper, each substation already has load, generation, and geographic coordinates determined from the previous step, and this data is available and used as the transmission line topology is generated.

The network designer will select one or more nominal voltage levels for the synthetic network. Most substations will contain buses at the lowest level, and a percentage of the substations will also contain a bus at higher voltage levels. The nominal voltage level selection and percentage of substations containing buses at each level vary widely among real power systems, and so these will often be chosen according to the desired characteristics of

a synthetic network. Typically, 10-20% of substations contain a bus at a higher system voltage level, and these may be chosen randomly with probabilities proportional to load. Large generating substations are significantly more likely to be at a higher voltage level as well. Transformers are added within substations to connect the nominal voltage levels. A substation's loads are usually connected to the lowest voltage level, and the generators are often connected to the highest voltage level at a substation, through a step-up transformer.

Once the substations are assigned buses at various nominal voltage levels, these voltage levels are connected with a set of transmission lines. An automated line placement process is necessary to build large synthetic systems in reasonable time. The approach of this paper is to match a set of network properties which characterize the network's graph topology, geometry, and power flow solution. These quantitative metrics can be used for building synthetic networks as well as for validation criteria of created networks.

A. Transmission Line Parameters

Transmission line electrical parameters required for ac power flow analysis include series impedance, shunt admittance, and MVA limits. Realistic per-distance parameters are assigned to synthetic lines based on datasheets and references, appropriate to the assigned nominal voltage level V_{line}, for both overhead lines and underground cables. For the line length l, the straight line distance between the assigned substation geographic coordinates is used. The actual right of way distance for a real transmission line is always longer than this, but the point to point distance suffices as an approximation.

The majority of actual lines in the transmission system are overhead lines, and therefore most synthetic lines are assigned parameters realistic to overhead lines. In building synthetic over-

head lines, first, a few candidate conductor types are selected typical to the assigned nominal voltage level, following the conventions of [18]. Data for each conductor includes outer radius r_o, geometric mean radius r', per-distance resistance r at 50°C, and the maximum current carrying capacity I_{max}. Transmission line capacities can be improved with conductor bundling, and this is common at higher voltage levels such as 345 kV and 500 kV. Synthetic lines at these voltage levels are given bundled conductors with b conductors per bundle, spaced at 18 inches. The final selection of conductor from the realistic choices is made based on estimated capacity requirements, as determined by the dc power flow in Section III-D. For the assigned voltage level, a list of candidate transmission tower configurations is also created using the variety of towers found in [19], [20]. One tower style is randomly selected for each line from the realistic candidates at that voltage level. From the tower data, the geometric mean distance GMD between phases is calculated, depending on the arrangement of the lines on the tower.

The overhead line parameters R, X, B, and MVA_{max} are calculated as follows [18]–[20]. f is the system frequency, 60 Hz for North America, and ϵ is the permittivity of free space, $1.42461 \times 10^{-8} \frac{F}{mi}$. For bundled conductors, r' and r_o will be replaced by equivalent bundle spacing values D_{SL} and D_{SC}.

$$R = \frac{r \times l}{b} \tag{2}$$

$$X = 2\pi f \times 0.0003218 \times \ln\left(\frac{GMD}{r'}\right) \times l \tag{3}$$

$$B = 2\pi f \times \left(\frac{2\pi\varepsilon}{\ln\left(\frac{GMD}{r_o}\right)} \times l\right)^{-1} \tag{4}$$

