Tractable Algorithms for Constructing Electric Power Network Models

By
Jonathan M. Snodgrass

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Electrical & Computer Engineering)

at the
UNIVERSITY OF WISCONSIN–MADISON
2021

Date of final oral examination: 08/24/2021

The dissertation is approved by the following members of the Final Oral Committee:

Christopher DeMarco, Professor Emeritus, Electrical & Computer Engineering
Bernard Lesieutre, Professor, Electrical & Computer Engineering
Line Roald, Assistant Professor, Electrical & Computer Engineering
Michael Ferris, Professor, Computer Sciences
Laurent Lessard, Associate Professor, Mechanical and Industrial Engineering,
Northeastern University

Abstract

Common in many aspects of power systems research is reliance on mathematical network models to conduct simulations and analyses of the electrical grid. The advancement of new algorithms for grid optimization, planning, and control rely heavily on large-scale, high-quality, realistic test data sets with which to evaluate such algorithms. Before 2001, the network models of the real-world electric transmission system were publicly available in the U.S., and these often served as the definitive benchmarks for algorithm development. However, security concerns have increasingly limited the availability of such real-world grid data, and requirements for rigorous, extensive algorithmic testing have expanded the need for a very wide range of data sets. These needs have driven emerging research into algorithms to create large-scale, realistic, synthetic power grid models.

Many traditional approaches for test network data set construction focus on taking a subsection of a real-world power system, and obscuring or anonymizing any protected information. However, data sets produced in this way are often small-scale, lack geographic information, and may omit many of the engineering parameters associated with each component. There have been growing efforts to improve upon traditional methods to create large-scale realistic synthetic power system networks. These more recent efforts typically fall into one of two categories: algorithms that focus primarily on characteristics of the graph associated with the transmission network, and algorithms that focus on the electrical performance of the power system. Graph-focused algorithms are attractive in their ability to create large-scale synthetic test systems at relatively low computational cost. However synthetic test systems produced in this fashion have often proven less realistic than initially supposed, and in particular, have failed to provide effective benchmark data sets for new power system analysis methods. Algorithms that model electrical performance of the network can create higher fidelity network models.

However, most of these algorithms rely on some form of mixed integer optimization, rendering them intractable when one seeks to create large-scale power system models.

To overcome the limitations of the existing literature, this dissertation creates novel algorithms that primarily focus on the electrical performance of the network, while employing formulations that yield tractable optimization computations (e.g., linear or quadratic programs). These formulation employ constraint equations that capture key electrical behavior of the power system, while using graph statistics as a feedback mechanism to iteratively improve the networks being constructed. The first set of algorithms modifies the classic network flow model ("KCL" constraints, in the terminology of circuit analysis), to form a linear programming-based method for constructing synthetic power networks. The second set of algorithms seeks to improve upon this first formulation by adding KVL constraints and constitutive relations for transmission circuit elements, drawing on a classical optimization-based minimal energy formulation of circuit equations that yields a successive quadratic programming calculation. Combined with sparsity-enhancement via a LASSO approach, these yield tractable network construction algorithms that can create large-scale, realistic, synthetic power system data sets. These algorithms are demonstrated through the creation of a large number of test data sets, producing models for medium to very large scale networks. In particular, the synthetic networks created for this dissertation range from 3,000 to 20,000 buses. Extensive empirical analysis in this thesis demonstrates that these synthetically produced data sets provide a good match to electrical performance characteristics and graph metrics

compiled from existing, real-world power systems.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Motivation

Since the introduction of the first computer, the power systems research community has relied on "realistic but not real" synthetic mathematical network models to test new algorithms for a variety of power system analyses. While the realism of the networks varies greatly, one thing all test cases have in common is that they are publicly available and contain no CEII (critical energy infrastructure information [1]), and thus can be published along with the proposed algorithms, research, or case studies [2]. This dissertation presents algorithms that create power system synthetic networks given an input set of buses (nodes) with associated geographic coordinates and net power injection scenarios.

However, a key difficulty for network creation algorithms is that when beginning with n given nodes, there exist up to $\binom{n}{2}$ potential transmission branches that could be added to interconnect these nodes. If all potential transmission line additions (also referred to

as "potential transmission line paths" or "potential transmission paths") were modeled in the network, this would result in $2^{\binom{n}{2}}$ possible combinations of "in-service" and "out-of-service" branches. Many existing approaches deal with the large polynomial number of branches and exponential number of network configurations by either substantially limiting the potential network connections [3, 4], clustering nodes to reduce the size of n [5, 6], or limiting the size of the network that can be modeled [7]. Thus, given the challenges posed by the electricity sector restructuring and the increased penetration of renewable energy, there is growing value in new classes of tractable network creation algorithms to create synthetic power system networks to model and study proposed changes to the electric grid.

Additionally, an inherent difficulty in designing a synthetic network creation algorithm is balancing graph theory and statistics from existing networks with the electrical performance of the synthetic networks being created. Researchers creating synthetic power system networks want to design large models that mimic the characteristics of the existing power grid. However, many of the synthetic network creation algorithms in the research literature fail to adequately model the electrical performance or behavior of the resulting power system. As explored later in the literature review in Chapter 2, many algorithms do not model the electrical characteristics or performance of the network at all, but instead create networks whose graph characteristics match statistics drawn from real-world power systems. Other network creation algorithms use approximations of the electrical flow in the network, while still others utilize DC power flow approximations coupled with graph theory metrics. What all of these algorithms have

in common is that they primarily rely on graph metrics or other heuristics that are not directly focused on the electrical performance of networks being constructed. Additionally, these algorithms rely on having existing power system network statistics, so these algorithms cannot be used to design a power system from the ground up since there is no existing grid from which to draw graph statistics. While the networks created by these graph-based algorithms initially appear to be realistic, further examination shows that their electrical behavior deviates from real power system networks. Thus, new classes of network construction algorithms are developed in this dissertation that primarily focus on the electrical performance of the networks being constructed while still utilizing graph theory and metrics to create realistic synthetic power systems.

These novel classes of algorithms draw on prior research from both the optimization and power systems communities. The first set of algorithms presented are motivated by the classical network flow problems and further inspired by early transmission expansion research, as further explored in Section 1.4.1. The second set of algorithms combines an optimization reformulation of DC power flow with a classic sparsity enforcing technique from the optimization community as detailed in Section 1.4.2.

1.1 Motivation for Creating Synthetic Power System Networks

Electric utility companies, regional transmission operators (RTOs), and Independent System Operators (ISOs) heavily rely on realistic models of the electric power system for the daily operation and short to long-term planning of the electric system. These models

are comprised of mathematical representations of network elements such as transmission lines, transformers, series and shunt devices, generators, and loads. The detail in these models varies greatly, from the DC power flow models used for steady-state planning to the dynamic cases utilized for transient stability studies. These mathematical models are utilized primarily because of their accuracy, but also because the alternatives such as conducting experiments on the actual power network, are not desirable. For example, determining the impact of a line contingency on the transmission grid by actually tripping the transmission line is not desirable, and also not allowed by many ISOs.

In the U.S. from 1935 until 1992, electric utilities were operated as vertically integrated monopolies where the same company controlled all aspects of the power system: generation, transmission and distribution. While this resulted in a lack of competition since there was only one utility company in each given geographic region, it enabled sharing of information within the utility company. This made forecasting load growth, generation and network expansion planning easier since data and information was freely shared among the different divisions of a utility.

However, in 1992 the National Energy Policy Act was passed, which created the outline for a wholesale generation market. Later in 1996, FERC issued order 888, which required utility companies to allow independent power producers (IPPs) to have non-discriminatory open access to the transmission system and established Independent System Operators (ISOs). Finally, Regional Transmission Operators (RTOs) were created in 1999 under FERC order 2000. These acts and orders resulted numerous states having

deregulated electricity markets where IPPs, RTOs, ISOs, and LSE (load serving entities) operate separately from each other. While the resulting electricity markets can help to decrease operating costs, the separation of utilities into different entities has resulted in a lack of information sharing. Thus, transmission expansion often has to be conducted using incomplete or imperfect information about future generation planning and load growth.

Additionally, the increased penetration of variable renewable power sources has resulted in an increased need for tractable large-scale network creation algorithms. In particular, the regions of the United States that have favorable conditions for wind power production are often are not electrically well connected to the rest of the country. Thus, many policy makers argue that substantial investment in transmission is needed to accommodate higher levels of renewable energy penetration.

Power system research heavily relies on performing simulations on realistic models of the electric power network. Historically, such models of the existing power system were easily obtainable in the United States, since they were publicly available and could be freely disseminated. However, shortly after the September 11, 2001 terrorist attacks, the Federal Energy Regulatory Commission (FERC) created Critical Energy Infrastructure Information (CEII) regulations that removed numerous documents regarding the energy and electric industry and utilities for the safety and security of the nation's infrastructure. As an illustration of the basis for this concern, [8] shows how attackers are able to produce more effective attacks with both topological and component parameter

data as opposed to either data independently. Restricted forms and data included the FERC Form 715 [9] and other databases that correlated power system network parameters with graph topological data. While researchers, energy market consultants, and other parties with a valid reason can request data labeled as CEII, the data and network models themselves cannot be published or released along with any accompanying reports or simulations. This resulted in the absence of large, realistic, publicly available power system network models for researchers to use. Given the age, small size, and lack of geographic and component data in the publicly available test cases [10], many have observed that there is a need for large, realistic, geolocated power system test systems. These models can be used to study a wide range of future power system algorithms and applications, such as optimal power flow [11], transmission expansion planning, and studying the impact of new technologies such as the unified power flow controller. Thus, algorithms that are useful both for creating synthetic power system cases and designing "greenfield" power system networks could be of particular value, and motivate this dissertation.

A key challenge in representing power system networks is the trade-off between accuracy and computational tractability. As later explored in this dissertation, algorithms for building models of power systems can vary tremendously in computationally cost, depending on the level of detail captured in the model. As reviewed in Chapter 2, many prior methodologies focus heavily on graph theory and graph-based metrics for constructing synthetic power system networks. Conversely, algorithms that incorporate both the electrical performance and graph metrics of the network usually significantly

limit the number of candidate branches considered to be added to the transmission network. Thus, a computationally tractable method that considers both power system performance characteristics and graph metrics without a substantial reduction in the number of candidate branch additions would be a significant contribution to the power system network construction literature. Such methods will be explored in this preliminary document, seeking to utilize the benefits of network graph metrics and the experience of graph theoretic research, as well as employing insights relating to electrical performance of the power system.

When creating synthetic electrical networks, work here will focus on algorithms that start with a network of buses with few to no existing network elements serving as edges in the graph. This results in a graph that is not connected, and typically causes electrical equations associated with the network to be infeasible. To address this difficulty, synthetic network creation algorithms often create initial networks targeting graph metrics drawn from real-world systems. However, it is difficult for these synthetic network creation algorithms (NCAs) to feasibly ensure system performance or reliability in the first steps of the algorithms. In contrast, power system network operators are concerned with system reliability, performance and cost, and typically do not seek designs that explicitly match a given node degree distribution, or create a small-world network, or otherwise target graph-based metrics.

As explored in later sections, these differing design methodologies and objectives between synthetic system creation and actual power system design make it difficult for any

algorithm to be used for both purposes. Purely focusing on the power system operational goals is impractical for synthetic network creation, since the resulting mixed-integer, non-linear program (MINLP) is computationally intractable at large scales. However, as later shown, creating power system networks purely using graph metrics derived from real-world power systems results in networks that are unrealistic. To address these issues, two classes of algorithms are created and presented in this dissertation that use both graph metrics and power system operational characteristics in network construction.

There are many proposed transmission expansion or network creation algorithms that address the need for additional transmission lines to accommodate increased renewable energy production, and some of the methods are summarized in [12]. However, as later detailed in Section 2.4, most of these methods make the planning process more tractable by removing a large number of potential transmission line paths before any detailed electrical analysis is performed. Oversimplifying slightly, the $2^{\binom{n}{2}}$ possible combinations of transmission line additions are a priori limited to some very modest number. Unfortunately, there is no guarantee that the resulting expansion plan is optimal, as a truly optimal expansion plan might have included some or many of the paths that were never considered in a more detailed analysis.

Thus, given the challenges posed by the electricity sector deregulation and the increased penetration of renewable energy, there is growing value in a new class of tractable network creation algorithms that can model the large number of potential transmission line paths and still produce a near-optimal expansion plan without complete information

regarding future generation expansion.

1.2 Network Construction Algorithms: Definitions and Frameworks

To guide the discussion regarding the existing network construction algorithms and those algorithms proposed in this dissertation, two frameworks are presented below. The first is a list of four steps commonly addressed by network construction algorithms, and will be utilized to classify and compare sets of network construction algorithms. Next, a high-level conceptual framework for implementing these four steps is developed, and will be utilized to compare the classes at and formulation types of the algorithms in the existing literature and this dissertation. The four steps are:

1. Choosing whether to add a new element on a given bus-to-bus path, i.e. assigning a value to the binary decision variable $\zeta_{i,j}$.
2. Assigning transmission element parameters such as per unit impedance and element capacity limit.
3. Determining the voltage levels of the new transmission elements and creating buses within substations that are connected by transformers.
4. Assigning the physical parameters of the transmission lines or other network elements such as conductor material type and size, conductor spacing, etc.

Depending on the nature of a specific algorithm, all four steps might be performed

multiple times, iterating between the different steps and updating values as needed. A separate problem that is not covered in this dissertation is the bus placement problem. This is the problem that creates, modifies or expands the set of vertices $\mathcal{V}$ (or power system buses), and chooses the location coordinate data for buses that are geolocated. Here we assume this information is specified at the outset.

This dissertation will first present network construction algorithms (NCAs) inspired by network flow-based algorithms such as the early work of [13]. In our framework, [13] is classified as a "KCL-compliant" power system construction algorithm. These algorithms perform step 1 using an optimization-based multi-level algorithm and sometimes implement steps 2-4 by utilizing heuristics to further modify the network. After the challenges and limitations of the KCL-only algorithms are explained, a novel set of 'KCL/KVL-compliant" NCAs are presented that model KCL, KVL and the network element constitutive relations. These novel algorithms implement all four steps in a multi-level optimization-based approach. Both the KCL and KCL/KVL-compliant NCAs implement step 4 by utilizing parameters and statistics from real world transmission lines and transformers assembled as described in Chapter 3.

1.3 A Conceptual Optimization Formulation for Network Construction

At the core of any network creation algorithm (NCA) is the process of assigning values to ζ, the network topology variable. To compare the multitude of different ways

to make these assignments, this work first introduces a high-level idealized optimization framework that is used throughout the remainder of this dissertation to compare the different types and classifications of network construction algorithms. This formulation is also used to guide the discussion regarding various relaxations and reformulations of the constraints in network construction algorithms.

The majority of network construction algorithms utilize the following inputs or a subset thereof: The set of nodes/buses (or vertices $\mathcal{V}$) with associated scenarios (S) for each point in the time series (T), each containing load power withdrawals (P_D) and specified generator power injections (P_G). An ideal network construction algorithm would solve the following optimization problem (1.1) with the full security constrained AC power flow equations being satisfied for each time step t in T. The network-related decisions variables ζ and b will be set for the entire time horizon, while the $P_{G,s}(t)$ variables change during each time step. Note that some simplifications appear even in this "umbrella" formulation, such as the choice not to represent inter-temporal constraints. This formulation is presented first so that the terminology established in this Section can be later utilized in the literature review in Chapter 2.

$$\min_{\zeta, b, P_{G,s}(t)} \quad \sum_{k \in 20 \text{ years}} \text{Production Cost}(\zeta, b, P_{D,s}(t), P_{G,s}(t)) + \text{Capital Cost}(\zeta, b) \tag{1.1a}$$

$$\text{s.t.} \quad \zeta \in \mathbb{Z}_2^m, \tag{1.1b}$$

Constraints for N-1 security constrained AC power flow feasibility, (1.1c)

for each time k, (1.1d)

i.e. satisfying KCL, KVL, and network element relations with, (1.1e)

Transmission line limits, (1.1f)

Bus voltage limits, (1.1g)

Generator limits, (1.1h)

System stability limits, (1.1i)

satisfied for each N-1 contingency and all scenarios (1.1j)

The authors in [4] have proposed a mathematical formulation that implements many of the aspects of the high-level "pseudo-formulation" in (1.1). The formulation by Tang and Ferris [4] is a multi-level hierarchical model, with the expansion model (1.2) being minimized over all scenarios considered in the equilibrium model (1.3-1.4). Equations (1.2)-(1.4) along with the nomenclature tables (Tables 1.1-1.3) are reproduced from [4]

In the framework of [4], the RTO (regional transmission organization) is solving the transmission expansion problem to minimize the total cost of operating the electric grid and the cost of investing in new transmission lines. The transmission line investment variables (x), load demand (d), LMP values (p) and generator unit commitment dispatches (u) are optimized for each OPF scenario $\omega \in \Omega$) and these values interact with the cost of expansion and operation in (1.2). Additionally, each IPP (Independent Power Producer, termed "Firm" in [4]) seeks to minimize the total cost of producing power ($\sum_{j \in G_f} u_j^{\omega} C_j(g_j^{\omega)}, y_j)$ and the cost of investment in upgrading existing generators or building new generation ($\Phi(y_j)$) subject to budgetary constraints ($H_{f(y)}$).

The connections between (1.1) and (1.2)-(1.4) are as follows. The transmission line investment variables (x) encompass both the topology variable ζ and the parameter variable b in the formulation in (1.1). The KCL constraint in (1.1e) is modeled using the power balance constraint (1.3b), while KVL and the constitutive relations are modeled using DC power flow in (1.3c). The generator limits (1.1h) and line limits (4.1c) are directly modeled in (1.3d)-(1.3e).

In [4], the authors chose to significantly limit the number candidate transmission paths considered in their expansion plan, thus limiting substantially the search space for the variable x. In particular the authors state: "we expect physical and infrastructural limitations as well as political and geographical complexities to severely limit possible layouts." Thus, to the extent that real-world geopolitical, topographical and environmental issues impact the problem, it is reasonable to assume that they primarily serve to

limit the potential number of transmission paths, rather than fundamentally changing the nature of the network construction problem. Thus, the work here will maintain the notation in (1.1), with the decision variable indicating network topology (ζ) separate from the decision variable modeling network branch parameters (b). Thus, if there are potential transmission line paths that are infeasible due to real-world constraints (crossing large bodies of water, environmentally sensitive regions, etc.), the corresponding $\zeta_{i,j}$ entries can be fixed at 0. However, to illustrate the computational tractability even in the case of large number of integer decisions, in this dissertation the complete graph of potential transmission paths is modeled, i.e. no ζ elements are set to 0, except for potential transmission paths that are clearly longer than is practically feasible (Section 3.2.5).

In addition, the differences between the two formulations are that (1.2)-(1.4) explicitly model generator expansion and investment, while (1.1) does not. Also, (1.1) includes the full security constrained AC power flow (in narrative form) as constraints, while (1.2)-(1.4) only includes the DC power flow model without N-1 contingency constraints. Thus, both formulations will prove useful throughout the dissertation when reviewing current network construction algorithms or presenting the novel algorithms in this dissertation.

Expansion Model (RTO)

$$\min_{x \in X} \quad \overbrace{\Psi(x)}^{\text{investment cost}} + \overbrace{\sum_{\omega \in \Omega} \pi_\omega \sum_{j \in N} \left(d_j^\omega p_j^\omega(x) + u_j^\omega(x) \hat{c}_j^\omega(x) \right)}^{\text{consumer cost}} \tag{1.2a}$$

$$\text{s.t.} L(x) \leq 0 \tag{1.2b}$$

Equilibrium Model

$\text{OPF}^{\text{UC}}(\forall \omega \in \Omega)$:

$$\min_{u,\, g,\, z,\, \theta} \quad \sum_{j \in G} u_j^\omega C_j \left(g_j^\omega, y_j \right) \tag{1.3a}$$

$$\text{s.t.} \qquad g_j^\omega - d_j^\omega = \sum_{i \in E(j)} z_{ij}^\omega \perp p_j^\omega \quad \forall j \in N, \tag{1.3b}$$

$$S_{ij} \left(\theta_i^\omega - \theta_j^\omega \right) = z_{ij}^\omega \quad \forall ij \in A, \tag{1.3c}$$

$$u_j^\omega \underline{g}_j \leq g_j^\omega \leq u_j^\omega \bar{g}_j \quad \forall j \in N, \tag{1.3d}$$

$$z_{ij}^\omega \in [-\bar{z}_{ij}(x), \bar{z}_{ij}(x)] \quad \forall ij \in A, \tag{1.3e}$$

$$\text{free}, \, u \in \{0, 1\} \tag{1.3f}$$

$\text{Firm}(\forall f \in F)$:

$$\min_{y} \quad \sum_{\omega \in \Omega} \pi_\omega \sum_{j \in G_f} \left(u_j^\omega C_j \left(g_j^\omega, y_j \right) + \Phi \left(y_j \right) \right) \tag{1.4a}$$

$$\text{s.t.} \quad H_f(y) \leq 0 \tag{1.4b}$$

Set	Description
$\omega \in \Omega$	Demand scenarios
$i, j \in N$	Network Vertices
$ij \in A$	Transmission lines, $A \in (N, N)$
$f \in F$	Firms
G	Generation nodes, $G \subseteq N$
$G_f \subseteq G$	Generation nodes belonging to firm f
$E(j) \in N$	Edges connected to node j

Table 1.1: Description of Sets

Model	Variable	Description
RTO	x	Investment in transmission line expansion
Firm	y_j	Investment in generator at j
ISO	z_{ij}^{ω}	Real power flowing along $i-j$ in scenario ω
	θ_j^{ω}	Voltage phase angle at j in scenario ω
	g_j^{ω}	Real power at j in scenario ω
	u_j^{ω}	Dispatch of generator at j in scenario ω

Table 1.2: Description of Variables

The high-level, conceptual formulation of (1.1) represents an idealized network creation algorithm (NCA) in the sense that its solution includes ζ (the choice of placement of transmission elements) and b (their electrical parameters). This model can be expanded by including unit commitment [14], generator expansion planning (GEP) [15], coupling the impact of the natural gas system to the electric transmission system [16,17], and/or modeling system dynamics [18,19]. The idealized network construction formulation could be further extended by allowing additional buses to be added to the network, i.e. expanding or modifying set $\mathcal{V}$. More detailed mathematical models of the short term network creation problem (also known as static network planning or static TEP [20]) are given in [21]. Several relaxations of (1.5) consisting of varying levels of tightness or strictness are presented in [22, 23].

Parameter	Description	Model
d_j^ω	Demand at j in scenario ω	All
$\Psi(x)$	Cost of line investment	RTO
$\widetilde{c}_j^\omega(x)$	Generator uplift costs	RTO
$p_j^\omega(x)$	LMP value	RTO
$L(x)$	Budgetary/engineering limitations	RTO
π_ω	Probability of scenario ω	RTO,Firm
S_{ij}	Susceptance of line $i-j$	ISO
$\left[\underline{g}_j, \bar{g}_j\right]$	Operating limits of generator at j	ISO
$\bar{z}_{ij}$	Capacity of transmission line $i-j$	ISO
$C_j\left(g_j^\omega, y_j\right)$	Cost function of generator at j	ISO,Firm
$\Phi(y)$	Cost of generator investment	Firm
$H_f(y)$	Budgetary constraints of firm f	Firm

Table 1.3: Description of Parameters

The network constraints, generator, bus, and transmission line limits all need to be satisfied (feasible) for each scenario s of each time step t. However, the transmission constitutive relations and line limits for potential transmission lines themselves are functions of the decision variable(s) b, further increasing the complexity and difficulty of the idealized NCA formulation. Also, if the NCAs being studied allow for existing transmission lines to be upgraded as part of the expansion process, the parameter decision variable b becomes expanded to include upgrades to existing transmission lines. This means the line limits of the existing transmission lines are also functions of b.

Transmission line design parameters (here denoted as b) can be represented using the lumped parameter π-model as explained in textbooks such as [24]. A mathematical formulation of the security constrained AC optimal power flow (AC SCOPF) equations can be found in [25]. The typical objective function of AC SCOPF is to minimize the operational cost of a given network with typical decision variables being generator

output power and voltage setpoints. This is the same objective function as the first term in (1.1a), and the same P_G decision variable as that appearing in (1.1). However, the objective function and constraints in [25] need to be extended to consider potential transmission line expansions represented by the variable ζ. For more information on security constrained optimal power flow, including terminology and post-contingency corrective actions, see [26, 27].

An ideal mathematical formulation of the network construction problem without system stability limits is presented in [28] and is utilized for transmission expansion planning. In [28], the authors Krishnan, Ho, Hobbs, et al. present an ideal version of the combined generation and transmission expansion problem. The authors list the criteria that a combined TEP/GEP formulation needs to address in order for it to satisfy the requirements of electric utilities, regional transmission operators (RTOs) and independent system operators (ISOs). The objective function contains the operational cost of the system and the capital cost of adding new lines to the network, consistent with the formulation here in (1.1a). However [28] expands its scope to include the cost of building new generators.

Additionally, [28] examines various formulations that differ in their modeling detail with regard to electrical network behavior, ranging from AC and DC power flow to the simple transportation model. It also includes several possible relaxations and approximations to the problem, including relaxing the integer decision variables to continuous variables and approximating the non-linear power flow equations by linear equations.

Additional modeling aspects include inter-temporal constrains, modeling uncertainties, and spacial granularity. In conclusion, [28] states that there are currently no proposed formulations that meet all of their listed criteria, This paper was published in 2015, indicating that there is still substantial research to be done in the field of transmission and generation expansion.

Thus, the full mixed integer non-linear programming (MINLP) formulation shown in both (1.1) and (1.2)-(1.4) are typically not computationally tractable when considering the full set of potential candidate branch additions for medium to large scale transmission networks with thousands of buses or more. Therefore, reformulations need to be performed on the objective function, transmission line and transformer modeling equations, power flow equations, security constraints, and/or decision variable constraints to design algorithms that are tractable.

1.3.1 Model Simplifications for Tractability

Since brute-force implementations of full mixed integer nonlinear programming (MINLP) implementations of the ideal conceptual network creation algorithms (1.1)-(1.4) are both impractical and intractable for medium to large scale networks, simplifications of (1.1)-(1.4) and (1.5) are needed. The algorithms presented in Chapter 4 may be viewed as severe simplifications through the use of the transportation model, which includes only the KCL constraints. Since the transportation model constraints (KCL equations and line limits) and objective function (minimize the weighted L1 norm of the element flows) are piecewise-linear and the decision variables (power flows on each network element)

are continuous, the transportation model NCA can be solved using linear programming. The resulting networks are further refined to improve their realism using the heuristic methods developed in Chapter 7, which include the use of the DC power flow equations (1.6). These linear network construction algorithms including KCL constraints are termed "KCL-compliant" or "KCL-only" NCAs (network construction algorithms) in future sections and chapters.

Next, Chapter 5 presents algorithms that model all three classes of constraints in (1.1e), i.e. KCL, KVL and the network element constitutive equations (albeit using simplified linear versions of the constitutive relations). Only the KCL constraints are directly modeled in the KCL-compliant NCAs. The KVL and element constitutive relations are captured in an expanded formulation through use of an L1 regularized loss minimization optimization problem. Instead of modeling the KVL and element relations as constraints, Section 5.1.1 shows how they can be recovered using a loss minimization objective function, that yields necessary conditions such that the element relations (Ohm's law) and KVL are guaranteed to be satisfied at an optimal solution. Since the constraints (KCL) are linear and the objective function is a polynomial with a quadratic and piece-wise linear term, an adaptation of the problem can be solved using a quadratic programming (QP) solver. This KCL constrained regularized loss minimization problem are the basis for the "KCL/KVL-Compliant" network construction algorithms (NCAs) developed in Chapter 5.

1.3.2 AC power flow constraints and DC power flow approximation

A more complete mathematical formulation of the constraints in (1.1) consistent with [21, 28], would require formulation of the AC power flow constraints in (1.1c). One modeling approximation that greatly aids in obtaining a tractable problem is the common linearization termed DC power flow. The DC power flow is utilized in both the heuristic additions to the KCL-only NCA (Chapter 7) as well as represented in the KCL and KVL NCA (Chapter 5).

The AC power flow constraints can be formulated in more detail as developed in a standard textbook description [29] as per (1.5). Note that P_i and Q_i are the net real and reactive power injections at bus i, respectively, where $P = P_G - P_D$. In the ideal NCA formulation, equations (1.5) are modeled at each bus i for each network element contingency, each time step t and each scenario s. Thus, P_i is the shorthand notation for the net power injection for the present contingency, scenario, and time step.

$$P_i = \sum_{j=1}^{n} |V_i||V_j|(G_{i,j}\cos(\theta_i - \theta_j) + B_{i,j}\sin(\theta_i - \theta_j)) \tag{1.5a}$$

$$Q_i = \sum_{j=1}^{n} |V_i||V_j|(G_{i,j}\sin(\theta_i - \theta_j) - B_{i,j}\cos(\theta_i - \theta_j)) \tag{1.5b}$$

However, as detailed in Section 2.4, many transmission network expansion algorithms do not use the full AC power flow equations due to the computational difficulties associated with such nonlinear, non-convex constraints in a large network. Thus, a simplified, lossless, linear approximation to the power flow equations (i.e. the DC power flow) is

often employed using the following assumptions [23]:

1. $R << X$ so the line admittance G is approximated to be zero
2. θ_i and θ_j are small, thus $\cos(\theta_i - \theta_j) \approx 1$ and $\sin(\theta_i - \theta_j) \approx (\theta_i - \theta_j)$
3. Per unit voltages are close to 1.0, so $|V_i| \approx 1$
4. Subject to assumptions 1 to 3, the right hand side of (1.5b) reduces to a constant, and thus (1.5b) is neglected.

These assumptions simplify (1.5a) to yield the linear DC power flow equations (1.6)

$$P_i = \sum_{\substack{i=1 \\ i \neq j}}^{n} B_{i,j}(\theta_i - \theta_j) \tag{1.6}$$

1.4 Motivations and Background for Novel Network Construction Algorithms

As explained in prior sections, modeling ζ as a symmetric matrix of binary decision variables directly as modeled in (1.1) results in an intractable problem for medium to large scale networks. Thus, a tractable algorithm that considers both the electrical performance of the network and the graph statistics and characteristics needs to be created.

For all types of network construction algorithms, a fundamental challenge in building a graph having n vertices is that there are up to $\binom{n}{2}$ potential edges (branches) that could be added to a new or existing network graph. If all $\binom{n}{2}$ potential branches (potential

transmission line paths) are modeled, this results in up to $2^{\binom{n}{2}}$ combinations of in-service and out-of-service branches, depending on the number of branches already present. A network construction problem can quickly become intractable, even for moderate sized systems. Thus, it is desirable to examine, modify and create classes of algorithms that include relaxations or reformulations of the ideal MINLP formulation of the power system network construction problem. To this end, this dissertation presents a family of problem formulations and algorithms that can be utilized for greenfield network design, synthetic network creation, and potentially be adapted for network expansion.

1.4.1 KCL-compliant Algorithm Motivation: Network Flow and AC Power Flow Relaxations

The simplest and most computationally tractable implementations of (1.1) or(1.2)-(1.4) employ linear approximations to the network electrical constraints. The simplest formulation of the ideal NCA (1.1) includes only the generator limits (1.1h), i.e. removing all constraints from (1.3) except 1.3d. Combining this with the total load power results in the simplest formulation: the copper plate approximation. The copper plate only includes the real generator output powers as the decision variables and requires that the total generation output power must match total load as the constraint. This is the network constraint used in the simplest formulation of the economic dispatch [30]. Also, as shown by [22, 23], the copper plate approximation can be viewed as the loosest relaxation of the AC power flow equations. In the algorithms in this dissertation, the

copper plate approximation is used as the constraints for the economic dispatch when determining the generator outputs for some of the net injection scenarios in Sections 4.4.1 and 5.2.3.

