Modeling Electric Grid Topology with Spatially-Aware Degree-Corrected Stochastic Block Model

Sanjana Kunkolienkar, Jonathan Snodgrass, Adam Birchfield, Thomas Overbye
Department of Electrical and Computer Engineering
Texas A&M University, College Station, TX, USA
{sanjanakunkolienkar, snodgrass, birchfield, overbye}@tamu.edu *

Abstract

The goal of this paper is to facilitate the development of synthetic electric grid topologies that replicate the structural properties of real-world power systems. This paper demonstrates that the topology of the North American transmission grid can be modeled using a Spatially-Aware Degree-Corrected Stochastic Block Model (SA-DCSBM), which captures three key features in real grids: modularity, heterogeneous node degree distributions, and distance-constrained connectivity. Once the model is fitted to the North American transmission network data, synthetic topologies (excluding electrical phenomena) are generated to demonstrate that they accurately reproduce real grid statistics across multiple structural dimensions, including modularity, edge length distribution, degree heterogeneity, and spectral robustness. The SA-DCSBM thus offers a modeling framework for creating high-fidelity synthetic electric grid topologies that preserve spatial and structural realism.

Keywords: Degree-corrected stochastic block model, electric grid, topology, vulnerability, complex networks.

1. Introduction

Electric grid research engineers are increasingly employing synthetic grid models to test algorithms and simulate planning scenarios. These models are especially important when real grid data are unavailable due to privacy concerns [1],[2]. A useful synthetic grid should resemble the real network in its underlying structure, that is, its connectivity patterns, modularity, and spatial organization.

While several synthetic grid models and test cases have been developed [3], [4], [5], [6], [7], this paper proposes a stochastic extension to existing synthetic grid generation algorithms to add modeling realism for intra-area and inter-area transmission lines.

1.1. Goal and Approach

The goal of this work is to support the creation of synthetic transmission networks that closely replicate the structural features of real-world electric grids. Specifically, this paper investigates whether networks in the North American Eastern Interconnect (EI) and the North American Western Interconnect (WECC) exhibit the structure implied by a Spatially-Aware Degree-Corrected Stochastic Block Model (SA-DCSBM), and whether such a model can be used to generate realistic grid topologies.

The Stochastic Block Model (SBM) [8] models network connectivity based on group membership, assuming all nodes within a group have identical degrees. However, real-world networks often exhibit both community structure and significant degree heterogeneity. To address the limitation of the lack of degree heterogeneity in SBMs, the Degree-Corrected Stochastic Block Model (DCSBM) [9] incorporates node-specific parameters that capture variations in degree, making it better suited for empirical networks. SBMs and their variations have been used to study community structures in social networks [10].

As described in [11], the *edge-probability matrix* $\mathbf{B}$ defines the probability of an edge existing between two nodes, based on the communities to which the nodes belong. That is,

$$P(\mathbf{A}_{ij} = 1) = \mathbf{B}_{z_i, z_j} \tag{1}$$

where $\mathbf{A}$ is the adjacency matrix. $\mathbf{B}_{z_i, z_j}$ is the edge-probability matrix where $i \in$ community z_i and $j \in$ community z_j. Then, intra-community edges have nodes within the same community z and connect with probability $\mathbf{B}_{zz} = \rho_{\text{intra}}$. The inter-community edges have nodes in different communities z, z' and connect with probability $\mathbf{B}_{zz'} = \rho_{\text{inter}}$. So,

$$\mathbf{B} = \begin{pmatrix} \rho_{intra} & \rho_{inter} \\ \rho_{inter} & \rho_{intra} \end{pmatrix}. \tag{2}$$

*This work is completed as part of the "Electric Grid Resilience" Project 60NANB24D210, funded by the Consolidated Appropriations Act, 2024, and administered through the National Institute of Standards and Technology of the US Department of Commerce.

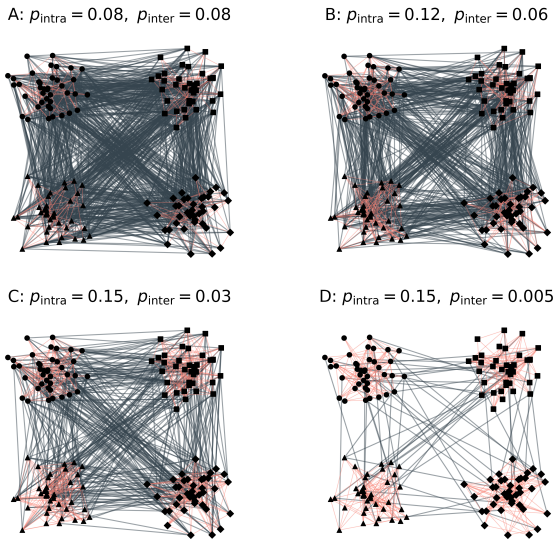

Figure 1. Effect of decreasing off-diagonal probabilities in a stochastic block model with four equal communities. All panels share the same node positions. As the inter-block probability p_{inter} decreases while p_{intra} increases and then stays constant, the network transitions from random-like connectivity (A) to well-separated communities (D).

The SA-DCSBM extends the classical Stochastic Block Model (SBM) [12] by incorporating three key features observed in electric grid networks: degree propensities that ensure nodes reproduce observed heterogeneity, block affinities that encode modular structure, and a spatial decay function that penalizes long-distance links. Together, these components yield networks that preserve the modular, heterogeneous, and spatially embedded character of real transmission grids.