$$\text{MVA}_{max} = \sqrt{3} \times I_{max} \times V_{line} \times b \tag{5}$$

Underground transmission lines are less common due to their high cost, but are sometimes used in urban settings where overhead right of ways are costly, unavailable, or are aesthetically problematic [21]. They are distinguished from overhead lines in their lower series impedance and much higher shunt charging admittance due to the tight bundling of the phase conductors. A few synthetic lines are assigned parameters as underground cables if their distance is short and they have a large load at either end, characteristics of a metropolitan setting. Similar to the overhead lines, a set of candidate conductors and cable configurations are derived from reference manuals [22], [23] for the associated voltage level. All cables are assumed to be cross-linked polyethylene (XLPE), also known as solid dielectric cables. Two cable material types are considered – copper conductor with lead sheath and aluminum conductor with aluminum sheath – and three conductor sizes for each material type. The final choice for conductor depends on the estimated capacity needed, and there is some randomization in the cable configuration. 75% of synthetic cables are configured to be pulled in a duct, while the rest are directly buried in the dirt [21]. Both installations assume earthing at a single point and a trefoil phase layout. Line parameters are calculated with equations similar to (2)–(5).
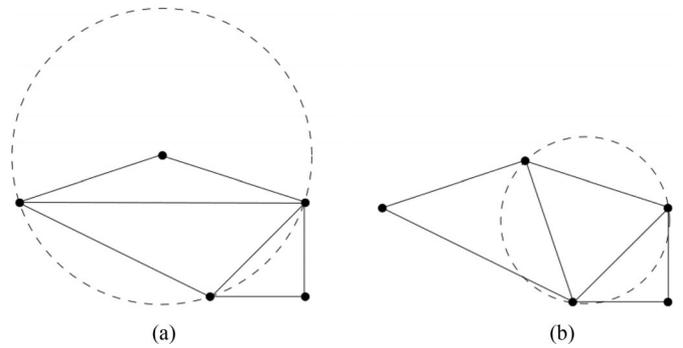

Fig. 3. Triangulation of a set of 5 points. (a) is not the Delaunay triangulation, because at least one triangle's circumcircle contains another point. (b) is the Delaunay triangulation of these points.

B. Basic Topological Characteristics

An electric power transmission system is represented by a connected graph with n substations as nodes and m transmission lines as edges. The approach of this paper is to build a connected graph at each voltage level, and for the combined graph at all voltage levels to remain connected even if one substation is removed. This property does not allow for radial substations and enhances contingency behavior. Many real power systems match these properties.

A previous analysis of the graphs of real North American interconnected power systems shows that n and m are highly linearly correlated [11]. The study observed that $m = 1.22\,n$ is a good approximation at any voltage level above 100 kV of the number of edges connecting substation nodes. Note that this figure does not include branches within substations, or multiple parallel lines between the same pair of substations. The figure is within the range given by other previous studies [3], [8], [9].

This paper proposes an algorithm which adds lines for all voltage level networks at the same time, iteratively inserting lines until there are $1.22\,n$ lines added at each nominal voltage level. A penalty-based system is used to rank the candidate lines during the selection process. As long as the connectivity requirements mentioned above are not yet met, heavy penalties are given to candidate lines which do not contribute to connectivity under a single node removal. These penalties dominate the early part of the line placement algorithm until these topological criteria are satisfied. The other properties described below increase the intelligence of the line selection process by adding other penalties, so that the final network will be more realistic and match multiple properties found on real power systems.

C. The Delaunay Triangulation and Minimum Spanning Tree

The Delaunay triangulation was applied to the synthesis of transmission networks in [11]. This graph is calculated from a set of coordinates, dividing the plane into triangles, in which no triangle's circumcircle contains another point [24]. For large sets the graph will contain approximately $6\,n$ line segments. The Delaunay triangulation is illustrated in Fig. 3. The Euclidian minimum spanning tree, which connects all points at the shortest possible distance with $n - 1$ segments, is guaranteed to be a subset of the Delaunay triangulation [24]. Previous work

TABLE III
CORRELATION BETWEEN DELAUNAY TRIANGULATION
AND ACTUAL TRANSMISSION LINES

Transmission Line Category	Average percentage for Eastern Interconnect	Average percentage for Western Interconnect
Minimum Spanning Tree	47.8%	44.3%
Delaunay Triangulation	75.6%	71.1%
Delaunay 2 neighbor	18.3%	21.5%
Delaunay 3 neighbor	4.6%	5.3%
Delaunay 4 neighbor	1.1%	1.4%
Delaunay 5+ neighbor	0.4%	0.7%

showed that actual power systems' scaling properties for average shortest path length and clustering coefficient are better matched by the Delaunay triangulation than a randomized approach such as using Small World networks [11].