The first equation in the ideal network construction algorithms (1.1)-(1.4) to be modeled is the KCL term in the first constraint (1.1e), i.e. (1.3b). This results in a tractable implementation of network flow as reformulated in [31] and further refined in Section 4.2. As explained in Section 4.2.2, the objective function in (1.1a) is reformulated as the minimization of a weighted L1 norm of the flows on all elements in the network. The constraints modeled are conservation of nodal flow (KCL) and element flow limits. This results in what some authors have termed the "transportation model" of a network, and is the basis for the algorithms developed in this thesis. Also, as demonstrated by [22, 23], network flow can be viewed as a relaxation of the AC power flow equations that is tighter than the copper plate relaxation.

If a different objective function is desired instead of a minimal weighted L1 norm of the line flows, either integer decision variables must be utilized or heuristics must be employed for selecting the lines to be added to the existing transmission system. For example, [32] minimizes the cost of new elements added to the power system network, but includes integer decisions variables indicating whether an element should be added between bus i and bus j. This integer decision variable limits the number of paths with non-zero power flows. However, since mixed-integer problems (MIP) are extremely computationally challenging for large networks, [32] also proposes relaxing the integer

decision variable to a continuous one and uses a sensitivity-based heuristic to solve the relaxed linear program. Another example is [33], which also minimizes the capital cost of new line additions, but adds DC power flow constraints with continuous variables. This incorporates the KCL, KVL and element constitutive relation constraints from (1.1e). However, given the large number of potential candidate expansion paths, [33] utilizes a iterative sensitivity-based heuristic for determining the paths to add to the existing network. The potential shortcoming in these approaches is that they both use heuristic approaches to limit the number of transmission expansion candidate paths instead using an optimization problem to choose which paths are included in the network.

This first set of NCAs that include the KCL constraints are tractable algorithms that put primary emphasis on an approximation of the network's electrical performance. However, the network flow relaxation tends to create networks that are less realistic electrically than networks created using more computationally expensive methods, as explored in Chapter 6. Thus, additional heuristic steps are added to modify the network after the optimization-based part of the algorithm terminates, as explored in Section 7. While these additions improve the realism of synthetic power networks built by the transportation model, they do not address all of the limitations and create some disadvantages of their own. Thus, a new reformulation is desired that maintains the computational tractability of the KCL-compliant transportation model but also includes the KVL and element relation constraints from (1.1e). This class of algorithms addresses many of the shortcomings of synthetic power systems built modeling only the KCL constraints, while avoiding the disadvantages of substantially relying on heuristics.

1.4.2 Motivation for the Creation of KCL/KVL-Compliant Algorithms

The next constraints from the conceptual formulation in (1.1) that should be implemented are the KVL and element constitutive equations from (1.1e). The hypothesis is that introducing bus voltage electrical constraints in a network construction algorithm will increase the realism of the network created.

The complexity and realism of algorithms including KCL, KVL and the element constitutive equations (1.1e) range from the linear DC power flow approximation (1.3.2) to the full security-constrained AC power flow based mixed-integer non-linear formulation in [28]. The difficulty in implementing optimization-based methods with full power flow and security constraints is their eventual intractability for large values of n. This is because the decision (or intermediate) variables include the binary decisions of whether to assign a transmission line to a particular bus-to-bus path ($\zeta_{i,j}$). Using a modeling method such as DC power flow models the branch flow as $\zeta_{i,j}x_{i,j}(\theta_i - \theta_j)$, which results in the multiplication of binary decision variable $\zeta_{i,j}$ and intermediate variables θ_i and θ_j. This multiplication results in a mixed-integer optimization problem which are in general still hard to solve and intractable for large n. Thus, optimization solutions to (1.1) use relaxations or approximations to the constraints and/or reformulations of the problem entirely. Or, as explained in [4], to make (1.2)-(1.4) tractable, the search space for x is significantly reduced from the complete graph of potential transmission path additions.

One possible formulation is to use one of the relaxations of the AC power flow

equations such as those in [22, 23, 34]. The problem with using such relaxations or reformulations is that the relationship between s (the vector of complex power flows on the network branches), $V\angle\theta$ (the bus voltages), and S (the bus net power injections), are explicitly defined by the AC power flow equations [24]. Without modeling ζ as binary, any loss-minimization based optimization formulation results in a solution that is one of the solutions to the AC power flow equations. While it might be possible to use relaxations of the AC power flow equations that use branch currents as decision variables, little work has been done on this to date.

Another potential formulation that includes the KCL and KVL constraints are the second-order cone problem (SOCP) relaxations of the AC power flow equations as presented in [22, 23]. The SOCP formulation is a tighter relaxation of the AC power flow equations than the network flow relaxation, and thus will produce more accurate solutions [22]. These relaxations and/or reformulations are often more realistic than the transportation model, but come at a cost of increased computational complexity.

Finally, an NCA modeling DC power flow could be utilized. Traditionally, the DC power flow equations $P = B_{\text{Bus}}\theta$ are modeled as constraints, with the bus admittance matrix B_{Bus} being a function of ζ, as in [33]. Unfortunately, this results in a non-linear problem since both ζ and θ are decision variables and if ζ are integer variables, the problem quickly becomes intractable. The problem with directly relaxing ζ to become a continuous variable is that this would require multiple heuristics or other methods to ensure a sparse solution for the ζ variable. To avoid these heuristics, instead of directly

relaxing the ζ variable, (1.1) is reformulated such that the decision variables become the line flows and a flow above a threshold on a line constitutes a assignment of 1 for the zeta variable.

To create such a network construction algorithm, a DC power flow approximation of 1.1 (i.e. (1.3c)) is reformulated such that the decision variables are the flow on a given branch. The problem with directly including DC power flow equations as constraints is that this results in only a single solution to the problem, resulting in a feasible space of a single point. Instead there needs to be a trade-off between satisfying the DC power flow equations and ensuring a sparse assignment to the ζ variable. Thus, an objective function is created whose necessary conditions at optimality result in the DC power flow equations, as further detailed in Chapter 5.

The relationship between a loss minimization optimization problem and KVL has long been recognized. One presentation may be found in Chapter 5 of the classic 1963 Guillemin textbook which offers a proof that KVL can be derived from the minimum energy optimization problem for a resistive circuit [35]. Other authors who have observed the relation between optimization problems and circuit constraints include [36–40]. However, these earlier works were primarily focused on the analysis of given networks, and do not suggest using the loss minimization optimization problem (i.e. minimum energy solution) for the network construction process or for solving DC power flow. While algorithms utilize quadratic programming to model polynomial generator cost functions or minimize system losses [41,42], the author could find no mention in the existing liter-

ature of using a quadratic loss minimization to model DC power flow. After a thorough search of the existing literature, this author is confident that use of loss minimization to solve DC power flow and for network creation represents a novel contribution.

To enforce sparsity for assignments to the ζ variable, an approach similar to the LASSO is utilized. As detailed in Section 5.1.2, a trade-off is introduced between the quadratic loss minimization term and the sparcifying effect of an L1 regularizer.

1.5 Dissertation Organization

The goal of this dissertation is to present two novel algorithms for the construction of synthetic power system models. Chapters 1-3 provide background information and motivations for these novel algorithms, while Chapters 4-5 present the optimization-based portions of the two algorithms. Chapter 6 demonstrates the effectiveness of the two novel algorithms by presenting several resulting synthetic networks networks. Since the focus of this dissertation is on the creation of the algorithms for developing synthetic power system networks, more detailed analysis of individual synthetic networks is left for future work, with an overview provided in Chapter 6. A more detailed outline of the dissertation is given below.

Chapter 1 has laid the foundational terminology and establishes the motivations for the proposed NCAs in future chapters. Sections 1.2 and 1.3 presented idealized, conceptual formulations for a power system network construction problem, which established a framework for the discussion in subsequent sections. Next, the motivations for this

dissertation were presented in Section 1.4, first establishing the reasons for creating synthetic power systems followed by the inspirations for the algorithms presented in this dissertation.

Next, Chapter 2 conducts a thorough review of the existing literature on synthetic power system network creation and to establish the shortcomings of the existing methods. Section 2.1 summarizes several previously proposed generic graph construction algorithms, some of which have been adapted to create power network graphs. Section 2.2 reviews the optimization-based network construction algorithms while section 2.3 explores approaches that primarily utilize heuristics. Section 2.4 summarizes some transmission expansion algorithms and draws the motivation for the KCL-compliant network construction algorithms. Since it is not possible to create large-scale synthetic power system models using the idealized formulation in Section 1.3, Section 2.5 explains how statistics and metrics are assembled from existing power systems as a method of comparing the networks created in this dissertation.

The input data required to create the synthetic power system networks described in this dissertation is discussed in Chapter 3.

Chapter 4 presents the KCL-compliant network construction algorithms that represent linear KCL equations using the edge-to-bus incidence matrix. A multi-level optimization-based formulation of the network construction problem is given in Section 4.1. Additional steps to improve these networks are explained in Chapter 7. These additions are more heuristic in nature and therefore treated in a separate chapter instead

of being included in Chapter 4.

Chapter 5 proposes a second class of network construction algorithms that seek to enhance the realism of network models by including both KCL and KVL constraints. A minimum resistive loss optimization formulation implemented using quadratic programming is presented in Section 5.1.1, followed by a derivation showing that the necessary conditions for optimality recover KVL. Section 5.1.2 shows how including a weighted L1 norm term in the objective function enforces sparsity in the solution of line power flows, with associated analysis of this formulation given in Section 5.1.4. Section 5.1.5 presents a circuit analogy of a resistor in series with a pair of back-to-back diodes, and reformulates the optimization problem to represent DC power flow. A multi-level optimization-based network construction algorithms presented in Section 5.3 that utilizes the derivations in the previous sections. Finally, additional modifications to the networks produced in this chapter are presented in Chapter 8.

Finally, Chapter 6 examines the results of both sets of network construction algorithms presented in this dissertation. Sections 6.5 and 6.4 present both quantitative and qualitative comparisons between the synthetic power system networks and existing power system networks, as well as establishing the superiority of the KCL/KVL-compliant algorithms over the KCL-only algorithms.

Chapter 2

Network Creation Algorithms

With the motivation for creating synthetic power systems established in Section 1.1, and the background and motivation for the two classes of algorithms presented in this dissertation established in Sections 1.4.1 and 1.4.2, this chapter provides further motivation by presenting a review of the existing literature. Balancing the realistic treatment of network graph topology and electrical performance characteristics in the construction of large-scale synthetic power system data sets has proven a challenging problem. As illustrated in the following literature review, the most computationally efficient existing algorithms tend to focus primarily on graph topology, with significantly less consideration given to ensuring realistic electrical performance in the resulting network. Conversely, more general formulations that fully consider network topology and power flow performance often limit the search space (e.g., [4]), in part because of computational challenges in the mixed integer nonlinear program that typically results.

This literature review is broken up into four sections, with each section covering a different class of network construction algorithms that could be utilized to construct power system networks. A brief review of some individual algorithms will be discussed, along with an explanation for why these algorithms, and the class of algorithms in general, are insufficient or inadequate to create realistic and functional power system networks. First, Section 2.1 presents a general overview of the larger category of graph creation algorithms, followed by subcategories that could be utilized for power system network creation. Section 2.2 then presents optimization based methods ranging from linear programming to mixed integer nonlinear programming. Section 2.3 summarizes additional algorithms that do not fall neatly into either of these two categories, such as anonymizing existing power system networks. Finally, Section 2.4 discusses historical power system network creation, which primarily relies on transmission expansion algorithms and heuristics.

Given the overview of the network construction problem in Section 1.2 and idealized formulation in Section 1.3, previously proposed network construction algorithms can be categorized in this context. Before introducing the novel network construction algorithms, a review of existing synthetic power system network creation algorithms (NCAs) is performed to show the shortcomings of these methods, and the benefits of the proposed algorithms.

2.1 General graph construction algorithms

Although the majority of this dissertation focuses on algorithms that create or expand power system networks, graph creation algorithms can be used to create a variety of graphs describing behavior across various sectors. Such graphs are used to model interactions between various entities and elements of organisms and entities such as the Internet, transportation systems, social networks, neuron networks or biological organisms. Thus, it is beneficial to give a brief overview of graph construction algorithms (GSAs) that can be used to create graphs that can be used for other applications, and highlight which algorithms can be utilized or adapted for power system network creation.

Using standard notation for a graph (see [43] for an example) defines a graph (G) to be comprised of a set of vertices ($\mathcal{V}$) connected by a set of edges (E). The graph can be undirected, meaning the edges have no polarity. Alternatively, the graph many be directed, indicating that each edge has an originating vertex and a terminating vertex. With the exception of some graphs discussed in Section 2.1, the graphs discussed in this dissertation are directed graphs. However, when considering graph metric analysis such as node degree or cycle basis, the graphs are treated as undirected, since power can flow in either direction in a transmission branch.

Graph creation algorithms (GCA) seek to build a graph by connecting the vertices (nodes) together using edges (branches). Some algorithms are expansion algorithms, which start from an existing graph and add additional edges and sometimes new nodes

to the existing graph. Other algorithms will be termed "greenfield" graph creation algorithms which construct a graph starting with a few to none of the nodes initially connected together. While all graph construction algorithms create or modify set E, some algorithms do not modify set $\mathcal{V}$ and thus require it as a fixed input. Other algorithms build up the set $\mathcal{V}$, either creating it "from scratch" or modifying a starting set $\mathcal{V}$. Unless otherwise specified, the algorithms proposed or presented in this dissertation require a fixed set of nodes specified as $\mathcal{V}$ as an input.

The work of this dissertation will focus on the creation of power system networks, which are graphs that contain added information to model the electrical performance of the network. To create these graphs and their associated electrical network parameters, a more general class of algorithms is developed and is termed network construction algorithms (NCA). While graph construction is fairly consistently defined in previous literature (see Section 2.1), this dissertation will use the term network construction to introduce added features not typically found in previous work. Here, network construction algorithms are defined as a set of algorithms that create a graph (G) with topology decision variable ζ along with associated network element parameters (b) and electrical circuit constitutive relations. Network construction algorithms (NCA), are similar to graph creation algorithms (GCA) in that they both can create or modify sets E and possibly V. However, network construction algorithms are a superset of graph creation algorithms because NCAs create edges that play the role of electrical network elements that have one or more associated parameters and voltage-current constitutive relations, whereas standard GCAs create graph edges with at most one scalar parameter, the edge

weight.

2.1.1 Categories of Graph Construction Algorithms

While graphs have numerous properties and classifications [44, 45], the work in [46] presents three characteristics that will be the focus of the following review of GSAs: randomness, heterogeneity, and modularity. The extent to which each graph exhibits each feature helps classify the graph into various categories. In a heterogeneous graph the number of connections each node has is more or less the same. The randomness property of a graph is the probability of any two randomly chosen nodes to be connected together. A graph with high randomness will have a low probability that two randomly chosen nodes will be connected. Finally, modularity describes groups of nodes that are more densely connected together than to the rest of the network.

The first class of graphs is regular graphs, include graph types such as meshes, trees, or polygons. Regular graphs typically have longer average paths and higher clustering coefficients. There are numerous methods for generating regular graphs, including deterministic, heuristic and optimization based methods. A summary of algorithms utilized to deterministically generate all of the graphs having a set number of vertices is presented in [47]. Minimum-weight spanning trees are a subset of regular graphs and there have been numerous algorithms proposed to generate the spanning tree graphs. The authors in [48] summarize several heuristic algorithms that span a range of time complexity generate minimum spanning trees (MSTs), while [49] presents a survey of integer programming formulations to solve the minimum spanning tree problem. Since

MSTs can be utilized as part of the power system network creation process, they will also be briefly discussed in Section 2.1.3.

The next type of graphs are random graphs, which can range from random ER (Erdos-Renyi) graphs to graphs with a specified node degree distribution but randomly assigned nodes. ER (Erdos-Renyi) graphs [50] typically have low heterogeneity and Gaussian distributions for their node degree distribution, and are created by picking pairs of nodes at random and connecting them. On the opposite spectrum, graphs can be randomly generated but do not necessarily have low heterogeneity or a Gaussian degree distribution. For example, both [51] and [52] present algorithms for generating random graphs that match a given node degree distribution. Finally, [45] presents a comprehensive review of modeling approaches for random graphs.

2.1.2 Algorithms Used for Graph Creation

A number of authors have asserted that power system networks have specific graph properties, such as fairly predictable node degree distributions, and that construction of network graphs that mimic these will yield realistic power system models [6, 53, 54]. However, the work in [55] shows the networks created using small-world node degree distributions are less realistic than first expected.

Examples of approaches that utilize graph metrics to create synthetic power systems include [56] which proposes creating synthetic power system network models by utilizing small-world graph properties. Additionally, the approach in [3] utilizes the placement

of nodes on the Delaunay triangulation to construct an initial synthetic power system network that is later modified using a DC power flow based heuristic. The authors in [57] propose a method of using a sensitivity analysis to add additional lines to account for N-1 contingency analysis, which is extended to the synthetic power system network design problem in [58].

A classical method of creating graphs is clustering, which either uses sets of existing vertices or modifies existing graphs to generate new graphs. The authors in [59] present an extensive review of various clustering techniques and how they can be utilized to create different graphs. Additionally, [60] summarizes and evaluates graph clustering algorithms. Finally, an type of clustering, the k-nearest neighbor algorithm, can be utilized to generate graphs as presented in [61].

More recent techniques for generating graphs tend to focus on optimization algorithms. Several such algorithms and techniques are presented in [62], and one specific optimization-based example is [63] which generates graphs by minimizing an energy-like function, and makes use of simulated annealing.

Other algorithms such as preferential attachment are explored in [64–66]. For a broad and extensive review of numerous algorithms used for graph construction, see [65].

2.1.3 GCAs useful for power system network creation

Several of the graph construction algorithms described above can be utilized to create power system networks. For example, one such algorithm include clustering algorithms,

which first cluster then connect smaller sets of buses (power system nodes) before connecting the clusters together. These cluster-and-connect style of algorithms are further explored in Section 2.3.1. They also make use of the next category of graphs and accompanying GSAs: small-world graphs.

Small-world graphs lie in the middle of the spectrum between regular and random graphs, with portions of the graph being highly clustered like regular lattice graphs yet the overall graph has smaller average path lengths, which is characteristic of random graphs [53]. These characteristics are often present in power system networks, since smaller local networks are often more densely connected, with a smaller number of long-distance links connecting the smaller clusters in the network. Authors such as those in [54] have made use of the small-world properties of power systems to create synthetic power systems.

Finally, the Delaunay triangulation [67] can be utilized to connect nodes or buses together into triangles where all of the resulting circumscribing triangles only contain three vertices. Work in [3] observes that in real world data sets studied by the authors, approximately 95% of transmission lines connect buses that are either first or second neighbors on the Delaunay triangulation, and uses this information to motivate algorithms for the creation of synthetic power system topologies.

2.2 Optimization based methods

The difficulty for optimization-based methods with full power flow and security constraints is their eventual intractability for large values of n. This is because the decision (or intermediate) variables include the integer decisions of whether to assign a transmission line to a particular bus-to-bus path ($\zeta_{i,j}$). Using a modeling method such as DC power flow models the branch flow as $\zeta_{i,j}\frac{(\theta_i-\theta_j)}{x_{i,j}}$, which results in the multiplication of integer decision variable $\zeta_{i,j}$ and intermediate variables θ_i and θ_j. This multiplication results in a mixed-integer optimization problem which are in general still hard to solve and intractable for large n. Thus, optimization solutions to (1.1) often result in one or more relaxations of the constraints and/or reformulations of the problem entirely. The complexity and realism of the algorithm implementations of (1.1) range from the copper plate relaxation [22] to the full mixed-integer non-linear implementation in [28]. A progression ranging from linear programming formulations to the full mixed-integer nonlinear programming model is presented in the subsections below.

2.2.1 Linear Programming Based Network Construction Algorithms

Linear programs require both their objective function and constraints to be linear. The constraints modeled can range from total system power balance (i.e. sum of generation must equal total load) in the copper plate formulation [22], to more complete node-by-node current or power balance constraint (KCL) [13], to modeling DC power flow on existing transmission lines [68] or linearized AC power flow constraints [69]. A

summary of other linear relaxations for transmission expansion is given in [70].

The first step in creating a network construction algorithm (NCA) is to include the KCL constraints for each node (1.1e), and such an NCA will be subsequently termed KCL-compliant algorithms. This results in the transportation or transshipment model as summarized in [71]. The transportation model is a subset of the larger class of network flow problems, with the most typical formulation minimizing a weighted L1 norm of the network branch flows subject to the node power balance constraints (KCL) and branch flow limit constraints, as presented in [31]. In this formulation, ζ is relaxed to become a continuous variable, and an assignment of 1 is represented by a non-zero flow on a branch.

One of the earliest approaches to construct power system networks utilizing an optimization based solution to the transportation problem is presented in [13]. Garver's method creates a complete graph of potential transmission lines (termed overload paths), and formulates an optimization problem minimizing a weighted L1 norm of the flow on the existing lines and overload paths (p). Although perhaps not recognized in the 1970 paper, Garver's use of the L1 regularizer in the form of the sum of the absolute value creates a sparse assignment of values to the vector p of decision variables [72]. Because Garver assigns the overload paths a weighting factor that is five times the weighting factor on existing transmission lines, the majority of the flow values on potential transmission line paths are 0, while the existing lines are more heavily loaded. While each instantiation of the linear program may be solved very efficiently, Garver's approach

requires solving many linear programs to arrive at a final network construction. To increase the accuracy of the power system networks, a hybrid method can be introduced where the flows on the existing transmission lines are modeled using DC power flow while the potential transmission line paths are represented using the transportation model [68, 73].

Finally, additional formulations of the network flow relaxation are given in [22, 23] that model real and reactive power flows and net injections, as well as the squared voltage magnitudes at each bus $(V_i)^2$. These relaxations and/or reformulations are often more realistic than the transportation model, but come at a cost of increased computational complexity. Thus, a new reformulation is desired that maintains the computational tractability of the KCL-compliant transportation model but also includes the KVL and element relation constraints from (1.1e).

Limitations of KCL-only linear models

Computational experience using KCL-only network construction algorithms indicates that they typically result in a network that is more radial than a realistic electrical grid (Section 4.4). Moreover, network element flows often substantially differ from the flows calculated using either DC or AC power flow.

With appropriate weights, the objective function employed in these KCL-compliant network construction algorithms may be interpreted as minimizing the total MW-miles of the network, i.e. the sum of the length of each line multiplied by the magnitude of flow on that line (Section 4.2). Although this results in a formulation that is computationally

tractable for creating large networks, the resulting network often displays a radial, tree-like structure. Also, since the network flow representation only models KCL and the line limits, there is no method of directly modeling the dependence of line flows on network parameters such as impedance. Additionally, there is no inclusion of the relationship between bus-based quantities (e.g. bus voltage angle) and branch-based quantities (e.g. branch power flow), which are governed by equations such as Ohm's Law or DC power flow. Instead, in the existing literature parameters are assigned to the edges created in the KCL-compliant based algorithms using heuristics that initially do not account for the resulting impact on line flow. As an example, [74] uses statistical data on line capacities from the existing power system to assign transmission line parameters. However, because such a method does not consider the characteristics of individual transmission lines, the method in this dissertation initially assigns voltage levels and transmission line parameters that best fit the line limits assigned during the KCL-compliant linear algorithm (Section 7.1.1).

The fundamental issue with all methods of assigning network branch parameters to the network graph edges is that the power flow resulting from full circuit analysis will significantly differ from the flow vector p resulting from the linear program. This is an inherent limitation of the network flow model, since the decision variables are the flows on the network branches (p) and the only constraints imposed are flow conservation at nodes (KCL) and line limits on existing lines. Thus, flows are free to be routed to avoid overloading any existing transmission lines, whereas the same lines modeled using DC power flow would result in significant overloads. Thus, a method needs to be imple-

mented to correct the overloaded and severely underutilized transmission lines. Several methods of selecting conductor parameters and capacity limits have been proposed, such as [75], in addition to algorithms for optimally upgrading existing transmission systems [76, 77]. However due to its tractability for large scale networks, a sequential DC power flow is utilized in this dissertation to incrementally update the transmission line parameters until the majority of the lines are no longer overloaded (Section 7.1.2).

While there are heuristic work-arounds to the shortcomings of KCL-only algorithms detailed above, they require additional steps in the overall network construction algorithms and have no guarantees of acceptable performance. These additional steps sometimes require human intervention, and still many not produce a network that satisfies all network constraints. Thus, it is desirable to develop a network model that overcomes the limitations of the KCL-only model while still remaining computationally tractable. Such a model that incorporates the KVL, KCL and network element relations (e.g. Ohm's law) is developed in Chapter 5.

2.2.2 Mixed integer linear programming

If a different objective function is desired instead of a minimal weighted L1 norm of the line flows, either integer decision variables must be utilized or heuristics must be employed for selecting the lines to be added to the existing transmission system. For example, [32] minimizes the cost of new elements added to the power system network, but includes integer decisions variables indicating whether an element should be added between bus i and bus j. This integer decision variable limits the number of paths

with non-zero power flows. However, since MIP problems are typically intractable for large networks, [32] proposes relaxing the integer decision variable to a continuous one and uses a sensitivity-based heuristic to solve the relaxed linear program. Another example is [33], which also minimizes the capital cost of new line additions, but adds DC power constraints with continuous variables. This incorporates the KCL, KVL and element constitutive relation constraints from (1.1e), i.e. includes constraints (1.3b-1.3e). However, given the large number of potential candidate expansion paths, [33] utilizes a iterative sensitivity-based heuristic for determining the paths to add to the existing network.

To increase the fidelity of the electrical performance of a network construction algorithm, ζ can be directly formulated as an integer decision variable. This allows for formulations to range from mixed-integer transportation model formulations [78] to DC power flow [79] or linearized AC power flow with either integer or binary decision variables for constructing new network elements [80]. Some examples of mixed-integer transmission expansion include [78], where the transportation model with integer decision variables is solved using a hierarchical branch and bound method. Once DC power flow or linearized AC power flow constraints are included in the optimization formulation, all of the constraints in (1.1e) have been modeled, i.e. KCL, KVL, and network element relations. However, as mentioned in earlier sections, mixed integer problems are difficult to solve, and typically intractable for large-scale networks. Thus, a network construction framework needs to be developed that has the scalability and tractability of the network flow model but the realism of the mixed-integer methods.

2.2.3 Mixed integer nonlinear programming

Finally, the most realistic and accurate network expansion algorithm is the mixed integer non-linear (MINLP) AC power flow formulation. MINLP formulations implement (1.1e)-(1.1h) from the ideal NCA representation in Section 1.3 or the full OPF formulation in (1.3), sometimes including the full AC power flow equations in (1.5). As mentioned in earlier sections, [28] presents the full, idealized AC optimal power flow (ACOPF) generation and transmission expansion planning (GTEP) formulation. Several relaxations and reformulations that make the problem tractable for systems of practical size are presented in [28]. Also, [21] presents the full AC transmission network expansion planning (AC TNEP) problem, accompanied by various reformulations and relaxations. The conclusion of papers such as these and [23] is that the AC OPF TNEP formulation is the most accurate, but is not suitable for systems larger than small-scale models.

However, some mixed integer non-linear formulations don't necessarily include the full AC power flow formulation. For example, [7] presents a multi-year planning horizon, multi-element objective function with both operational and investment costs. However, [7] differs from the proposed algorithms in this dissertation since it uses the DC power flow model and implements a genetic algorithm to solve the network expansion problem. Additionally [4] presents a hierarchical method utilized integer decision variables for transmission and generation expansion and investment, and the DC power flow to equations to model the electrical performance of the network. To ensure the

problem can be solved, the number of candidate transmission paths modeled by the variable x is significantly limited. As observed in Chapter 1, this decision is justified by the authors by noting that "physical and infrastructural limitations as well as political and geographical complexities" have the potential to substantially limit the number of candidate transmission lines considered [4].

Additionally, decomposition methods could be utilized that break the full optimization problem into smaller scale sub-problems that are solved separately. For example, [81] utilizes decomposition to solve the network construction problem as applied to transportation networks [82]. However, the difficulty in using decomposition methods is properly modeling the interconnection between the sub-problems, and the geographic "edge effect." Whenever a network is constructed, unless the entire geographic footprint of the interconnection is modeled, special care needs to be taken to properly model the effects of the network, generation and load in the adjacent geographic regions that aren't included in the model. Usually this is accomplished by including import or export buses to represent the adjacent regions which models the inter-area flows as determined by RTOs or ISOs.

If these import or export buses are not modeled, computational experience with NCAs suggest that buses at the geographic edge of a network can contain abnormalities such as a high number of transmission lines connected to generator or large load buses. One may hypothesize that this is because the power output from the generator is unable to be routed to the adjacent regions and instead must be transmitted to a smaller

geographic region. This can be observed in the results of some of the systems in Chapter 6. For the large networks instructed in this dissertation, the ratio of the geographic edges to the remaining interior of the geographic footprint is relatively small. However decomposition methods create multiple sub-problems with a much higher geographic edge to interior ratio, thus exacerbating the geographic edge problems.

2.3 Heuristic-Based Power System Network Creation Algorithms

2.3.1 Cluster and connect style algorithms

Several authors utilize a cluster and connect style algorithm where smaller networks are first created using one of several methods, and then connected together to form a large network. For example, [83] uses a multi-level clustering approach based on several graph properties of power system networks. Additionally, [84] utilizes a cluster-and-connect style algorithm that reconnects smaller fragments of existing power systems to create a larger synthetic power system network model. Finally, [6] utilizes a "cluster and connect" style algorithm which targets specific node degree and line length distributions. However, the overall problem with "cluster and connect" algorithms is that it is difficult to ensure realism of both the individual clusters and the overall system. Cluster and Connect algorithms also have the shortcomings of the general graph theory-based algorithms discussed above.

2.3.2 Anonymizing existing network models

Some of the earliest work synthetic power system models were created by using equivalent representations of sections of the existing network. The real geography and locations of the network were somewhat obfuscated since portions of the system were reduced to a single equivalent bus, generator, load or transmission line. These early synthetic test cases can be found at [10]. However, these earliest test cases are small, unrealistic and do not contain geographic coordinates for the buses.

Thus, additional synthetic test cases have been created since the original models were released in the 1970's. One method employed in [85, 86] involved anonymizing existing networks from RTE France. Additionally, the method utilized in [84] utilizes small portions of existing power system networks, but anonymizes them by connecting the fragments together in ways that do not reveal the actual topology of the real power system. Additional "cluster and connect"methods include generating the clusters synthetically [6, 83]. While anonymizing networks can create realistic synthetic power system networks, there is a trade-off between ensuring that no CEII (critical energy infrastructure information) is released in the networks, and ensuring the realism of the synthetic network.

2.4 Insights From Transmission Network Expansion Algorithms

Another subset of network construction algorithms are expansion algorithms that take a connected (or mostly connected) network graph and add additional elements to

the network. The additions are usually a small percentage of the total elements in the network. The primary challenge for network expansion algorithms is accurately simulating large scale networks that require nonlinear, non-convex equations.

One class of network expansion algorithms of interest in this dissertation is that of transmission expansion planning (TEP). This problem seeks to add additional element(s) to an existing power system transmission network, ideally at the lowest possible operating and capital costs while meeting constraints that ensure reliable electrical performance. Most transmission expansion algorithms rely on the pre-existence of a connected transmission system to compute the power flow on the existing elements [15]. This is one feature that precludes the majority of industry focused TEP algorithms from being directly used to create synthetic power system networks, since the synthetic power system creation algorithms must start with no existing transmission lines in the model. Or, if an existing network is not required, modeling the (ζ) decision variables as binary or integer makes using TEP algorithms intractable for creating medium to large scale "greenfield" power systems or synthetic networks.

Transmission expansion planning (TEP) approaches can be broken down into three main categories: industry design methodologies, mathematical optimization models and heuristic methods [20]. Several systematic reviews of previously proposed transmission expansion include [15, 87]. Additionally, meta-heuristic, global optimization methods and optimization solvers and tools are are summarized in [87].