1.2. Importance of Spatial Realism and Existing Grid Creation Approaches

Spatial realism is essential because many grid risks are geographically driven [13]. Large-scale weather events span multiple states, so synthetic networks with misrepresented distances cannot recreate correlated outages. Geomagnetic disturbance studies rely on line length and orientation, which govern induced currents and transformer stress. Cascading-failure simulations are likewise sensitive to physical layout, as long-distance ties can accelerate overload propagation [14]. To support credible planning and analysis, a synthetic grid topology must reproduce actual line lengths and inter-regional connections.

Progress toward modeling a realistic transmission network is attributed to several model families. One of the first spatial graph generators was introduced in [15], where substations were placed using Poisson or uniform sampling, and edge creation was limited by distance. A two-stage growth process was proposed in [16]: first, a cost-efficient backbone satisfying demand was laid out; second, additional lines were inserted to match empirical degree, path-length, and clustering statistics, all while respecting node coordinates. Metrics such as spatial densities, line lengths, and degree histograms from a target grid are measured in [17]. Then, edges are rewired until those distributions are replicated.

Anonymized real systems are used in [4] by fragmenting the networks, reassembling a user-defined mix, and adjusting limits through AC optimal power-flow studies to pass N-1 security tests. In [18],[19], candidate lines are selected from Delaunay neighbors; an iterative penalty–reward search then adds or removes branches, balancing line-length cost against contingency sensitivities. A historical-evolution model in [20] is optimized for construction cost and robustness, and anchors all decisions to actual geography.

Existing grid creation algorithms provide valuable test cases and have advanced the generation of realistic topologies. Building on this foundation, our approach embeds regions as explicit structural constraints during network construction by enforcing inter-area sparsity. This feature, observed in real-world transmission grids as a reduced density of ties between regions [21], is a structural property that the proposed algorithm is designed to preserve.

1.3. Contributions

This paper makes the following contributions:

1. It shows that transmission grids in the EI and the WECC are well-approximated by SA-DCSBM;
2. It uses the fitted models to generate new synthetic topologies that preserve node counts and spatial coordinates while approximating observed edge-level and spectral metrics;
3. It establishes SA-DCSBM as the next step in improving the realism of synthetic grid topologies.

2. Electric grid topology and why it fits SA-DCSBM

Electric grids are naturally compatible with the Spatially-Aware Degree-Corrected Stochastic Block Model (SA-DCSBM). Specifically, electric grids show three important features that align well with this model: (1) wide variation in how many connections each node has in every region (degree heterogeneity), (2) modular

structure based on regions, and (3) edge formation constrained by geographic distance. While the formal SA-DCSBM formulation is introduced in the next section, this section focuses on why electric grids are a good fit.

To motivate this, the electrical grid is represented as a graph, $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ is the set of nodes representing substations and $\mathbf{E}$ is the set of edges representing transmission lines. Each node $v \in \mathbf{V}$ is assigned to one of k regions, which are treated as communities in the model. The key notation is summarized below:

n: Total number of nodes (substations),

$i, j \in \mathbf{V}$: Nodes in the graph,

k: Total number of communities (regions),

$g_h, g_k \in \{1, \ldots, K\}$: Community labels,

$\mathbf{A}_{ij}$: Adjacency matrix entry (1 if an edge exists between i and j; 0 otherwise).

Electric grids are operationally partitioned into regions defined by control and planning boundaries. When viewed as graphs, these regions often correspond to topological clusters. The extent to which these clusters form well-separated communities is quantified by the modularity score Q [22],

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(g_i, g_j). \tag{3}$$

Q in Eq. 3 compares the observed number of intra-community edges to the expected number under a random graph with the same node degrees. The value of Q may lie in the range $[-1, 1]$, but in real-world networks it typically falls between 0 and 1. A high Q indicates that most edges fall within communities rather than between them, reflecting strong internal cohesion and weak external connectivity. A low Q would imply that the community partition is either arbitrary or that the network is densely interconnected across groups.

When operational boundaries are treated as community labels, $Q = 0.898$ for the EI and $Q = 0.872$ for the WECC suggest these regions align with meaningful topological separations in the grid. Figure 2 shows the edge count matrix for the top 10 communities in the WECC (left) and the EI (right). The strong diagonal dominance visually confirms that intra-area connectivity dominates, i.e., most edges are concentrated within individual communities. Off-diagonal entries are sparse, indicating a relatively small number of inter-area links. This connectivity pattern is captured by Q, which

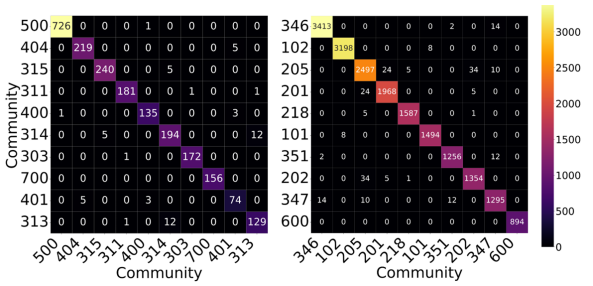

Figure 2. Edge count matrix showing the top 10 regions/communities in the WECC (left) and the EI (right). Strong diagonal dominance confirms that most edges are intra-area, consistent with modular operational structure.

helps explain the high Q scores observed in both interconnections.

In addition to modularity, real grids exhibit degree heterogeneity both globally and locally, within specific regions. Figure 3 illustrates the degree distribution across a few regions in the WECC (top) and the EI (bottom). Each box plot represents the distribution of node degrees within a given area. The consistent presence of long whiskers and outliers, even in regions with relatively low median degree, indicates that some substations have many more connections than others.

