Toward Realism: Sparse Matrix Statistics in the North American and Synthetic Electric Grid Models

Sanjana Kunkolienkar, *Graduate Student Member, IEEE*, Farnaz Safdarian, *Member, IEEE*, Thomas J. Overbye, *Fellow, IEEE*

***Abstract*—This letter uses sparse matrix statistics to highlight a structural gap between North American grid models and the synthetic cases commonly used in research. North American grids exhibit area-based modularity, characterized by dense intra-area links and limited inter-area connectivity. By contrast, synthetic grids often over-mesh across areas. Analysis shows that these surplus tie lines increase fills in the Jacobian factorization, which reduces realism and computational efficiency. The observations confirm that inter-area connectivity influences sparse matrix behavior. This work emphasizes the importance of incorporating modular structure into synthetic grid design.**

***Index Terms*—Fills, grid topology, large-scale electric grids, power system modeling, sparse matrices, spatially embedded synthetic power grid models.**

I. Introduction

The structural properties of electric power grids are central to accurate modeling, scalable computation, and realistic simulation. As grids grow more complex with renewable integration and inter-regional coordination, the need for realistic large-scale models has become critical. Such models support research and education [1], and their value depends on reproducing the structural features of the actual transmission networks. A key feature is the non-uniform sparsity of grid topology, shaped by historical and operational factors. In North America, transmission systems developed as independent operating areas, densely interconnected internally yet joined by relatively few tie lines [2]. These ties enable inter-area transfers but are lightly loaded under normal conditions, leading to strong intra-area and comparatively weak inter-area connectivity. In this letter, this property is referred to as area-based modularity.

Despite its significance, area-based modularity has not been explicitly incorporated into the construction or validation of spatially embedded transmission network models. Using topology, sparse matrix statistics, and comparisons with synthetic cases, this letter shows that the actual grids consistently exhibit denser intra-area than inter-area connectivity. In terms of sparse matrix statistics, this structure reduces fill-in, shortens factorization paths, and improves computational efficiency, whereas synthetic grids lack area-based modularity, diverging in both structural and numerical behavior. Establishing area-based modularity as a fundamental property provides a path to building more realistic, tractable, and practically relevant spatially embedded synthetic grid models for research.

Authors are with the Department of Electrical and Computer Engineering, Texas A&M University, 400 Bizzell St, College Station, TX 77843 USA (e-mails: {sanjanakunkolienkar, fsafdarian, overbye}@tamu.edu). This work was conducted as part of the 'Electric Grid Resilience' Project (60NANB24D210), funded by the National Institute of Standards and Technology (NIST) of the U.S. Department of Commerce.

II. Evidence for Area-Based Modularity in Electric Grids

Area-based modularity is the structural property that grids exhibit higher intra-area connectivity than inter-area connectivity. In this context, an *area* denotes an operationally defined region of the transmission system, such as a balancing authority, control area, planning zone, or a similarly designated operational footprint used in real-world grid operation and planning. These areas are externally defined and are not inferred algorithmically from network topology. Instead, they reflect long-standing organizational, regulatory, and operational boundaries in transmission system management.

Within a given operating area, substations are typically interconnected through relatively dense local transmission infrastructure to support regional power balancing, voltage control, and reliability requirements. In contrast, connections between different operating areas are limited in number and are realized through a small set of inter-area tie lines designed primarily for bulk power exchange and contingency support. This organization reflects historical development and operational practice, where local balancing is prioritized before long-distance transfers.

Although recognized in practice, area-based modularity (previously termed area sparsity [3]) has not been validated as a quantitative network feature. To provide such validation, the metrics introduced in [3] were applied to both North American and synthetic grid models [4]. These metrics quantify how frequently substations separated by a geographic distance x are directly connected by transmission lines. The resulting function, denoted $S_a(x)$, can be interpreted as a distance-dependent connection density (Eq. 1), measuring the fraction of all substation pairs at distance x that are electrically connected.

