Topological Graph Metrics for Detecting Grid Anomalies and Improving Algorithms

Jonas A. Kersulis*, Ian A. Hiskens*, Carleton Coffrin†, and Daniel K. Molzahn‡
*Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI, USA
†Advanced Network Science Initiative
Los Alamos National Laboratory, Los Alamos, NM, USA
‡Energy Systems Division
Argonne National Laboratory, Argonne, IL, USA

***Abstract*—Power grids are naturally represented as graphs, with buses as nodes and power lines as edges. Graph theory provides many ways to measure power grid graphs, allowing researchers to characterize system structure and optimize algorithms. We apply several topological graph metrics to 33 publicly-available power grids. Results show that a straightforward, computationally inexpensive set of checks can quickly identify structural anomalies, especially when a broad set of test networks is available to establish norms. Another application of graph metrics is the characterization of computational behavior. We conclude by illustrating one compelling example: the close connection between clique analysis and semidefinite programming solver performance. These two applications demonstrate the power of purely topological graph metrics when utilized in the right settings.**

***Index Terms*—Complex networks, network theory (graphs), power grids.**

I. Introduction

The buses and transmission lines of a power grid translate naturally to the nodes and edges of a graph. This connection has been recognized for many years, and numerous graph structural properties have been studied with various power systems applications [1], [2]. Graph-theoretical methods have been used to identify system vulnerabilities [3]–[5], detect structural anomalies [6], generate and validate synthetic grids [7]–[10], create meaningful visualizations [11], [12], and perform partitioning [13], [14]. The graph analysis methods employed may be divided into two categories: weighted, where electrical information is embedded in the graph, and unweighted, where only topology is considered. Though there is a fundamental difference between a power system's topology and its electrical structure [4], [8], [15], unweighted graph analysis is ideal for quickly detecting unusual connectivity patterns. The topological algorithms used in this paper are computationally inexpensive, and numerical results are relatively easy to interpret. Consequently, it is feasible to scan large power grids to check for structural anomalies.

Efficient detection of structural anomalies calls for two ingredients. The first is a large collection of test networks from which a notion of "normal" topology may be derived. NESTA [16] is well-suited for this purpose. The second ingredient is a combination of straightforward, computationally inexpensive graph metrics capable of highlighting deviations from the norm. One of the main contributions of this paper is the selection of several topological metrics (along with discussion of less useful ones that were also considered), their application to NESTA, and revelation of several structural anomalies. As will be shown, our methods efficiently reveal highly-connected subnetworks in all four PEGASE project [17] topologies, in addition to highlighting the unusual density of the 162-bus IEEE dynamic test case.

The second portion of this paper expounds on the connection between maximal cliques and a recently-developed semidefinite programming (SDP) optimal power flow (OPF) algorithm [18]. Because SDP solver time increases sharply with the size of the semidefinite constraint matrix, decomposition into a set of constraints with smaller submatrices can significantly improve solver performance. According to a matrix completion theorem, any valid decomposition must consist of combinations of maximal cliques. We study the greedy clique merge algorithm proposed in [18], which merges overlapping cliques in search of the decomposition that minimizes SDP solver time. Our results show that solution time is not minimized at the point one might expect, suggesting that topological considerations merit further study in the SDP OPF context.

The paper proceeds as follows. Section II introduces graph notation and defines various topological graph metrics. Section III applies these metrics to NESTA networks, characterizing typical grid topology and highlighting unusual connectivity patterns. Section IV contains our discussion of maximal cliques and SDP OPF performance, and conclusions are presented in Section V.

II. Metric definitions and intuition

A. Power grids as complex networks

A power grid consists of electrical generators and loads joined by power lines and transformers. The structure of a power grid may be represented as a graph, which consists of a set of nodes $\mathbf{N}$ connected by a set of edges $\mathbf{E}$. Connections may be encoded in an adjacency matrix A, where each element A_{ij} is 1 if nodes i and j have an edge between them, and 0

This work was supported by the ARPA-E GRID DATA program.

otherwise (diagonal elements are all 0). The translation from power grid to graph involves a few modeling choices. We consider only undirected, unweighted graphs, where edges have no orientation or other properties. In the test networks we consider, each substation is one node. Though the data has this limitation, our ideas are applicable to high-fidelity models of real grids, which often have multiple buses within a substation [12]. We omit parallel edges between vertices. Finally, because some public test networks use transmission line objects to model transformers (and vice versa in rare cases), we allow both power lines and transformers to be edges in our graph representations.[1]