This heterogeneity in node degree can be quantified using the degree Gini coefficient, adapted for graph networks in [23]. For a network with n nodes and degrees, $\deg(i)$, sorted in non-decreasing order $\deg_{(1)} \leq \deg_{(2)} \leq \ldots \leq \deg_{(n)}$, the degree Gini coefficient is defined as:

$$G(\mathbf{G}) = \frac{n + 1 - 2\dfrac{\sum_{i=1}^{n} (n + 1 - i) \deg_{(i)}}{\sum_{i=1}^{n} \deg_{(i)}}}{n - 1}. \tag{4}$$

The Gini coefficient ranges from 0 to 1, where 0 indicates perfect equality (i.e., all nodes have the same degree), and 1 indicates maximum inequality (i.e., one node has all the connections, and the rest have none).

Figure 3 illustrates degree heterogeneity across regions, showing variability within and between regions. Figure 2 highlights strong intra-area connectivity, reinforcing the modular character of the grid. These two properties, degree heterogeneity and modularity, are central to the SA-DCSBM.

To encode them formally, the SA-DCSBM builds on the classical DCSBM [9]. In the DCSBM model, the probability of an edge between nodes i and j for graph $\mathbf{G}$ is:

$$\Pr(\mathbf{A}_{ij} = 1) = \boldsymbol{\theta}_i \boldsymbol{\theta}_j \mathbf{B}_{g_i g_j}. \tag{5}$$

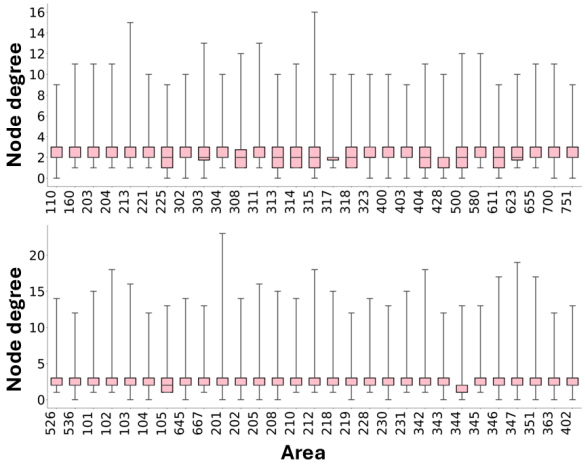

Figure 3. Degree distribution for selected regions in the WECC (top) and the EI (bottom).

Where:

$\mathbf{A}_{ij}$: Adjacency matrix,

$\boldsymbol{\theta}_i$: Vector of node-specific degree parameters.

g_i, g_j: Community labels for nodes i and j,

$\mathbf{B}_{g_i g_j}$: Connection probability between communities.

This model captures degree heterogeneity through the $\boldsymbol{\theta}_i$ parameters, and modular structure through the matrix $\mathbf{B}$. Figure 1 shows how varying the off-diagonal values of $\mathbf{B}$ shifts the model from clearly separated communities to a nearly random structure. When extended to account for spatial constraints, it becomes useful in modeling electric grid topologies.

Electric grids exhibit modularity that is both structural and functional in nature. The functional aspects of these communities have been redefined in [24] and [25], introducing metrics such as electrical coupling strength to indicate strong intra-community connections and weak inter-community connections, as demonstrated in systems with fewer than 500 buses. Additionally, as shown in [26], nodes within topological communities often synchronize together, and their community consistency is correlated with dynamical stability. This provides strong evidence that community-aware generative models are well-suited for producing realistic synthetic grid topologies that preserve both structural and operational properties.

3. Data and Model Fitting

This section describes the datasets used, the voltage configurations, and the methodology for assigning

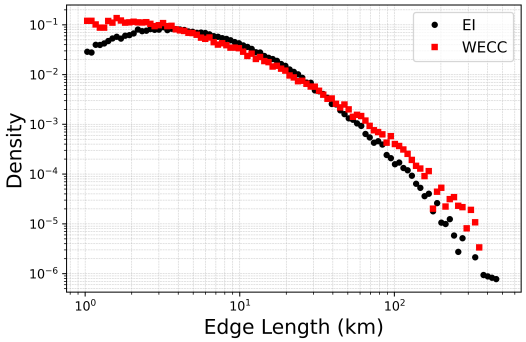

Figure 4. Distribution of edge lengths in the WECC and the EI.

communities that align with regions in the electric grid. The North American transmission grid data is used. This data includes node locations and voltage levels. Each network is converted into an undirected, unweighted graph $\mathcal{G}$, where nodes are substations and edges are transmission lines. Node coordinates are spatial embeddings, and edge lengths are calculated using the haversine formula. Each node is also assigned to an operational area, which serves as its community label. The final graphs capture the hierarchical and modular structure of real power systems.

To examine how structural properties vary with voltage level, the network is analyzed under three configurations: 345kV, 500kV, and 765kV transmission networks, each combined with the Low Voltage (LV) network. The LV network is considered under 300kV. This results in layered variants with identical node sets but different edge sets, allowing a controlled comparison of how high-voltage overlays affect inter-area connectivity and influence spectral, topological, and modular characteristics.

Community labels for SA-DCSBM fitting are taken directly from regions in the data, such as control regions or balancing authorities. These regions are treated as fixed blocks g_i, with each node assigned a label based on the area metadata of its substation. Unlike unsupervised methods that detect communities using spectral patterns or modularity optimization [27], this approach preserves geographic boundaries. As a result, the fitted edge probability matrix $\mathbf{B}$ retains real-world meaning by quantifying how strongly different operational zones are connected.