For the graphs of transmission line networks of various utilities in the Eastern and Western interconnects in North America, the transmission lines connect geographically located substations. A comparison was made for each real transmission line to the Delaunay triangulation of the substations at that voltage level. Table III shows the results averaged over 28 Eastern networks and 15 Western networks, each over 100 kV with more than 80 substation nodes. The percentages indicate how many transmission lines fall on the minimum spanning tree or Delaunay triangulation. For those that fall on neither, the neighbor distance is given along the segments of the Delaunay triangulation. For a line that is a Delaunay 3 neighbor, the shortest path between the line endpoints along the Delaunay triangulation is three segments. The minimum spanning tree segments are also on the Delaunay triangulation. This shows that overlap with the Delaunay triangulation and its near neighbors is a good indicator of real power systems.

A synthetic power system's correlation with the Delaunay triangulation at each voltage level can be an important indicator of how realistic the transmission line topology is. It can easily be integrated into the penalty-based iterative line placement algorithm by setting a quota of lines which may be added from each category: minimum spanning tree, Delaunay, Delaunay 2 neighbor, and so forth. To ensure adequate progress toward the quotas, during the line placement process penalties are added to candidate lines in categories that exceed the proportion allowed. Each nominal voltage level will have its own set of Delaunay categories with its own set of quotas.

In fact, using the Delaunay triangulation can be a great aid to building synthetic transmission line topologies. There are approximately n^2 possible lines connecting a set of n substations. Restricting candidate lines to the Delaunay 3 neighbors and closer cuts the number down to about $23 n$, dramatically reducing the number of candidate lines to analyze, especially for networks with more than 1000 substations. This first-level pruning eliminates many candidate transmission lines which are statistically very unlikely due to geographic constraints.

D. DC Power Flow Analysis to Aid Convergence

Real power system models have convergent ac power flow solutions. Using a dc power flow solution as part of the iterative line placement algorithm can increase the solvability of cases, since lines are placed in locations that facilitate connecting generation with load. The standard dc power flow [25] solves for a vector of system bus voltage angles, $\boldsymbol{\theta}$, with known real power injection $\boldsymbol{P_{inj}}$ and system susceptance matrix $\boldsymbol{B}$. It is solved as

$$\boldsymbol{\theta} = -\boldsymbol{B}^{-1}\boldsymbol{P_{inj}}. \tag{6}$$

As long as the system is connected, meaning $\boldsymbol{B}$ is not a singular matrix, the dc power flow will have a solution. Any candidate line can then have an estimated power flow value P_{est}, provided the line length and per-distance impedance d_{21} and x_l are known:

$$P_{\text{est}} = x_l \cdot d_{21} \cdot (\theta_2 - \theta_1) \tag{7}$$

where θ_1 and θ_2 are the voltage angles of the buses this line would connect. A negative penalty is given to candidate lines, proportional to P_{est}. This incentivizes power flow corridors in the automatic line placement algorithm, subject to the other system constraints.

To ensure a connected network before enough transmission lines have been added, the dc power flow step adds an additional set of temporary impedances corresponding to lines forming the substations' Euclidian minimum spanning tree. These impedances are in addition to the actual transmission lines as they are added, and the impedance magnitudes are increased throughout the line placement process until the graph is connected by actual transmission lines, at which point these temporary impedances will be removed. The temporary impedances serve to connect the graph while the in-progress transmission line topology is still disconnected, allowing dc power flow solutions even in the earliest stages.