B. Topological graph metrics

1) Degree distribution: In a graph, each node's degree is the number of nodes to which it is connected. Mathematically, the degree of node i in a graph with $|\mathbf{N}|$ nodes is given by

$$k_i = \sum_{j=1}^{|\mathbf{N}|} A_{ij} \ . \tag{1}$$

The node degree distribution $\mathrm{P}(k)$ is a meaningful and concise metric based on this property:

$$\mathrm{P}(k) = \frac{|\mathbf{N}_k|}{|\mathbf{N}|} \ , \tag{2}$$

where $\mathbf{N}_k$ is the set of nodes with degree k [8]. Degree distribution has system reliability implications [4], [5], [9], [10], [12] and reveals some useful information at a glance. A long tail with high-degree nodes indicates one or more strongly-connected hub nodes (typically at high voltage levels [12]), which make the system more vulnerable to targeted attack [3], [5]. A high relative density of degree-1 and degree-2 nodes suggests a grid with long paths or radial components, and low meshing [5]. We have found that maximum, mean, and median values sufficiently portray most interesting features for NESTA networks. These values are defined, respectively, as:

$$k_{\max} = \max k : |\mathbf{N}_k| > 0 \tag{3a}$$

$$\bar{k} = \frac{1}{|\mathbf{N}|} \sum_{k=1}^{k_{\max}} k \cdot |\mathbf{N}_k| \tag{3b}$$

$$k_{\text{med}} = x : \sum_{k=1}^{x} \mathrm{P}(k) \geq \frac{1}{2} \text{ and } \sum_{k=x}^{k_{\max}} \mathrm{P}(k) \geq \frac{1}{2} \ . \tag{3c}$$

Considerations of flexibility, reliability, and physical space keep $k_{\max}$ from growing too large in real power grids. In the North American power grid, the probability of a substation having more than x transmission lines decreases exponentially with x [5]. Thus, high $k_{\max}$ hints that network reduction or perhaps even a modeling error has occurred. The mean distribution value $\bar{k}$ is related to how meshed the system is, and lies between 2 and 3 for most grids. Median node degree is difficult to find in the literature, and is less useful for characterizing system structure. Though some small, dense networks have median degree 3, the vast majority have $k_{\text{med}} = 2$ (see Section III-A).

2) Degree assortativity: The degree assortativity coefficient r measures the extent to which nodes of like degree connect to each other. It is equivalent to the Pearson correlation coefficient of node degrees [19]. If $\mathbf{K} = \{k : |\mathbf{N}_k| > 0\}$, then r is given by

$$r = \frac{1}{\sigma_k^2} \left[\sum_{x,y \in \mathbf{K}} xy \left(\frac{|\mathbf{E}_{xy}|}{|\mathbf{E}|} - \frac{|\mathbf{E}_x|}{|\mathbf{E}|} \frac{|\mathbf{E}_y|}{|\mathbf{E}|} \right) \right] . \tag{4}$$

In (4), x and y represent any two node degree values, $|\mathbf{E}_{xy}|/|\mathbf{E}|$ is the fraction of all edges connecting a node with degree x to a node with degree y, $|\mathbf{E}_x|/|\mathbf{E}|$ is the fraction of all edges that start or end at nodes with degree x, and σ_k is the standard deviation of the node degree distribution $\mathrm{P}(k)$. Values of r range from -1 to 1, indicating perfect disassortativity and assortativity respectively [19]. Power system topologies tend to be slightly disassortative, as mentioned by [8], [10] and confirmed in Section III-B.

3) Rich-club coefficient: High degree assortativity suggests a strongly-interconnected set of high-degree nodes. The rich-club coefficient [20] detects this structure, sometimes referred to as a "hubs of hubs" or "rich club," directly. Let $\mathbf{N}_{>k}$ be the set of nodes with degree greater than k, and $\mathbf{E}_{>k}$ the set of edges between those nodes. Then the rich-club coefficient is computed for each degree k as

$$\phi(k) = \frac{2|\mathbf{E}_{>k}|}{|\mathbf{N}_{>k}| \left(|\mathbf{N}_{>k}| - 1 \right)} \ , \tag{5}$$

which is the fraction of all possible edges that exist among nodes with degree greater than k. If there is some degree $x : \phi(k) = 1 \ \forall \ k > x$, then ϕ is said to "saturate" [6], and the graph is said to exhibit the "rich-club effect" [20]. Only two nodes and one edge are required to make this happen: if the two highest-degree nodes in a graph each have degree y and are connected, with the next-highest node degree being x, then $\phi(k) = 1 \ \forall \ k \in [x, y-1]$. It is therefore important to consider the number of nodes involved in a rich club to ensure one is not simply observing two highly-connected nodes.

4) Cliques: A fully-connected subgraph is called a clique. In graph notation, a clique $\mathbf{C}$ satisfies

$$\mathbf{C} \subseteq \mathbf{N} : A_{ij} = 1 \ \forall \ i, j \in \mathbf{C} \ . \tag{6}$$

Although each subset of a clique is technically also a clique, we restrict our focus to those *maximal* cliques which cannot be expanded by adding other nodes. Determining the maximal cliques of a general undirected graph is an NP-hard problem, but a memory-efficient adaptation of the Bron-Kerbosch algorithm [21] identifies the maximal cliques of our NESTA power grid graphs in a reasonable time.