A brief background and historical overview of utility industry transmission planning

is found in [88]. First, a reliability assessment is conducted to determine the areas of the system where the reliability criteria are in danger of being violated [89, 90]. Once the potentially difficult areas are identified, the regional transmission operator (RTO) or independent system operator (ISO) will follow their defined methodology for planning new transmission expansion projects [91, 92]. For example, for reliability transmission expansion projects in MISO, the transmission planners evaluate the solutions proposed by both the RTOs and MISO itself and pick the transmission lines with the best performance. For projects proposed for their economic benefit, the line or lines with the benefit-to-cost ratio are selected for construction [93]. The transmission lines are then designed using a set of criteria and guidelines, with an example design manual found in [94]. In industry practice, the voltage level of a new transmission line is determined a priori based on the voltage levels present at the target substations or the availability of space to expand a substation to include another voltage level. Finally, some examples of transmission expansion projects from the Upper Midwest region are given in [95] and from Europe in [96].

However, not much mention is given to the selection of the potential transmission expansion corridors to address the areas with potential violations of the reliability criteria. From the industry guides, it appears that the current method relies on an a priori removal of a large number of the potential transmission corridors from the transmission expansion study. This leaves a much smaller, and therefore more manageable, number of potential transmission line paths to study. Numerous simulations are run using a variety of scenarios to determine the best transmission line to build. It appears that

RTOs and ISOs choose potential transmission paths based more on geopolitical considerations than mathematical and operational constraints. Thus, it is beneficial to create algorithms that are able to model all potential transmission expansion paths without the a priori removal of many paths.

2.5 Comparison Metrics Drawn from Existing Power System Networks

In an ideal world, one would compare networks created utilizing full MINLP implementations of (1.1) with the networks generated by the algorithms reviewed above and the two classes of algorithms proposed in this dissertation. Thus, future work might include the creation of small to moderate scale networks generated using (1.2)-(1.4) from [4]. These "ideal" networks could then be compared against both the real-world networks and the synthetic networks created using the algorithms in this dissertation. One may hypothesize that the even these "ideal" networks will not exactly match characteristics of existing networks, since the existing real-world networks have evolved over the last century and include numerous generator retirements and additions, in addition to the changing boundaries of utilities service territories. These, and other geopolitical considerations, have resulted in network design choices that would be seen as "sub-optimal" when purely considering the cost functions presented in (1.1a) and (1.2a).

However the AC power flow equations are nonlinear and nonconvex, and currently there does not exist a solver capable of solving large-scale mixed integer nonlinear non-

convex problems on practical timescales. Thus, for the analysis in this dissertation, instead of comparing networks generated using the ideal NCA with synthetic networks, the synthetic networks are instead compared to existing power system networks in the United States, while respecting the CEII requirements associated with the real world cases. Chapter 3 describes the sources of network data for the two largest interconnections in the United States, the Eastern Interconnect and the Western Interconnect.

There are numerous sources of so-called "realism metrics" that seek to determine whether a synthetic power system is "realistic", i.e. accurately reflects one or more existing power system networks. These metrics fall into categories of graph theory [3, 53, 97], statistics of power system network parameters and devices [98], and a combination of the two [99, 100]. However, these comparisons have significant shortcomings as true statistical tests, since there are only three samples of real world networks in the United States, the Eastern and Western Interconnects and the ERCOT network. Whereas statistical tests require a large number of samples to create a reference distribution, and the statistical test typically estimates the likelihood of the given test subject (i.e. the synthetic network) being drawn from the distribution of real-world examples.

Thus, the following performance metrics are employed with the more modest of evaluating the "closeness" of a synthetic power system to the small number of real samples in existence. These metrics fall into two categories, graph statistics (Node degree distribution and cycle basis) and electrical performance metrics (DC and AC power flows and histograms of MVA limits). The following subsections describe how these metrics

were developed and how the benchmark statistics were computed utilizing the real world FERC Form 715 cases described in Section 3.2.2.

Additionally, future work includes a more thorough comparison of the networks generated in this dissertation with the other networks created as part of the ARPA-E GRID DATA program [101]. The metrics will include both those presented in this dissertation (node degree, cycle basis) and will most likely include those in [58, 98, 100].

2.5.1 General Graph Metrics

As further explained in Section, graphs can be classified by their 2.1.1 randomness, heterogeneity, and modularity, and these categories can be used to compare synthetic power system networks with existing networks. Additionally, traditional graph metrics such as average shortest path length and betweenness centrality can be used to compare real-world power system graphs to synthetic network models [58]. Finally, a comparison has been performed between real world networks and the synthetic Western system generated using the algorithms described in Chapters 4 and 7. The authors utilize three categories of metrics: purely topological measures, electrical component based measures, and metrics that quantify both the topology and electrical behavior of power networks.

Metrics in the first category include the node degree centrality, Eigenvector centrality, spectrum of the normalized Laplacian, and an analysis using the Fiedler eigenvector. A more general node degree analysis is performed in Section 2.5.2. The length of trans-

mission lines at each voltage level is examined as the second category of metrics, and these statistics are presented in Section 6.5.1. Finally, the hybrid metrics include the Topological and Impedance Element Ranking (TIER): TIER analysis (first presented in [99]), and the Singular Value Decomposition of the Power Flow Jacobian. Since these metrics have been calculated in [100] for one of the synthetic power system networks presented in this dissertation, the calculation and comparison of the metrics for the remaining networks is left for future work.

2.5.2 Node Degree Distributions

The first of the two graph metrics utilized to compare the synthetic systems generated in this dissertation with the real networks is the node degree distribution of the network. The degree of a vertex or node is computed by counting the number of branches that are connected to that node. The degree of each node is computed for a network, and then plotted on a histogram graphs using integer sized bins. For example, figure 2.1 shows the node degree graph for the Eastern Interconnect network from the form 715, where the horizontal axis is the node degree, and the vertical axis is the number of nodes in the network having this node degree. The node degree plot illustrated in 2.1 illustrates the challenge in the use of standard statistical tests for comparison of such data between networks. In viewing the histogram of this figure, it is tempting to describe it as a "distribution." However, to rigorously consider this histogram as a probability distribution, one should carefully consider the underlying probability model; i.e. what is the sample space, and what is the nature of the function(s) on its elements that then

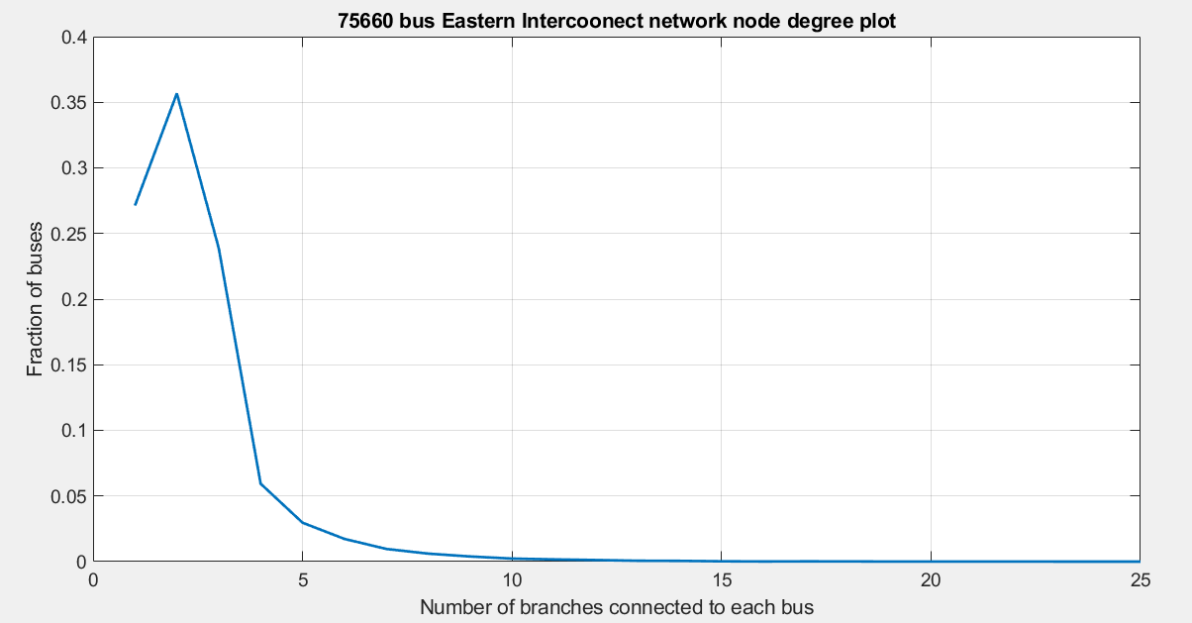


Figure 2.1: Node degree plot of 75660 bus Eastern Interconnect network

define random variables? In the context here, it is reasonable to assume that the sample space is composed of instances of power system network models. The random variable of node degree would then be a vector-valued function of a given network. Therefore, a histogram such as that in Figure 2.1 is **not** displaying characteristics of a large number of samples of some scalar random variable. Instead, it displays the behavior of the components of a **single sample** of a vector-valued random variable.

To compare networks of different sizes, figure 2.2 is normalized by the total number of nodes in the network to produce a node degree distribution as shown in figure 2.2. Notice that the vertical axis is a log scale plot of the decimal percent of the total number of buses at each node degree.

As mentioned earlier, true statistical tests cannot be performed to compare synthetic networks to the actual networks, because there are only three separate power networks

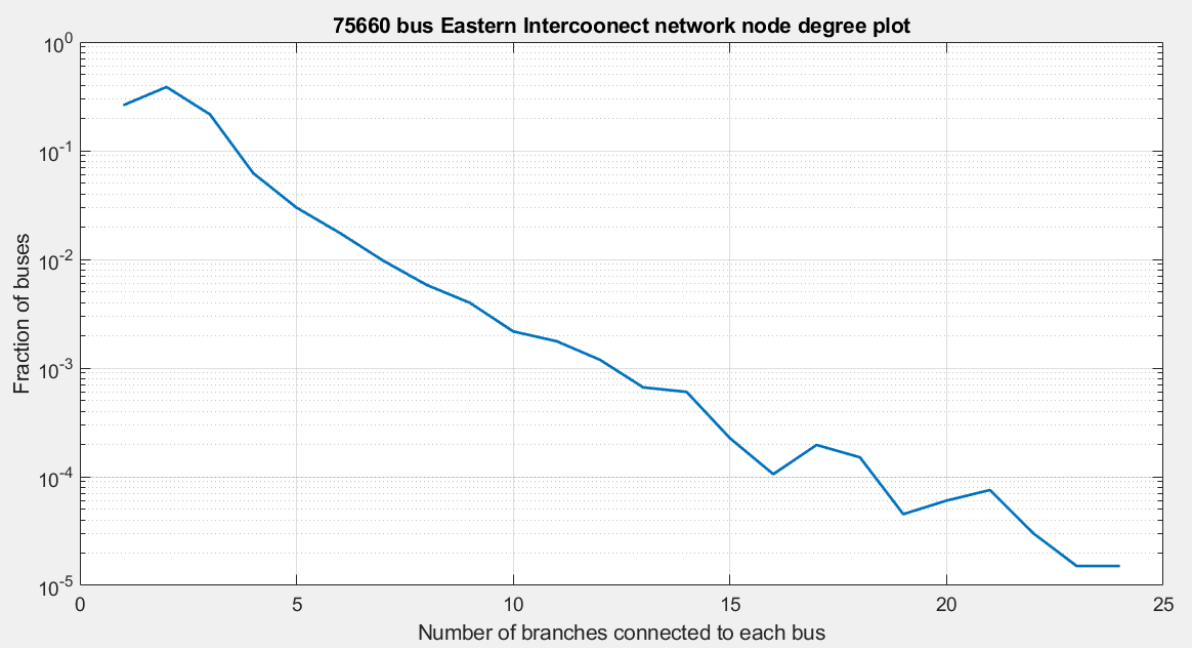
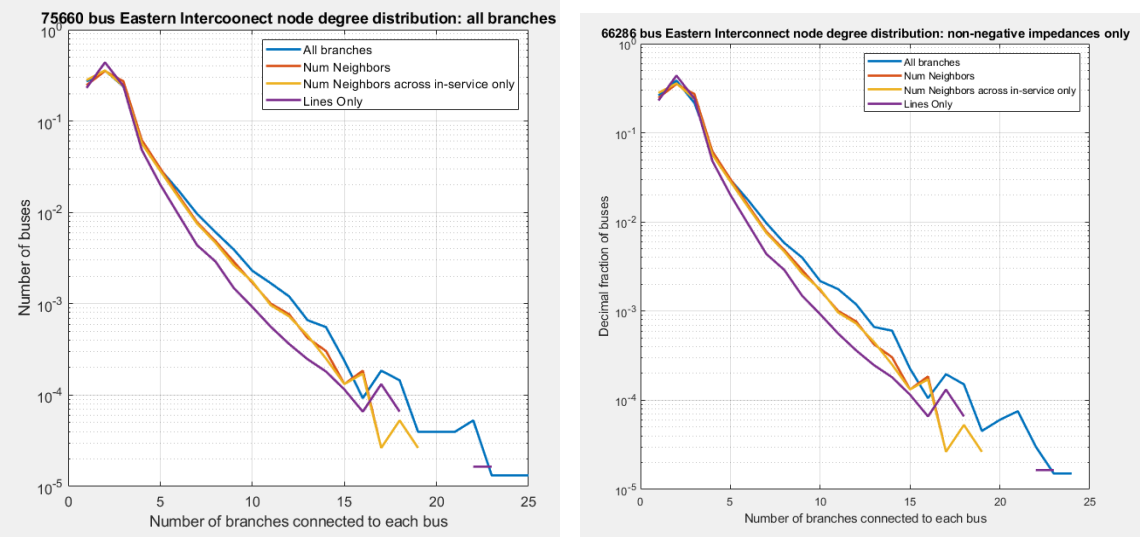


Figure 2.2: Node degree distribution of 75660 bus Eastern Interconnect network

in the United States. Instead, the general shapes of the node degree distributions of synthetic networks are compared to the general shapes of different interpretations of the node degree distributions from the Eastern and Western interconnects. Additionally, an affine function is fit to the log-linear plot of the node degree distribution and a percent error is computed between the affine fits for the synthetic network and the real network containing the same geographic footprint.

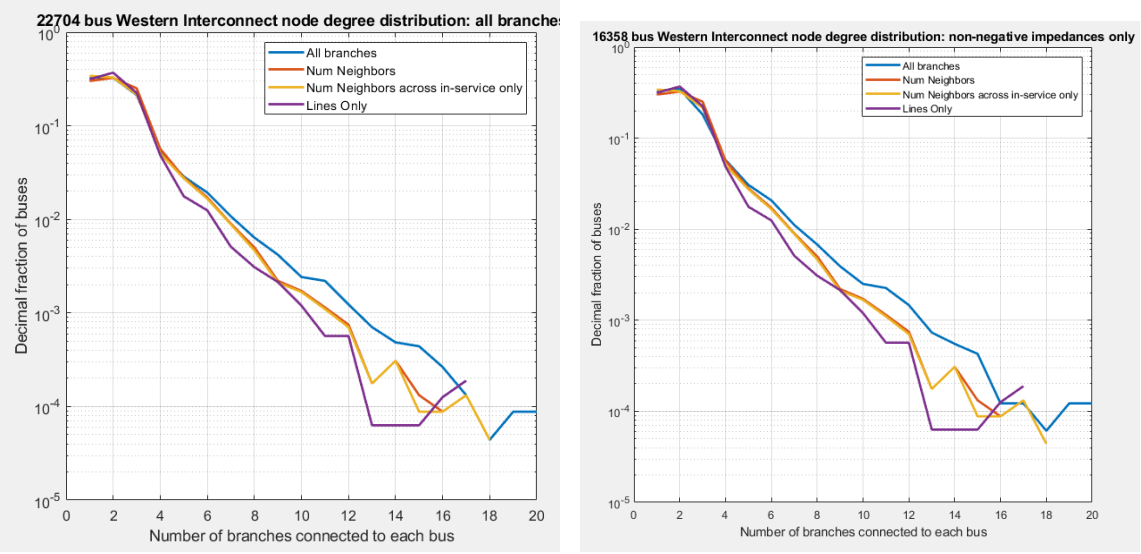
However, even this comparison metric is not without its drawbacks. As shown below in figures 2.3 and 2.4, there are many ways to interpret the node degree distribution of a real power system. All cases were created by taking the Eastern interconnect summer 2022 peak planning case from the FERC Form 715 report and cleaning up the data. For all cases, out of service buses were removed along with all out of service branches. Additionally, only the largest island in each case was kept, which involved removing all



(a) Including all branches (b) Including only positive reactances

Figure 2.3: Eastern Interconnect Node Degree Distributions

buses and branches in the smaller islands that were not connected to the largest island. The figures title "all branches" indicate that these models were created by keeping all network branches, even negative resistances and reactances that model equivalent areas and three-winding transformer windings. This resulted in network models that had more buses than the next set of network models labeled "nonzero impedances" in the graph legends. These networks were created by removing all branches with negative resistances or non-positive reactances, which resulted in some buses becoming disconnected or island in from the main network, thus causing the number of buses in the final case to be lower than the "all branches" cases. Additionally, the node degree distribution plots with "num neighbors" in the legend used the number of connected neighbors as the node degree metric. This considered only unique connections to neighboring buses and thus eliminated parallel branches, since the synthetic networks created in this dissertation have few to no transmission corridors with multiple lines. The plots titled "number



(a) Including all branches (b) Including only positive reactances

Figure 2.4: Western Interconnect Node Degree Distributions

neighbors across in-service only" only counted neighboring buses that were connected by a branch that is in-service.

Upon examining figures 2.3 and 2.4, there is a noticeable difference between the node degree distributions when comparing the networks created using the different methods described above. As expected, the networks containing all of the buses and all of the branches in the largest island had a higher fraction of buses with larger node degrees, and the networks examining only the number of lines connected to each node has the smallest fraction of buses with higher node degrees. It is also to be expected that the node degree distributions when examining the number of neighbors, i.e. only examining unique connections between buses, had a node degree distribution between the "all branches" and "lines only" cases.

Although they are subtle, there are differences between the node degree distributions

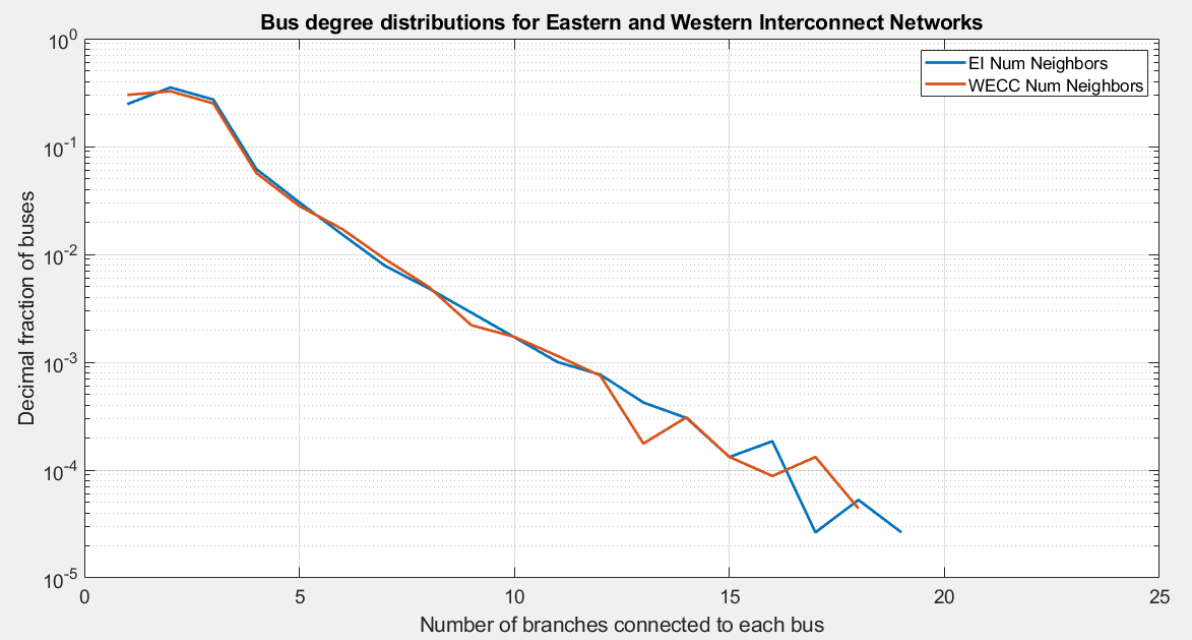


Figure 2.5: Node degree distributions for Eastern and Western Interconnects: All Branches

of the Eastern and Western interconnect cases, as seen in figure 2.5. Thus, when examining the node degree of synthetic networks, the general shape of the distribution is the primary feature to be analyzed, since slight deviations are expected. To quantify the shapes of the node degree distributions and to better compare real to synthetic networks, an affine function was fit to the log-linear node degree distribution plots, as shown in figure 2.6. These curve fits allow differences in area to be computed, thus establishing a "difference metric" for the comparison of real and synthetic networks.

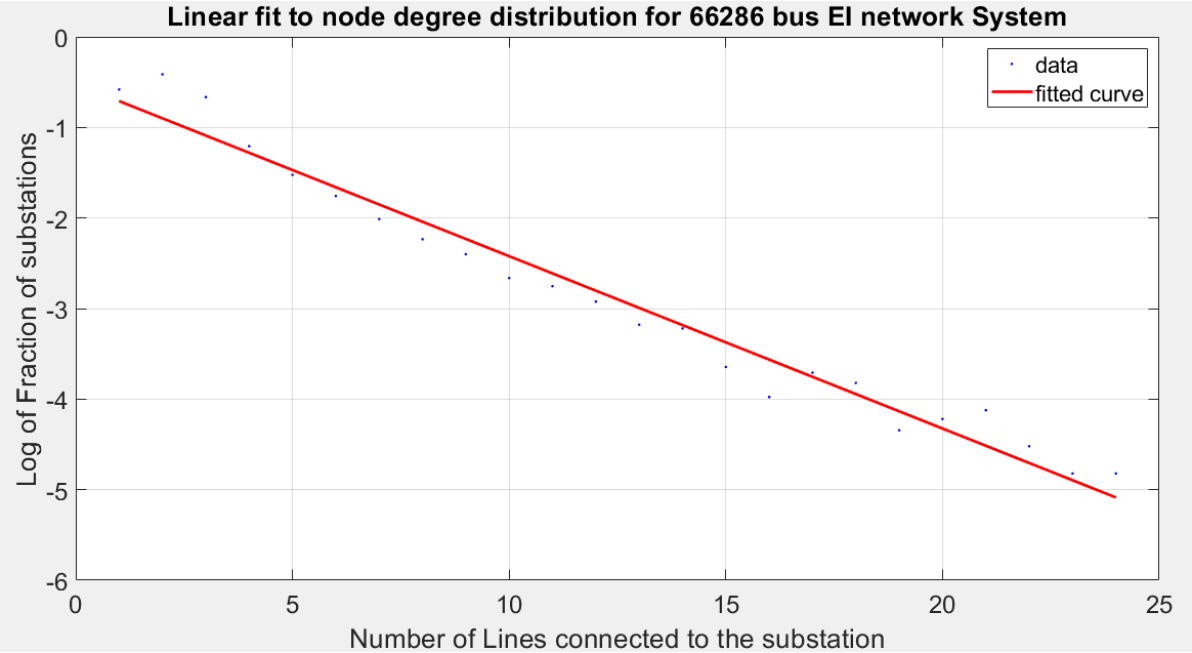


Figure 2.6: Affine function curve fit for Easter Interconnect node degree distribution

2.5.3 Graph Mesh Analysis

What is missing from the research literature on power system network graph analysis is a metric indicative of the degree of "meshing" in a graph. Such a metric or set of metrics could be useful for both comparing synthetic power system networks with existing real networks, as well as resilience or robustness analysis of existing networks.

A cursory examination of the existing literature shows that the better established analysis methods such as the meshedness coefficient or counting the number of faces on a graph are limited to planar graphs. An additional measure of the level of messiness in a graph can be determined by examining the fundamental loops in a graph. While the number of fundamental loops is a known quantity and equal to $m - n + 1$, there are multiple cycle bases that could be constructed from the loops in a graph.

Initial results and analysis have been performed using the cycle basis and are presented in Appendix D. However, given the limitations explained below, additional research is necessary to evaluate the usefulness of the cycle basis as a means to compare power system graphs.

The limitations of the analysis in Section D are multifaceted. First, is that the algorithm utilized by MATLAB [102] to compute the cycle basis makes a somewhat arbitrary selection among cycle bases. The algorithm *attempts* to minimize the sum of the lengths (number of branches per cycle) of the cycles in cycle basis, but even the example presented in the MATLAB documentation for the cyclebasis function [103] does not produce a minimum cycle basis. This issue can be remedied by implementing

a different algorithm that finds a true minimum cycle basis, such as [104–106].

Next, there is the challenge of comparing histograms of cycle basis lengths across graphs of different sizes. Graphs with the same number of nodes and branches can have different numbers of cycles in the cycle basis depending. Thus, the question arises of how to normalize the results of the cycle basis analysis. The most straightforward normalization is to divide by the number of cycles in the cycle basis, but additional analysis is needed to verify the validity of this approach.

Additionally, the comparison of the resulting histograms of the lengths of the cycle basis needs to be improved. Currently Section D.1 makes a simple closed-form functional fit (either exponential, power, or Weibull) to the histogram of the length of the cycle basis and then compares the area under these curve fits for the synthetic versus real networks. However better methods of comparing histograms will likely prove more effective, such as Quantile-Quantile plots or the Kantorovich–Rubinstein metric.

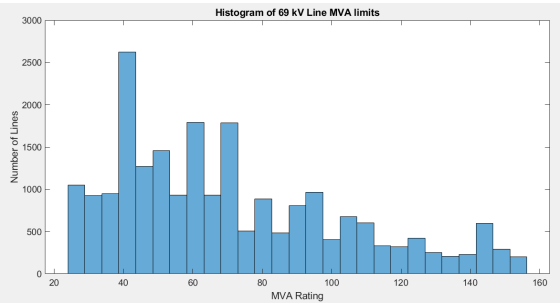
There are standard methods such as the Kolmogorov-Smirnov test and other analysis methods described in [107]. However, one must be careful when applying such measures such as these which conduct tests or analyses on probability distributions, as noted previously for node degree. In order to construct a probability density function (pdf), one needs to establish the event space for the probability model, which requires either a closed form expression for the distribution, or many samples to construct a distribution. Unfortunately, there are only three existing real-world power system networks, the Eastern, Western and ERCOT Interconnect models, which is an insufficient number of

samples to fully construct the event space of power system networks. Thus, one must exercise caution when treating histograms of quantities obtained from a single power system network as probability distribution functions.

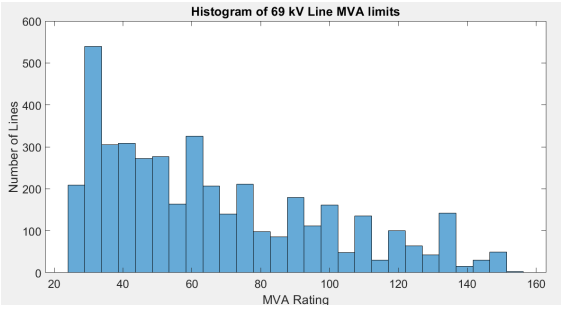
Finally, there is the issue with interpreting the results of the cycle basis analysis. Section D.1 presents histograms of the lengths of the cycles returned from the non-unique MATLAB algorithm, and compares the general shape of these histograms between the synthetic and real networks. But the impact of the cycle basis lengths on electrical performance characteristics such as network resilience to transmission line outages is not clear. Additionally, ongoing questions remain such as how much variation can there be a network construction and topology while still yielding the same minimal cycle basis. These and other questions need to be answered in future research before the cycle basis can be a consistent measure of network meshness.

2.5.4 Histograms of Line MVA Limits

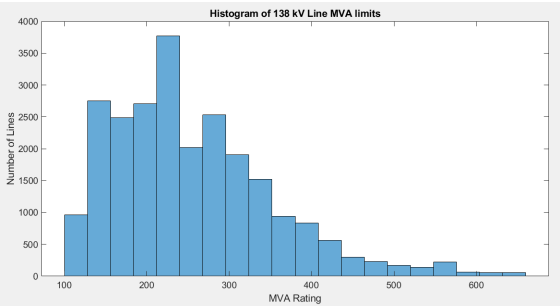
Another metric for comparing synthetic power system network models to their corresponding real networks on the same geographic footprint is to examine the MVA limits of the transmission lines in the networks. As later explained in Section 3.2.6, the FERC Form 715 rate A MVA limits and the inferred thermal MVA limits from the FERC Form 1 are utilized to create an upper and lower bound on transition line MVA limits allowed at each voltage level. However, there is a substantial variation of distributions of MVA limits that could be created within this range. Thus, one method of comparing systems is to examine the histograms of the MVA limits at each voltage level in the network.



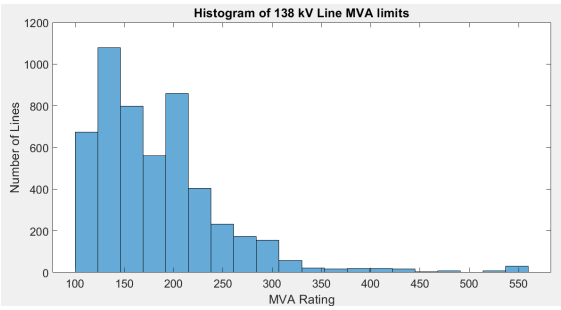
(a) EI 69 kV



(b) WECC 69 kV



(c) EI 138 kV



(d) WECC 138 kV

Figure 2.7: Eastern vs Western Interconnect Branch MVA Limit Histogram Plots Part 1

First, obvious outliers are discarded such as MVA limits of zero, or MVA limits outside of the 95th to the 99th percentile of the real-world data. These MVA limits do not correspond to physically realizable transmission lines, but instead are usually used to model equivalent areas or connections to adjacent networks.

While no synthetic system will result in a perfect match of its MVA limit histograms with the real world networks, this comparison is still useful to examine the general shape of the histograms, and use the discrepancies to inform modifications to the synthetic networks. Also notice that there are differences between the histograms of the Eastern and Western interconnect networks as shown below in Figures 2.7 and 2.8. In general

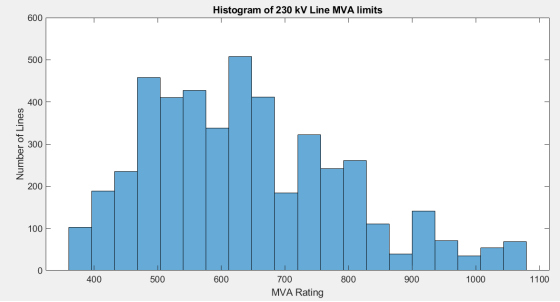
notice that the histograms are left heavy, with the number of lines at higher MVA limits dropping off. These characteristics are later used to compare real world MVA limit histograms against synthetic power system networks.

2.5.5 DC and AC Power Flow Statistics

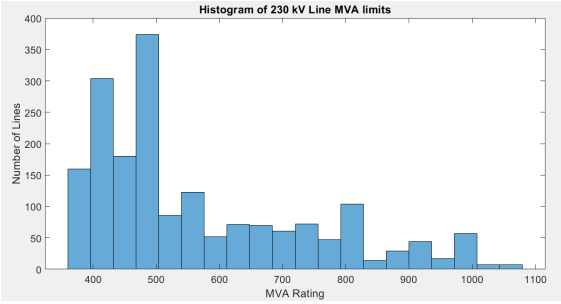
In addition to the graph metrics described in the preceding sections, different quantitative measures of electrical performance are also developed and utilized to both compare real-world systems to synthetic systems as well as comparing synthetic systems generated using different classes of algorithms. Since since in the theory real power system networks are operated to be N-1 secure, a comparison between real and synthetic systems using most of the metrics listed below is not meaningful. Thus, the primary metric used to compare electrical performance between real and synthetic systems is the percent losses in the network at various operating points. Real power systems usually have losses between 3-7%, so this is used as a basic benchmark to assess the "realism" of synthetic power system networks.

The following list of quantitative measures are used to compare the DC power flow results of synthetic power system networks generated using different sets of algorithms.