E. Geographic Line Intersections

Sometimes, two transmission lines of the same voltage level will intersect, or pass by one another, external to a substation. These can be found by simple geometry if the substation coordinates are known at the line endpoints. If all the lines come from the Delaunay triangulation as in [11], there can be no intersections. However, matching the 2 neighbors and 3 neighbors of the Delaunay triangulation, as is done in this paper, allow for lines to intersect one another.

An analysis of line intersections in high-voltage networks of the Eastern and Western Interconnects in North America indicates that line intersections do occur, however they are uncommon, especially for higher voltage levels. The line algorithm will often intrinsically produce line intersection rates larger than is realistically seen, so a penalty is added to candidate lines that intersect existing lines, so that the intersection rate can be made to match that of actual cases.

IV. ADDITIONAL POWER SYSTEM COMPLEXITIES

Once a test case has an ac power flow solution with buses, generators, loads, transformers, and transmission lines, additional complexities can be added to improve the realism of the case and include data necessary for various types of studies.

Voltage control of real power systems involves a variety of reactive power compensation and other devices: shunt capacitor banks, transformer taps, static var compensators, and synchronous condensers. A recent Eastern Interconnect model contained about 19 000 buses where the voltage was regulated by a device such as a generator, tap-changing transformer, or switched shunt. Of these, about 1100 devices regulate a bus other than the one to which they are connected. Over 500 buses are regulated by multiple devices, with 12 regulated by eight or more devices. A simplified method for synthetic voltage control, used in the example of this paper, is to identify buses in the created network which have per-unit voltage magnitudes outside the target range of 0.97-1.05. Iteratively, manually add shunt capacitor banks or shunt reactors on some of these buses to adjust the voltage to the desired level.

A variety of specialized studies such as transient stability can be done with added data. An illustrative example of adding transient stability devices would be to model each synthetic synchronous generator with a classical model, as defined in [26], with parameters for inertia H and transient impedance X'_d assigned based on a random distribution for real generators of that fuel type. Further research can expand this framework to properly configure the dynamic models for synchronous machines, turbine-governor systems, excitation systems, wind turbines, complex loads, and system stabilizers.

Other types of specialized studies require additional data. Economic studies require cost and pricing, which are often quite confidential. Synthetic cases with realistic cost data can be used for optimal power flow and unit commitment studies. Reliability studies require more line and generator information about contingency limits and failure rates, among other information. Geomagnetic disturbance analysis requires dc resistance data, transformer configuration, and other additional device parameters. The authors' previous work describes applying synthetic networks to geomagnetic studies [11]. Each of these specialized applications can augment the basic set of buses, generators, loads, and branch topology to meet these particular needs on real geographic footprints with large, realistic, freely-available test cases.

V. EXAMPLE: 2000 BUS CASE

The preceding methods, depicted in Fig. 4, are applied in this section to build an example 2000-bus test case. The geographic footprint selected is the part of the U.S. state of Texas served by the Electric Reliability Council of Texas (ERCOT), which comprises most of the state. Eight geographic areas are selected, and the substation count is specified as 1500. Two nominal voltage levels are used, 345 kV and 115 kV.

The substation placement process began with the 2300 postal codes within the ERCOT service territory, and the 500 generators reported for this area [14], [17]. For 1500 substations, the postal codes and generators are clustered into substations as described in Section II, with $1500\,(N_s) = 1333\,(N_l) + 83\,(N_g) + 84\,(N_b)$, corresponding to the fractions of these types of substations found in real cases. For each substation, an area-dependent power consumption per capita is assigned, as shown in Table IV, with the total system load of 49,776 MW.
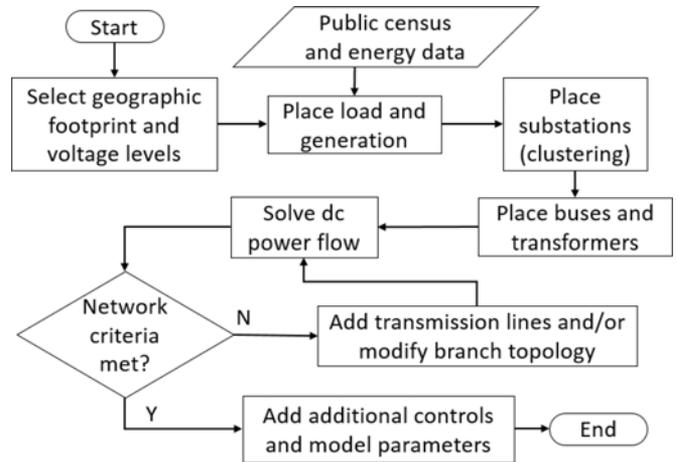