4. Methodology

4.1. Spatially-Aware DCSBM

Fitting the DCSBM model to a network involves optimizing the degree correction parameters ($\boldsymbol{\theta}_i$), the

block affinity matrix ($\mathbf{B}$), and the block memberships (g_i). In electric grids, block structure is aligned with operator-defined regions that correspond to regulatory, operational, or geographic boundaries, which improves computational efficiency by fixing block memberships a priori, eliminating the need for their inference. Using predefined blocks ensures results remain operationally relevant.

To optimize only the degree correction parameters ($\boldsymbol{\theta}_i$) and the block affinity matrix ($\mathbf{B}$) while fixing g_i to predefined regions, a maximum likelihood approach is employed. The model parameters for classical DCSBM shown in Eq. 5 are optimized by maximizing the log-likelihood of the observed graph:

$$\mathcal{L}(\boldsymbol{\theta}, \mathbf{B}) = \sum_{i,j} \mathbf{A}_{ij} \log \left(\boldsymbol{\theta}_i \boldsymbol{\theta}_j \mathbf{B}_{g_i g_j}\right) \; - \; \sum_{i,j} \boldsymbol{\theta}_i \boldsymbol{\theta}_j \mathbf{B}_{g_i g_j}. \tag{6}$$

The first term accounts for the likelihood of observed edges, and the second term penalizes the expected number of edges given the model parameters, ensuring a normalized probabilistic formulation. To account for geographic distance in edge formation, the proposed spatially-constrained DCSBM modifies the edge probability as follows:

$$\Pr(\mathbf{A}_{ij} = 1) = \boldsymbol{\theta}_i \boldsymbol{\theta}_j \mathbf{B}_{g_i g_j} e^{-d_{ij}/\lambda} \qquad i \neq j. \tag{7}$$

where d_{ij} denotes the haversine distance between nodes i and j, and $\lambda > 0$ is the spatial decay parameter. This form penalizes long-distance edges and encourages local connectivity, in line with the physical constraints of electric grids.

Among candidate heavy-tail kernels, the single-parameter exponential $e^{-d_{i,j}/\lambda}$ captures the prevalence of very short links (maximal at $d = 0$) and provides a smoothly decaying long-distance tail, without the extra tuning required by two-parameter forms such as Weibull, log-normal, or power-law. With only a single scale parameter(λ), the exponential kernel is interpretable, statistically tractable, and aligned with the link-length distributions observed in transmission networks. For these reasons, exponential decay is adopted as the baseline kernel. However, while it is effective for most line lengths, it may underrepresent the heavier tails associated with rare backbone transmission lines.

The Bernoulli log-likelihood under the spatial model then becomes,

$$\mathcal{L}(\boldsymbol{\theta}, \mathbf{B}, \lambda) = \sum_{i<j} \Big[\mathbf{A}_{ij} \ln \left(\boldsymbol{\theta}_i \boldsymbol{\theta}_j \mathbf{B}_{g_i g_j} e^{-d_{ij}/\lambda}\right) + (1 - \mathbf{A}_{ij}) \ln \left(1 - \boldsymbol{\theta}_i \boldsymbol{\theta}_j \mathbf{B}_{g_i g_j} e^{-d_{ij}/\lambda}\right) \Big]. \tag{8}$$

Initial parameter estimates are computed as:

$$\theta_i^{(0)} = \frac{k_i}{\sum_{j; g_j = g_i} k_j}, \quad B_{rs}^{(0)} = \frac{m_{rs}}{\sum_{i \in r, j \in s} e^{-d_{ij}/\lambda}} \tag{9}$$

where m_{rs} is the number of observed edges between blocks r and s. The initial value for λ is set at the mean length of the observed voltage layer network.

The goal is to refine the model parameters such that the predicted edge structure closely matches the observed network. The node-level parameters $\boldsymbol{\theta}_i$ reflect each node's relative importance in forming connections, while the matrix $\mathbf{B}$ captures the inter-block interaction tendencies. Once initialized, the parameters are updated iteratively to maximize the log-likelihood. The contribution of each observed link to the likelihood is computed, and the sum is used to assess the overall model fit.

The optimization is subject to three constraints. To resolve scale ambiguity, $\boldsymbol{\theta}$ is constrained such that the sum within each block equals the total degree:

$$\sum_{i \in g} \theta_i = k_G. \tag{10}$$

All $\mathbf{B}_{g_i, g_j}$ values must remain non-negative to ensure valid probability estimates. Block memberships g_i are predefined and not optimized during the fitting process.

Gradients are then derived to support iterative likelihood maximization:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_i} = \sum_{j \neq i} \left[\frac{\mathbf{A}_{ij}}{\boldsymbol{\theta}_i} - \frac{\boldsymbol{\theta}_j \, \mathbf{B}_{g_i g_j} \, e^{-d_{ij}/\lambda}}{1 - \boldsymbol{\theta}_i \boldsymbol{\theta}_j \mathbf{B}_{g_i g_j} e^{-d_{ij}/\lambda}} \right] \tag{11}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{B}_{rs}} = \sum_{\substack{i: g_i = r \\ j: g_j = s}} \left[\frac{\mathbf{A}_{ij}}{\mathbf{B}_{rs}} - \frac{\boldsymbol{\theta}_i \boldsymbol{\theta}_j e^{-d_{ij}/\lambda}}{1 - \boldsymbol{\theta}_i \boldsymbol{\theta}_j \mathbf{B}_{rs} e^{-d_{ij}/\lambda}} \right] \tag{12}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i<j} \left[\frac{\mathbf{A}_{ij} - \boldsymbol{\theta}_i \boldsymbol{\theta}_j \mathbf{B}_{g_i g_j} e^{-d_{ij}/\lambda}}{1 - \boldsymbol{\theta}_i \boldsymbol{\theta}_j \mathbf{B}_{g_i g_j} e^{-d_{ij}/\lambda}} \right] \frac{d_{ij}}{\lambda^2}. \tag{13}$$

After each iteration, $\boldsymbol{\theta}_i$ is re-normalized to satisfy the constraint in Eq. 10. The stopping criterion is when the relative change in log-likelihood falls below 10^{-6}. The optimal fitting takes 1,500 iterations to converge for the low-voltage network and 100-200 iterations for the different high-voltage networks.