5) Chordal graph extensions: A graph is chordal if every cycle of at least four nodes has a chord, which is an edge between nodes that are non-adjacent in the cycle. One can obtain a chordal extension (not unique) with few additional edges via sparsity-preserving Cholesky factorization [22]. There is a

[1] This can add a substantial number of "leaf nodes," but we repeated all analysis with generator step-up transformers excluded to verify that results are not qualitatively affected.

linear-time algorithm for identifying the maximal cliques of a chordal graph [23].

6) Adjacency spectral radius: The eigenvalue spectrum of the adjacency matrix A is relatively expensive to compute, especially for larger networks. While this metric yields interesting results, computation time makes it less useful for quickly revealing structural anomalies. The magnitude of the largest eigenvalue (also known as adjacency spectral radius) is considered in Section III-E, where we show that a combination of other metrics reveals similar information more quickly.

III. Structural Anomalies in NESTA Networks

Of the thirty power system graph analysis papers reviewed in [2], nearly 85% use one of the IEEE Literature-Based Power Flow Test Cases (e.g. the IEEE 14-Bus System). NESTA [16] includes these networks, RTE and PEGASE systems [17], [24], and many other publicly-available test networks. This makes NESTA an ideal proving ground for graph metrics. Because the archive includes multiple versions of some topologies (e.g. several Polish grid operating conditions), we focus on the subset of 33 representative NESTA networks shown in Figure 1 to ensure no single topology is over-represented. In the remainder of this section, we apply the metrics described in Section II to this collection of networks. Any graph analysis that was not hand-coded was performed by NetworkX [25].

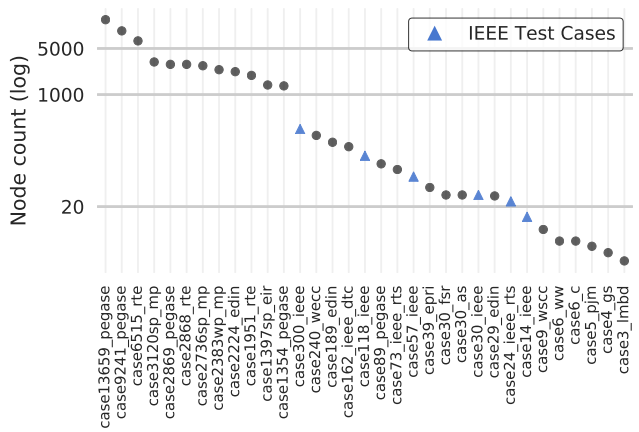

Figure 1. Semilog plot of node counts, illustrating size categories.

A. Degree distribution

A graph's maximum node degree is the highest degree observed among its nodes (3a). Figure 2 shows that this property scales roughly linearly with the log of network size for NESTA networks. Although the trend makes it appear that maximum node degree can grow arbitrarily large, it is important to note that real substations are limited by the design constraints discussed in Section II-B1. The three labeled outliers in Figure 2 are all PEGASE networks, and the three points with maximum degree 9 in the 1k-5k size range are the three Polish grid variants included in our NESTA sample. The unusually high maximum degree of the large PEGASE topology has been pointed out [6], but to our knowledge no prior work has shown how unusual the 89-bus PEGASE system is for its size. This network's highly-connected hub component is apparent from visual inspection of a 2D graph layout, but maximum node degree and other metrics discussed later on can identify this abnormal structure more quickly and efficiently.

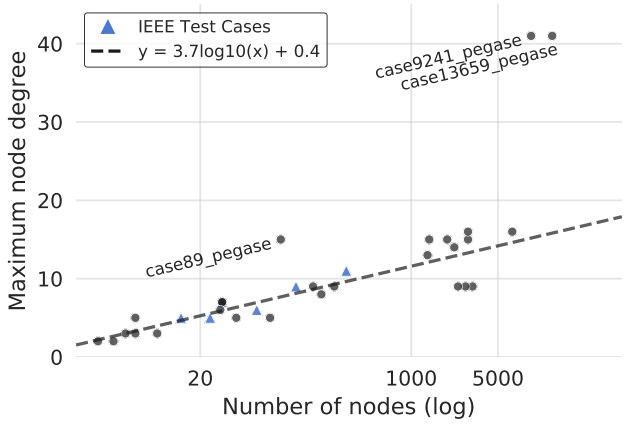

Figure 2. Semilog plot of maximum node degree vs. number of nodes. The dashed least-squares fit line excludes the labeled outliers.

The average mean node degree for our NESTA sample is 2.70, with a standard deviation of 0.53. Table I contains the data, along with a "rugplot" illustration of the distribution of values. Note the tight groupings about 2.4 (RTE and Polish networks) and 2.7 (PEGASE and IEEE networks, highlighted pink and blue respectively). Highly-meshed transmission networks like the 162-bus IEEE dynamic test case may naturally have high mean degree, but the PEGASE 89-bus network is clearly unusual. On the other end of the distribution, case189_edin appears to have long paths and low meshing.