Fig. 4.   Flowchart of network synthesis process.

TABLE IV
POWER CONSUMPTION PER CAPITA BY AREA

Area	Per-capita consumption, kW	Area	Per-capita consumption, kW
COAST	2.362	NORTH	1.859
EAST	1.563	SCENT	2.045
FWEST	1.854	SOUTH	1.843
NCENT	2.438	WEST	2.368

TABLE V
PENALTY STRUCTURE FOR LINE PLACEMENT IN 2000 BUS CASE

Criterion	Amount of penalty for each candidate line segment
Distance	$+2$ per mile
Power Flow	$-0.5 \times P_{est}$ based on (7)
Category	$+200$ if ahead of quota for the segment's category: minimum spanning tree, Delaunay triangulation, or 2 or 3 neighbor of the Delaunay triangulation
Voltage level connectivity	-300 if segment contributes to connectivity at single voltage level, or connects to a radial substation
Overall connectivity	-1000 if segment contributes to full system connectivity under single node removal
Intersections	$+500$ if segment intersects with existing line

Load power factor is fixed at 0.96. By area, generators are given an initial dispatch proportional to load.

All 1500 substations are given a 115 kV bus, and 225 (15%) are given a 345 kV bus. These are selected randomly, with larger generators and load given higher probability. Transformers connect these buses within each substation. Generators are connected from a 13.8 kV generator bus to their substation's highest voltage bus through a generator step-up transformer. Loads are connected to the lowest system voltage bus in the substation. Per-distance parameters are determined for typical 345 kV and 115 kV overhead and underground transmission lines, with several variations and conductor upgrades. Lines which are shorter than 8 miles and connect at least 200 MW of load are considered underground.

The transmission line placement process adds 287 lines at 345 kV and 1813 lines at 115 kV. Of these lines, 231 are added as double circuits, 42 are added as triple circuits, and 22 are
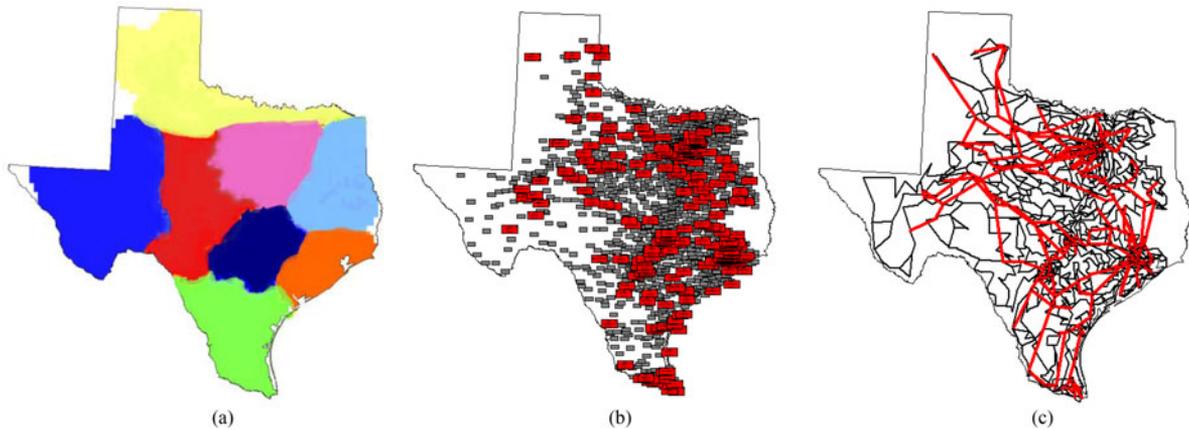
Fig. 5. The development of the 2000-bus case. (a) Eight areas, color-coded. (b) Substation placement. (c) Transmission lines added. Red substations and transmission lines are 345 kV; gray substations and transmission lines are 115 kV.