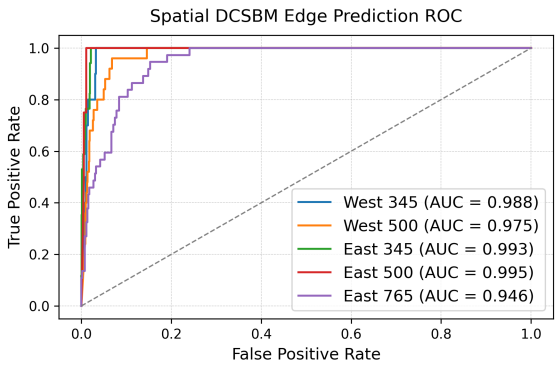

Figure 5. Link prediction using the fitted SA-DCSBM.

Once the spatial DCSBM parameters are fitted, synthetic graphs are generated by sampling edges based on the learned parameters $\boldsymbol{\theta}$, $\mathbf{B}$, and λ. These generated graphs preserve both the modular and spatial structure of the original grid and are validated against real network statistics in Section 5.1.

4.2. Convergence and Goodness-of-Fit

The receiver operating characteristic (ROC) curve and its area under the curve (AUC) provide a measure of goodness of fit for the SA-DCSBM because they directly assess the model's ability to distinguish real edges from non-edges based on learned edge probabilities[28]. Since SA-DCSBM defines a probabilistic distribution over edges incorporating spatial decay, modular structure, and degree correction, a high AUC indicates that the model consistently assigns higher scores to true edges than to false ones. This threshold-independent evaluation is particularly useful for power grid topologies, which are sparse and lack a natural cutoff for edge presence. Figure 5 shows the ROC plot for the proposed model fit.

For the WECC at 345 kV and 500 kV, AUC values of 0.984 and 0.975, respectively, demonstrate separation between existing and non-existent edges. Even higher AUCs are achieved for the EI at 345 kV and 500 kV (0.993 and 0.995), indicating good prediction, while the 765 kV layer shows slightly reduced Q (0.946), reflecting greater spatial heterogeneity at that voltage level. Overall, the near-unity AUCs indicate that the model score provides a good fit to the topology of each layer.

4.3. Grid topologies from the fitted model

The fitted SA-DCSBM synthesizes grid topologies using each node's geographic coordinates, voltage level, and community label as input. The model parameters include node-specific degree factors $\boldsymbol{\theta}$, the inter-block connectivity matrix $\mathbf{B}$, and spatial decay constants λ, which are loaded separately for the high-voltage (HV: 345kV, 500kV, and 765kV) and low-voltage (LV: less than 300kV) layers.

Nodes are first divided into high-voltage (HV) and LV sets based on a specified voltage threshold. Edges are then sampled from the fitted SA-DCSBM model, yielding an expected total edges that is higher than the fixed edge budget, typically set as $E_{\text{tar}} = [1.22n]$ following empirical estimates of grid density [3]. To ensure spatial plausibility and improve computational efficiency, candidate edges within each voltage group are filtered using a K-nearest-neighbor (KNN) search. Each HV node retains its 5 nearest neighbors, while each LV node retains its 15 nearest neighbors. This KNN filtering step reduces the sampling space while preserving the intended probabilistic structure of the SA-DCSBM.

Edge probabilities are computed using the fitted model parameters, combining degree correction factors, inter-block affinities, and the spatial decay $p_{uv} = \theta_{g_u}\theta_{g_v}B_{g_ug_v}e^{(-d_{uv}/\lambda)}$. Additional penalties reduce the probability of forming edges between high-degree nodes or between LV nodes that do not share a common neighbor. These weights are normalized, and edges are sampled without replacement to match the typical edge budget.

Since generative models may not produce a fully connected graph, a final patching step is used to connect any isolated components. This is performed using a Delaunay-based minimum spanning tree that identifies spatially short links between disconnected parts. This step modifies only a small portion of the graph (fewer than 0.1% edges are added) and does not interfere with the original sampling process.

Figure 6 shows the resulting synthetic topology of the SA-DCSBM model on the EI footprint. Most edges remain on land because each HV and LV bus considers only its nearest 4 and 15 neighbors, respectively, while the exponential kernel $e^{-d/\lambda}$ suppresses longer candidates. The affinity matrix $\boldsymbol{B}$ assigns higher probabilities to intra-area than to inter-area links, reinforcing local connections and discouraging water-spanning shortcuts. The resulting topology preserves the spatial, modular, and voltage-layer statistics observed in the real grid.