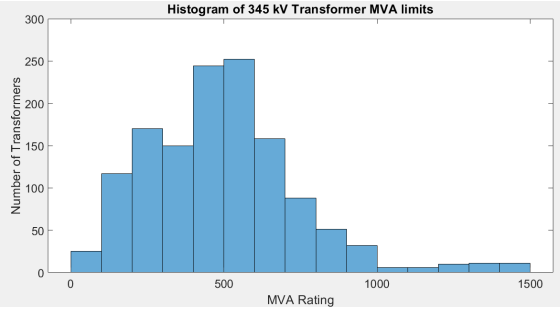
- Angle difference across branches: max, mean, standard deviation
- Losses: percent
- Percent MVA loading on branches: mean, standard deviation



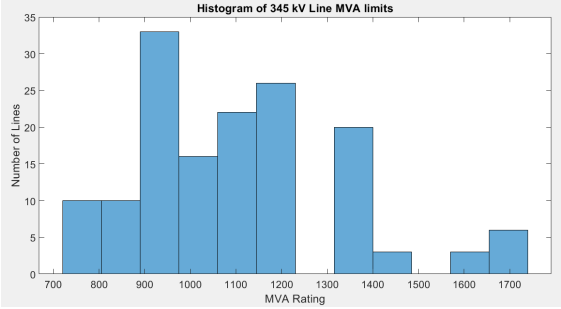
(a) EI 230 kV



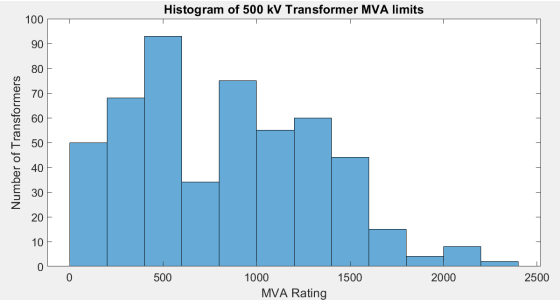
(b) WECC 230 kV



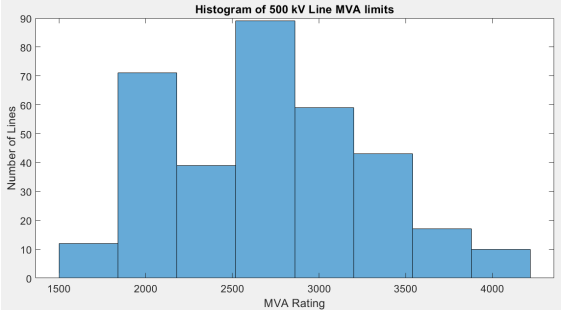
(c) EI 345 kV



(d) WECC 345 kV



(e) EI 500 kV



(f) WECC 500 kV

Figure 2.8: Eastern vs Western Interconnect Branch MVA Limit Histogram Plots Part 2

- Overloaded transmission branches: number of branches overloaded and % of total
- Overloaded branches as a % of MVA rating: max, mean and standard deviation

Additionally, a DC optimal power flow (OPF) is computed for all net injection scenarios for each synthetic network. If the DC OPF converges without modifications, it is noted in the metrics. If the cases infeasible, the branch MVA limits are relaxed to the maximum percent overload from the DC power flow, then successively lowered until the case no longer is feasible. The percentage increase in the relaxed MVA limits compared to the original base case is noted as a metric as well, in addition to the statistics listed above.

Chapter 3

Statistical Studies of Real-World Data Sets: Methods to Identify Realistic Parameters

3.1 Data Inputs Required by Algorithms

A network construction algorithm can be thought of as a procedure that takes the input data described in this chapter and produces a transmission network graph with the associated design parameters (b) as an output. However, the level of modeling detail in the inputs and outputs can vary tremendously. This detail can range from merely giving the topology of the network, i.e. an assignment of the ζ variable, to a detailed model that includes the construction types and GPS coordinates of each transmission tower along with the distributed parameter electrical representation of transmission lines. For example, the linear programming portion of the NCAs presented in Chapter 4 model the

transmission design parameters (b) as only the line limits, and thus the resulting networks require subsequent computations and modifications to extend the transmission system models to the necessary π-equivalent circuit model. The KCL/KVL NCAs proposed in Chapter 5 directly model the b parameters as the line impedance (X), requiring fewer steps after the termination of the inner loop of the algorithm to determine the remaining parameters of the transmission line model. However, per [28], even industry transmission expansion algorithms rarely model all of the engineering parameters that coule enter the b vector in full detail.

In Section 1.3, an optimization formulation of an ideal network construction algorithm was presented with inputs of real and reactive load power ($P_{D,s}$, $Q_{D,s}$), the number of buses in the network (n), and the lengths of each potential connection in the network (matrix D). As an intermediate step, the real power output of the generators (P_G) are computed using the economic dispatch (as described in Appendix B.1) that utilizes the peak of the real load power (P_D) and the generator cost curves. The outputs of all algorithms presented in this dissertation are the decision of which bus-to-bus paths are assigned transmission lines (ζ) and the engineering design parameters for each transmission line (b). These output values are briefly summarized in Section 3.3, and in more detail in the following chapters.

For the algorithms described in this dissertation, two classes of input data are required: given substation and candidate branch data. At a minimum, substation level data must include the set of substations to be connected, the lengths of each potential

transmission line path in the complete graph connecting all the substations, and a set of net injection scenarios. The minimum branch level data required is a set of per unit length impedance values with associated MVA limits that can be assigned to network branches.

To meet this requirement, the substation level input data supplied to the algorithms in this dissertation are: the geographic location of each substation and a set of generators and time varying load curves associated with the substations. Geographic coordinates are utilized to compute the distance of each path in the complete graph, and the generator cost curves and power limits are combined with the load curves to produce a set of net injection scenarios. The Substation level input data is described briefly in [108] and processes of creating the intermediate data are explained in more detail in subsequent chapters.

The requirement for the branch level input data is met by a "lookup" table of possible transmission line conductors that can be assigned to each branch, along with the associated per unit length impedance values and MVA limits. The process of creating this table is included in Section 3.2, with the table itself being presented in Section 3.2.9.

A set metrics were developed in order to quantify the similarity of the synthetic networks created in this dissertation to real-world power systems. Section 2.5 details how statistics and data separate from the input data are collected to determine the realism of the synthetic networks in this dissertation.

3.1.1 Input Data for Existing Network Branches

The first required input for the network construction algorithms presented in this dissertation is information regarding existing network elements, such as transmission lines, along with sets of parameters b that can be assigned to transmission lines added to the network work. Note that the list of existing transmission lines can be empty, which indicates that either a greenfield network or synthetic network will be constructed. Ideally, both the KCL-only and KCL/KVL-compliant algorithms might use information on the cost of construction for each potential new transmission line that could be added to the network. However, such computations can be quite complex since the cost is impacted by transmission tower construction configuration, tower and conductor materials used, the route of the transmission line, labor rates, etc. Thus, the bus-to-bus distance for each potential transmission path is used as a simple surrogate for the cost of new transmission, since the cost of new transmission lines is a function of the length of the line [12]. See Sections 4.2.2 and 5.2 for an explanation of this process.

In addition to the expansion costs, the KCL-only algorithms require each existing transmission line should include line flow limits in MVA or MW, and the associated cost with transporting power on each existing transmission line. Since detailed cost information is not easily quantifiable, the length of each existing transmission line is again utilized as a surrogate for the weighting factors that set the operational cost of delivering power along each line (see Section 4.2.2). Also, the existing transmission line power limit in either MVA or MW is required to avoid unrealistically large flows being

routed on existing transmission lines.

The KCL/KVL-compliant algorithms require input data for the existing transmission network elements. First, the impedance data (X) for the existing network elements is required, since the algorithms use X to compute the power flowing on existing and potential network elements. The power limits for existing transmission network elements that can be upgraded is also required, as this will allow the algorithms to determine if an element is overloaded and in need of an upgrade. An optional, but useful input is the list of possible transmission line conductors and tower configurations that can be used in a network design. This information allows for a more realistic updates to the X matrix during the middle and outer loop steps (Sections 5.3.2 - 5.3.3).

The proposed network construction algorithms also require a list of existing buses in the network, with associated power injections or withdrawals. To specifically create synthetic power system networks, the algorithms utilize a list of buses with GPS coordinates, and provided synthetic load data as described in [108]. The net power injection to each bus is computed utilizing the methods described in Sections 4.4.1 and 5.2.3. Finally, the target total lengths of the transmission lines and other target metrics, along with the synthetic transmission line parameters were determined using the approach in Section 3.2. Further discussion regarding validating the networks created using the proposed algorithms is found in Chapter 6.

3.2 Data Sources Describing Real-World Power System Networks

Real-world data from the North American Eastern Interconnect and Western Interconnect systems were utilized to create the branch-level input data as well as generate statistics to test and validate the realism of the synthetic power system networks created in this dissertation.

The branch level data consists of three categories of information: physical characteristics such as conductor size, phase-to-phase spacing, line lengths and voltage levels, modeling parameters such as transmission line impedance values, and quantities that are a combination of the two such as line MVA limits. The three primary sources of information utilized in this dissertation are the publicly available FERC Form 1 and DOE State risk profiles, and the confidential CEII FERC Form 715. As shown below in table 3.1, each of the three of input data sources supplies some, but not all, of the required input data. Additional input data and information was utilized to calculate realistic parameters for the transmission lines and transformers in the network, as shown in Table 3.1 and explained in more detail later in this section.

The remainder of this chapter is organized as follows. Sections 3.2.1-3.2.3 give background information about the three primary input data sources. Sections 3.2.4-3.2.7 explain how statistics were calculated for the physical data such as geographic region, line length and phase to phase spacing values. Section 3.2.8 Details how the physical data were used to calculate modeling parameters such as impedance values. Finally, Section 3.2.9 presents the methodology for assembling conductor lookup tables at each

Input Data		Sources		
Category	**Data**	**Form 1**	**Form 715**	**DOE**
Line Lengths	Maximum at each voltage level	Yes	Yes*^	No
	Totals at each voltage level	Yes	No	Yes**
Voltages	kV Values present in each region	Yes	Yes^	Yes**
MVA Values	Range for each voltage level	Yes*	Yes^	No
Transformer Impedances	Typical X/R ratios at each MVA class	No	Yes	No
	Typical %Z values for each primary-secondary voltage class	No	Yes	No

* These quantities were calculated or inferred from the original source
^The FERC Form 715 Does not have explicit geographic information included in the data. Thus, statistics were assembled on either a whole network level, or an ISO level for entity such as ISO-New England or New York ISO
** The DOE state risk profile only includes aggregate transmission line lengths in two categories, low voltage (< 230 kV) and high voltage (≥ 230 kV). However, this data was used to double check the total line lengths calculated from the Form 1.

Table 3.1: Input Data Sources for Network Level statistics

Data	**Source**
MVA Limits	[109]
Per unit length reactance values	[110]
Reactance Spacing Factors	[110]

Table 3.2: Input Data: Transmission Lines Impedance Parameters

voltage level for each synthetic geographic region.

3.2.1 FERC Form 1

The FERC Form 1 [111] contains information regarding the assets possessed by investor owned utilities and some rural electric co-ops or municipal utilities. The Form 1 contains transmission line design choices and some geographic information, but does not contain any topological information or impedance values. The geographic information included is the utility or entity that owns the power system equipment, and a separate spreadsheet that includes the location of the headquarters of the utility. This information is used to approximately determine the state where the transmission line exists. The Form 1 was created to ensure that investor-owned utilities were accurately reporting

ID	Utility	State	From	To	Voltage (kV)	Material Type	Tower Type	Length (mi)	Cond/ Phase	Size (kcmil)	Conductor Type
275	American Transmission Company LLC	WI	Glenview	Shoto	138	Wood	Single Pole	11.19	1	477	ACSR
275	American Transmission Company LLC	WI	Freeman	Presque Isle	138	Wood	Single Pole	7.77	1	605	ACSR
275	American Transmission Company LLC	WI	Canal	Rosiere	138	Wood	H-Frame	17.08	1	477	ACSR
275	American Transmission Company LLC	WI	Highway V	East Krok	138	Wood	H-Frame	20.92	1	477	ACSR
121	Northern States Power Company (Wiscor	WI	(W3102) ROCKY RU	EAU CLAIRE	345	Wood	H-Frame	78.43	1	795	ACSR
275	American Transmission Company LLC	WI	Pleasant Prairie SW	Arcadian	345	Steel	Tower	0.18	2	2156	ACSR
275	American Transmission Company LLC	WI	Pleasant Prairie SW	Zion	345	Steel	Single Pole	0.19	2	2156	ACSR
121	Northern States Power Company (Wiscor	WI	Summary of 69 kV	System	69	UG	UG	1071.35	1	810	XLPE
275	American Transmission Company LLC	WI	Oak Creek SW YD	Bluemound	230	Steel	Tower	0.16	1	1033.5	ACSR
275	American Transmission Company LLC	WI	Oak Creek SW YD	Bluemound	230	Steel	Tower	0.19	2	1033.5	ACSR
275	American Transmission Company LLC	WI	Oak Creek SW YD	Bluemound	230	Steel	Tower	0.29	2	1033.5	ACSR
275	American Transmission Company LLC	WI	Oak Creek SW YD	Bluemound	230	Steel	Tower	0.29	2	1033.5	ACSR
275	American Transmission Company LLC	WI	Oak Creek SW YD	Bluemound	230	Steel	Tower	0.35	2	1033.5	ACSR
275	American Transmission Company LLC	WI	Oak Creek SW YD	Bluemound	230	Steel	Tower	1.25	2	1033.5	ACSR
275	American Transmission Company LLC	WI	Oak Creek SW YD	Bluemound	230	Steel	Tower	1.28	2	1033.5	ACSR
275	American Transmission Company LLC	WI	Oak Creek SW YD	Bluemound	230	Steel	Tower	2.43	2	1033.5	ACSR
275	American Transmission Company LLC	WI	Oak Creek SW YD	Bluemound	230	Steel	Tower	6.2	2	1033.5	ACSR
275	American Transmission Company LLC	WI	Oak Creek SW YD	Bluemound	230	Steel	Tower	13.14	2	1033.5	ACSR
275	American Transmission Company LLC	WI	Oak Creek SW YD	Bluemound	230	Steel	Tower	14.57	2	1033.5	ACSR
275	American Transmission Company LLC	WI	Oak Creek SW YD	Elm Road 345	230	Steel	Single Pole	0.19	1	2156	ACSR
275	American Transmission Company LLC	WI	Oak Creek SW YD	Bluemound	230	Steel	Tower	0.49	2	2156	ACSR

Figure 3.1: Excerpt of the FERC Form 1 for utility companies in Wisconsin

their power system assets and investments. This information is useful for synthetic power system creation, because the Form 1 contains construction information about the majority of transmission lines in the existing network, including line lengths, voltages, number of conductors per phase and conductor size in kcmil, transmission structure construction material and type (e.g wood H-frame). The FERC Form 1 is able to be publicly available because it contains no actual transmission line lumped parameter model impedances or topological information about the power system. However, using the information available in the Form 1, approximate transmission line impedances can be calculated. Instead of directly back calculating possible impedance values for the transmission lines in the FERC Form 1, the construction information was utilized to create sets of possible line designs for different geographic sub regions across the United States.

For each of the synthetic regions listed in Section 3.2.4, the following data was gathered from the Form 1: individual line lengths, voltage levels, transmission structure construction type and material, number of conductors per phase, and size of conductors in kcmil. In the following sections this information was used to calculate or to infer the following quantities: transmission line MVA thermal limits phase-to-phase spacing of transmission lines, and approximate lumped parameter impedance values.

Data Processing for FERC Form 1

The challenge in utilizing the FERC Form 1 is that there is no standard method of representing the data requested from each of the utilities. Thus, a large amount of work was necessary to standardize the entries to allow statistics to be calculated from the date in the Form 1. The author gratefully acknowledges the contributions from undergraduate researchers Tyler Whitmore, Michael Zupan and Andrew Bryce in cleaning up the Form 1 so it could be utilized for the research in this dissertation. Fig. 3.1 Shows an excerpt from the Form 1 for several utility companies in Wisconsin.

3.2.2 FERC Form 715

Originally publicly available, the FERC Form 715 was classified as CEII (Critical Energy Infrastructure Information) in 2001 following the terrorist attacks on September 11. The Form 715 contains a large amount of data, including steady-state power flow cases of the Eastern interconnect system, and various one lines and node breaker schematics. Without substantial analysis and modification, little to no reliable geographic informa-

tion is included in the Form 715. Thus, the geographically based information in the Form 1 is combined with the transmission line statistics from the Form 715 to create inputs for creating synthetic power system networks.

For the FERC Form 715 Eastern and Western interconnect systems, only the largest connected island was kept for the analysis. Two versions of the system were created. One was utilized to compute statistics about the set of branches in the network, such as line lengths and per unit impedance values. The second was used to compute graph-based statistics. The first system used to compute branch statistics had all negative impedance and negative resistance branches removed, which corresponded to interconnections modeling equivalent areas, and the conversion of the 3-winding transformers to sets of 2-winding transformers. The second system used to compute graph theory metrics kept all branch elements, but removed all buses and connected elements that were not part of the largest connected island.

The following aggregate statistics were collected from the Form 715: distributions and curve fits of line MVA limits, impedance data to calculate line limits (Section 3.2.2), and typical X/R ratios and %Z values for transformers. The X/R ratios were assembled by discretizing the MVA limits into 50 MVA blocks and computing the average X/R ratio for each of these blocks. The percent the values were calculated by taking every combination of voltage levels in each of the in the synthetic geographic regions. For each of these primary-secondary voltage pairs, an average per unit impedance (%Z) of the transformers was computed.

Inferred Line Length Calculations

To calculate the line lengths for the transmission lines provided in the FERC Form 715, the transmission line wave propagation equations were utilized. Recall that electromagnetic waves propagate in any medium less than the speed of light.

Also recall that the per unit length capacitance and per unit length impedance are interrelated as seen in the following equations,

$$l = \frac{\mu_0}{2\pi} \ln(\frac{D}{R}) \tag{3.1a}$$

$$c = \frac{2\pi\epsilon_0}{\ln(\frac{D}{R})} \tag{3.1b}$$

Where l and c are the per unit length inductance and capacitance respectively, R is the effective conductor radius, and D is the conductor-to-conductor spacing. Note that in reality, $\mu_r \times \mu_0$ is utilized, but μ_r is close to 0. Thus, for a lossless transmission line,

$$\frac{1}{\sqrt{lc}} = \frac{1}{\sqrt{\mu_0\epsilon_0}} = c \tag{3.2}$$

Where c is the speed of light. Since,

$$X = 2\pi f \ell l \tag{3.3a}$$

$$B = 2\pi f \ell c \tag{3.3b}$$

then (3.2) can be arranged as follows,

$$\ell = \frac{c\sqrt{XB}}{2\pi f} \tag{3.4}$$

Where ℓ is the line length, f is the system frequency, X and B are the per unit reactance and charging susceptance, respectively. Thus, (3.4) can be utilized to infer the lengths of the transmission lines provided in the FERC Form 715, assuming the lines do not contain series compensation that would cause $X \neq 2\pi f l \ell$.

3.2.3 DOE State Risk Profiles

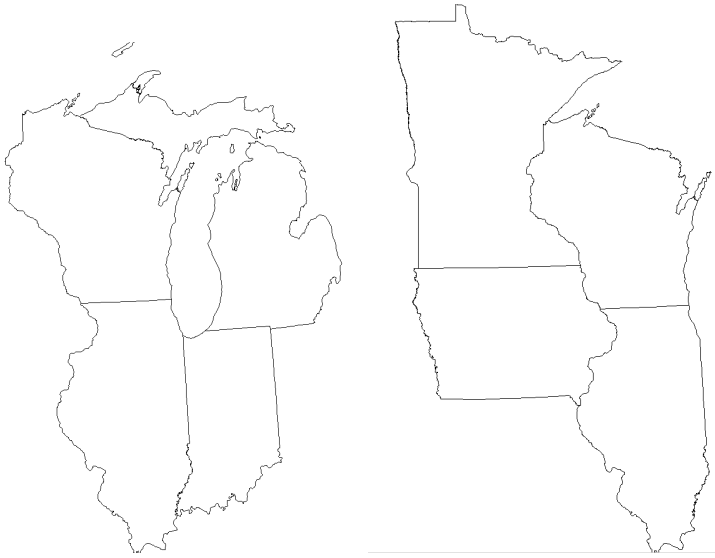
The Department of Energy has published a energy sector risk profile report for each of the 50 states [112]. While the primary purpose of this report is to determine the relative magnitude and probability of risks that a states energy infrastructure my encounter, this report also contains a high level map of the transmission grid in each state, as well as aggregate total statistics of transmission line mileages. As shown in table 3.1, the DOE data contains the least granular data, since it only contains aggregate statistics for high voltage (>= 230 kV) and low voltage (< 230 kV) transmission lines. However, since not all rural electric co-ops are required to submit the FERC Form 1, the DOE data can be utilized to fill missing line mileage totals in the Form 1 data.

3.2.4 Synthetic Regional Areas

Since creating a synthetic power system network that mimics the complete Eastern interconnect is computationally intractable without access to a large compute cluster, 13 geographic footprints were created to design regional networks much like the networks that the current ISOs, RTOs, balancing authorities, or municipal utilities operate. These regional areas were designed to balance the competing objectives of matching the existing EPA eGRID or NERC regions, while also maximizing the "convexity" of the region, as explained below.

Since the algorithms presented in this dissertation do not directly model the cost of routing or constructing a transmission line based on the topography of the land under the line, the algorithms are agnostic to the real geography of the land. Thus, if a region is constructed with a large body of water present in the region, then the algorithms will construct lines over this lake without realizing that transmission lines cannot be built across large bodies of water. Thus, a region such as the four upper Midwestern states shown below in Fig. 3.2a would not be chosen since many transmission lines would be routed across Lake Michigan. Instead, an upper Midwest region that did not include Lake Michigan was created as shown below in Fig. 3.2b.

Additionally, the regions were chosen such that a minimal number of transmission lines would be routed across adjacent states that were not part of the model. For example instead of creating a synthetic Southeastern region as shown below in Fig. 3.3a, instead the Southeastern synthetic area shown in Fig. 3.3b was used. In Fig 3.3a, numerous

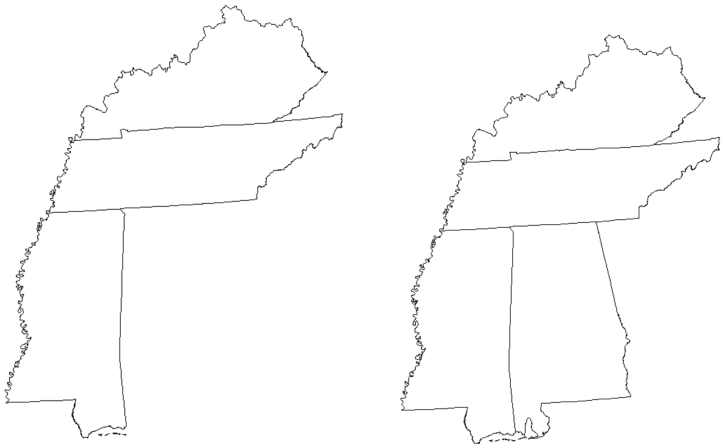


(a) Including Lake Michigan (b) Excluding Lake Michigan

Figure 3.2: Possible upper Midwest geographic footprints

transmission lines would've been routed from Mississippi to Tennessee and cross the area where Alabama would have been if it were included. This would create a higher number of long transmission lines then would be otherwise realistic, since there would be no intermediate substations in Alabama for these lines to connect to. Note that Georgia could have been added to this synthetic Southeast region, although this would have caused Florida to be its own region, and would've created a fairly small synthetic case.

Fig. 3.4 shows a list of the 13 synthetic regions, and the states contained in each region. In addition to creating more "convex" areas, and avoiding encompassing large bodies of water, the regions were created such that the total number of synthetic substations ranged between 3000 and 9000. Three of the regions (NEISO, NYISO, CAISO) are almost exactly matching the service territory of ISOs, while the synthetic MISO region



(a) Without Alabama (b) With Alabama

Figure 3.3: Possible southeast geographic footprints

roughly approximates the upper portion of the MISO system. The synthetic SPP region includes service territory of SPP and ERCOT combined. While some might protest not modeling ERCOT as its own independent interconnection, one could think of this synthetic region as a hypothetical situation where ERCOT annexed SPP. Finally, the remaining eight regions are based off of the EPA eGRID subregions, since the subregions were roughly within the target range of the number of buses desired, and this creates a connection to a real-world entity and regional divisions.

Additionally, the FERC Form 1 and DOE state risk profiles were utilized to determine voltage levels for each region that best approximated the most commonly utilize voltage levels by the utilities in that region. While each state often had some transmission lines at either five or six of the voltage class is utilized in this dissertation, the most common 2 to 3 high voltage levels (230 kV+) were chosen for each of the 13 synthetic regions. Some of the choices were fairly straightforward, for example all of the subregions in WECC utilize 230, 345 and 500 kV, while there was much more overlap in the Eastern interconnect

Region	Substations	States
AZNM	4700	Arizona, Colorado, Nevada, New Mexico, Utah
CAISO	8400	California
FRCC	7500	Florida, Georgia, South Carolina
MISO	7000	Illinois, Iowa, Minnesota, Wisconsin
NEISO	3500	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
NWPP	3200	Idaho, Montana, Oregon, Washington, Wyoming
NYISO	5000	New York
RFCE	5400	New Jersey, Pennsylvania
RFCW	7400	Indiana, Michigan, Ohio
SPP	8500	Kansas, Nebraska, North Dakota, Oklahoma, South Dakota, Texas
SRMV	3300	Arkansas, Louisiana, Missouri
SRTV	4600	Alabama, Kentucky, Mississippi, Tennessee
SRVC	6500	DC, Delaware, Maryland, North Carolina, Virginia, West Virginia

Figure 3.4: List of the states contained in the 13 geographic footprints

regions. Regions such as NEISO (synthetic ISO-NE) for more straightforward sense to 30 and 345 KV were the predominant voltage levels in that region. However, some regions such as NYISO (synthetic New York ISO) were slightly more complicated. Only looking at the Form 1 data for NYISO would indicate that only 230 and 345 KV were utilized. However examining the transmission maps in the DOE state risk profile from New York revealed that utilities also utilize 765 kV.

In real-world power system design, utility engineers make the choice of which of the highest 3 voltage levels (345, 500 or 765 kV) to utilize is mostly based on previous designs and prior experience. For example, utilities such as AEP have utilized 765 kV, while utilities across the Southeast prefer 500 kV and utility companies in ERCOT only build 345 KV as their highest voltage level.

3.2.5 Transmission Line Length Statistics for Input Data

As seen in table 3.1, differing granularity of information is available in the FERC Form 1, Form 715, and the DOE state risk profiles. Since the purpose of the Form 1 is to report total length of transmission lines owned by utilities, the total quantities reported for each they are fairly reliable, as long as the majority of utilities or co-ops in that state are required fill out the Form 1. However, as seen in Fig. 3.1, the individual lengths of the lines in the Form 1 are only somewhat reliable, with some utilities reporting segments of lines that go from once a substation to the other, as seen at the end of the table. Thus, the Form 1 was utilized to compute approximate total transmission line lengths at each voltage level for each of the 13 synthetic regional areas described in Section 3.2.4. Additionally, the Form 1 was also utilized to determine maximum realistic transmission line lengths at each voltage level. Since the Form 715 includes all buses, including intermediate switching stations with no reactive power compensation, the maximum line length in the Form 1 is a more accurate representation of the total transmission line lengths between one full substation and another.

The Form 1 already had line lengths for each individual transmission line or line segment, but the data was not well organized, so further data analysis was performed on the Form 1, as explained in later sections in this chapter. Fig. 3.5 shows the results of one of the analyses performed on the FERC Form 1 data, and Table 3.3 gives an example of the total transmission mileage in two states (Pennsylvania and New Jersey). Note that the total mileage at 69kV aren't accurate, since FERC only requires reporting

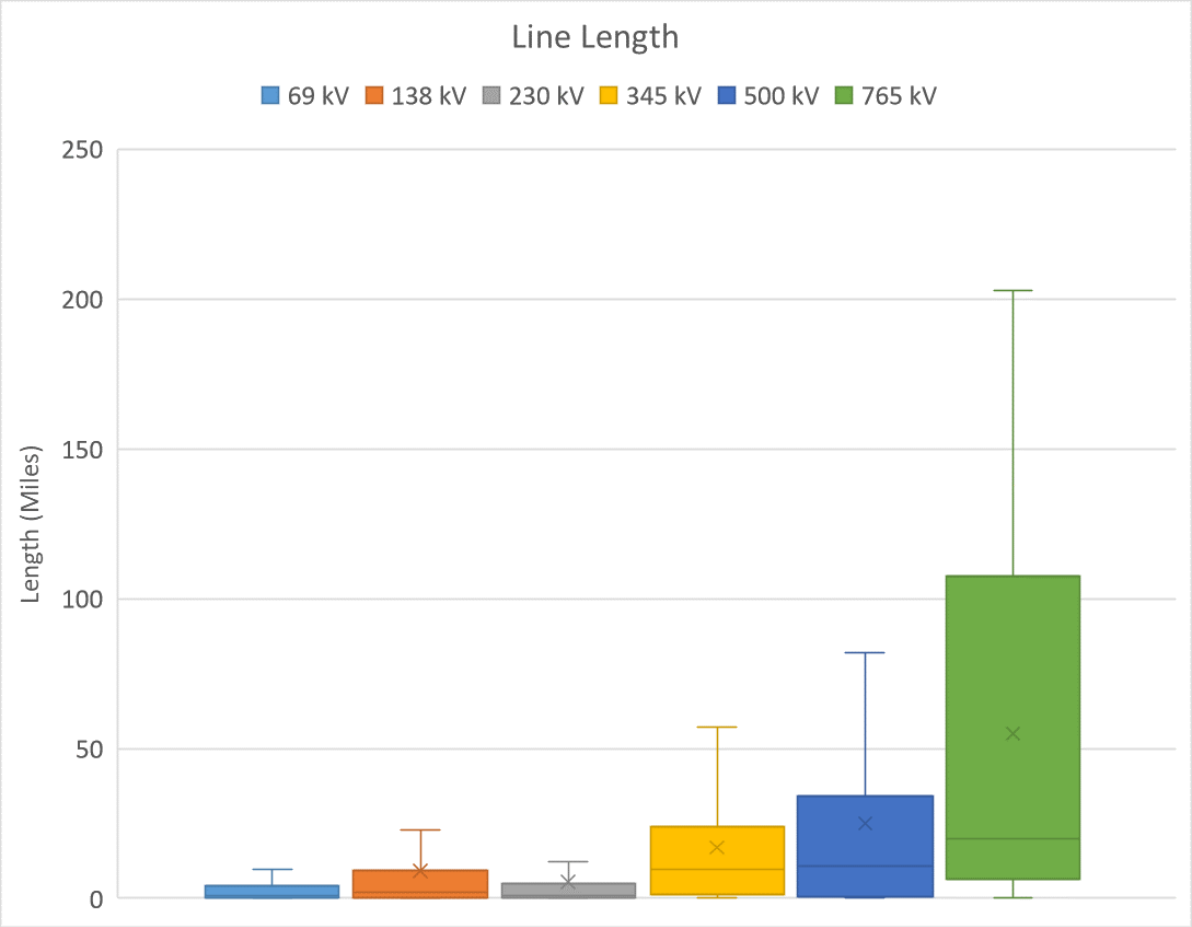


Figure 3.5: Transmission line lengths by voltage level for a subset of PJM (Pennsylvania and New Jersey)

of equipment installed on the bulk electric system, which has been traditionally defined as 100kV and higher voltages. This is seen in Table 3.1, which only lists total mileage values for 69 kV and below.