TABLE VI
STRUCTURAL STATISTICS OF THE 2000-BUS CASE

Structural Statistic	Target Quota	345-kV network	115-kV network
n	- -	225	1500
m	- -	287	1813
m/n	1.22	1.276	1.209
Minimum spanning tree	50%	49.1%	50.6%
Delaunay Triangulation	70%	69.7%	70.9%
Delaunay 2 neighbor	25%	24.7%	24.7%
Delaunay 3 neighbor	5%	5.6%	4.4%
Line intersections	< 20%	6.5%	9.1%

added as quadruple circuits. The lines are all added from the set of line segments including the Delaunay triangulation and any pair of substations separated by three or fewer Delaunay segments. At each of 420 iterations, each of these candidate lines is ranked according to a set of penalties as shown in Table V, corresponding to the observations in Section III of this paper, and the five highest-ranked segments are added to the network. Based on the dc power flow results, conductors were selected for transmission lines, and some lines were upgraded to multiple circuits. Fig. 5 shows the 2000-bus case at different stages of development. The computation time for building the line topology of this case was about 5 minutes.

The synthetic network has no relation to the actual grid in Texas, except that the load and generation profiles are similar. The transmission line topology meets validation criteria described in Section III in the form of structural statistics gathered from the Eastern Interconnect. The combined graph of both voltage levels is fully connected and remains so under a single node removal. Each individual voltage level network is fully connected. The number of transmission lines in proportion to the number of substations is close to the 1.22 target for both networks. Table VI shows structural statistics for this case, including the proportion of lines at each voltage level which are from the substations' minimum spanning tree, Delaunay triangulation, and the Delaunay neighbors with distance 2 and 3. Comparing these proportions to those from Table III shows that
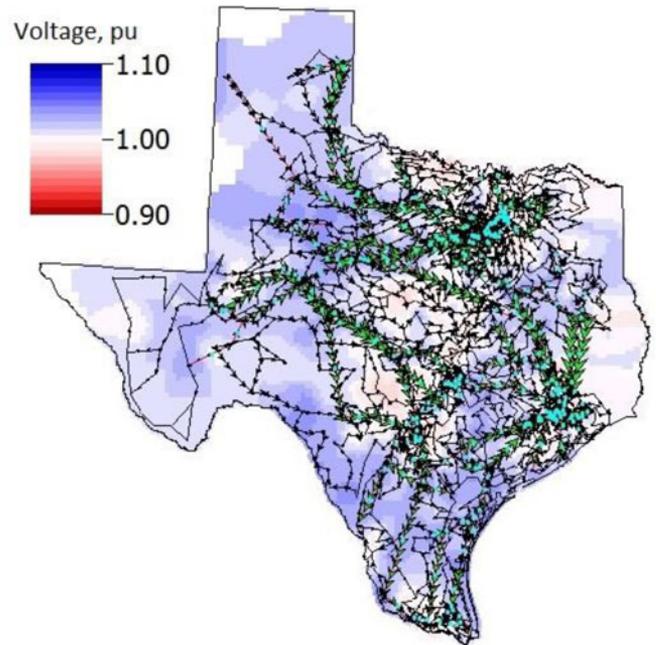

Fig. 6. Voltage profile contour of the 2000-bus case.

the proportions of each category are close to target. Table VI also shows the line intersection rate, which the intersection penalty has brought below the maximum thresholds normally seen in actual cases.