5. Results

5.1. Evaluation Metrics

Structural metrics are grouped into size, degree distribution, connectivity, and spatial span. As observed in Table 1, node counts match by design, and edge counts differ by at most 7%. The mean degree stays within ±0.2

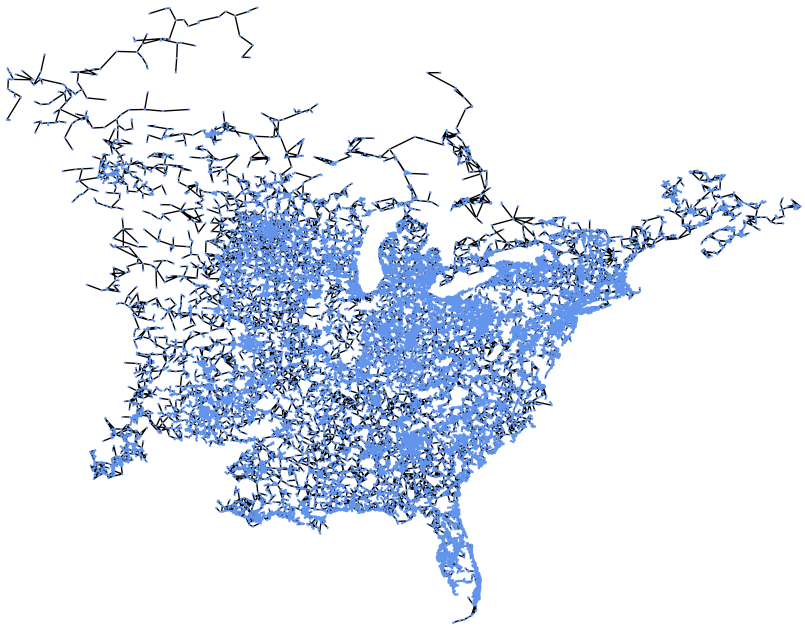

Figure 6. Visualization of SA-DCSBM generated synthetic transmission topology for the EI footprint.

of the real value, and the degree Gini is unchanged, so the overall degree spread is retained.

Table 1. Size and degree distribution metrics for Real and SA-DCSBM grids, averaged across all voltage layers.

Metric	WECC		EI	
	Real	*Model*	*Real*	*Model*
Number of nodes	11630	11630	32913	32913
Number of edges	14835	15711	45585	43886
Mean degree	2.56	2.70	2.77	2.67
Degree Gini	0.30	0.29	0.29	0.29

In graph theory, spectral metrics[29] connect the structure of a network to the eigenvalues of its Laplacian matrix, offering insights into connectivity and robustness. The spectral radius is defined as the largest eigenvalue of the adjacency matrix. It is closely linked to the capacity and density of loops and paths in the network. [29] also highlights that the spectral radius serves as a lower bound for assessing network resilience.

Algebraic connectivity[30] quantifies how well-connected a network is as a whole. A higher algebraic connectivity implies that more edges must be removed to disconnect the system into isolated parts. It is sensitive to the presence of rare but crucial long-distance transmission corridors that bridge otherwise weakly linked regions. In power grid networks, the overall algebraic connectivity is typically low or close to 0.

In Table 2, the spectral radius is reproduced to within 10%, indicating that the model matches the overall scale of the dominant flow paths in the real system. However, the algebraic connectivity is two orders of magnitude lower in WECC topology, revealing that the model underrepresents the long-distance ties that maintain global coherence. These sparse high-capacity corridors bridge distant regions and enhance the system's resilience to fragmentation. Their underrepresentation means that while local clusters remain well-connected, the global backbone network is comparatively weaker in the SA-DCSBM model.

Table 2. Connectivity metrics for Real and SA-DCSBM grids, averaged across all voltage layers.

Metric	WECC		EI	
	Real	*Model*	*Real*	*Model*
Algebraic connectivity	1.13×10^{-4}	1.39×10^{-6}	8.85×10^{-7}	6.52×10^{-7}
Spectral radius	6.31	5.89	5.67	5.46
Clustering coefficient (C)	0.052	0.060	0.066	0.061

The clustering coefficient C, calculated using the formulation in [31], remains of the same order of

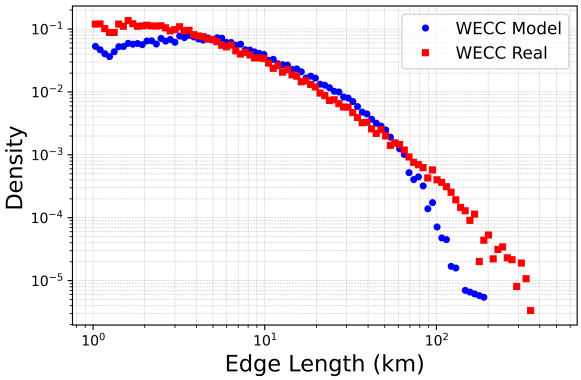

Figure 7. Edge length distribution for the model topology on the WECC footprint.

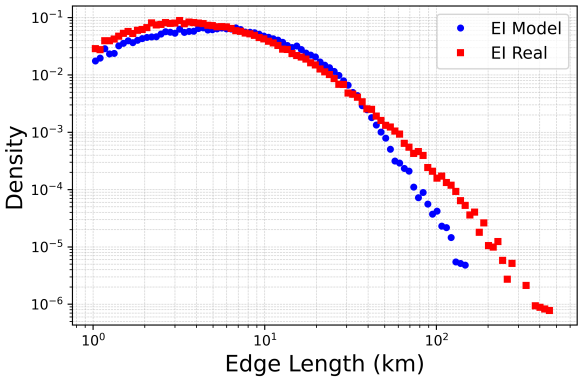

Figure 8. Edge length distribution for the model topology on the EI footprint.

magnitude in the synthetic networks as seen in Table 2. In both the WECC and the EI models, C stays close to the real-world values, indicating that the model preserves local loop density and short-range meshing typical of real transmission grids. A slight overestimation in the WECC and a slight underestimation in the EI suggest that the SA-DCSBM model slightly biases toward more local clustering in compact regions, while underrepresenting meshing in more expansive regions. Mean edge length matches within about a kilometer across layers. Yet, its standard deviation is always smaller, confirming that very long spans are hardest to reproduce, as shown in Figure 7 and Figure 8.