This notion is further decomposed according to operating area membership. The intra-area metric $S_{\text{intra}}(x)$ measures the connection density between substations belonging to the same operating area, while the inter-area metric $S_{\text{inter}}(x)$ measures the corresponding density for substations in different operating areas. Large separations between $S_{\text{intra}}(x)$ and $S_{\text{inter}}(x)$ indicate strong area-based modularity, reflecting a predominance of local, within-area connectivity over long-distance inter-area ties.

The metrics are defined as

$$S(x) = \frac{N_{\text{conn}}(x)}{N_{\text{tot}}(x)}$$
$$S_{\text{intra}}(x) = \frac{N_{\text{conn}}^{\text{intra}}(x)}{N_{\text{tot}}^{\text{intra}}(x)}, \qquad S_{\text{inter}}(x) = \frac{N_{\text{conn}}^{\text{inter}}(x)}{N_{\text{tot}}^{\text{inter}}(x)}. \tag{1}$$

Results for EI and WECC show that $S_{\text{intra}}(x)$ consistently exceeds $S_{\text{inter}}(x)$, confirming dominant intra-area connectivity in North American grids. Synthetic grids lack this separation, with $S_{\text{inter}}(x)$ often approaching or exceeding $S_{\text{intra}}(x)$, indicating unrealistically dense inter-area ties. As shown in Fig. 1, intra-area connectivity (red) dominates in North American grids, whereas synthetic grids exhibit elevated inter-area connectivity (blue). This establishes area-based modularity as a measurable transmission grid property and shows that synthetic models do not reproduce it without area-level constraints. Because these connectivity patterns determine Jacobian sparsity, the resulting structural differences directly affect numerical performance, as examined next using sparse matrix statistics.

III. Sparse Matrix Statistics

In steady-state power system analysis, the operating point is obtained by solving the nonlinear power balance equations $f(x) = 0$, where $x = [\theta, V]$ contains bus voltage angles and magnitudes [5], [6]. The Newton–Raphson (NR) method linearizes these equations as $J(x^k)\Delta x^k = -f(x^k)$, where $\mathbf{J} = \partial f/\partial x$ is the Jacobian. Because transmission networks are sparsely connected, $\mathbf{J}$ is also sparse, and this structure can be exploited to reduce factorization cost [7].

Sparse matrix statistics provide insight into structural properties [8]. The *number of nonzeros*, $\text{nnz}(\mathbf{J})$, captures direct electrical couplings; modular networks with dense intra-area and limited inter-area connections yield smaller values. The *fill-ins*, $\text{nnz}(L+U) - \text{nnz}(\mathbf{J})$, measure the additional nonzeros introduced during factorization, which grow with surplus inter-area ties (shown with an example matrix in Fig. 2). Finally, the *factorization path length*, given by the maximum depth of the elimination tree of $\mathbf{J}$, reflects the number of forward–backward substitution steps; area-sparse networks produce shorter paths by limiting long-range dependencies.

The work done in [9] quantified how computational effort in power system analysis scales with system size, showing that the number of nonzeros (fills) grows approximately as $n^{1.2}$, leading to factorization and repeat-solution costs that scale as $n^{1.4}$ and $n^{1.2}$, respectively. Building on this, [8] extended the analysis to real-world and synthetic grids with up to 110,000 buses, empirically finding that fills grow as $n^{1.07}$, factorization time as $n^{1.38}$, and forward/backward substitution as $n^{1.17}$, thereby confirming and refining the scaling behavior of sparse matrix computational complexity in large-scale power systems.