The Form 715 transmission line parameters are accurate enough to perform a steady state AC power flow, thus the inferred transmission line lengths as calculated in Section 3.2.2 should be fairly reliable. However as mentioned in Section 3.2.2, the Form 715 does not contain geographic information that is correlated with the power flow cases, thus the Form 715 was limited in the information that could be utilized. The maximum

State	Voltage Class (kV)	Length (miles)
NJ	69	212
	138	2789
	230	1211
	345	17
	500	1094
PA	69	489
	138	14590
	230	4550
	345	2033
	500	2147

Table 3.3: Total transmission line miles by voltage class for Pennsylvania and New Jersey

Voltage (kV)	1%	10%	50%	90%	95%	97%	98%	99%	99.5%	99.9%	Max
69	0.4	1.4	4.2	11.3	14.3	16.8	18.7	22.3	26.3	45.1	224.7
138	1.0	1.7	5.2	18.1	24.2	29.4	33.6	41.3	51.1	71.9	170.6
230	1.1	2.3	8.8	36.0	50.0	61.3	74.4	92.6	109.9	148.4	161.6
345	1.5	4.7	21.7	68.2	86.5	101.9	117.5	146.7	210.4	274.4	302.0
500	1.5	7.9	32.5	93.9	113.2	121.8	132.8	178.7	183.3	192.3	194.6
765	2.1	15.5	82.8	160.7	167.4	172.1	178.8	243.8	252.9	252.9	258.4

Table 3.4: Quantiles of the Form 715 transmission line lengths (in miles) per voltage level

transmission line lengths and ranges of MVA limits at each voltage level were utilized as input data. As explained later in Section 3.2.6, Distributions of the MVA values at each voltage level were utilized as comparison metrics for the synthetic cases.

The DOE state risk profiles were utilized as a double check of the total line lengths at each voltage level in each synthetic region from the Form 1. Since some states or geographic footprints have a high number of rural electric co-ops that do not have to submit the Form 1 report, the total length of transmission lines for that state can be

Voltage (kV)	1%	10%	50%	90%	95%	97%	98%	99%	99.9%	99.9%
69	1.0	1.5	5.4	28.9	57.5	125.5	210.2	294.5	432.9	45.1
138	1.0	1.6	6.3	31.3	51.0	80.3	110.5	133.4	409.6	71.9
230	1.1	1.8	8.0	38.2	57.2	77.6	93.9	133.3	303.6	148.4
345	1.1	2.8	15.9	59.8	88.9	114.3	126.9	176.6	337.8	274.4
500	1.3	3.3	32.2	114.4	133.4	179.2	224.5	289.1	289.1	192.3
765	1.6	4.6	34.1	135.6	164.8	190.6	190.6	190.6	190.6	252.9

Table 3.5: Quantiles of the Form 1 transmission line lengths (in miles) per voltage level

under-counted. In those cases, the Department of Energy data is utilized to fill in the missing totals from the Form 1, and the Form 1 is utilized when the DOE data has unrealistically low total length.

When designing transmission lines, power system engineers often do not create transmission lines beyond certain lengths for each voltage class for a variety of practical reasons. First, due to system stability limits, the maximum angle difference across a transmission line is typically constrained to be less than an upper bound of 30° to 45°. In a DC power flow model, the power flowing across a transmission line is proportional to the angle difference divided by the series impedance of the line, so the maximum amount of power flowing on a line is constrained by the angle stability limit. The transmission line power limit is selected as the lesser of the line's thermal limit and the stability limit. Thus, as the length of the line increases, the impedance increases which reduces the maximum flow across the transmission line allowed by the stability limit. Thus, each voltage has a length beyond which it is not practical to build transmission lines.

To capture these phenomena in the networks to be constructed in this dissertation,

Voltage (kV)	69	138	230	345	500	765
Maximum Length (mi)	25	65	90	130	150	180

Table 3.6: Maximum transmission line length by voltage level

data from the Form 1 and 715 were utilized to determine maximum allowable line length at each voltage level. Table 3.4 and 3.5 statistical analysis of the length values individual line lengths from the Form 1 and the inferred line lengths from the form 715. The maximum length values in table 3.6 were chosen as the maximum allowable length with a synthetic systems in the distribution system. The maximum lengths were mostly based on a range between the 98th and 99th percentile in the inferred lengths from the FERC Form 715 data. However, the lower voltage lines (69 and 138 kV) were allowed a maximum length that was closer to the 90th to the 97th percentile in the Form 1 data. This is because the Form 1 data sometimes excludes intermediate switching stations that do not have reactive power compensation, and thus the effective length of the line is longer than inferred from the form 715 data.

3.2.6 Inferred Transmission Line MVA Limit Statistics for Input Data

After synthetic regional areas and associated voltage levels were created, and the transmission line length statistics were generated for the input data, a range of allowable MVA limits needed to be decided for each of the voltage levels. This process was more straightforward than determining statistics about the line lengths, since the MVA values were directly given in the FERC's form 715. However as discussed in Section 3.2.2, reliable geographic information is not readily available in the form 715, thus thermal

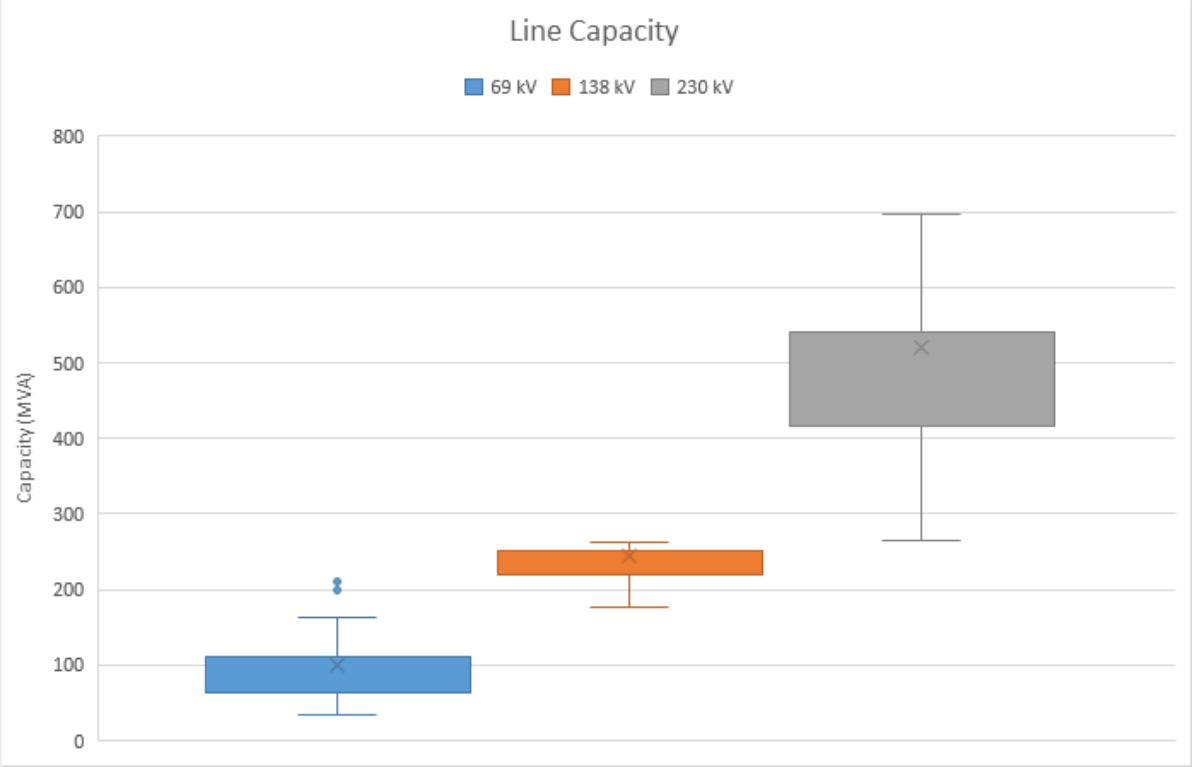


Figure 3.6: Range of inferred MVA Limits by voltage level for subset of PJM (Pennsylvania and New Jersey), lower voltages

MVA limit values were inferred from data provided in the Form 1.

As seen in Fig. 3.1, Each of the transmission lines in included in the Form 1 includes a number of conductors per phase, a size in kcmil and a conductor type (ACSR, ACSS, etc). This information can be utilized to determine approximate thermal limits in MVA using transmission conductor manufacture data sheets, such as [109]. Once this MVA information is added to the Form 1 data, plots of the ranges of the MVA values can be determined from this data, as shown below in Fig. 3.6 and 3.7. As explained in Section 3.2.9 below, these ranges were utilized to assemble the first set of conductor lookup tables.

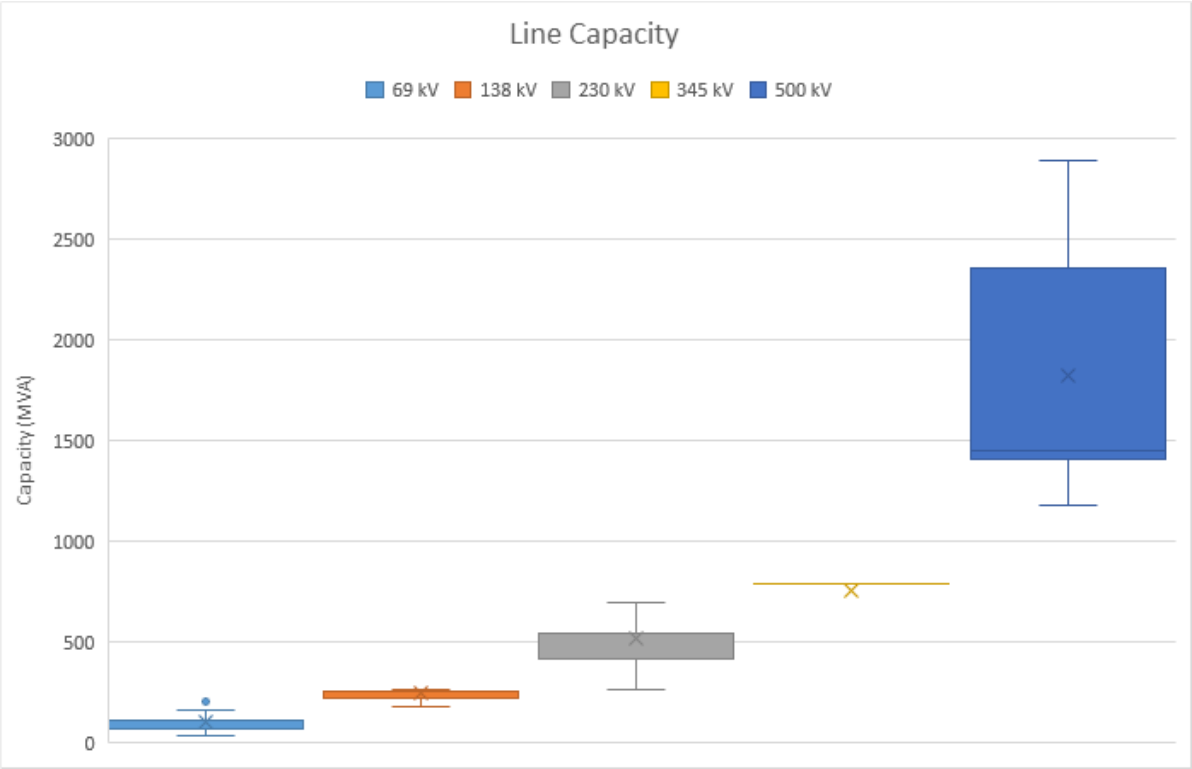


Figure 3.7: Range of inferred MVA Limits by voltage level for subset of PJM (Pennsylvania and New Jersey), full range

Voltage (kV)	1%	5%	10%	50%	90%	95%	99%	99.50%
69	14	20.84	28	62	126	149	233	256
138	67.5	101	120	225	375	424	569.1	632.2
230	239	320	398.3	627	940	1125.3	1370	1406.978
345	442.36	717	872.8	1381.5	1922.06	2080	2546.33	2749.47
500	1200	1732	1732.1	2750	3984	4257.26	5195.19	5196
765	1550.59	3975	4433.2	6955	9871.7	9871.7	9871.7	9871.7

Table 3.7: MVA limit statistics for the Form 715 Data

To create the second set of conductor Lookup tables as described in Section 3.2.9, statistical analysis was performed on the MVA limits in the form 715. Table 3.7 shows the quantile values for the limits included in the form 715. Values in the 5th to the 10th percentile were used as the lower bound on the MVA limit range, while values in the 90th the 95th percentile were utilized as the upper bound on the MVA limit range.

However, for 138 kV the upper range of the MVA limits were extended by utilizing the 90th the 95th percentile in the inferred Form 1 MVA limits shown in table 3.8. Since parallel paths in the same transmission corridor were not modeled in the KCL-only multilevel algorithm, the 138 KV transmission line MVA limits were expanded to include two conductors per phase, effectively modeling a transmission tower with two circuits. This allowed for the correction of otherwise overloaded transmission lines once voltage levels had been established, as discussed later in Section 7.2.2.

3.2.7 Phase-To-Phase Spacing Values

The goal of creating lookup tables of conductor parameters is to calculate impedance parameters for synthetically created transmission lines. As explained further in Section

Voltage (kV)	1%	5%	10%	50%	90%	95%	99%	99.50%
69	22	33	38	64	142	162	194	202
138	49	128	142	219	439	459	708	776
230	265	318	321	473	834	1024	1086	1293
345	594	918	918	1093	1623	1623	1670	1687
500	795	863	879	1406	2056	2811	2938.44	3180
765	1216	1216	1320	1632	5279	6530	6530	6530

Table 3.8: MVA limit statistics for the Form 1 data

3.2.8, both the geometric mean radius (GMR) of a conductor, and the geometric mean distance (GMD) between the phases of a transmission line are required to calculate the impedance of that line [24, 110]. The GMR values were directly supplied by the conductor manufacturers such as [109], and can be utilized to directly determine the corresponding X_a values as explained in Section 3.2.8. However, determining realistic ranges of GMD values is a bit more involved.

First, the Form 1 data was cleaned and simplified as follows: the multiple transmission structure construction types were reduced to the following five options: Single Pole, 2 Pole, 3 Pole, Lattice Tower, H-Frame. The construction materials were also reduced to two options: would and steel. Transmission structures with a construction type or material different than the owes listed earlier were converted to the closest approximation of the actual structure.

For these five construction types and two materials, ranges of GMD values were computed using typical schematics of transmission structures found in [110, 113–115]. These ranges were subsampled to three points: compact average and wide. These values

were either taken directly from the computed ranges or by extrapolating to create smaller or larger GMD values, or interpolating between voltage levels for missing data. This data is shown below in Table 3.9.

Using this table of possible GMD values across all voltage levels and possible transmission structure constructions, potential GMD values were inferred for all transmission lines in the FERC Form 1. This was accomplished by matching the tower type, material, and voltage level given in the Form 1 to values in Table 3.9. Next, a distance weighted distribution of the GMD values was determined for each of the voltage levels for each of the 13 synthetic regions.

After creating a range of GMD values for each transmission structure and material at each voltage level, a distribution of GMD values was created at each voltage level using the data in the FERC Form 1, as shown in table 3.10. First, a list of all inferred GMD values (in feet) was assembled for each voltage level. For example, column one of table 3.10 lists all of the GMD values of the 138 kV transmission lines in Florida, Georgia and South Carolina. The total lengths in miles of all transmission lines having this inferred GMD are calculated and reported in column two. Using the first two columns, a percentage of total mileage at the given voltage level is computed and listed in column three. Finally, column four lists the cumulative percentage of the given row and all preceding rows, similar to a cumulative distribution function (CDF).

To assign a GMD value to a synthetic transmission line, a uniformly distributed random number ranging from 0 to 1 is utilized. The numerical value in column 4 that

			GMD (Feet)			
voltage	**Material**	**Construction**	**Compact**	**Average**	**Wide**	**Source**
69	Steel	Single Pole	7	9	12	5
69	Steel	Tower	15	15	15	1
69	Wood	H-Frame	8	11	14	1,5
69	Wood	Single Pole	4	8	14	1,5
138	Steel	Single Pole	10	13	15	5
138	Steel	Tower	16	18	26	1
138	Wood	3 Pole		23		4
138	Wood	H-Frame	18	20	22	1,4
138	Wood	Single Pole	8	13	15	4,5
230	Steel	Single Pole	18	23	28	Extrapolation
230	Steel	Tower	27	31	35	4
230	Steel	3 Pole		34		4
230	Wood	H-Frame	21	25	29	4
230	Wood	Single Pole	14	18	22	Extrapolation
345	Steel	Single Pole	29	32	34	2
345	Steel	Tower	30	37	42	2
345	Wood	H-Frame	33	38	44	2
345	Wood	Single Pole	20	24	28	2
500	Steel	Single Pole	31	35	39	2
500	Steel	Tower	32	40	48	2
500	Steel	H-Frame	36	42	44	2
765	Steel	2 Pole	47	50	53	2
765	Steel	Tower	50	53	56	2
765	Steel	H-Frame	60	63	66	2

Table 3.9: Transmission line phase to phase spacing values (GMD) by voltage level and transmission structure construction

GMD (Feet)	Total Miles	Percent	Cumulative Percent
8	36	0.5%	0.0051
10	358	5.1%	0.0559
13	547	7.8%	0.1335
15	1498	21.3%	0.3460
16	1349	19.1%	0.5375
18	976	13.9%	0.6760
20	727	10.3%	0.7792
21	1	0.0%	0.7793
22	1425	20.2%	0.9816
26	130	1.8%	1.0000

Table 3.10: Total mileage weighted GMD percentage values at 138kV for synthetic FRCC region

is closest to, but less than the random number is selected as the GMD for the synthetic transmission line. For example, if the GMD random variable (also termed the "GMD index") is assigned a value of 0.1, then the GMD of that transmission line would be 10 feet. These phase to phase spacing values are then used below in Section 3.2.8 to compute the transmission line impedance values.

3.2.8 Calculating Transmission Line Impedance Values

Recalling the per unit length inductance and capacitance equations from Section 3.2.2, the following modifications were made for a 3-phase transmission line with stranded conductors:

$$l = \frac{\mu_0}{2\pi} \ln(\frac{GMD}{GMR}) \tag{3.5a}$$

$$c = \frac{2\pi\epsilon_0}{\ln(\frac{GMD}{GMR})} \tag{3.5b}$$

To aid in computing per unit length impedance values (inductive and capacitive), the approach in [110] was utilized that split each equation in (3.5) into two terms: one corresponding to the geometric mean radius (GMR) of the conductors (X_a or X'_a), and one corresponding to the geometric mean distance (GMD) between the conductor bundles in each phase (X_d). These equations are shown below:

$$X_L = X_a + X_d \tag{3.6a}$$

$$X_C = X'_a + X_d \tag{3.6b}$$

Where X_L and X_C are the the inductive and capacitive reactances per unit length, and X_a, X'_a, and X_d are defined in Appendix 3.2 of [110]. Since each of the transmission lines in the FERC Form 1 contains conductor size in kcmil, conductor type (ACSS, ACSR, etc) and voltage level, reference guides such as [110] can be utilized to look up the X_a and X'_a values of each transmission line in the Form 1. Next, the inferred GMD values (described above in Section 3.2.7) are utilized to determine the corresponding X_d value for each transmission line in the Form 1. Finally these values are utilized to

calculate possible transmission line impedances for every transmission line in the FERC Form 1.

Since the Form 1 is publicly available and contains no identifiable geographic or topological information, these potential impedance values from the Form 1 are purely synthetic and do not correspond directly to any real impedances in the real world transmission systems. However, these impedance values were inferred using engineering first principles, and are thus realistic potential impedance values that could be assigned to the real lines in the Form 1. Thus, the process of calculating these impedance values can be utilized to determine impedance values for synthetic power system networks, or to calculate impedance values for real world cases that have existing topology and geographic information but no line impedance parameters.

3.2.9 Assembling the Conductor Data Look-Up Tables

As explained in Section 3.2.4, 13 sub-regions were created based on the NERC regional entities, Independent system operator service territories, and the EPA eGRID subregions. Using the information for each of the synthetic regions as described in the preceding sections, two sets of data tables were assembled for potential transmission line conductors that could be installed in each state or regional area. There are two different sets of data tables created, one utilized for determining the voltage levels present at each substation (Sections 7.1.2 and 5.3.3) and one for resizing conductors within a voltage level (Sections 7.1.3 and 5.3.2).

When analyzing the FERC Form 1 database, it was observed that utility companies did not utilize all of the transmission conductor sizes available in [109] at each voltage level. Instead, utility companies tended to select at most two high voltage levels above 300 kV, as discussed in Section 3.2.4. Also, utility companies tended to increase the minimum size conductor utilized as the voltage level increased. To capture these phenomena, the first set of data tables created using the FERC Form 1 database recreates the range of the conductor sizes employed by utility companies within the geographic footprint of the network being constructed. The smallest and largest conductor used by the utility companies set the lower and upper bounds for the data table, and common conductor types are used to create a discretized table of conductors spanning the line limit range. The number of conductors in this range is often larger than the number of unique conductors in the Form 1 database. This is to prevent large gaps in the MVA limits between conductor sizes and allows for a more gradual and smooth transition between entries in the table.

The second table has a separate sub-Section for each voltage level in the transmission system region, and was created using statistics and data ranges from the FERC Form 715 [9]. This data table has a wider range of conductor ratings (including multi-conductor bundles) to accommodate upgrades to a transmission line within a voltage class. This data table is also useful for transmission expansion planning research, since it gives a wider range of potential realistic transmission line conductors. See table 3.11 below for an excerpt from the 138kV Section of the second transmission line conductor data table.

The R_a, X_a and X'_a values were taken from the 345 KV transmission line reference book [110]. The range of number of conductors bundled per phase increased as the voltage level increased, beginning with one conductor per phase for 69 kV and increasing to 2-4 conductors per phase at 500-765 kV.

As explained further in Sections 4.3 and 5.3.2, these conductor lookup tables were used to compute the per unit impedance values for each transmission line in the network. This is accomplished in two steps: first finding the conductor in the table at the specified voltage level that is closest to the desired MVA limit (Column 1, table 3.11). Then, the last three columns in Table 3.11 are utilized to compute the per unit impedance values for each branch, as described in Sections 3.2.8, 7.1.1, and 5.2.1.

In the early stages of the network construction algorithms in this dissertation, a substation can contain buses of any voltage level, i.e. transmission lines can be assigned any voltage subject to the length constraints discussed in Section 3.2.5. Thus, the first conductor lookup table is utilized for this portion of the network construction algorithms, to avoid assigning too many transmission lines to a lower voltage that would otherwise be assigned a higher voltage based on the MVA rating of the transmission line. Once the voltages of the transmission lines were determined, the second conductor Lookup table is utilized to allow resizing of transmission lines within its specified voltage level. The larger range of MVA values in each voltage level is utilized to model double circuited lines at lower voltage levels, and to allow transmission lines to be resized to correct overloads.

Capacity (MVA)	Voltage (kV)	conductor size (kcmil)	conductors per phase	Ra	Xa	Xa'
205	138	715	1	0.128	0.399	0.092
219	138	795	1	0.119	0.403	0.0917
230	138	900	1	0.106	0.399	0.0907
238	138	954	1	0.099	0.395	0.0897

Table 3.11: Excerpt from the 138 kV transmission conductor data table

3.3 Outputs for each algorithm class

As explained later in Chapter 4, the output of the multilevel KCL-only algorithm is a topology (an assignment to the Zeta variable), and associated transmission line limits, which is the only data included in the b set. Subsequent additions to the KCL-only NCA create transmission line parameters, buses, and transformers, and modify the network to satisfy the majority of line limit constraints when running a DC power flow, as explained in Chapter 7. While this approach allows for the tractable creation of large-scale realistic synthetic power systems, it requires an algorithm with multiple steps in loops, and relies substantially on heuristics.

In contrast, the KCL/KVL-compliant algorithm as described in Chapter 5 includes a model of DC power flow, thus the output of the multilevel algorithm is a network topology (and assignment to the Zeta variable), and a b set that includes real world conductor assignments to the transmission lines. These real-world conductors include parameters such as voltage level, thermal limits in MVA, impedance values (R,X,B), and design choices such as conductor size, number of conductors per phase, and phase to phase spacing. The benefit of this KCL/KVL algorithm is that by directly modeling DC

power flow, it substantially reduces the complexity of the full network creation algorithm and significantly reduces the reliance on heuristics.

Chapter 4

KCL-Compliant Network Construction Algorithms via Linear Programming

The first step to creating a tractable implementation of the ideal network construction formulations in Section 1.3 is to model the flow conservation constraint in (1.1e) or (1.3b). This constraint, Kirchoff's Current Law (KCL), requires that the sum of current or power entering a node (bus) must equal zero and is the first step in representing constraints associated with electrical performance of the network. The proposed network construction algorithms (NCAs) utilize the economic dispatch or variants thereof as a decoupled determination of the production cost term of the objective function (1.1a). Thus, the remaining term in the objective function only depends on the new transmission lines being constructed (ζ) and the engineering design parameters of the new line(s) (b), i.e. the variable (x) in (1.2)-(1.3).

To establish additional terminology utilized in this and future chapters and associated sections, the following nomenclature is presented and adds to the nomenclature ending on page xiii.

Chapter 4 Nomenclature

Subscript $_e$	Will denote a preexisting line in a network, or a line added during the network creation process, e.g. d_e is the node-to node distance of an existing transmission line
Subscript $_p$	Will denote potential transmission line paths of candidate lines in the network creation process
Subscript $_{p\|\|e}$	Will denote the potential transmission path that is in parallel to an existing transmission line

4.1 Problem Formulation and Multi-Level Optimization Framework

The set of buses with associated data is passed to the multi-level algorithm that determines the topology of a power system network graph by making an assignments to the binary decision variable ζ. The inner loop of the algorithm solves a linear program that determines the power flows p on the existing lines and potential transmission line paths. This information is passed to an middle loop that determines which new transmission line(s) to add to the network or existing lines to upgrade and calculates their line capacity limits. Once the middle loop determines that no lines should be added or upgraded, control is passed to the outer loop, which updates the net injection vector by considering different generator dispatch scenarios and N-1 contingencies. The full multi-level algorithm terminates when the target branch to bus ratio is met, or when all of the generator and N-1 scenarios have been run. This process is shown in the block

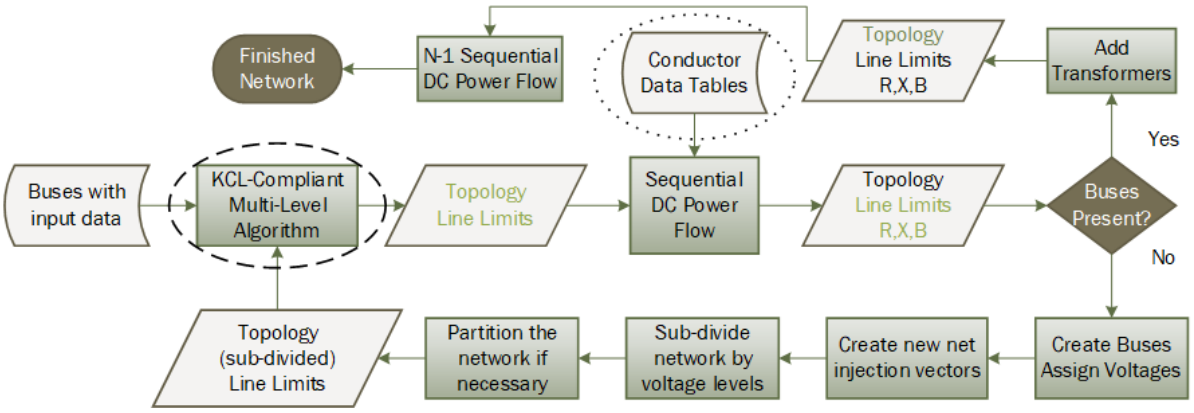


Figure 4.1: Flowchart of the full KCL-compliant NCA

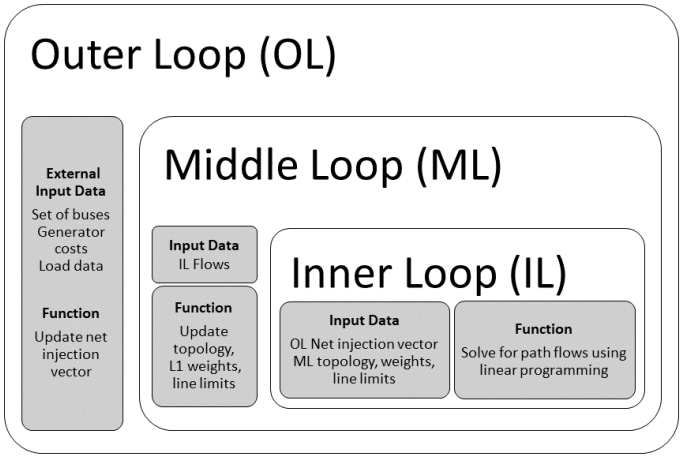


Figure 4.2: Block diagram of the multi-level optimization based algorithm

diagram in Fig. 4.2

A flowchart showing the full KCL-compliant NCA is shown in Fig 4.1. This chapter presents the multi-level optimization algorithm contained in the dashed line, which is used to create (or modify) the network topology. Chapter 7 will present the remaining portions of the full NCA, which improve the topologies presented in this chapter by modifying line parameters and voltage levels. The "Conductor Data Tables" input with a dotted ellipse is discussed in Chapter 3.

4.2 Inner Loop: Linear Programming Optimization Formulation

The core of the inner loop is the optimization problem in (4.1), which uses the input data of nodal power injections, P (Section 4.4.1), objective function coefficients c (Section 4.2.2, and incidence matrix A_c(Section 4.2.1). The inner linear program is solved sequentially and the resulting line flows p of each solution are passed to the middle loop (Section 4.4) after each successful solution of the linear program. Note that constraint (4.1c) only applies to existing transmission lines. Also note that for the implementation of the KCL-only algorithms in this dissertation, P is created using a peak load scenario and generator economic dispatch, which ensures that $\sum_{i=1}^{n} P_i = 0$, i.e. (4.1d) is satisfied.

$$\min_{p} \quad |c^T p| \tag{4.1a}$$

$$\text{s.t.} \quad Ap = P, \tag{4.1b}$$

$$|p_j| \leq \text{line thermal limit } \forall j, \tag{4.1c}$$

$$\sum_{i=1}^{n} P_i = 0 \tag{4.1d}$$

For the algorithms presented in this chapter, both the capital and operational costs from Section 1.3 are very roughly represented by the weighted L1 norm the flow on the potential and existing elements, respectively. The portion of the cost vector c corresponding to existing lines (weighing factor of 1) serves as a very rough surrogate for

the portion of the operational cost due to losses. In addition, the c vector for potential transmission line paths (weighted by a factor of 5 relative to existing lines) is a surrogate for the capital cost of building new transmission lines.

The work for this chapter implemented the standard form optimization problem using MATLAB [116] and solved the problem using CPLEX [117] through its MATLAB connector. The solution produces a vector p of flows on the existing transmission lines and the potential transmission line paths in the modified complete graph. Since minimizing the L1 norm of an objective function tends to produce a sparse solution [118], the majority of the line flows on the potential transmission line paths in the modified complete graph are to numerical accuracy, zero [13].

4.2.1 Incidence Matrix Creation

To allow separate representations of flows on known, existing transmission lines verses flows on potential transmission line paths, incidence matrices associated with two graphs were created. One incidence matrix represents the sub-graph of existing transmission lines in the network (A_e) and the other incidence matrix represents the complete graph of potential transmission line paths connecting any two buses in the network [13] (A_p). All existing transmission lines in the network have corresponding $\zeta_{i,j}$ variables set to 1. If A_e is empty, this indicates that a greenfield or synthetic network will be created. The two incidence matrices are then combined to form a single composite incidence matrix A_c. Note that if the incidence matrix A_e is non-empty, the resulting matrix A_c will have some identical duplicate columns, one column representing the existing transmission line,

and a duplicate entry in A_p representing the potential transmission line path in parallel to the existing path. While this choice introduces duplicate columns in the combined A_c matrix column rank deficient, it is necessary to allow the linear program to enforce the power limits on the existing lines, while allowing the flow on the potential transmission line paths to be unconstrained.

$$\begin{bmatrix} A_{\mathrm{c}} \end{bmatrix} = \begin{bmatrix} A_{\mathrm{e}} & A_{\mathrm{p}} \end{bmatrix} \tag{4.2}$$

Where A_e is the graph incidence matrix for existing transmission lines, A_p is the graph incidence matrix for potential transmission paths, A_c is the complete incidence matrix, and

$$A_{i,j} = \begin{cases} 1 & \text{when branch } j \text{ originates at bus } i \\ -1 & \text{when branch } j \text{ terminates at bus } i \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

A final incidence matrix, A (4.2.1), was constructed by removing unrealistically long potential transmission line paths from A_c, which results in setting the corresponding $\zeta_{i,j}$ variables to 0. Statistics on transmission line lengths were collected from the FERC Form 1 database [111] and inferred from the FERC Form 715 database [9]. These statistics followed a roughly power law distribution, so the maximum lengths were capped for each voltage level to exclude the outliers in the tail of the distribution. It was also observed from the FERC databases that two generation plants were typically not directly

connected by transmission lines longer than approximately 100 miles. This is most likely due to the fact that the angle difference between the generators can create abnormally high power flow on transmission lines with large impedances due to long lengths. Thus, the bus-to-bus distance matrix D informed removal of potential transmission line paths with lengths beyond these lengths. This resulted in a maximum potential transmission line path distance of 250 miles for extra high voltage lines and a maximum generator-to-generator potential transmission line path length of 100 miles.