Median node degree data are less interesting. Of our NESTA sample networks, 76% have $k_{\text{med}} = 2$, while remaining values are 3. Unlike mean degree, k_{med} is not sensitive to whether generator step-up transformers are modeled, at least for NESTA networks. Dense networks like case_162_ieee_dtc (which has been described as "quite robust" [26]) may have a median node degree of 3, but this is unusual for networks larger than a few hundred nodes. The largest NESTA network with $k_{\text{med}} = 3$ is the IEEE 300-bus test case.

B. Degree assortativity

Figure 3 suggests that transmission systems tend to be slightly disassortative. Over half of the NESTA sample networks have $|r| < 0.2$, and one standard deviation within the mean corresponds to roughly $(-0.3, 0.15)$. The two largest PEGASE networks are highly assortative, as first shown in [6]. The 13k-bus variant is less assortative due primarily to the thousands of generator step-up transformers it includes, which are omitted from the 9k-bus network. On the other end of the spectrum, case9_wscc is highly disassortative. This tiny network consists of a ring of six nodes, three of which have leaf nodes attached. The unusual disassortativity arises

TABLE I
MEAN NODE DEGREE VALUES FOR REPRESENTATIVE NESTA NETWORKS.

	Mean node degree
case89_pegase	4.629
case6_ww	3.667
case162_ieee_dtc	3.457
case29_edin	3.448
case9241_pegase	3.075
case118_ieee	3.034
case73_ieee_rts	2.959
case240_wecc	2.900
case14_ieee	2.857
case24_ieee_rts	2.833
case2869_pegase	2.766
case57_ieee	2.737
case30_ieee	2.733
case30_as	2.733
case30_fsr	2.733
case13659_pegase	2.727
case300_ieee	2.727
case1354_pegase	2.526
case2224_edin	2.522
case6515_rte	2.488
case1951_rte	2.435
case2383wp_mp	2.422
case2868_rte	2.421
case5_pjm	2.400
case2736sp_mp	2.385
case1397sp_eir	2.368
case3120sp_mp	2.362
case39_epri	2.359
case6_c	2.333
case189_edin	2.148
case4_gs	2.000
case9_wscc	2.000
case3_lmbd	2.000

from the fact that no edge joins nodes with matching degree. Although highly disassortative networks tend to be tiny, and highly assortative networks are typically in the "medium" range, there is no evidence of a size-related trend overall.

C. Rich-club effect

Table II lists all groups of three or more nodes in NESTA networks that are connected by at least 80% of their potential edges (we omit "clubs" consisting of just two nodes; see Section II-B3). The first row reads: "At least 80% of all possible edges exist between the 28 nodes in case9241 with degree greater than 25." The first two rows describe rich clubs first identified in [6], and the next two rows also concern PEGASE networks. The brevity of Table II is perhaps more significant than its contents. In every other case where the rich-club coefficient $\phi(k)$ reaches 0.8 for a NESTA network, the "rich club" consists of just two nodes. For a dozen networks, there

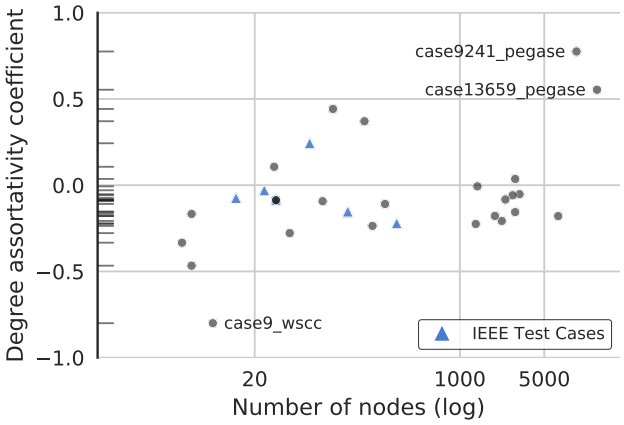

Figure 3. Semilog plot of degree assortativity vs. network size.

is no k for which $\phi(k)$ reaches 0.8. Mean node degree makes one PEGASE network stand out, degree assortativity draws attention to two, and three PEGASE networks have unusually high maximum node degree. The rich-club coefficient is the first metric that separates all four PEGASE networks from the rest of NESTA.[2]

TABLE II
ALL NESTA RICH CLUBS WITH MORE THAN 80% OF POTENTIAL EDGES AND AT LEAST 3 NODES.

	Nodes involved	Degree
case9241_pegase	28	>25
case13659_pegase	19	>29
case89_pegase	11	>11
case2869_pegase	5	>13
case3_lmbd	3	>0

D. Cliques

The vast majority of power grid maximal cliques contain just two nodes. Of the 124,211 maximal cliques in our representative NESTA networks, the mean size is 2.085. This leads to a linear relationship between the number of maximal cliques in a power system and the number of nodes. While the number of maximal cliques grows linearly with the number of nodes, the size of the largest clique (or "maximum clique") does not. Figure 4 plots maximum clique size against network size for our sample networks. As with the rich-club coefficient, this metric makes the four PEGASE networks stand out. Setting aside these networks, the least-squares fit would effectively be a horizontal line. Although PEGASE networks have unusually large cliques, roughly 99% consist of just two or three nodes.