Algorithm 1 Synthetic Grid Modification

for each transmission voltage level v **do**
  Extract the subgraph.
  Rank inter-area candidate lines by geographic length and PTDF sensitivity.
  while $S_{\text{inter}}^{(v)}(x) \geq S_{\text{intra}}^{(v)}(x)$ for any distance bin x **do**
    Tentatively remove the selected inter-area candidate line at voltage level v
    if network connectivity at voltage level v is violated **then**
      Restore the removed line and mark it as non-removable.
    end if
  end while
end for
Output: Modified synthetic grid with enforced area-based modularity across voltage levels.

To evaluate the hypothesis in this work, sparse matrix statistics were computed for cases in EI and WECC, as well as for their synthetic counterparts developed using work done in [4], [10]. Each case was solved from a flat start, the Jacobian at the base operating point was extracted, and factorization with sparsity-preserving ordering was applied to obtain fill-in and path statistics, summarized in Table I. Existing synthetic grid models were modified to reduce inter-area lines using Algorithm 1. The code used is available at [11].

Area-based modularity affects the behavior of sparse matrices by shaping the Jacobian sparsity pattern. Dense intra-area connectivity forms a block-diagonal structure, whereas excess inter-area ties introduce long-range couplings that increase the number of nonzeros, fill-ins, and factorization path length. As a result, variations in $S_{\text{intra}}(x)$ and $S_{\text{inter}}(x)$ are directly reflected in nnz, fill-ins, and factorization statistics.

The comparisons reveal that North American grids consistently produce fewer fill-ins and shorter factorization paths than unmodified synthetic grids of similar size and connectivity. For example, at 70,000 buses North American grid models yield 201,520 fills and a longest path of 253, whereas unmodified synthetic grids produce 376,438 fills and a path of 624. Modifying the existing synthetic grid models, by reducing inter-area tie lines, reduces synthetic statistics to 210,690 fills and a path of 372, much closer to North American grids. At 80,000 buses, modified models record 245,966 fills compared to 262,256 in actual grids, while unmodified models reach 424,988. These reductions confirm the computational benefits and realism of enforcing area-based modularity.

Figs. 4 and 3 illustrate this effect. The unmodified synthetic Jacobian shows widespread fill-ins, while the modified version reveals localized, block-like regions aligned with intra-area groupings. Scaling analysis further emphasizes the difference: actual grids follow sublinear growth of approximately $n^{0.99}$ (black), modified synthetic grids track even closer at $n^{0.94}$ (red), and unmodified synthetic grids grow faster at $n^{1.16}$

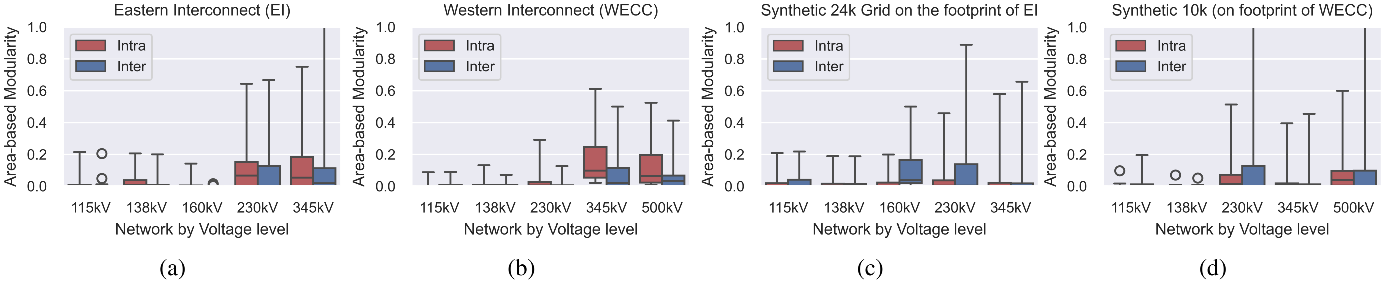

Fig. 1: Area-based modularity across real and synthetic North American transmission networks, resolved by voltage level. Panels (a) and (b) report results for EI and WECC, while panels (c) and (d) show corresponding synthetic grid models constructed on the EI and WECC geographic footprints. Box plots summarize the distribution of the modularity metric evaluated across operating areas or regional partitions at each voltage level.