While the exponential kernel offers a straightforward way to bias sampling toward short links and maintain a gentle long-distance tail, it still tends to understate the true variance of the observed distribution when constrained to a fixed edge budget. Since $e^{(-d/\lambda)}$ decays monotonically from $d = 0$, it over-weights short and medium edges relative to long-range edges. As a result, the synthetic network's spread of distances remains tighter than in reality. In practice, recovering the full empirical variance will require adding a heavier-tailed component.

Table 3. Edge length statistics for Real and SA-DCSBM grids, averaged across all voltage layers.

Metric	WECC		EI	
	Real	*Model*	*Real*	*Model*
Edge length mean	13.90	13.87	13.23	12.81
Edge length std	24.95	15.11	17.49	10.83

Thus, the SA-DCSBM preserves node-level heterogeneity, realistic geographic scale, and modularity. However, it does not reproduce the longest ties that enhance algebraic connectivity, which reflects the limitation of the modeling choice. These characteristics can inform future refinements of the SA-DCSBM to capture grid topology parameters better.

5.2. Intra- vs. Inter-Area Edge Patterns

To evaluate how well the SA-DCSBM reproduces modular structure in electric grids, key metrics including modularity Q, intra-area edge fraction ϕ_{intra}, and inter-area cut ratio ϕ_{inter} are computed for both real and synthetic networks across multiple voltage layers and interconnections using,

$$\varphi_{\text{intra}} = \frac{m_{\text{intra}}}{m}, \quad \varphi_{\text{inter}} = \frac{m_{\text{inter}}}{m}, \tag{14}$$

where m is the total number of edges.

Figure 9 shows results for the WECC at 345 kV and 500 kV, and the EI at 345 kV, 500 kV, and 765 kV. In the WECC system, the synthetic graphs closely match the real network. Modularity values differ by about 0.04 (e.g., $Q = 0.893$ vs. $Q = 0.846$ at 345 kV), and intra-area edge fractions remain high ($\phi_{\text{intra}} = 0.976$ real vs. 0.958 synthetic). Although the SA-DCSBM model slightly overestimates the inter-area cut ratio (0.042 vs. 0.024), the overall structural pattern is preserved.

In the EI system, across all three voltage layers, modularity of synthetic topologies is within 0.02 to 0.03 of the real grid ($Q = 0.868$ to 0.872 synthetic vs. 0.874 to 0.877 real), and intra-area edge fractions exceed 0.90 in both real and synthetic cases. Inter-area cut ratios remain consistent, with real values around 0.091 to 0.092 and synthetic values between 0.092 and 0.096. These results confirm that the SA-DCSBM captures both modular structure and geographic cohesion present in the North American transmission networks. Thus, SA-DCSBM enables forward generation of network topologies with

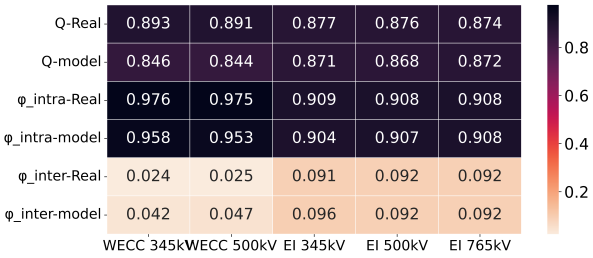

Figure 9. Comparison of modularity Q, intra-area edge fraction ϕ_{intra}, and inter-area cut ratio ϕ_{inter} across real and synthetic topologies for the WECC and the EI footprint at various voltage layers.

explicit spatial and modular constraints, supporting its use in the synthesis of structurally realistic test systems.

6. Conclusion

SA-DCSBM has been shown to generate large-scale synthetic transmission topologies that reproduce key structural features of actual grids. Geographic constraints are incorporated through an exponential distance-decay kernel, while degree correction and community affinities capture heterogeneity within the network. Together, these elements recover observed patterns of community structure, clustering, and spatial span.

SA-DCSBM remains a structural model; electrical aspects such as power flows, reactive support, voltage magnitudes, and frequency dynamics are not represented. An extension is to couple the topology generation process with power-flow feasibility and dispatch constraints, thereby integrating structural fidelity with system performance. Comparative studies against alternative spatial generators will further clarify the value of SA-DCSBM across broader suites of topological and operational metrics.

Although the model reproduces spatial and community structure at scale, it currently underrepresents long-distance, high-capacity connections. These rare but essential links contribute to overall robustness. Potential extensions include the use of hierarchical spatial penalties to differentiate intra-area distances from long-haul connections. Such refinements would enable the framework to capture both local clustering and the few long-range corridors vital for bulk system resilience.

References

[1] U.S. Federal Energy Regulatory Commission. (2025) Ferc critical energy infrastructure information (ceii). [Online]. Available: https://www.ferc.gov/ceii

[2] Council of the European Union, "Council directive 2008/114/ec of 8 december 2008 on the identification and designation of european critical infrastructure," Official Journal of the European Union, Tech. Rep. L 345/75, 2008.