The resulting case has an ac power flow solution without any manual intervention. Some low-voltage areas benefit from 36 manually added shunt capacitor banks. The generator setpoints were modified manually as well for some units, and five shunt reactors were added in areas that had overvoltages. The final case has no lines loaded more than 85%, nor voltages outside of the 0.97-1.05 per-unit range. Fig. 6 shows the ac power flow solution with the system voltage profile contour [27]. The case specifications and benchmark power flow results are publicly available online [12].

VI. CONCLUSION AND FUTURE WORK

Synthetic networks built with the methodology of this paper can equip power systems research with high-quality public test cases which match the size and complexity of actual grids but do not pose data security concerns. The substation placement locates the synthetic grid on familiar geography, with load corresponding to population and generation plants coming from public records. The clustering process combines load and generation information into a set of substations with proper proportions of generation-only, load-only, and mixed substations, subject to realistic considerations with regard to a generator's fuel type. The variety expressed in these synthetic substations add to the realism of these cases.

The transmission line topology extends previous work, adding a variety of structural criteria met by both real networks and the product of the proposed methodology. The use of the Delaunay triangulation greatly reduces the number of candidate lines which need to be analyzed and ensures the synthetic networks will reflect real systems' geographic constraints. A method which iterates over a dc power flow solution encourages lines that will contribute to convergence. Line intersections are also considered.

The paper also introduces additional power system complexities which can be addressed in future work to augment the power flow models of synthetic grids with components critical to specific research areas. Cases suitable for voltage control, transient stability, economics, reliability, and geomagnetic disturbance analysis are all possible in this framework.

The 2000 bus case illustrates the capabilities of the network synthesis methodology. Complete with branch parameters, geographic coordinates, load and generation profiles, and ac power flow benchmark solution results, this case is suitable for cross-validating research studies on a larger scale than most existing public test cases.

REFERENCES

[1] Q. Zhou and J. W. Bialek, "Approximate model of European interconnected system as a benchmark system to study effects of cross-border trades," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 782–788, May 2005.

[2] Z. Wang, R. J. Thomas, and A. Scaglione, "Generating random topology power grids," in *Proc. 41st Annu. Hawaii Int. Conf. Syst. Sci.*, 2008, pp. 183–183.

[3] 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*, vol. 1, no. 1, pp. 28–39, Jun. 2010.

[4] Z. Wang, A. Scaglione, and R. J. Thomas, "The node degree distribution in power grid and its topology robustness under random and selective node removals," in *Proc. 2010 IEEE Int. Conf. Commun. Workshops*, 2010, pp. 1–5.

[5] Z. Wang, S. H. Elyas, and R. J. Thomas, "A novel measure to characterize bus type assignments of realistic power grids," in *Proc. 2015 IEEE Eindhoven PowerTech*, 2015, pp. 1–6.

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

[7] M. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, Jan. 2003.

[8] E. Cotilla-Sanchez, P. D. H. Hines, C. Barrows, and S. Blumsack, "Comparing the topological and electrical structure of the North American electric power infrastructure," *IEEE Syst. J.*, vol. 6, no. 4, pp. 616–626, Dec. 2012.

[9] G. A. Pagani and M. Aiello, "The power grid as a complex network: A survey," *Phys. A, Statist. Mech. Appl.*, vol. 392, no. 11, pp. 2688–2700, Jun. 2013.

[10] K. M. Gegner, A. B. Birchfield, T. Xu, K. S. Shetye, and T. J. Overbye, "A methodology for the creation of geographically realistic synthetic power flow models," in *Proc. 2016 IEEE Power Energy Conf. Illinois*, Champaign, IL, USA, Feb. 2016, pp. 1–6.

[11] A. B. Birchfield, K. M. Gegner, T. Xu, K. S. Shetye, and T. J. Overbye, "Statistical considerations in the creation of realistic synthetic power flow models for geomagnetic disturbance studies," *IEEE Trans. Power Syst.*, to be published.