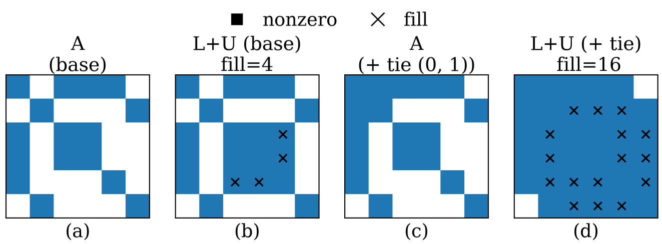

Fig. 2: Illustration of fill-in generated during LU factorization. Panels (a)–(b) show the matrix sparsity pattern and corresponding factorization for a base case, where limited fill entries appear (marked with ×). Panels (c)–(d) show the effect of adding one inter-area tie line producing 16 fill entries. Blue squares denote original nonzero entries; black crosses denote fill-in.

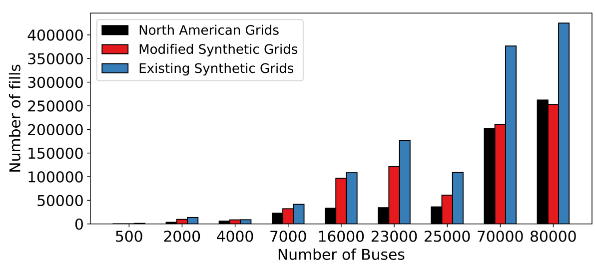

Fig. 3: Comparison of fill-ins after sparse LU factorization for North American grid models, existing synthetic grids, and modified synthetic grids across increasing system sizes.

(blue). This contrast highlights how redundant inter-area ties inflate structural complexity and confirms that area-based modularity has measurable consequences for power system computation. While reducing inter-area connectivity improves structural realism and numerical behavior, such modifications entail trade-offs regarding transfer capability, congestion, and security, and should not be interpreted as an operational recommendation.

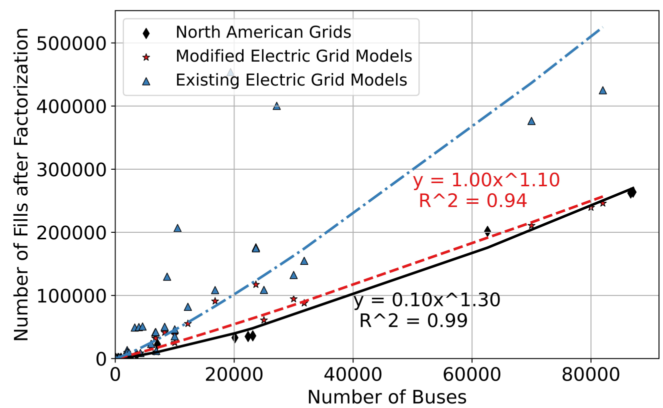

Fig. 4: Number of fill-ins after sparse LU factorization as a function of system size for North American, existing synthetic, and modified synthetic grid models.

IV. Implications for Power Grid Modeling

The findings of this study highlight important implications for synthetic grid modeling. Area-based modularity indicates that regional clustering and limited inter-area connectivity are fundamental features of North American grids. Building on this observation, two implications emerge. First, validation should extend beyond nodal degrees, impedance ranges, and geographic placement. While existing graph-theoretic and spectral metrics capture important aspects of network structure [10], they do not explicitly encode predefined operating area boundaries or the geographically constrained nature of inter-area transmission planning. Metrics that capture intra-area clustering and inter-area modularity are therefore necessary to ensure structural fidelity.