[3] A. B. Birchfield, E. Schweitzer, M. H. Athari, T. Xu, T. J. Overbye, A. Scaglione, and Z. Wang, "A metric-based validation process to assess the realism of synthetic power grids," *Energies*, vol. 10, no. 8, p. 1233, 2017.

[4] S. J. Young, Y. Makarov, R. Diao, R. Fan, R. Huang, J. O'Brien, M. Halappanavar, M. Vallem, and Z. H. Huang, "Synthetic power grids from real world models," in *2018 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2018.

[5] M. H. Mohammadi and K. Saleh, "Synthetic Benchmarks for Power Systems," *IEEE Access*, vol. 9, pp. 162 706–162 730, 2021.

[6] S. Kunkolienkar, F. Safdarian, J. Snodgrass, A. Birchfield, and T. Overbye, "A Description of the Texas A&M University Electric Grid Test Case Repository for Power System Studies," in *2024 IEEE Texas Power and Energy Conference (TPEC)*. IEEE, 2024, pp. 1–6.

[7] S. Hamidreza and W. Zhifang, "Autosyngrid: A matlab-based toolkit for automatic generation of synthetic power grids," *International Journal of Electrical Power & Energy Systems*, vol. 118, p. 105757, 2020.

[8] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[9] B. Karrer and M. E. Newman, "Stochastic blockmodels and community structure in networks," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 83, no. 1, p. 016107, 2011.

[10] G. Curato and F. Lillo, "Optimal information diffusion in stochastic block models," *Physical Review E*, vol. 94, no. 3, 2016.

[11] B. Karrer and M. E. J. Newman, "Stochastic blockmodels and community structure in networks," *Physical Review E*, vol. 83, no. 1, Jan. 2011.

[12] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas, "Comparing community structure identification," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 09, pp. P09 008–P09 008, 2005.

[13] A. B. Birchfield and T. J. Overbye, "A review on providing realistic electric grid simulations for academia and industry," *Current Sustainable/Renewable Energy Reports*, vol. 10, no. 3, pp. 154–161, 2023.

[14] A. Asztalos, S. Sreenivasan, B. K. Szymanski, and G. Korniss, "Cascading failures in spatially-embedded random networks," *PLoS ONE*, vol. 9, no. 1, p. e84563, 2014.

[15] Z. Wang, R. J. Thomas, and A. Scaglione, "Generating random topology power grids," in *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS)*. IEEE, 2008, pp. 1–10.

[16] A. Schmuck, O. Rennert, G. Matthiae, P. C. Müller, and A. Schnettler, "A complex-network approach to the generation of synthetic power transmission networks," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1662–1673, 2014.

[17] S. Soltan and G. Zussman, "Generation of synthetic spatially embedded power grid networks," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE, 2016, pp. 1–5.

[18] A. B. Birchfield and T. J. Overbye, "Planning sensitivities for building contingency robustness and graph properties into large synthetic grids," in *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 2020, pp. 3167–3171.

[19] A. Birchfield, H. Xu, K. Gegner, and T. J. Overbye, "Building synthetic power transmission networks of many voltage levels spanning multiple areas," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2020, pp. 1–5.

[20] R. Espejo, S. Lumbreras, and A. Ramos, "A complex-network approach to the generation of synthetic power transmission networks," *IEEE Systems Journal*, vol. 13, no. 3, pp. 3050–3058, 2019.

[21] S. Kunkolienkar, F. Safdarian, J. Snodgrass, and T. J. Overbye, "Quantification of area sparsity in large-scale electric grids," in *2023 IEEE Kansas Power and Energy Conference (KPEC)*. IEEE, 2023.

[22] M. E. Newman, "Modularity and community structure in networks," *Proc Natl Acad Sci U S A*, vol. 103, no. 23, pp. 8577–82, 2006.

[23] C. Domicolo and H. Mahmoud, "Degree-based gini index for graphs," *Probability in the Engineering and Informational Sciences*, vol. 34, no. 2, pp. 157–171, 2020.

[24] X. Wang, F. Xue, S. Lu, L. Jiang, E. Bompard, and M. Masera, "Understanding communities from a new functional perspective in power grids," *IEEE Systems Journal*, vol. 16, no. 2, pp. 3072–3083, 2022.

[25] M. E. Eddin, M. Massaoudi, H. Abu-Rub, M. Shadmand, and M. Abdallah, "Novel functional community detection in networked smart grid systems-based improved louvain algorithm," in *2023 IEEE Texas Power and Energy Conference*. IEEE, 2023, Conference Proceedings.

[26] H. Kim, S. H. Lee, and P. Holme, "Community consistency determines the stability transition window of power-grid nodes," *New Journal of Physics*, vol. 17, no. 11, p. 113005, 2015.

[27] M. Guerrero, R. Baños, C. Gil, F. G. Montoya, and A. Alcayde, "Evolutionary algorithms for community detection in continental-scale high-voltage transmission grids," *Symmetry*, vol. 11, no. 12, p. 1472, 2019.

[28] N. Stanley, T. Bonacci, R. Kwitt, M. Niethammer, and P. J. Mucha, "Stochastic block models with multiple continuous attributes," *Applied Network Science*, vol. 4, no. 1, p. 54, 2019.

[29] J. Wu, M. Barahona, Y.-J. Tan, and H.-Z. Deng, "Spectral measure of structural robustness in complex networks," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 41, no. 6, pp. 1244–1252, 2011.

[30] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak mathematical journal*, vol. 23, no. 2, pp. 298–305, 1973.

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