4.2.2 Objective Function Weighting Factors

In order to determine the least expensive transmission expansion plan, the capital cost of building new transmission lines should be reflected in the objective function of the optimization problem. Since the cost of adding a new transmission line is approximately proportional to the length of the path, the KCL-compliant NCA utilize a weighted upper-triangular bus-to-bus distance matrix D as an estimate for the cost of a line's construction, following the suggestion in [13]. The matrix D is also represented as a vector of bus-to-bus distances d. In the initial implementation, the cost of transmission line construction per mile is assumed uniform for all voltage levels, but this can be refined as desired.

In many scenarios, the bus-to-bus distance matrix will be computed from the given GPS coordinates for each bus using the Haversine formula [119]. If GPS coordinates are not available an approximate bus-to-bus distance matrix can to be directly supplied as input data. The bus-to-bus distance matrix is reorganized as a column vector and used

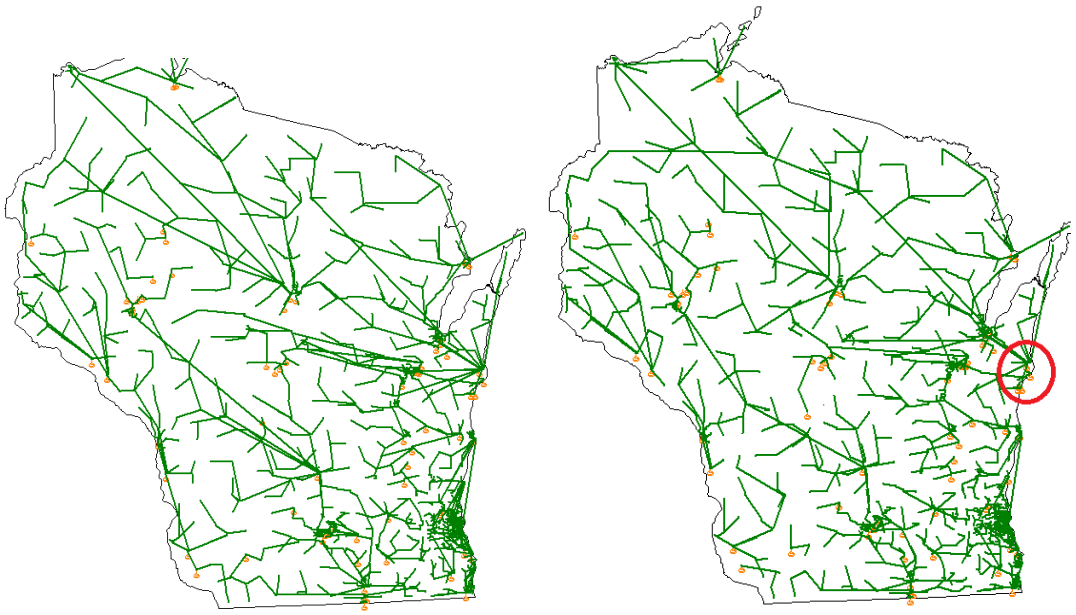
to calculate the scaling factor vector c in the objective function. However, if no network geographic information or locations are available, the cost vector c needs to be directly supplied to the algorithm in Section 4.1.

For network construction that proceeds from a given base network, existing transmission lines are given a cost of 1 per unit length ($1d$) [13]. If each transmission line is assumed to have the same per unit resistance and the short line model of a transmission line is used, the losses are proportional to square of the flow times the length of the line. Thus minimizing the weighted L1 norm of flows on existing lines is a rough surrogate for minimizing the losses in the existing transmission system, which serves an surrogate of the operational cost of the existing transmission system. Potential transmission line paths that are in parallel to existing transmission lines are assigned a cost of $3d$, which models the cost of upgrading the conductor on an existing transmission line. potential transmission line paths with no parallel existing line are given a cost of $5d$ [13], which approximates the cost per mile of constructing a new transmission line. While this approximation replicates the cost dependency of transmission lines on the length of the line, it is not a true model of transmission line cost which does not depend on the amount of power flowing on the line compared to the formulation in (4.1).

$$\begin{bmatrix} c \end{bmatrix} = \begin{bmatrix} d_{\mathrm{e}} \\ 3d_{\mathrm{p}\|\mathrm{e}} \\ 5d_{\mathrm{p}} \end{bmatrix} \tag{4.4}$$

Where d_e are the node-to-node distances of existing transmission lines, $d_{p||e}$ are the node-to-node distances of potential transmission line paths that are in parallel with existing transmission lines, and d_p are the node-to-node distances of potential transmission line paths without an existing line in parallel.

An intermediate cost of $3d$ was assigned to potential transmission line paths in parallel to existing transmission lines, where d is the length of the line. This value corresponds to the cost associated with replacing circuit conductor(s) or upgrading the voltage on an existing transmission line. After testing values ranging from 2 to $5d$ using 2-3 test cases, $3d$ was empirically determined as the best value to achieve the target node degree distribution, and matches both the upper range of [120] and the lower range of [121]. A one-line diagram illustrating an example network resulting from the use of this intermediate cost weighting factor is shown in Fig. 4.3b. The impact is somewhat subtle, but can be seen when carefully comparing the bus near Point Beach generating station in eastern Wisconsin near Lake Michigan, which is circled in red in Fig. 4.3b. In Fig. 4.3a, the Point Beach bus has a higher node degree than the same bus in Fig. 4.3b, since adding an intermediate weighting factor causes more lines to be upgraded instead of building an excessive number of new lines.



(a) Tree-structured network (b) Reduced cost of line upgrades

Figure 4.3: One-Line Diagrams of WI network showing impact cost weighting c

4.3 Middle Loop: Sequential Updates to Optimization Parameters

The middle loop uses the values of p that are computed from the inner loop to add or update the line limits (4.1c) and the cost function parameters c (4.1a). The line limits are used in subsequent work to determine initial per unit impedance values. Control is then passed back to the inner loop to calculate new line flows p using the updates from the middle loop. This proceeds as follows.

After successfully solving each iteration of the inner linear program, the values of the power flows on lines, p, are used to determine which potential transmission line paths are assigned a line MVA limit and added to the system to become existing lines

for the next iteration of the inner loop. Modifying the method in [13], the k potential transmission line paths with the largest flows are assigned line limits and c values, where $k \in 1, 2, ..., 20$. The value of k within this 1 to 20 range is heuristically chosen based on the number of nodes in the network.

The line ratings for the newly added transmission lines are assigned either a value of 125% of the flow on the corresponding path, or the MVA rating of the smallest physical transmission line modeled in the system, whichever is larger. The newly added lines are also given cost weights c corresponding to the node-to-node distance (see Section 4.2.2). Finally, cost weights c for the potential transmission line paths in parallel to the newly added transmission lines are reduced from $5d$ to $3d$. This choice was a empirically-based heuristic decision that was found to avoid having buses with an unrealistically high node degree in the resulting network. This reduction in the cost weights c roughly represents the lower cost associated with future capacity upgrades, i.e. replacing the conductor on a transmission line. The rationale for this choice is further explained in Section 4.2.1. If there is a non-zero flow value on a potential transmission line path that is in parallel to an existing transmission line (either preexisting or added in a previous iteration of the middle loop), the limit on the existing line is increased by 150% of the flow on the potential transmission line path. This value creates some margin in the power limit of the lines.

The process of sequentially solving the inner linear program and performing the middle loop updates to line limits and c values is continued until the inner loop solution p

results in no additional transmission lines added to the system or line limits upgraded. At this point, the network is passed to the outer loop which updates the net injection vector. To avoid multiple small sequential increases in line limits that prolong the run-time of the algorithm, once the middle loop has stopped adding transmission lines to the existing network, the middle loop is terminated when the largest upgrade to an existing transmission line limit is less than or equal to 15 MVA. Due to the sparsity of the solution vector p, all of the potential transmission line paths in the modified complete graph (Section 4.2.1) are included in the intermediate solutions until after final iteration of the outer loop. After this final iteration, all of the potential transmission line paths in the complete graph are removed from the model, leaving only paths that have been added to the existing network. The $\zeta_{i,j}$ variable associated with each path is set to 1, indicating that a transmission element is to be constructed along this path. Computational experience has shown that after the linear programming-based inner and middle loop formulation first terminates, it typically creates more radial, tree-structured networks than are found in existing power system networks. This is remedied in the following section, which adds loops to the network by updating the net injection vector.

4.4 Outer Loop: Generation and Load Scenarios

The primary purpose of the outer loop is to compute or update the net injection vector P. In order to model the load and generation scenarios discussed in Section 1.3, multiple sets of net injection vectors (P) are created. To compute the net power

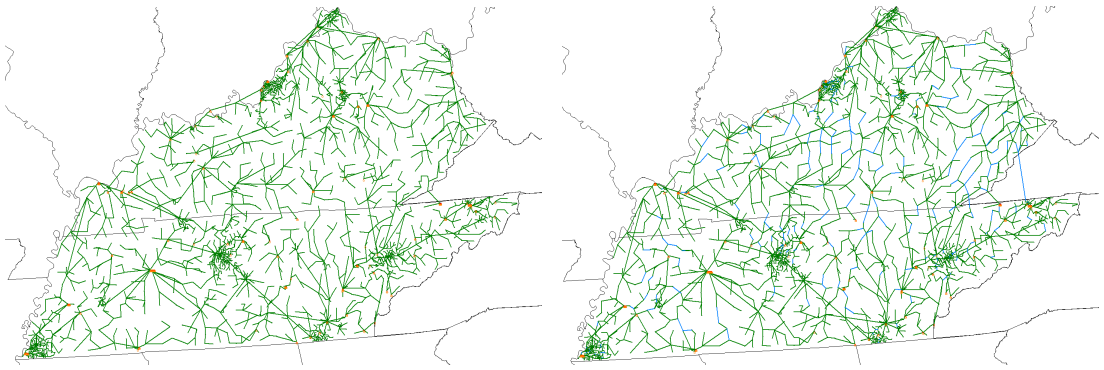
injection or withdrawal, either real or synthetic data [108] is required for both load values and generator dispatch points. The following sign convention is utilized: loads, or power sinks, are denoted as negative power injections to a bus while generators, or power sources, are represented as providing positive power injections to a bus.

4.4.1 Net Injection Scenarios

Since the network flow problem is highly dependent on the location of the generation sources, varying the output of the generators creates variations in the network topology. Thus, multiple sets of net injection vectors were created to make the resulting networks more robust to operating conditions with widely varying generation and load. This is later illustrated using four plots of a network created in the geographic footprint of Tennessee and Kentucky.

As is typical for FERC planning cases, a forecast peak load was utilized to create the net injection scenarios. This peak load scenario was created by using the concurrent peak from the synthetic load data described in [108]. The first net injection scenario was created using the concurrent peak of the synthetic load data and the generator economic dispatch (Appendix B.1.

After the middle loop terminates using this first net injection scenario (Section 4.4), the net injection scenario is updated, and the network topology generated from this first iteration of the outer loop, (i.e. assignment of the ζ variables) is passed back to the middle loop until it again terminates. This second net injection scenario uses the



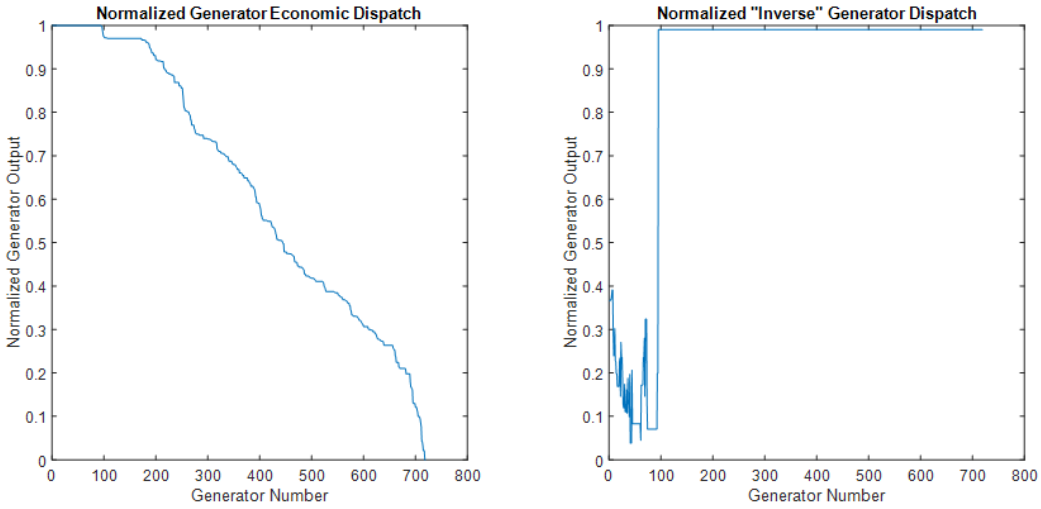
(a) Network created using the economic dispatch(b) Additional lines added using the complementary dispatch

Figure 4.4: One-Line Diagrams on Tennessee and Kentucky Network showing effect of updating the net injection scenarios

complementary dispatch (defined below) and modified system load (if necessary). This complementary dispatch net injection vector, along with the existing lines created in the economic dispatch-based iterations of the middle loop, result in a more robust network with additional lines added relative to the base case.

Fig. 4.4a shows the resulting network using the economic dispatch net injection vector. The generators are drawn using small orange circles. Notice that the network is largely radial and "tree-structured", with the generator buses having the highest node degrees. Next, Fig. 4.4b plots the new lines added from the complementary dispatch step in blue and shows the existing lines from the economic dispatch step in green. Extensive computational experience demonstrates that running the linear program with an additional net injection vector typically results in substantial addition of new lines that reduce the average cycle length.

After the economic dispatch (Fig. 4.5a) is performed, the resulting generator output



(a) Economic dispatch (b) Complementary dispatch

Figure 4.5: Normalized Generator output plots

power assignment is used to compute an alternative dispatch, termed the complementary dispatch. Any generator with real power output $P_{G,i}$ less than 85% of the corresponding maximum power output $P_{g,i}^{\max}$ is constrained to be at it's maximum output. Then, the lower bounds of the generators not set to their maximum output are set to 0 to allow the complementary dispatch to be feasible. Finally, the optimization problem is solved, and a dispatch for all of the generators is computed.

However, if the total system load is below the sum of the modified lower bounds for the generators, the total load is proportionally scaled to be equal to the sum of the modified lower bounds. This ensures that the union of the generator power outputs from the economic dispatch and complementary dispatch results in each generator being at or above 85% of its maximum power output. This allows for the networks created to have transmission capacity to successfully deliver power from any of the generators in

the network.

The result of the complementary dispatch is shown in Fig. 4.5b. First, the normalized result of the economic dispatch is plotted in Fig. 4.5a. The generators are sorted by their normalized real power output, and assigned a reference number to generate the plot of the complementary dispatch. Notice in Fig. 4.5a, approximately 100 generators are dispatch at or near their maximum output, with the rest of the generators are dispatched at a lower power output corresponding to their respective costs. Thus, as shown in Fig. 4.5b, the complementary dispatch sets generators numbered 100 and onward equal to their maximum output. The remaining generators (numbered 1-100) are given a new lower bound of 0, and their output under the complementary dispatch is shown in Fig. 4.5b.

Note that this method can be parallelized by simultaneously creating networks using the economic dispatch net injection and complementary net injection and taking the union or some other combination of the two networks for the final network.

4.4.2 Generator N-1 contingencies for network robustness

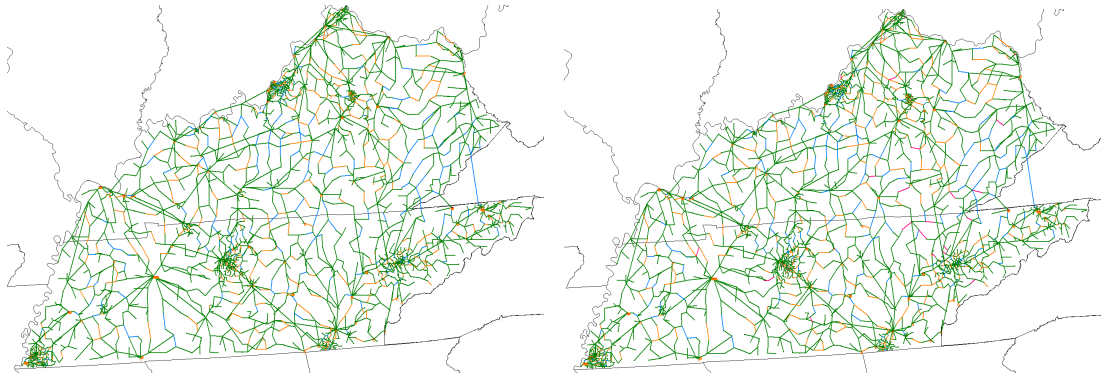
After creating the economic dispatch and complementary dispatch net injection vectors and running two iterations of the outer loop, additional iterations of the outer loop are run using injector vectors created by considering generator N-1 contingencies.

Starting with the generator substation having the largest total maximum output, all generators at that substation are removed from the network. If the removal of this

generation capacity causes the total system load to be higher than the total network generation capacity, the load is scaled down proportionally. Then, the power output of the remaining generators is calculated using the economic dispatch methodology and a corresponding net injection vector is calculated. The inner loop linear program is repeatedly solved until the middle loop terminates, and the resulting network topology (ζ) is used as the input for the next iteration of the outer loop.

The removed generation is placed back in service and the generators at the next largest capacity substation are removed and a new net injection vector is calculated. The process in the preceding paragraph is repeated until the net injection vectors from every generating substation contingency have been calculated and used in the middle loop, or until a target bus to branch ratio is reached. Fig. 4.6a shows the result of the N-1 simulation for generator buses. The lines from the first outer loop iteration are shown in green, the lines added using the complementary dispatch are plotted in blue, and he lines added as a result of the net injection vectors from the generator N-1 contingencies are shown in orange.

The process in the previous 2 paragraphs is then repeated, but the complementary dispatch methodology is utilized to compute each net injection vector instead of the economic dispatch. This generator contingency methodology with complementary dispatch is also continued until all generating substations have been sequentially disconnected or until a target bus to branch ratio is reached. Fig. 4.6b plots the lines in red that are added from the generator N-1 analysis using the complementary dispatch.



(a) Network created using the economic dispatch with generator N-1 contingencies

(b) Additional lines added using the complementary dispatch with generator N-1 contingencies

Figure 4.6: One-Line Diagrams on Tennessee and Kentucky Network showing effect of Generator N-1 contingencies

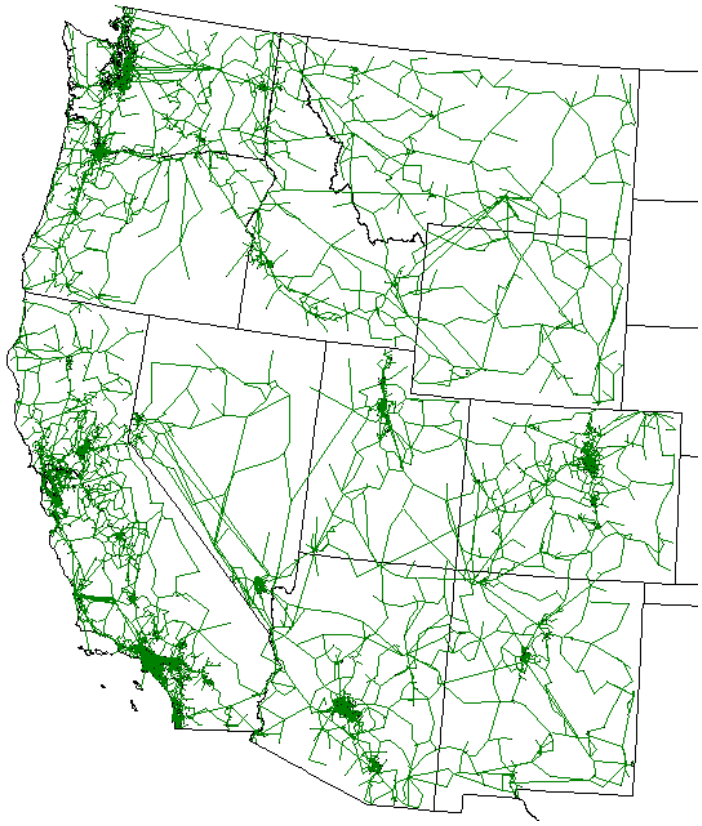


Figure 4.7: One-Line of the full 10,000 bus WECC model

4.5 Chapter summary

After exploring the benefits of modeling physical systems using network graphs, an ideal network construction algorithm (NCA) was presented in Section 1.3. Since this ideal formulation is intractable for networks of practical size, a review of previous reformulations and relaxations of the ideal problem were further explored in Section 1.4.1. However, the majority of the algorithms reviewed either did not model the physical properties of a network with sufficient accuracy, or were still intractable for realistically sized systems. Thus, a KCL-complaint reformulation of the ideal NCA was created and presented in Section 4.1. As further detailed in [100], the topologies generated by the multi-level KCL-compliant network creation algorithm are fairly realistic.

However, after the termination of the outer loop of the KCL-compliant multi-level algorithm the network models only include a network topology (i.e. assignments to the ζ variable) and line thermal limits. Thus, further steps are necessary to assign impedances to lines, create buses with associated voltage levels, and update the network to have a feasible AC power flow. This requires additional steps in a KCL-compliant NCA that assign network elements with associated parameters and constitutive equations to each network branch, and are the portions of Fig. 4.1 that are not in a dashed or dotted ellipse. These additional steps are discussed in Chapter 7 since they are heuristic in nature and do not directly modify the multi-level optimization formulation presented in this chapter. These heuristics modify the initial network created in this Section (Fig. 7.3) to become a power system network model whose DC power flow solution satisfies

the line limits.

Computation experience using KCL-compliant network construction algorithms indicates that they typically result in a network that is more radial than a realistic electrical grid. Moreover, network element flows often substantially differ from the flows calculated using either DC or AC power flow. While there are heuristic modifications to the KCL-compliant algorithm that can address this problem, a more robust solution is sought. The next chapter will present set of novel NCAs that combine the first three general NCA steps introduced in Section 1.3 into one multi-level optimization-based algorithm. This is accomplished by modeling all 3 elements of (1.1e) (KCL, KVL and the network element constitutive relations) in the inner loop, and performing steps 1-3 all within an optimization-based multi-level algorithm without additional heuristic steps.

Chapter 5

KCL and KVL-Compliant Network Construction Algorithms via Quadratic Programming

Recalling the ideal network construction frameworks presented in Section 1.3 (1.1)-(1.4), the first electrical performance constraint to be implemented in any NCA is the KCL constraint in (1.1e), i.e. (1.3b). Chapter 4 explored KCL-compliant NCAs that can create realistic networks with the addition of multiple heuristic steps. Thus, to reduce the reliance on heuristics and address the other issues mentioned in Section A.1.3, a novel class of network construction algorithms (NCAs) is created and presented in this chapter.

The next set of electrical performance constraints that can be included in an NCA

are KVL and the element constitutive equations (1.1e). These bring added realism to the NCA by enforcing the DC power flow constraints in (1.3c). Thus this new class of algorithms will model KCL, KVL and the element constitutive relations (e.g. Ohm's Law) and is termed a KCL/KVL-compliant NCA. Sections 5.1.1 - 5.1.3 shows how KCL is modeled directly using equality constraints while KVL and Ohm's law are represented using an objective function that minimizes losses in a power system network.

However, since a complete graph of all potential transmission expansion paths is modeled in this new KCL/KVL NCA, a weighted L1-norm of the network element flows is used as a regularizer to enforce sparsity on the resulting solution of flows on the network. Section 5.1.2 gives a more detailed review of regularization and how it can be used to create a sparse solution and Section 5.1.2 shows the optimization problem reformulation that includes the L1 regularizer. The KCL/KVL NCA can also be used to model DC power flow by using a change of variable explained in Section 5.1.3.

5.1 Optimization Formulation Derivations

5.1.1 Loss Minimization Subject to Constraints

As observed in the literature review in Section 1.4.2, Ohm's law and KVL can be recovered via the KKT conditions on minimum-energy objective function (5.1a), while KCL is modeled using constraint (5.1b). For simplicity of the following derivations, the formulation minimizing losses in a resistive circuit is used in (5.1), but later sections change the notation to represent the analogous DC power flow. R is a diagonal matrix

of the network element resistances, i and I are the vectors of branch and bus current respectively, and A is the network incidence matrix. For the derivations in the following section, A is assumed to be the incidence matrix relating to the existing elements in a network, but in Section 5.2.2 the incidence matrix is expanded to include potential network transmission line expansion paths.

At first glance, it might appear that the objective function does not properly model Ohm's Law, but as shown in Section 5.1.1, the KKT conditions associated with (5.1) result in Ohm's law for resistive circuits or DC power flow for a power system network. Thus, the following standard form quadratic program with linear constraints is formulated in (5.1),

$$\min_{i} \quad \frac{1}{2} i^T R i \tag{5.1a}$$

$$\text{s.t.} \quad I^0 - Ai = 0 \tag{5.1b}$$

Equivalence of KKT Conditions to Ohm's Law

The following equations are derived to support the proposition from Section 1.4.2 that Ohm's law or DC power flow can be modeled using an optimization formulation minimizing resistive losses. First, the following Lagrangian formulation is obtained from the optimization formulation in (5.1),

$$\min_{i,\lambda} \quad L(i,\lambda) \tag{5.2a}$$

$$\text{s.t.} \quad L(i,\lambda) = \frac{1}{2} i^T R i + \lambda^T (-Ai + I^0) \tag{5.2b}$$

The necessary conditions for optimality from the partial derivative of the Lagrangian are as follows,

$$\frac{\partial L}{\partial i} = i^T R - \lambda^T A = 0 \tag{5.3a}$$

$$\frac{\partial L}{\partial \lambda} = I^{0T} - i^T A^T = 0 \tag{5.3b}$$

Notice that the necessary condition (5.3b) directly recovers KCL, replicating the original constraint in (5.1b). Hence the analysis of this and succeeding sections will focus on (5.3a) and subsequent reformulations. In (5.3a), the dual variables λ play the role of node voltages. Therefore, denoting the dual variables as V, (5.3a) becomes,

$$i^T R + V^T A = 0 \tag{5.4}$$

Rearranging, (5.4) while remembering R is a diagonal matrix results in,

$$A^T V = Ri \tag{5.5}$$

From (5.5), one can see that the dual variables V are indeed the vector of bus voltages.

Additionally, A^TV are the branch element voltages v, where each entry in the vector v is the voltage drop across the element connecting bus i and j, i.e. $v_k = V_i - V_j$. Thus, $A^TV = Ri$ is the matrix form of Ohm's Law, i.e. $v_k = R_k i_k$, and $A^TV = v$ represents KVL [122]. Thus it can be seen that the necessary condition of the loss minimization optimization problem for a resistive circuit recovers Ohm's Law for each resistive branch element, and simultaneously enforces KVL.

5.1.2 Using an L1 Regularizer to Enforce Sparsity

Given that Ohm's law can be represented using a resistive circuit loss minimization formulation, the formulation in (5.1) can be used as the inner loop in a multi-level network construction algorithms that model KCL, KVL, and the element constitutive relation of Ohm's Law. Such a formulation would follow the pattern of Section 4.1, with the line flows being calculated in an optimization based-inner loop, and sequential updates to the network being performed in an outer loop.

However, creating a network graph requires modeling all of the potential elements that could be added to the network. This requires a method to discourage solutions in which current or power flows in every potential element in the network. As explored in Section 2.2, most existing algorithms include binary decision variables for the potential elements (or paths) that will be constructed. This allows for only a small subset of the potential elements to carry power or current, and thus be added to the existing network. However, this mixed integer formulation results in a bi-linear optimization problem caused by the multiplication of two decision variables, which can become com-

putationally intractable for large networks.

One approach is to relax the binary decision variable into a continuous variable as suggested by [28]. However, a method is needed to select only a small subset of the potential transmission line paths to build as existing lines. One such method utilizes an L1 regularizer to reduce the number of "active" potential transmission line paths by enforcing sparsity on the solution to the optimization problem.

Tibshirani's 1994 work [72] presents the Least Absolute Shrinkage and Selection Operator (LASSO). The LASSO was originally presented as a regression method that reduced the number of active predictor variables using an L1 norm. While the original formulation of the LASSO includes an L1 norm as a constraint, [118] reformulates the LASSO to include a weighted L1 norm in the objective function. Using tuning parameter γ, the weight can be varied between the L1 norm and the quadratic regression term. The work in [118] also explains how the L1 regularizer in the LASSO is utilized to create a sparse solution, with the γ parameter changing the magnitude of the sparsity. Thus, the optimization problem from (5.1) is reformulated to include a weighted L1 regularizer on the current flow decision variable vector i, as shown in (5.6). Note that j is taken to be the indicator of bus numbers, since i is the vector of branch current flows.

$$\min_{i} \quad \frac{1}{2} i^T R i + \gamma |c^T i| \tag{5.6a}$$

$$\text{s.t.} \quad I^0 - Ai = 0, \tag{5.6b}$$

$$\sum_{j=1}^{n} I_j^0 = 0 \tag{5.6c}$$

Here γ is a scalar tuning parameter that shifts the weight between the quadratic term and the L1 regularizer in the objective function. The scalar weight γ can be increased from 0 to create a trade-off between the quadratic loss-minimization term capturing DC power flow and the L1 regularizer enforcing sparsity. Additionally, it should be noted that $|i|$ is the L1 norm of the current vector i, which is the sum of the element-wise absolute value function. Finally, c is the vector of weights for the L1 cost function, as further explained in Section 5.2.2.

5.1.3 Modeling DC Power Flow

Earlier it has been established that Ohm's law can be represented by a quadratic loss minimization problem, and an L1 regularizer can be added to ensure sparsity of the solution vector of current flow i on the branches. Next, a relabeling of quantities is implemented to demonstrate that the loss minimization problem is perfectly analogous to the DC power flow. This involves changing the decision variables from branch currents to branch power flows p and changing the input vector from bus net current injections to bus net power injections P^0. Thus, the diagonal matrix of line resistances R is replaced

by a diagonal matrix of line reactances X and used to find the minimal "energy" solution in a "lossless" purely reactive circuit in 5.7.