Second, the absence of modularity permits excessive tie-line transfers. Standard network modularity or community-detection-based measures may identify statistically dense subgraphs, but the resulting partitions do not necessarily correspond to operational areas such as control regions, utilities, or balancing authorities. This weakens the credibility of studies on congestion, markets, and stability that depend on realistic

TABLE I
Comparison of sparse matrix statistics in the North American grid models and the synthetic grid models on the same footprint.

	North American Grid Models				Existing Synthetic Grid Models				Modified Synthetic Grid Models			
No. of buses	nonzeros	fills	Ave FP	longest FP	nonzeros	fills	Ave FP	longest FP	nonzeros	fills	Ave FP	longest FP
500	1,782	340	22	33	1,802	1,166	22	40	1,320	384	16.7	28
2,000	8,180	3,504	33	55	7,334	13,504	84	114	7,106	9,590	75.4	114
4,000	11,839	5,990	36	54	14,020	8,782	56	95	14,011	8,490	53.4	87
7,000	27,773	22,674	94	146	22,471	12,134	58	83	23,804	32,064	110.1	182
23,000	80,710	36,046	75	112	86,865	174,602	211	360	85,648	121,150	139.3	218
70,000	227,933	201,520	172	253	238,932	376,438	312	624	233,966	210,690	188	372
80,000	311,438	262,256	205	302	284,552	424,988	281	624	278,605	245,966	171	372

flow patterns. By explicitly linking area-level structure to sparse matrix behavior, area-based modularity complements existing validation metrics rather than replacing them. Taken together, these insights suggest that area structure should be embedded into synthetic grid creation from the outset.

V. Conclusion and Future Work

This letter demonstrates that area-based modularity is a measurable structural property of transmission networks. Incorporating area-based modularity as a validation criterion improves both the structural realism and computational efficiency of synthetic grids. Future work will focus on constructing new synthetic grid models that explicitly enforce area-based modularity and on studying how this structure influences the propagation of cascading failures in large-scale power systems.

References

[1] 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.

[2] J. Cohn, "When the grid was the grid: the history of north america's brief coast-to-coast interconnected machine [scanning our past]," *Proceedings of the IEEE*, vol. 107, no. 1, p. 232–243, 2019.

[3] S. Kunkolienkar, F. Safdarian, J. Snodgrass, and T. Overbye, "Quantification of area sparsity in large-scale electric grids," in *2023 IEEE Kansas Power and Energy Conference (KPEC)*. IEEE, 2023, pp. 1–6.

[4] 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 *2016 IEEE Power and Energy Conference at Illinois (PECI)*. IEEE, 2016, pp. 1–6.

[5] N. Sato and W. Tinney, "Techniques for exploiting the sparsity or the network admittance matrix," *IEEE Trans. on Power App. & Syst.*, vol. 82, no. 69, pp. 944–950, 1963.

[6] W. Tinney, V. Brandwajn, and S. Chan, "Sparse vector methods," *IEEE Trans. on Power App. & Syst.*, no. 2, pp. 295–301, 2007.

[7] W. F. Tinney and J. W. Walker, "Direct solutions of sparse network equations by optimally ordered triangular factorization," *Proc. of the IEEE*, vol. 55, no. 11, pp. 1801–1809, 1967.

[8] F. Safdarian, Z. Mao, W. Jang, and T. Overbye, "Power system sparse matrix statistics," in *2022 IEEE Texas Power and Energy Conference (TPEC)*. IEEE, 2022, pp. 1–5.

[9] F. Alvarado, "Computational complexity in power systems," *IEEE Trans. on Power App.& Syst.*, vol. 95, no. 4, pp. 1028–1037, 1976.

[10] A. B. Birchfield and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic transmission grid models," *IEEE Trans. on Power Syst.*, vol. 32, no. 4, pp. 3258–3265, 2017.

[11] S. Kunkolienkar, "Synthetic grid connectivity," 2025, gitHub repository, accessed December 30, 2025. [Online]. Available: https://github.com/SanjanaKunkolienkar/SyntheticGridConnectivity