E. Adjacency matrix spectrum

Figure 5 illustrates adjacency spectral radius versus number of nodes. As with the rich-club coefficient, this metric casts

[2]Table II also lists case3_lmbd, but this "rich club" contains the entire network: three nodes in a ring.

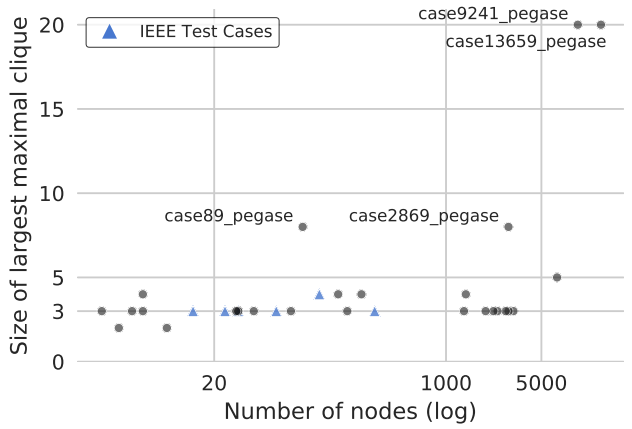

Figure 4. Maximum clique size versus network size.

the four PEGASE networks as outliers. All other networks lie close to the trendline, which increases gradually with the log of network size. While the isolation of all PEGASE networks is compelling, previously-mentioned metrics provide similar information with significantly less computation.

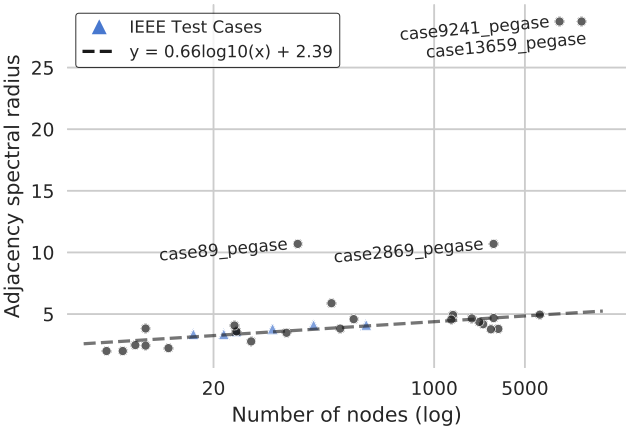

Figure 5. Semilog plot of largest adjacency matrix eigenvalue versus number of nodes. The dashed fit line excludes the labeled outliers.

IV. Cliques and semidefinite programming

A. Background

Advances in semidefinite programming have drawn the attention of the power systems community, specifically in connection with optimal power flow (OPF). Though not exact for all OPF problems [27], [28], the semidefinite relaxation finds global optima for many test cases [29], [30]. Since SDP solvers work best with semidefinite constraint matrices no larger than a few dozen elements to a side, the decomposition outlined in [18] is essential for solving larger SDP OPF problems [31]. The positive semidefinite matrix completion theorem in [32] governs valid decompositions [22]: suppose we have an undirected chordal graph and an associated incomplete matrix A. Then A can be completed to a positive semidefinite matrix if and only if all submatrices associated with the graph's maximal cliques are positive semidefinite. Thus, decomposition involves 1) forming a chordal extension of the power grid graph (see Section II-B5), 2) identifying chordal extension maximal cliques (see Section II-B4), and 3) replacing the semidefinite constraint matrix with a set of constraints corresponding to maximal clique submatrices. While the maximal clique decomposition results in the smallest possible submatrices, it also introduces linking constraints due to overlap between cliques.

A single node may belong to many maximal cliques. Each of these cliques has its own copy of that node's two variables (real and imaginary complex voltage phasor components in the OPF setting), but each physical parameter must ultimately have one value. Thus, maximal clique decomposition introduces a number of linking constraints that varies with the amount of overlap between cliques. Since SDP solvers like Mosek apply primal-dual methods, a primal linking constraint corresponds to a dual variable, and changes in the number of constraints and variables have similar impacts on solver performance [18]. If merging two cliques eliminates enough linking constraints to offset the resulting increase in submatrix size, the sum of variables and linking constraints is reduced overall, and SDP performance should be improved. For this reason, [18] proposed a greedy clique merging algorithm that repeatedly combines the pair of cliques whose merger yields the greatest reduction in the sum of variables and linking constraints. Suppose the algorithm stops when a specified number of submatrices, designated L, is reached. As shown in [18], SDP solver time gradually decreases as L shrinks (i.e. as more cliques are merged). If L becomes too small, however, the submatrices grow large enough to outweigh the benefits of linking constraint elimination. This trade-off is closely related to chordal extension maximal cliques, which are relatively inexpensive to obtain. The remainder of this section expounds on the relationship between clique characteristics, clique merge algorithm behavior, and SDP OPF performance.