Note that the the susceptance for line k that is used in the DC power flow calculation is chosen to be $b_k = \frac{1}{x_k}$. This choice of parameters differs from the choice of line susceptances in the AC power flow calculations, where $b_k = \Im(\frac{1}{r_k + jx_k})$. Also note that X^{-1} is also a diagonal matrix with $X^{-1}_{i,i} = (X_{i,i})^{-1}$. Thus $X^{-1} = B_{\text{Line}}$ where B_{Line} is a diagonal matrix of line series susceptances, which is termed the primitive line susceptance matrix, B_{prim}

$$\min_{p} \quad \frac{1}{2} p^T X p + \gamma c^T |p| \tag{5.7a}$$

$$\text{s.t.} \quad P^0 - Ap = 0, \tag{5.7b}$$

$$\sum_{j=1}^{n} I_j^0 = 0 \tag{5.7c}$$

As before, the Lagrangian function is developed as follows,

$$L(p, \lambda) = \frac{1}{2} p^T X p + \gamma c^T |p| + \lambda^T (Ap - P) \tag{5.8}$$

With resulting necessary condition,

$$\frac{\partial L}{\partial p} = p^T X + c^T U - \lambda^T A = 0 \tag{5.9}$$

Where the U term is developed as in (5.17), but with respect to the flow of power instead of the flow of current.

$$U_{i,j} = \begin{cases} -1 & \text{if } i = j \text{ and } p_i < 0 \\ 1 & \text{if } i = j \text{ and } p_i > 0 \\ 0 & \text{if } i \neq j \text{ and } i_k \neq 0 \\ [-1, 1] & \text{if } i = j \text{ and } i_k = 0) \end{cases} \tag{5.10}$$

Instead of using V as before, the dual variables λ can instead be relabeled as the vector θ, so (5.9) becomes,

$$Xp - A^T\theta + Uc = 0 \tag{5.11}$$

Which can be rearranged to solve for the line power flows as follows,

$$p = X^{-1}(A^T\theta - Uc) \tag{5.12}$$

Thus, (5.12) can be rearranged into the following (modified) DC power flow equation,

$$P = AB_{\text{prim}}(A^T\theta - Uc) \tag{5.13}$$

Remembering that $AB_{\text{prim}}A^T = B_{\text{Bus}}$, (5.13) can be used to solve for θ,

$$\theta = B_{\text{Bus}}^{-1}(P + AB_{\text{prim}}Uc) \tag{5.14}$$

5.1.4 Proofs and Derivations: Existence and uniqueness of solution for QP with L1

Now that the optimization formulation of the KCL and KVL compliant network construction algorithm has been established, properties of the problem need to be examined. Using standard textbook proofs and results [123], it can readily be established that an optimal solution to (5.6) exists and is unique. This is because the objective function is strongly convex and the constraints are convex, so the solution exists and is unique. A brief summary of the observations leading to this conclusion are given below.

The R matrix is diagonal with all positive entries, so the quadratic term of the objective function is convex. The absolute value function is also convex, thus the addition of the two convex terms in the objective functions is also convex. The Hessian of the objective function is positive definite at all points except $i = 0$. However, assuming $I^0 \neq 0$, the KCL equality constraints remove $i = 0$ from being a solution to the problem. Thus, since the solution of all zero flows is not feasible, the equality constrained optimization formulation in (5.6) has a unique optimum point. For a more rigorous proof of the existence and uniqueness of the solution to the minimal resistive loss quadratic program (QP) with an L1 regularizer, see [124]

5.1.5 Circuit Interpretation of Necessary Conditions for $\gamma \neq 0$

It is beneficial to examine the effect of the L1 regularizer on the necessary conditions of (5.6) and to interpret these from a circuit analysis perspective. Following the same

approach as in Section 5.1.1, the Lagrangian is constructed in (5.15).

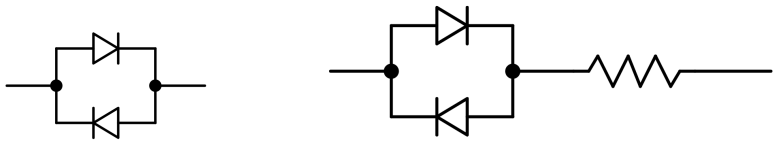
$$L(i, \lambda) = \frac{1}{2} i^T R i + \gamma c^T |i| + \lambda^T (Ai - I^0) \tag{5.15}$$

The first necessary condition of the Lagrangian formulation of the optimization problem is calculated in (5.16). Remember that $\frac{\partial L}{\partial \lambda}$ directly recovers KCL, so analysis of this necessary condition is omitted for future sections. Additionally, since $\frac{d}{di}|i|$ is not defined at $i = 0$, the piecewise function U (5.17) is used to represent $\frac{d}{di}|i|$.

$$\frac{\partial L}{\partial i} = i^T R + \gamma c^T U - \lambda^T A = 0 \tag{5.16}$$

$$U_{i,j} = \begin{cases} -1 & \text{if } i = j \text{ and } i_k < 0 \\ 1 & \text{if } i = j \text{ and } i_k > 0 \\ 0 & \text{if } i \neq j \text{ and } i_k \neq 0 \\ [-1, 1] & \text{if } i = j \text{ and } i_k = 0) \end{cases} \tag{5.17}$$

Note that for the last condition of (5.17) ($i_k = 0$), $U_{k,k}$ can be set to any sub-gradient of $|i_k|$ since the current value itself is 0. However, due to numerical tolerances of floating point arithmetic and to simplify further discussion regarding the necessary conditions, hence the choice is made to set $U_{k,k}$ to 0 when $i_k = 0$. Following the method in Section



(a) Pair of back-to-back diodes

(b) Back-to-back diodes with series resistor

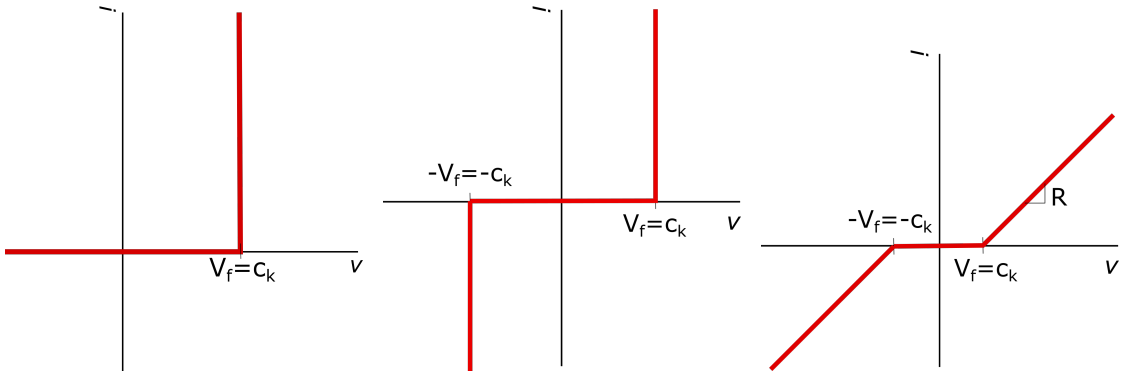
Figure 5.1: Circuit diagrams for the following diode configurations

5.1.1, λ can be relabeled V so (5.16) can be rearranged to become,

$$A^T V + \gamma U c = R i \tag{5.18}$$

Since (5.18) has the same form as (5.5), KVL is still captured by $A^T V = v$. However, the Uc term acts as a "voltage offset" that modifies the voltage across each network element, making the voltage difference between buses i and j greater than the the voltage drop across the resistive element connecting the two buses. This serves to reduce the current flow on some resistive elements to enforce sparsity in the solution of i. As explored further in Section 5.1.5, the effect of Uc on the element current flow can be described using a circuit analogy of a pair of back-to-back diodes in series with every resistive network element.

Recall from circuit analysis that an ideal diode conducts no current when the diode is reverse biased (i.e. when the voltage drop is below the diode's threshold voltage) but behaves as a short circuit able to conduct current when the voltage drop reaches the threshold voltage, as shown in Fig. 5.2a. To allow for reverse current flow in a network path, a reverse polarity ideal diode then is placed in parallel with the first ideal to form



(a) An ideal diode (b) Pair of back-to-back diodes (c) Back-to-back diodes with series resistor

Figure 5.2: Voltage vs. current curves for the following diode configurations

the back-to-back diode pair shown in Fig. 5.1a. This diode configuration has the voltage vs. current curve shown in Fig. 5.2b. Finally, since (5.18) includes a resistive circuit element in series with the Uc offset term, a resistor is included in series with the back-to-back diode pair as shown in Fig. 5.1b. This circuit behavior of a pair of back-to-back ideal diodes in series with a resistor is illustrated by the piecewise curve in Fig. 5.2c. In this and future sections, the combined resistive element (or reactive element for DC power flow) and pair of back-to-back diodes in series is treated as a single equivalent network element that has the voltage vs current curve of Fig. 5.2c.

This behavior of a pair of back-to-back ideal diodes simulates the effect of the Uc term on the current flow in (5.18), with the c_k term acting like the threshold voltage of an ideal diode. When there is no current flowing in a circuit element there is no voltage drop across the element from Ohm's law, and $U_{k,k} = 0$ from (5.18). When the voltage difference between bus i and bus j is greater than zero but lower than c_k, then no current flows in branch k. This is because the "threshold voltage" from the c_k term has not been

exceeded, much like the threshold voltage threshold in an ideal diode. However, when $|V_i - V_j| > c_k$ then current begins to flow in branch k according to a modified version of Ohm's law, i.e. $i_k = R_{k,k}^{-1}(V_i - V_j + c_k U_{k,k})$, as illustrated in Fig. 5.2c.

However, later results show that the numerical values for current or power determined by QP solvers are never identically zero, but instead have a small numeric value assigned. Thus, experimental results have suggested that per unit current or power flow values below $2e^{-}2$ can be considered numerically zero. This behavior may be viewed as analogous to the small but non-zero current flow in a non-ideal diode before it reaches its threshold voltage.

For the Loss minimization problem modeling DC power flow, in (5.11) the term $B_{\text{prim}}A^T\theta$ represents the power flowing in the transmission lines in a DC power flow solution. Thus, the term $B_{\text{prim}}Uc$ is an angle "offset" that is present in every network path. These angle offsets can be thought of as an ideal phase shifting transformer (PST) that is connected to every existing element and potential transmission line path in the network. γc can be thought of as the "tap" controlling the PST and increasing or decreasing the magnitude of the angle offset. However, unlike a physical PST as used in practical operations, here the angle offset always "opposes" the angle difference between buses i and j and moves the angle difference across the transmission line closer to 0. Thus, if $|\theta_i - \theta_j| < c_j$ then $p_j = 0$, where p_j is the power flowing on the line connecting buses i and j. These circuit analogies of a resistive circuit with back-to-back diodes or DC power flow with ideal phase shifting transformers can be used to verify the results

computed using QP solvers.

5.2 Input and Calculated Data

Similar to the KCL-only NCA, the required input data is the set of buses with associated geographic and power flow data. However in the KCL-KVL NCA, initial values for the line reactances (X) must be provided. As detailed in later sections, the same set of input buses and associated load scenarios is used as the input for the KCL-KVL algorithm.

The input data of number of buses (n), list of existing transmission lines (if applicable), L1 norm cost weighting factors (c) and sets real power injections (P) are the same as in Sections 3.1 and 4.2. As in Section 4.2.2 and further explained in Section 5.2.2 below, c can be calculated using the bus-to-bus distance matrix D. However, for the KCL/KVL-compliant NCA, the per unit impedances (X) of existing transmission lines are also required. In addition, initial impedance values must be assigned to all potential transmission line paths in the network, which can be accomplished using criteria-based methods listed in Sections 5.2.1 and 5.2.1, or directly supplied to the KCL/KVL NCAs. If only line power limits are given for the existing network, a method similar to that in Section 5.2.1 can be used to assign approximate line impedance values to the existing transmission elements.

To inform comparisons of network metrics, the input data also includes the following quantities: the voltage levels that will be utilized in the network being created , the

maximum length of each transmission line at each voltage, and the total length in miles of transmission lines at each voltage level. As explained further in Section 3.2, this data is compiled from the publicly available FERC Form 1 and DOE state risk profiles, with optional input of statistics gathered from the FERC Form 715. As documented in Section 2.5, additional data from the FERC Form 1 and 715 is utilized to validate the realism of the networks created in this dissertation.

5.2.1 Initial Assignments

After constructing the modified complete graph as described above in Section 5.2.2, initial transmission line parameters are assigned using the method described in this section. The following quantities are initially assigned to each potential transmission line path and are further updated in the multilevel algorithm described below: voltage level, number of conductors per phase, conductor size in kcmil, phase to phase spacing (GMD), and the associated lumped parameter impedance values (R,X,B) and thermal limits in MVA that are computed from the previous quantities.

First, a range of permissible voltage levels need to be assigned each potential transmission line path in the network. The minimum voltage level is determined using the length of the potential transmission line path, and the maximum allowable voltage level is determined by the input data to the algorithm (Section 3.2). The maximum allowable line length at each voltage level present in the network as described in Section 3.2.5 is used to pre-assign voltage levels to each potential transmission line path. For example, Table 5.1 shows the maximum lengths are shown for the 7000 bus 4-state upper Midwest

system. As an example, lines that are between 100 and 130 miles are initialized at 345 KV and are given a minimum allowable voltage of 345 KV. Lines between 75 and 100 miles have a minimum allowable voltage of 230 KV and are initialized to the voltage level, and so on until all of the potential transmission line paths in the network have a assigned minimum allowable voltage and a preassigned initial voltage level.

Next, each potential transmission line path is assigned a numerical value that is used to determine the phase to phase spacing for transmission lines created in this path. As explained in Section 3.2.7, this value, called the GMD index, it is a random number chosen on the distribution of phase to phase spacing values. As the voltage level of the potential transmission line path or assigned existing transmission line changes, the GMD value is updated from the new distribution of the new voltage level using the same index value.

Finally, the average MVA capacity conductors at each voltage level are chosen from the conductor lookup tables as the initial assignments for the potential transmission line path impedance parameters. The conductor data lookup tables described in Sections 7.1.1 and 3.2.9 were utilized to calculate the per unit length impedance parameters, which were multiplied by the path length to determine the potential transmission line path reactances (X). These reactance values are needed for the inner loop calculations as described below in Section 5.3.1, along with the cost function terms described above in Section 5.2.2. As discussed later in Section 5.3.2, these initial voltage and line parameter assignments are updated in the middle loop of the multilevel algorithm as described in

Voltage Level (kV)	69	138	230	345
Maximum Length (miles)	30	75	100	130

Table 5.1: Maximum allowable lengths for transmission lines by voltage level

Section 5.3.2.

5.2.2 Incidence matrix and cost function calculations

To design a synthetic power system with no existing transmission lines, a complete graph is created that connects every pair of buses in the network. This initial complete graph is then modified by removing all transmission lines that are beyond the specified maximum length at each voltage level, using the method described in Section 4.2.1. However, unlike the incidence matrix used in the KCL-only NCA (Section 4.2.1), the incidence matrix in the KCL-KVL algorithm does not have potential transmission line paths in parallel with existing lines. As described later in Section 5.3.3, the outer loop modifies the network topology by converting potential transmission line paths into existing lines, rather than adding an existing line in parallel with a potential transmission line path.

If the KCL-KVL network construction algorithm is used for transmission expansion or for synthetic network generation with a initial condition set of existing lines, the complete graph of potential transmission line paths is created such that potential transmission line paths are not placed in parallel with existing lines. If the transmission expansion algorithm allows for upgrading of existing transmission lines, the flow on the existing transmission lines can be used to inform line upgrades as described later in

Section 5.3.2.

Initially, the incidence matrix is created by constructing a complete graph of potential transmission line paths connecting every pair of buses in the network. If existing line limits are not enforced in the inner loop, the potential transmission line paths are removed for corridors with existing transmission lines, so that at most a single branch is represented between any two nodes. However, if the line limits on existing transmission lines will be enforced in the inner loop optimization problem and the existing transmission lines are allowed to be upgraded, then there needs to be a potential transmission line path with no flow limit in parallel with each existing transmission line with a line limit to be enforced. If flow limits on existing transmission lines are enforced in the inner loop, and existing transmission lines are not allowed to be upgraded, then the potential transmission line paths in parallel with existing lines are removed.

Thus, unlike in the KCL-compliant formulation (Section 4.2), modeling both KVL and KCL allows the incidence matrix to be updated to only include the modified complete graph and not duplicate paths. Additionally in this KCL/KVL-compliant NCA, the reactive network element and voltage offset (represented using an ideal phase shifting transformer in Section 5.1.3) are represented as a single combined network element when constructing the incidence matrix. This eliminates the need for intermediate buses in the incidence matrix formulation.

In the KCL-compliant NCA formulation, the weighting factors c served multiple purposes including serving as a rough surrogate for the operational cost of the existing

lines ($1d$ term) and modeling the construction costs for upgrades on existing network elements ($3d$ term) or new element construction ($5d$ term). For further explanation of the cost surrogates and justifications for these choices see Sections 4.2.2 and A.3. However, in the KCL-KVL NCA formulation, the $p^T Xp$ term models the operational portions of the NCA by representing DC power flow as explained in Section 5.3.1. Thus, the weighted L1 regularizer $c^T|p|$ only needs to be used to model the cost of expansion or line upgrades.

A general form of the L1 weighting function c is given below, where α and β are scaling parameters. In this dissertation, both α and β are set to 1, implementing the standard MW-mile representation in the cost function, i.e. a power systems version of the LASSO-based NCA. However, (5.19) can be modified by including scaling factors for different voltage levels, and by reducing α to account for the reduced cost of upgrading existing transmission lines as explained in the KCL Section A.3.

$$\begin{bmatrix} \\ c \\ \\ \end{bmatrix} = \begin{bmatrix} \alpha d_{\mathrm{e}} \\ \\ \beta d_{\mathrm{p}} \end{bmatrix} \tag{5.19}$$

5.2.3 Net Injection Scenarios

One of the inputs required by the multilevel KCL-KVL algorithm is a set of net injection vectors P. These scenarios can be created multiple ways, for example by using a set of load and generation scenarios provided by an RTO or ISO. However for the research in this dissertation, two methods were used to create realistic-but-not-real net

injections scenarios. Both methods utilized the synthetic load data and generator cost curves created as part of the EPIGRIDS project and briefly described in [108]. The first method creates a sinusoidally varying 24 hour load curve using the non-concurrent peak from the synthetic load data. A sinusoidal curve is created that ranges from 0.4 to 1.0, and load scenarios are created by multiplying every load value by this sinusoidal curve. This load data then uses a rudimentary unit commitment algorithm described below to dispatch the generators using the provided cost curves.

A method also utilized in this dissertation was to pick a subset of the year-long hourly load values created from the EPIGRIDS project. While there are several different methods to pick these values, a minimum subset was chosen to span the range of the aggregate yearly minimum and yearly peak load value. This is accomplished by finding the maximum peak and minimum values for the year and choosing 2 1/2 days on either side of that peak or minimum to include in the set of load scenarios.

For both of the methods described above, a rudimentary unit commitment algorithm was created that utilized the economic dispatch (Section B.1) as the initial condition. If the sum of the minimum MW constraints of the generators is higher than the aggregate load value in a scenario, the minimum MW value was temporarily set to 1 for all generators. After an initial dispatch was calculated, the generator with highest marginal cost was decommitted and the process was repeated. At every iteration, the generator MW lower bounds were tested to see if the original minimum MW constraints could be reinstated. This process was continued until a preset "spinning reserve threshold percentage"

was reached. For this dissertation that percentage was taken to be 10%.

Since the economic dispatch based unit commitment algorithm did not commit some of the more expensive generators for any of load scenarios, an "un-economic dispatch" based unit commitment algorithm was created that dispatched the most expensive generators. First, an economic dispatch was utilized to determine the cheapest generators and those generators were than de-committed. This was accomplished by computing the generator with the lowest average cost per MW and decommitting that generator. This process was repeated until the spinning reserve threshold percentage was met. As in the economic dispatch based unit commitment algorithm, if the sum of the generator minimum MW values was higher than the load at that scenario, the minimum MW value was set to one for all generators until the original MW lower bound could be reinstated.

Originally, the two unit commitment algorithms were run on an entire network being constructed. However, due to large intra-area flows, the unit commitment algorithms were modified to dispatch the generators within regional areas independently, mimicking the process performed by the balancing authorities or ISOs in the real network. This was done to prevent the network construction algorithm from creating a large number of parallel lines to transfer large amounts of power across the network from a collection of large capacity generators to larger load centers.

Since the methods described above created approximately 480 load scenarios, this large lumber of load scenarios was sub-sampled to reduce the computational burden

earlier in the KCL-KVL algorithm. The sub-sampling was accomplished by choosing load scenarios that represented unique unit commitment patterns. Further sub-sampling could be accomplished by only taking snapshots that modeled multiple generating plants being committed or de-committed.

5.3 Multi-Level KCL/KVL-compliant Network Creation Algorithm

In order to construct power system networks using the loss minimization formulation described in previous sections, a multilevel algorithm was created consisting of three levels, as shown in Fig. 5.4. The inner loop is an optimization based algorithms that sequentially solves 5.7 for each of the net injection scenarios, increasing γ in a similar fashion as the LASSO. The middle loop uses the resulting flows on each path for each scenario to update the branch parameters and associated optimization parameters. The outer loop uses the same flow information to make topology changes the network by eliminating a subset of the potential transmission line paths. The outer loop terminates when the desired graph-based metrics are met, such as average branch to bus ratio, average node degree and/or average length of the loops in the cycle basis.

Unlike the KCL-compliant network construction algorithm described in Chapter 4 which required multiple heuristic additions and instantiations of the multilevel algorithm, the full KCL-KVL compliant network construction algorithm is substantially simpler, as seen in the flowchart in Fig. 5.3. Most notably, the KCL-KVL compliant network construction algorithm models DC power flow directly in the inner loop, thus substantially

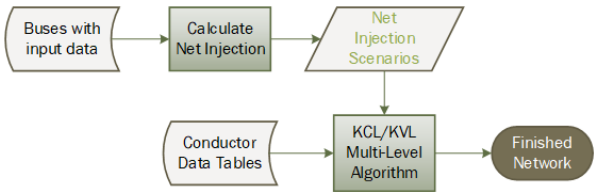


Figure 5.3: Flowchart of the full KCL/KVL network construction algorithm

increasing the realism of the electrical performance of the network at the very beginning stage of the algorithm. In comparison, the KCL-compliant network creation algorithm uses the network flow model to determine the topology, and only models DC power flow after the topology has been established. A comparison of the networks produced using the two algorithms is found in Chapter 6.

Also, the addition of the middle and outer loops in the multilevel algorithm do not guarantee that the network created will be globally optimal, i.e. be an optimal solution to (1.1). While Section 5.1.4 shows that the optimization problem in (5.20) will yield a globally optimal solution, the methods of removing potential transmission paths in Section 5.3.3 are heuristic in nature and not an optimization problem. However, Chapter 6 shows at the networks created using this multilevel algorithm still there many similarities with the existing power system networks.

5.3.1 Inner Loop Optimization

The core function of the inner loop is to solve for p^{ω}, the set of network branch flows for both the potential transmission line paths and existing lines for each of the net injection scenarios ($\omega \in \Omega$ from Table 1.1), and to pass this data to the middle and outer loops. As shown in Fig. 5.4 and described in Section 5.2 above, the input data required

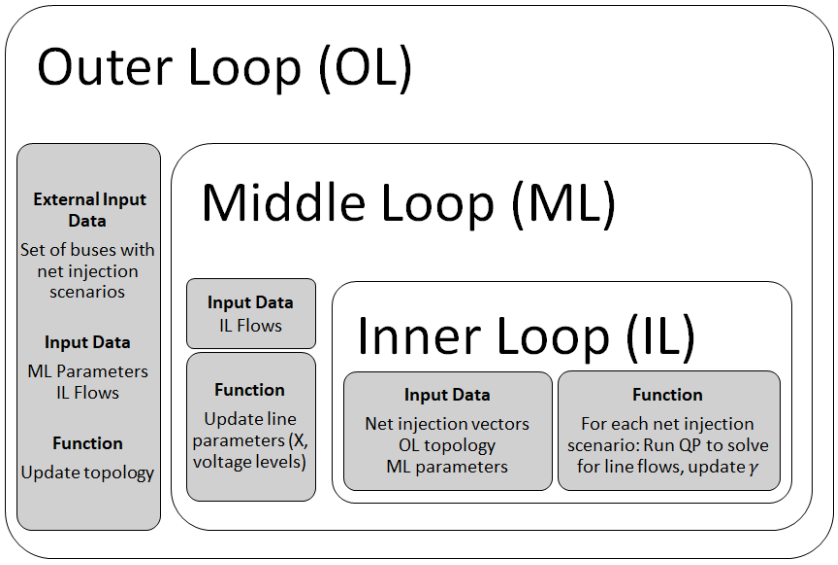


Figure 5.4: Block Diagram of the multi-level KCL/KVL-compliant algorithm

by the inner loop is a network topology with associated parameters and a set of net injection scenarios. At the core of the inner loop is the loss minimization optimization problem derived above in sections 5.1.1-5.1.3, shown in (5.20).

For each of the net injection scenarios ($\omega \in \Omega$), a DC power flow with no L1 regularizer is first solved and is used to resize the transmission lines in the middle loop (Section 5.3.2). Next, (5.20) is solved using a method similar to the LASSO. Both α and β in (5.19) were set to 1, corresponding with the minimal MW-mile cost function explained in Section 5.2.2. γ is initialized to a small value such that only a few percent of the network paths have flows that are numerically zero, and a QP solver is used to compute the optimal value for p. As in the LASSO regression method, γ is increased in value until a set percentage of the branch flows are numerically zero, in this case 20%. Once this point is reached, the inner loop is terminated and the next net injection scenario is solved.

This process of increasing γ creates a trade-off between the smoothing effect quadratic term representing DC power flow in the optimization problem, and the sparsifying effect of the L1 regularizer. The goal of this trade-off in the resistive circuit analogy is to balance the competing objectives of modeling DC power flow, which results in many paths having nonzero flows, and constructing a realistic power system network, which would require that the majority of the potential transmission line paths in the complete graph to have zero flow. Another way to view this trade-off is increasing the value of the cutoff voltage of the back-to-back diodes as explained in Section 5.1.5.

Relating this to the high-level conceptual formulation in Section 1.3, the operational costs (1.1a) for both existing and potential transmission lines (paths) are approximated by the quadratic term ($p^{\omega^T} X p^{\omega}$) of (5.20) while the capital cost is approximated by a weighted vector of costs, γc. Existing transmission lines, either preexisting lines or lines labeled as "keep" in the outer loop have their L_1 weighting factor set to 0 ($c_j = 0$), while potential transmission line paths are assigned a cost term c as explained in Section 5.2.2

The NCA algorithms proposed in this dissertation utilize the economic dispatch [30] or variants thereof, to determine the production cost term of the objective function (1.1a). Thus, the objective function of this novel loss minimization model using quadratic programming only depends on the new transmission lines being constructed (ζ) and the engineering design parameters of the new line(s) (b), modeled as the line reactances (X).

Since the constraints (KCL) are linear and the objective function is composed of a quadratic and a piece-wise linear term, an adaptation of the problem can be solved using a quadratic programming (QP) solver. For this dissertation, Gurobi package is utilized as the QP solver [125] through its MATLAB [126] connector. However, the addition of the L1 regularizer creates a discontinuity in the gradient of the objective function, presenting difficulties for some solvers, as further explored in Section E.1.

$$\min_{p^\omega} \quad \frac{1}{2}{p^\omega}^T X p^\omega + \gamma c^T |p^\omega| \tag{5.20a}$$

$$\text{s.t.} \quad P^\omega - A p^\omega = 0, \tag{5.20b}$$

$$|p^\omega| \leq \text{line limits}, \tag{5.20c}$$

$$\sum_{i=1}^{n} P_i^\omega = 0 \tag{5.20d}$$

As an optional preprocessing step, the net injection scenarios can be scaled up uniformly to avoid numerical issues. At the beginning of the algorithm, the network is nearly a complete graph, causing the net power injections to each bus to "spread out" over a large number of potential transmission line paths. Thus, each potential transmission line path carries a numerically small flow, oftentimes approaching the value of the numerical cutoff threshold discussed in Section E.1. These numerical issues, a "pre-solve" of the inner loop can be conducted by expanding the range of gamma and running the inner loop for one net injection scenario. The resulting flows from this solution are used

to uniformly scale up the set of net injection scenarios such that the largest flow is equal to the maximum MVA limit in the conductor lookup table.

To decrease the computational time for the inner loop, a further sub-sampled set of net injection vectors can be utilized when the network is still dense. As more lines are removed in the outer loop and certain targets are met (for example branch to bus ratio's or average node degree), a more granular set of net injection scenarios can be utilized. Also each of the inner loop scenarios can be run in parallel since each scenario is independent from the other scenarios and the output data for each scenario concatenated into a matrix and sent to the middle and outer loops for analysis and modifications to the network.

5.3.2 Middle Loop Parameter Updates

As shown in Fig 5.4, the main function of the middle loop is to take the flow values from the inner loop and update the network branch parameters. This is performed using three steps, first using the maximum flows across the scenarios to determine the target MVA limits of the branches, then assigning new conductors to the paths based on the updated MVA limits, and finally calculating the lumped parameter impedance values (R,X,B) for the updated transmission conductors.

The maximum flow across each scenario and the existing thermal limits in MVA were used to determine the utilization percentage for each path in the network. Then, a specified percentage (in this case 10%) of the lines that were overloaded and/or under-

utilized were resized. The thresholds for overloaded and underutilized lines were chosen to be 100% and 50%, respectively. To create a network that is more robust for load growth, margin may be added to the percentage threshold for a line being considered overloaded, i.e. a line could be resized if the maximum flow across all scenarios is greater than 90% of the line rating. Likewise, the threshold percentage for the line being considered underutilized can be reduced, but since lines are size based on the maximum flow across all the scenarios, 50% was chosen to avoid having too many lines assigned to an unnecessarily high voltage level. To avoid making drastic changes the network, lines were resized by increasing or decreasing their capacity by a percentage of the line's specified MVA rating, in this case 30%.

Once a line's MVA rating has been updated, a new conductor from the conductor lookup table needs to be assigned to each line with an updated rating. This is done by taking the conductor with the closest MVA rating to the new rating assigned in the first step, and having the MVA rating updated to match the new conductor assigned. This update of conductor might change the associated voltage level of the line. Currently, the algorithm allows voltage levels to be changed during all middle loop iterations, but future work includes modifications where the middle loop is changed to fix the voltage level of lines after a certain point in the algorithm.

Once the new conductor has been assigned to the potential transmission line path or existing transmission line, the impedance parameters (R,X,B) are calculated using the values in the conductor lookup tables. As further explained in Section 3.2.9, each

entry in the conductor lookup table is a transmission line conductor with associated parameters such as voltage level, per unit length inductive and capacitive reactance parameters, and thermal limit ratings in MVA. The inductive and capacitive reactances at one-foot spacing (X_a and X'_a, from the lookup table) are used along with the inductive reactance spacing factor (X_d, calculated using the GMD) to calculate the impedance of each transmission line, as further explained in Section 5.2.1.

Finally, whether the middle loop upgrades and/or downgrades transmission lines is dependent on the progression of the full multi-level algorithm as well as the flow values. Early in the algorithm, the network is very dense and thus lines are not resized. This is because either the path flows are unrealistically small because each line carries a very small percentage of the net injection at each bus, or the net injection scenarios have been scaled up linearly as described in Section 5.3.1. In either case, it does not make sense to resize the transmission lines based on these unrealistically small or artificially scaled flow values. Once the number of paths in the network has decreased to a more reasonable level, branches can begin being resized. In this algorithm, once the branch to bus ratio has decreased below 20, branches that are overloaded are increase in capacity, but only potential transmission line paths that have been converted to existing lines (and had their corresponding L1 regularizer removed) are downsized at this point. Once the network drops below a branch to bus ratio of 12, then both overloaded and underutilized potential transmission line paths and existing line are resized. However, at each of these points in the algorithm, only one iteration of the middle loop is run for each iteration of the outer loop. This is because the network is still far from being finalized, and and

only a single "step" should be taken in resizing the lines. Additionally, when the network has a branch to bus ratio greater than six, the inner loop flows for each net injection scenario are used to determine the maximum flow for line resizing.

Once the network drops below a branch to bus ratio of 6, the DC power flow computed in the inner loop is utilized to compute the maximum flow across each scenario. This change occurs because the network is starting to approach completion, and the flows returned from the inner loop optimization calculation differ from the DC power flow calculations. Recall from Section 5.3.1 that the inner loop uses a loss minimization to model DC power flow, but this is regularized using the L1 norm to enforce sparsity in the solution to p. The effect of the L1 regularizer can be thought of as a phase shifting transformer offsetting the voltage difference across a transmission line and modifying the power flowing through that line. Thus, the flows resulting from the inner loop do not exactly match the DC power flow, which is desirable to drive some of the flows on the lines to zero. However, once the network nears completion, the lines are resized using the DC power flow values so that the final network will be "DCOPF feasible", i.e. all of the transmission lines will be within their MVA limits for all of the net injection scenarios, except for those transmission lines that are at the maximum allowable size and need to be double circuited.