[12] (2016). [Online]. Available: http://icseg.iti.illinois.edu/synthetic-power-cases/texas2000-june2016/

[13] Z. Wang and R. J. Thomas, "On bus type assignments in random topology power grid models," in *Proc. 2015 48th Hawaii Int. Conf. Syst. Sci.*, 2015, pp. 2671–2679.

[14] U.S. Census Bureau. 2010 Census Gazetteer Files: ZIP Code Tabulation Areas. (2010). [Online]. Available: https://www.census.gov/geo/maps-data/data/ gazetteer2010.html

[15] R. Xu and D. Wunsch, "Partitional clustering," in *Clustering*. Hoboken, NJ, USA: Wiley, 2009, pp. 67–72.

[16] Electric Reliability Council, Texas, Inc., "Hourly load data archives." (2016). [Online]. Available: http://www.ercot.com/gridinfo/load/

[17] U.S. Energy Information Association, Form EIA-860, 2014. [Online]. Available: http://www.eia.gov/electricity/data/eia860/index.html

[18] J. D. Glover, T. J. Overbye, and M. S. Sarma, *Power System Analysis and Design*, 6th ed. Boston, MA, USA: Cengage Learning, 2017.

[19] General Electric Company. *EHV Transmission Line Reference Book*. New York, NY, USA: Edison Electric Institute, 1968.

[20] Electric Power Research Institute (EPRI). *Transmission Line Reference Book, 345 kV and Above*. Palo Alto, CA, USA: Elect. Power Res. Inst., 1975.

[21] Public Service Commission of Wisconsin, "Underground electric transmission lines," Public Service Commission of Wisconsin, Madison, WI, USA, 2011.

[22] Silec, "High and extra-high voltage underground solutions," General Cable Technol. Corp., Highland Heights, KY, USA, 2013.

[23] ABB Inc. High Voltage Cables, "XLPE AC land cable systems user's guide," ABB Inc., Huntersville, NC, USA, 2012.

[24] F. M. Preparata and M. I. Shamos. *Computational Geometry: An Introduction*. New York, NY, USA: Springer-Verlag, 1985.

[25] A. J. Wood, B. F. Wollenberg, and G. B. Sheble, "The 'DC' or linear power flow," in *Power Generation, Operation, and Control*, 3rd ed. Hoboken, NJ, USA: Wiley, 2014, ch. 6, p. 277.

[26] P. W. Sauer and M. A. Pai, *Power System Dynamics and Stability*. Champaign, IL, USA: Stipes Publishing, 2007, p. 103.

[27] J. D. Weber and T. J. Overbye, "Voltage contours for power system visualization," *IEEE Trans. Power Syst.*, vol. 15, no. 1, pp. 404–409, Feb. 2000.

Adam B. Birchfield (S'13) received the B.E.E. degree from Auburn University, Auburn, AL, USA, in 2014. He is currently working toward the M.S. degree in electrical and computer engineering at the University of Illinois at Urbana-Champaign, Urbana, IL, USA.

Ti Xu (S'13) received the B.S. degree from Tsinghua University, Beijing, China, in 2011, and the M.S. degree from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, in 2014, where he is currently working toward the Ph.D. degree in electrical and computer engineering.

Kathleen M. Gegner (S'12) received the B.S. degree in electrical engineering from the University of Nebraska-Lincoln, Lincoln, NE, USA. She is currently working toward the M.S. degree in electrical and computer engineering at the University of Illinois at Urbana-Champaign, Urbana, IL, USA.

Komal S. Shetye (S'10–M'11) received the B.Tech. degree from the University of Mumbai, Mumbai, India, in 2009, and the M.S. degree in electrical engineering from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, in 2011. She is currently a Senior Research Engineer in the Information Trust Institute, University of Illinois at Urbana-Champaign.

Thomas J. Overbye (S'87–M'92–SM'96–F'05) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison, Madison, WI, USA. He is currently the Fox Family Professor of Electrical and Computer Engineering with the University of Illinois at Urbana-Champaign, Urbana, IL, USA.