B. Clique merge implementation

The first step towards implementing clique merge is to obtain a chordal extension of the power grid graph (which is not already chordal in general) using the procedure described in Section II-B5. Next, the maximal cliques of the chordal extension are found, and a clique graph is formed. Each node in the clique graph represents a maximal clique. Each edge lies between a pair of cliques that share nodes, and has weight equal to the number of nodes shared. At each step, the greedy clique merge algorithm eliminates an edge of the clique graph, merging its endpoints into a single group of nodes. Since the algorithm looks for the greatest overlaps between pairs of clique graph nodes, it will only eliminate edges belonging to the maximum-weight spanning tree of the clique graph, which can be obtained with Prim's algorithm [33]. The effective number of SDP optimization variables may be obtained from this spanning tree, which we denote $T = \{\mathbf{N}_T, \mathbf{E}_T\}$. If a node $\mathbf{n}_i \in \mathbf{N}_T$ represents a chordal extension maximal clique with $|\mathbf{n}_i|$ buses (each of

which has real and imaginary voltage phasor components), then the corresponding OPF semidefinite constraint matrix is square with $2|\mathbf{n}_i| \times 2|\mathbf{n}_i|$ elements, and $|\mathbf{n}_i|(2|\mathbf{n}_i|+1)$ unique optimization variables in the upper triangle. Similarly, an edge $e_i \in \mathbf{E}_T$ with weight w_{e_i} represents a clique overlap of w_{e_i} buses, which introduces $w_{e_i}(2w_{e_i}+1)$ linking constraints by duplicating their variables. Thus, the effective number of SDP OPF variables corresponding to T may be computed as

$$v(T) = \sum_{\mathbf{n}_i \in \mathbf{N}_T} |\mathbf{n}_i|(2|\mathbf{n}_i|+1) + \sum_{e_i \in \mathbf{E}_T} w_{e_i}(2w_{e_i}+1). \quad (7)$$

From here, implementation of the greedy clique merge algorithm is straightforward: at each step, compare $v(T)$ with $v(T_{-e_i})$ for each edge $e_i \in \mathbf{E}_T$, where T_{-e_i} is obtained by merging the endpoint cliques of e_i. The merge corresponding to the greatest reduction in effective number of SDP OPF variables is implemented, and T is updated. The algorithm may stop once $|\mathbf{N}_T|$ reaches a predefined threshold, once subsequent merges begin increasing $v(T)$, or according to some other rule.

C. Clique merge behavior for NESTA networks

We applied greedy clique merge to chordal extensions of our representative NESTA networks. Throughout algorithm execution, we tracked the number of variables and linking constraints, the size of the largest and smallest groups of nodes, and Mosek SDP solver times. These results are shown for the IEEE 300-bus test case in Figure 6. The size of the largest group (top solid greed line) is initially equal to the maximum clique size. This value rises sharply at first, then remains flat until the last few merges.[3] The size of the smallest group (bottom green line) remains at 2 until the last remaining clique is merged with another group during merge 212. As the greedy algorithm reduces the number of submatrices and eliminates linking constraints, there are fewer and fewer variables and linking constraints. The effective number of variables reaches a minimum (indicated by the "X" on the Variables curve) after roughly half of all merges have occurred. Although subsequent merges increase the effective number of variables, Mosek solver times continue to drop until around merge 200. Interestingly, solution time tends to be minimized closer to when the last clique is merged into another group.

Other NESTA networks behave similarly to the IEEE 300-bus case. There are a few quantities of interest for all networks. These include the fraction of merges until $v(T)$ is minimized, the fraction of merges until the last clique is merged into another group, and the fraction of merges until solution time is minimized. We estimated this last value by performing 50–250 Mosek trials[4] to obtain a spread of solver time data for each merge index, then discarding points in each spread more than two standard deviations from the mean, and averaging remaining trials. This computationally intensive process was impractical for larger networks. Figure 7 illustrates results for many networks in our NESTA sample. It typically takes around 50% of merges to minimize the number of variables, and roughly 75% of merges before the last unmerged clique is combined with another group. The gap between these points is wider for larger systems, and solver time tends to be minimized closer to where the last maximal clique is merged into a different group.

[3]Many cliques must be merged into medium-sized groups of nodes before it makes sense to merge these larger clusters together.

[4]Smaller networks required more trials due to increased variance relative to mean solver time.

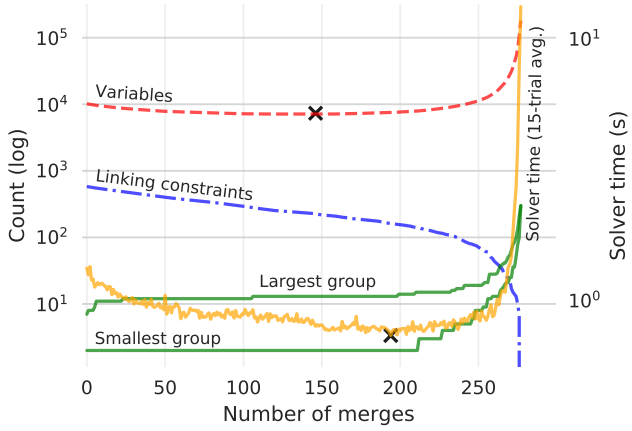

Figure 6. Semilog plot illustrating clique merge behavior for the IEEE 300-bus test case. Solver time is plotted against the right axis. Minimum variable count and minimum solution time are each indicated with an "X".

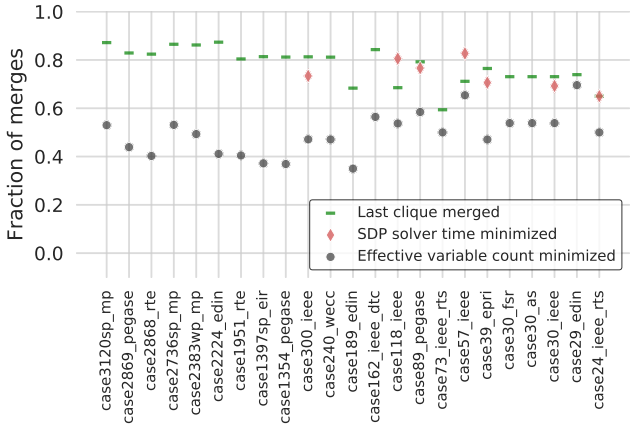

Figure 7. Merges required to reach important points of clique merge algorithm execution.

V. Conclusions

A number of topological graph metrics were applied to 33 networks in NESTA, and several structural anomalies were identified. Our metrics drew attention to PEGASE networks: each metric highlighted at least one of these networks, but three metrics (rich-club coefficient, maximum clique size, and adjacency matrix spectral radius) separated all four from the rest of NESTA. Our results indicate that each PEGASE network contains a relatively small but highly dense subnetwork,

which appears to be the result of kron network reduction or some similar technique. A combination of degree distribution, degree assortativity, and rich-club coefficient metrics is highly effective at identifying this structure within a large graph, while remaining computationally inexpensive. These metrics may be used to quickly scan new grid models for potential modeling or data issues.

The paper also focused on the connection between cliques and SDP OPF performance. Though electrical characteristics govern numerical behavior of algorithms, topological metrics can prove useful when partitioning or decomposition are involved. The semidefinite constraint that forms a bottleneck for SDP solvers may be decomposed into smaller constraints, according to maximal cliques of a chordal extension of the power grid graph. The authors of [18] showed that the maximal clique decomposition of a chordal extension obtained via Cholesky factorization introduces too many linking constraints due to overlap; merging cliques can significantly reduce solver time. While the heuristic employed in [18] achieves most of the benefit of clique merge, Mosek solver times typically continue to diminish after further merges take place, even though the effective number of variables increases. Our results supplement those of [18], as several additional test networks were studied, and Mosek was used in place of SeDuMi. Further study may lead to an improved clique merge algorithm that enables an SDP OPF routine to take full advantage of the positive semidefinite matrix completion theorem.

REFERENCES

[1] S. Seshu and M. B. Reed, *Linear Graphs and Electrical Networks*, first edition edition ed. Addison-wesley, 1961.

[2] G. A. Pagani and M. Aiello, "The Power Grid as a complex network: A survey," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 11, pp. 2688–2700, Jun. 2013.

[3] 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 *2010 IEEE International Conference on Communications Workshops*, May 2010, pp. 1–5.

[4] P. Hines, S. Blumsack, E. C. Sanchez, and C. Barrows, "The Topological and Electrical Structure of Power Grids," in *2010 43rd Hawaii International Conference on System Sciences*, Jan. 2010, pp. 1–10.

[5] R. Albert, I. Albert, and G. L. Nakarado, "Structural Vulnerability of the North American Power Grid," *Physical Review E*, vol. 69, no. 2, Feb. 2004, arXiv: cond-mat/0401084.

[6] P. Cuffe, "Assortativity Anomalies in a Large Test System," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 4169–4170, Sep. 2016.

[7] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid Structural Characteristics as Validation Criteria for Synthetic Networks," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–1, 2016.

[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 Systems Journal*, vol. 6, no. 4, pp. 616–626, Dec. 2012.

[9] A. B. Birchfield, K. M. Gegner, T. Xu, K. S. Shetye, and T. J. Overbye, "Statistical Considerations in the Creation of Realistic Synthetic Power Grids for Geomagnetic Disturbance Studies," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1502–1510, Mar. 2017.

[10] 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 2008)*, Jan. 2008, pp. 183–183.

[11] P. Cuffe and A. Keane, "Visualizing the Electrical Structure of Power Systems," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1–12, 2015.

[12] J. Zhao, H. Zhou, B. Chen, and P. Li, "Research on the Structural Characteristics of Transmission Grid Based on Complex Network Theory," *Journal of Applied Mathematics*, vol. 2014, p. e261798, Apr. 2014.

[13] J. Guo, G. Hug, and O. K. Tonguz, "Intelligent Partitioning in Distributed Optimization of Electric Power Systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1249–1258, May 2016.

[14] R. J. Snchez-Garca, M. Fennelly, S. Norris, N. Wright, G. Niblo, J. Brodzki, and J. W. Bialek, "Hierarchical Spectral Clustering of Power Grids," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2229–2237, Sep. 2014.

[15] P. Cuffe, E. S. Marin, and A. Keane, "For Power Systems, Geography Doesn't Matter, But Electrical Structure Does," *IEEE Potentials*, vol. 36, no. 2, pp. 42–46, Mar. 2017.

[16] C. Coffrin, D. Gordon, and P. Scott, "NESTA, The NICTA Energy System Test Case Archive," *arXiv:1411.0359 [cs]*, Nov. 2014.

[17] C. Josz, S. Fliscounakis, J. Maeght, and P. Panciatici, "AC Power Flow Data in MATPOWER and QCQP Format: iTesla, RTE Snapshots, and PEGASE," *arXiv:1603.01533 [math]*, Mar. 2016, arXiv: 1603.01533.

[18] D. K. Molzahn, J. T. Holzer, B. C. Lesieutre, and C. L. DeMarco, "Implementation of a Large-Scale Optimal Power Flow Solver Based on Semidefinite Programming," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3987–3998, Nov. 2013.

[19] M. E. J. Newman, "Mixing patterns in networks," *Physical Review E*, vol. 67, no. 2, Feb. 2003.

[20] J. J. McAuley, L. d. F. Costa, and T. S. Caetano, "The rich-club phenomenon across complex network hierarchies," *Applied Physics Letters*, vol. 91, no. 8, p. 084103, Aug. 2007, arXiv: physics/0701290.

[21] E. Tomita, A. Tanaka, and H. Takahashi, "The worst-case time complexity for generating all maximal cliques and computational experiments," *Theoretical Computer Science*, vol. 363, no. 1, pp. 28–42, Oct. 2006.

[22] M. Fukuda, M. Kojima, K. Murota, and K. Nakata, "Exploiting Sparsity in Semidefinite Programming via Matrix Completion I: General Framework," *SIAM Journal on Optimization*, vol. 11, no. 3, pp. 647–674, Jan. 2001.

[23] R. Tarjan and M. Yannakakis, "Simple Linear-Time Algorithms to Test Chordality of Graphs, Test Acyclicity of Hypergraphs, and Selectively Reduce Acyclic Hypergraphs," *SIAM Journal on Computing*, vol. 13, no. 3, pp. 566–579, Aug. 1984.

[24] S. Fliscounakis, P. Panciatici, F. Capitanescu, and L. Wehenkel, "Contingency Ranking With Respect to Overloads in Very Large Power Systems Taking Into Account Uncertainty, Preventive, and Corrective Actions," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4909–4917, Nov. 2013.

[25] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring Network Structure, Dynamics, and Function using NetworkX," in *Proceedings of the 7th Python in Science Conference*, Pasadena, CA, USA, Aug. 2008, pp. 11–15.

[26] P. Cuffe, "A Comparison of Malicious Interdiction Strategies Against Electrical Networks," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 7, no. 2, pp. 205–217, Jun. 2017.

[27] B. C. Lesieutre, D. K. Molzahn, A. R. Borden, and C. L. DeMarco, "Examining the limits of the application of semidefinite programming to power flow problems," in *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2011, pp. 1492–1499.

[28] C. Coffrin, H. L. Hijazi, and P. V. Hentenryck, "Strengthening the SDP Relaxation of AC Power Flows With Convex Envelopes, Bound Tightening, and Valid Inequalities," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3549–3558, Sep. 2017.

[29] J. Lavaei and S. H. Low, "Zero Duality Gap in Optimal Power Flow Problem," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 92–107, Feb. 2012.

[30] E. Dall'Anese, H. Zhu, and G. B. Giannakis, "Distributed Optimal Power Flow for Smart Microgrids," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1464–1475, Sep. 2013.

[31] R. A. Jabr, "Exploiting Sparsity in SDP Relaxations of the OPF Problem," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 1138–1139, May 2012.

[32] R. Grone, C. R. Johnson, E. M. S, and H. Wolkowicz, "Positive definite completions of partial Hermitian matrices," *Linear Algebra and its Applications*, vol. 58, no. Supplement C, pp. 109–124, Apr. 1984.

[33] E. W. Dijkstra, "A Note on Two Problems in Connexion with Graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.