Once the network is near completion (i.e. a branch to bus ratio of less than 3), the middle loop is run until only a few of the lines are overloaded. In order to run multiple iterations of the middle loop, the DC power flow is computed for each net injection

scenario after some of the transmission lines have been resized in order to determine how many lines are still overloaded. If a transmission line has been increased to the maximum size in the conductor lookup table, that line is flagged for manual intervention and no longer counts towards the total number of overloaded paths. Future work will include updating the algorithm to assign multiple circuits to the same potential transmission line path, i.e. routing multiple lines along the same transmission corridor.

The termination criteria for the middle loop are as follows. When the network has a branch to bus ratio greater than 20, the middle loop does not run at all. Once the branch to bus ratio drops below 20 but as above 6, the middle loop only completes one iteration per outer loop iteration. However, once the network is below a branch to bus ratio six, the middle loop terminates once five or fewer lines are overloaded. A small number of lines are allowed to be overloaded to avoid convergence issues where increasing the capacity of a line causes other lines in the system to become underutilized, and decreasing the capacity of those lines causes the first set of lines to become overloaded, and this "cycling" continues indefinitely.

Once the middle loop terminates, the path flows from the inner loop and the updated path and transmission line parameters are passed to the outer loop below in Section 5.3.3 which updates the topology of the network.

5.3.3 Outer Loop Topology Modifications

As shown in Fig 5.4, the main function of the outer loop is to take the flow values from the inner loop and the updated path and transmission line parameters from the middle loop and update the topology of the network by removing a percentage of the potential transmission line paths. This is accomplished by normalizing the flow values to better compare flow values across voltage levels, ranges of net injections, and lengths of lines, and removing the percentage of lines with the lowest normalized flow.

The core idea of the outer loop is to combine the sparcifying effect of the L1 regularizer with the mean and standard deviation of the flows to determine which potential transmission line paths to remove from the network. It might initially appear that one could remove the transmission lines with the lowest average (or maximum) flow across the scenarios. However computational experience has shown that this approach favors removing short, low voltage transmission lines and leaving a unrealistically large network of long high voltage transmission lines. Thus, the following normalizations were added to the flows to mimic the "per unitization" techniques elsewhere utilized in power systems.

First the mean of the flows from the inner loop is normalized by the length of the potential transmission line path. This step was added to, in a sense, "per unitize" the mean flow from the inner loop, and allow an accurate comparison of flow values across lines at different lengths, and thereby voltage levels. Before the addition of this normalization, absolute flow values were directly compared across all lines at all voltage

levels and lengths; for example the larger flow value in a long, high voltage 500 kV potential transmission line path would be compared directly to the much lower flow in a low voltage 69 KV sub transmission path. Without the normalization, the outer loop would remove the vast majority of the shorter, low voltage paths, leaving a network with an unrealistically high percentage of long, high voltage transmission lines. However the addition of this normalization substantially remedied this problem. Once the total length of high voltage lines in the network was below the target length specified in the input data (Section 3.2.5, the flow values were no longer normalized by the length of the lines.

Additionally, the resulting mean flows from the inner loop are further normalized by the net injection associated with each potential transmission line path, which is defined as the larger of the two net injections at the substations at either end of the potential transmission line path. Next, the mean flow was normalized by the node degree of each potential transmission line path, which is defined as the larger of the two node degrees of the substations at either end of the potential transmission line path. These heuristic normalizations were added because preliminary results showed that the outer loop otherwise reduced most of the buses to the node degree of one or two, while leaving a small number of very high node degree buses.

Next, the number of "low flow" scenarios was determined using the results from the inner loop. Since the inner loop terminates after driving the flows on 20% of potential transmission line paths to (numerically) zero, this information can be utilized to de-

termine the number of net injection scenarios where a potential transmission line path had "zero flow." This was accomplished by counting the number of inner loop scenarios where the flow value was below the numerical cutoff value (see Section E.1 for more information about the numerical cutoff).

The normalized flow values and the number of low flow scenarios for each potential transmission line path was utilized to determine which paths to remove. This is accomplished by classifying each potential transmission line path into one of four categories: definitely remove, possibly remove, possibly keep, definitely keep. Paths classified as "definitely remove" were deleted from the network at the end of each outer loop iteration, while lines labeled as "definitely keep" were upgraded to existing transmission lines. Existing lines in the network have their L1 regularizer terms removed by setting the corresponding c_j term to zero. This prevents the L1 regularizer from modifying the flow on existing lines, and thus the flow on the line is only determined by the angle difference across the line and the impedance of the line. Paths in the "possibly remove" category are assigned a probability of being removed in the outer loop, and paths in the "possibly keep" category are not removed in the current iteration of the outer loop, but could be removed in future iterations of the outer loop.

To ensure that the network remains connected and A remains full row rank, only potential transmission line paths that do not result in a disconnected network or isolated bus or buses are removed. First, radial all potential transmission line paths were assigned to the "definitely keep" category, preventing them from being removed and isolating the

bus at the end of the line. Additionally, if all of the lines connected to a single bus were marked for deletion, then all of these lines were moved to the "definitely keep" category and upgraded to existing transmission lines. Typically, only a few lines at each bus were categorized for removal each iteration of the outer loop, so when all of the lines at a bus were categorized as remove, it was a good indicator that the bus has a low node degree and thus no lines should be further removed from this bus. Additionally, if the removal of a potential transmission line path would cause the network to become disconnected, that potential transmission line path was moved to the "definitely keep" category and prevented from being removed in future iterations of the outer loop.

Paths that met a less stringent criteria were added to the "possibly remove" category, while paths that met a more restrictive criteria were directly added to the "definitely remove" category. For a potential transmission line path to be added to either of the remove categories, the number of low flow scenarios needed to be at or above the 3rd quartile of the set of nonzero low flow scenarios. Additionally, potential transmission line paths are added to the set of "possibly remove" paths if they have normalized mean flows below the median of all normalized mean flows, and the standard deviation of their inner loop flows is at or below the third quartile of all standard deviations. Paths with normalized mean flows at or below the first quartile, an inner loop flow standard deviation below the mean of all standard deviations, and meeting the low flow scenario criteria are directly added to the "definitely remove" category. potential transmission line paths with the highest 5% counts of total low flow scenarios, but otherwise not assigned to the "possibly remove" category were automatically added to this category

regardless of any other criteria, albeit with a lower chance of being removed.

At the end of the outer loop, up to a set percentage of the potential transmission line paths were removed from the network, in the case of this dissertation 5% was chosen. Experimental results show that removing 10% or more of the potential transmission line paths each outer loop iteration creates unrealistic networks with too many buses having too high of a node degree. First, all of the potential transmission line paths in the definitely remove category were removed from the network. If the number of paths in this category was less than 5% of the total potential transmission line paths, then lines from the "possibly remove" category were than removed until 5% of the potential transmission line paths had been removed from the network. The stochastic element of the "possibly remove" lines was determined using uniform random variables between 0 and 1. Each potential transmission line paths in the "possibly remove" category was assigned a random variable, and a "threshold value" was chosen for each outer loop iteration. potential transmission line paths whose random variable had a value above the "threshold value" were then removed from the network until 5% of the paths have been removed in the current outer loop iteration.

After the specified potential transmission line paths have been removed, the updated network topology was passed to the inner loop to again compute the flows using this new network. The random variables were reset and the flow normalization is performed at the beginning of each outer loop iteration. The outer loop terminates when a set termination criteria is met. Currently the outer loop terminates when the average node degree of the

network is below that of the Eastern Interconnect. However other termination criteria can be included, such as the length of the cycle basis, node degree distribution, etc.

Chapter 6

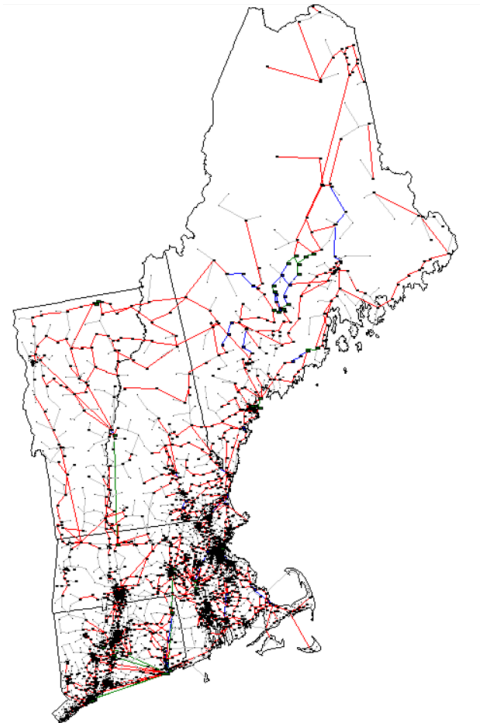
Empirical Results: Synthetic Power Networks as Constructed via KCL-Compliant and KCL/KVL-Compliant Algorithms

To test the performance of the multilevel algorithms presented in chapters 4 and 5, synthetic networks were generated for the 13 geographic footprints discussed in Section 3.2.4. In addition, three larger networks were created by combining the input data from multiple adjacent geographic footprints and creating larger contiguous systems. Two of the larger systems were constructed using footprints similar to PJM and WECC, with the third system covering ISO-NE, NYISO, and the eastern portion of PJM.

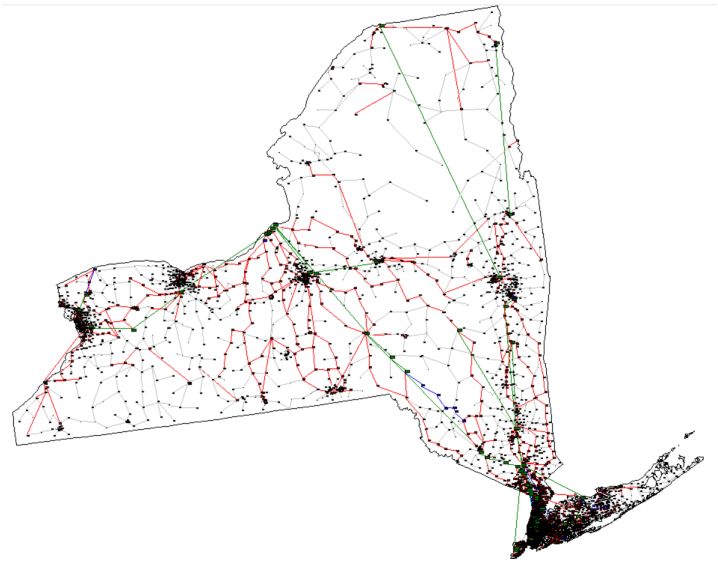
The synthetic network models for six of the geographic footprints are examined in more depth in the following sections, while quantitative statistics are assembled and

analyzed for all 13 networks. Visualizations of these results are provided via network one-line diagrams, and detailed plots of various graphs and electrical performance metrics. The performance of the network construction algorithms varies with the input data, which is highly dependent on the geographic region over which the network is constructed. Thus, it should not be surprising that the quality of the network construction algorithms' results can vary when applied over different footprints.

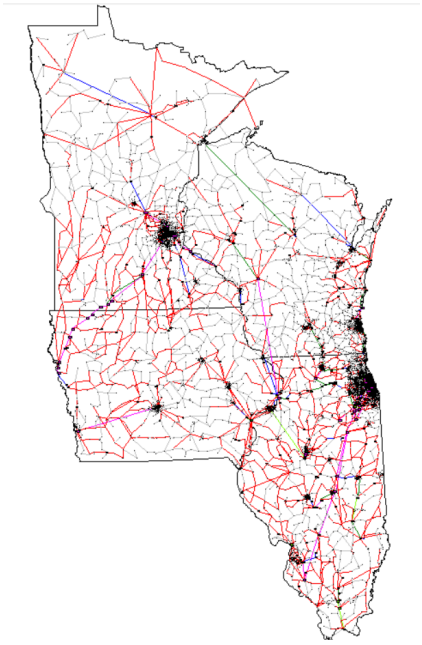
Since the location of the synthetic power system buses was determined using the centroids of the US census tracts, this results in some geographic areas having a much higher density of buses than other areas. The placement of the buses, along with other input data such as the loads, generators, and choices of possible transmission line and tower construction types can have a substantial impact on the resulting network as seen in the upcoming sections. The six networks are chosen to highlight some KCL-KVL compliant networks that showed close similarity to existing power systems, and to examine areas for improvement in the KCL/KVL algorithm by analyzing networks that had a somewhat lower degree of similarity to the existing power system. Thus, the networks examined in depth in this chapter show both the "best case" and "worst case" of the networks that are created by the KCL/KVL-compliant network construction algorithms.



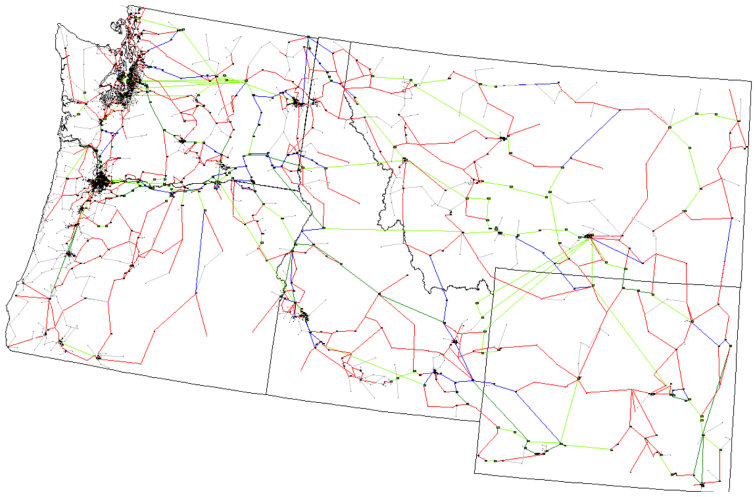
(a) NEISO



(b) NYISO

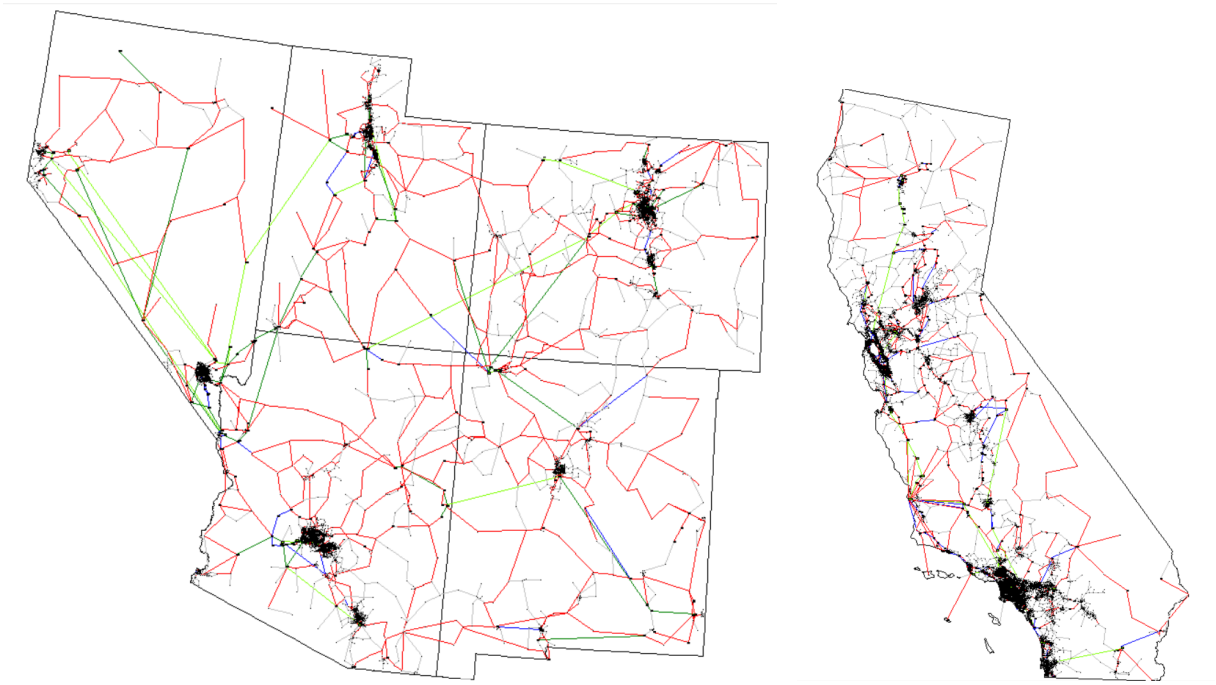


(c) MISO



(d) NWPP

Figure 6.1: One-Line Diagrams of KCL-compliant synthetic power system networks



(a) AZNM (b) CAISO

Figure 6.2: One-Line Diagrams of KCL-compliant synthetic power system networks cont.

6.1 KCL-compliant Network Creation Algorithm Results

Figures 6.1 and 6.2 show one-line diagrams for six synthetic networks created using the KCL-Only multi-level algorithm. The figures use the following color-coding for voltage levels: grey-69kV, red-138kV, blue-230kV, dark green-345kV, light green-500kV, purple-765kV. Both the KCL-only and KCL/KVL multilevel algorithms assigned voltage is directly to transmission lines and not to substations, thus all transmission lines at each substation are directly connected to a single per-unitized "bus" representing all voltage levels. Treating the transmission lines at each voltage level as separate "subnetworks", one can see that these subnetworks need not be fully connected to ensure the full network remains connected. Section 7.2 describes how the network models described above are

modified to include buses and transformers at each substation.

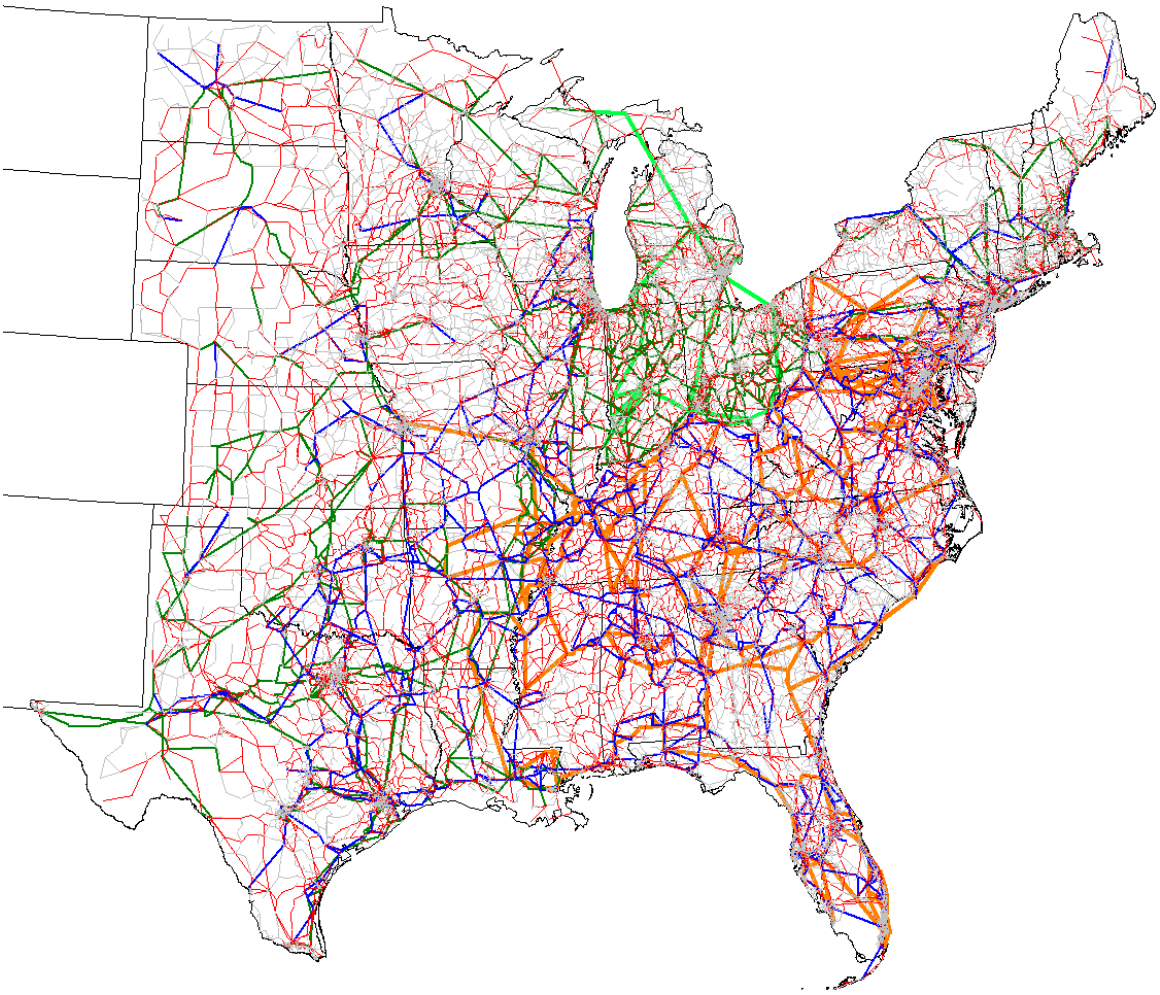


Figure 6.3: One-Line of a synthetic KCL-Compliant Eastern Interconnect model

6.1.1 KCL-Compliant Large Interconnect-Scale Networks

Figures 6.4 through 6.5 show one-line diagrams of synthetic Eastern and Western interconnect models, as well as a synthetic model of the central United States. These models were created using two or more geographic footprints and following the method described in Section 7.3.

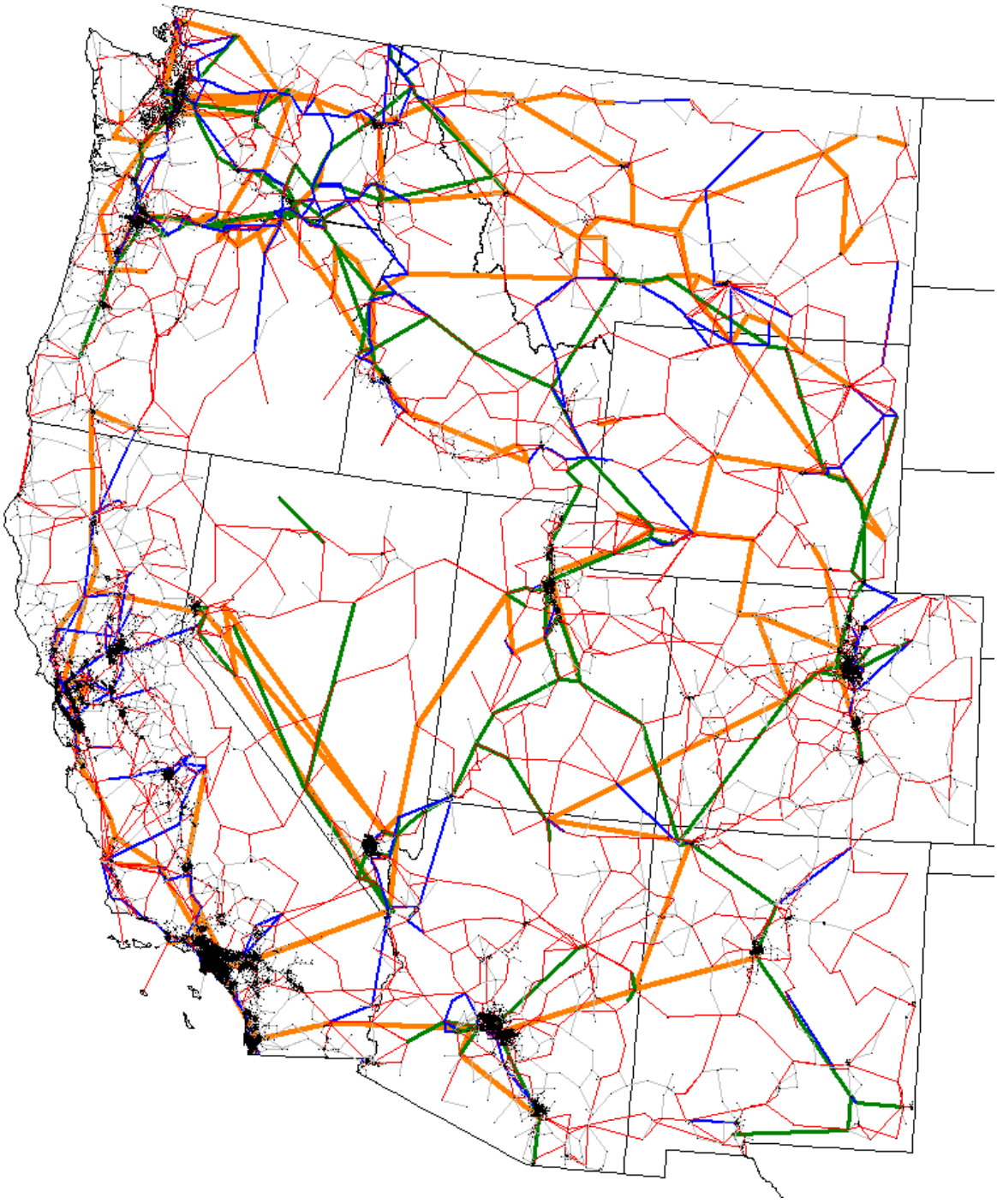


Figure 6.4: One-Line of a synthetic KCL-Compliant Western Interconnect model

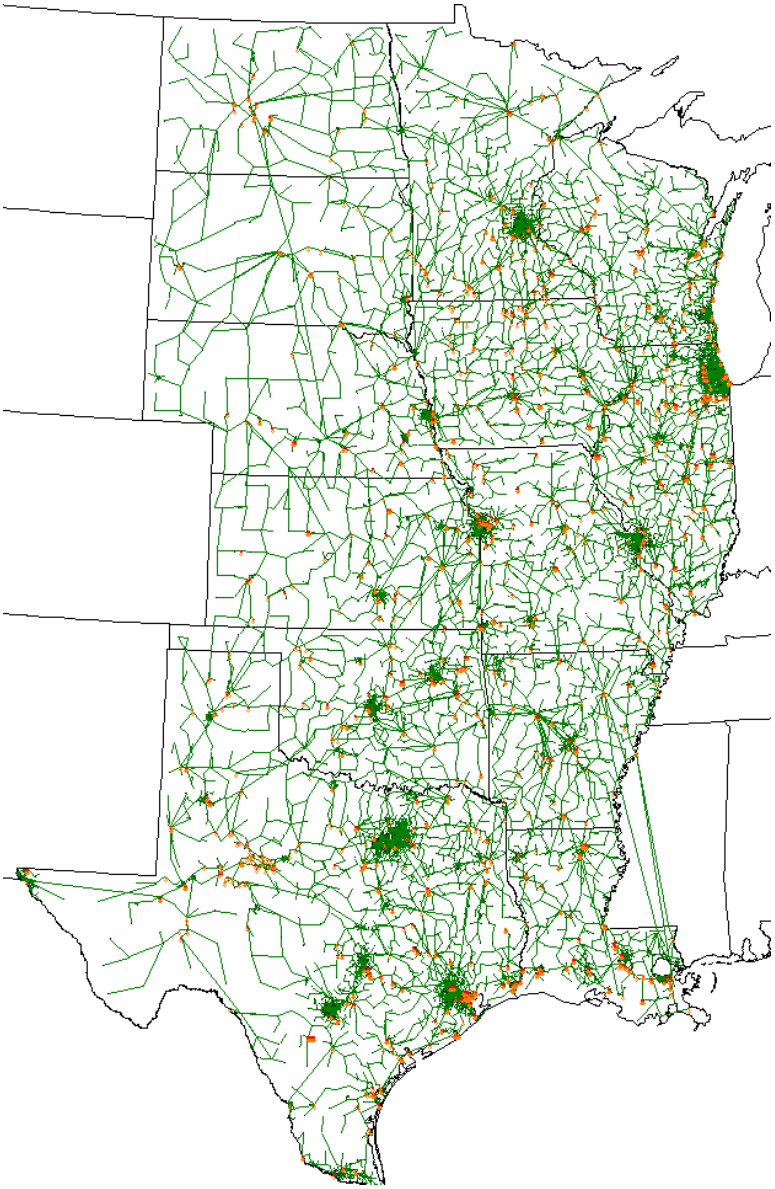
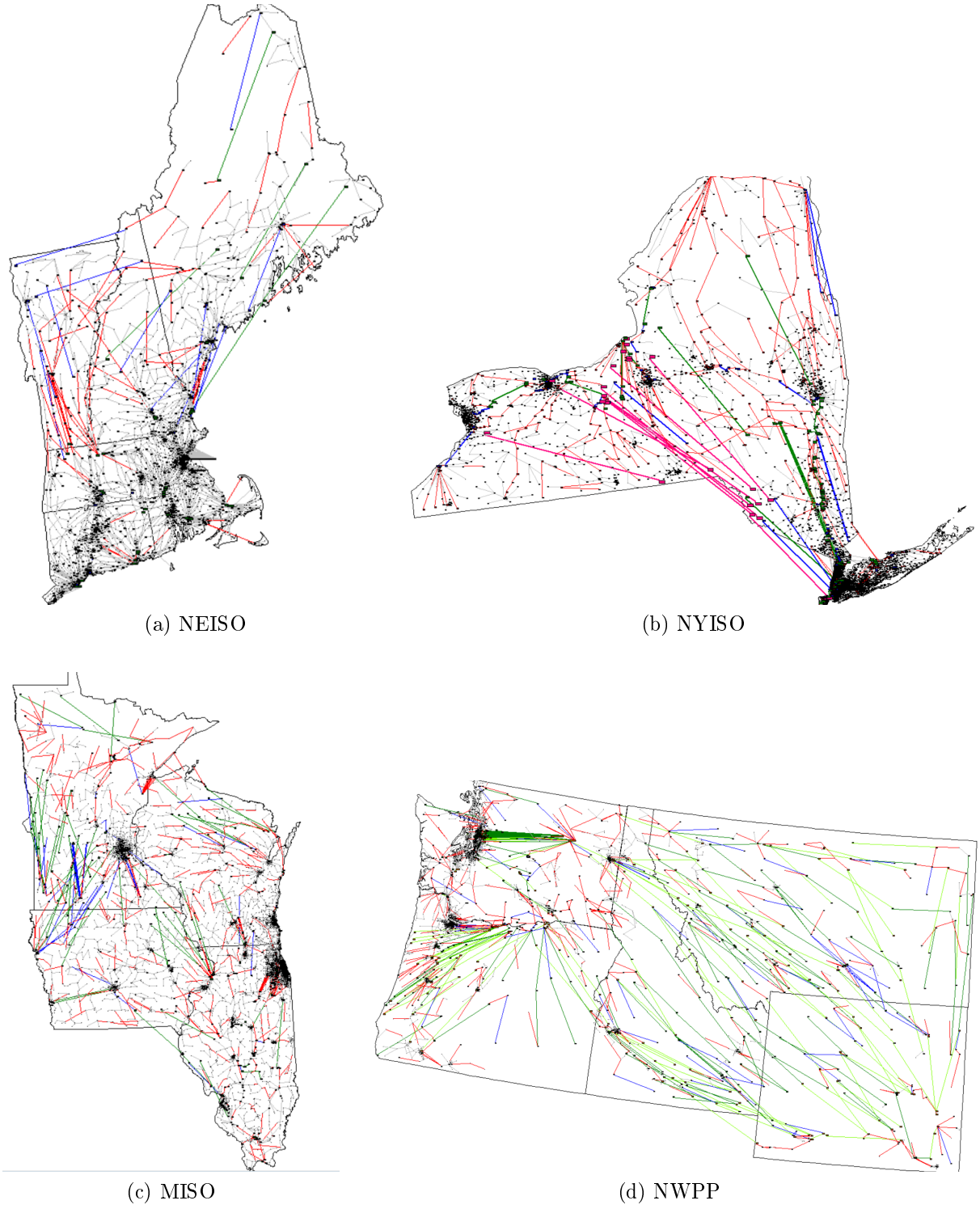


Figure 6.5: One-Line of a synthetic KCL-Compliant network model of the Central USA

6.2 KCL/KVL-Compliant Network Creation Algorithm Results

Figures 6.6 and 6.6 show the one-line diagrams of six synthetic networks created using the KCL/KVL multilevel algorithm. These six networks utilize the same bus-level input data as the KCL-only generated networks in Section 6.1, and follow the same voltage level color scheme.



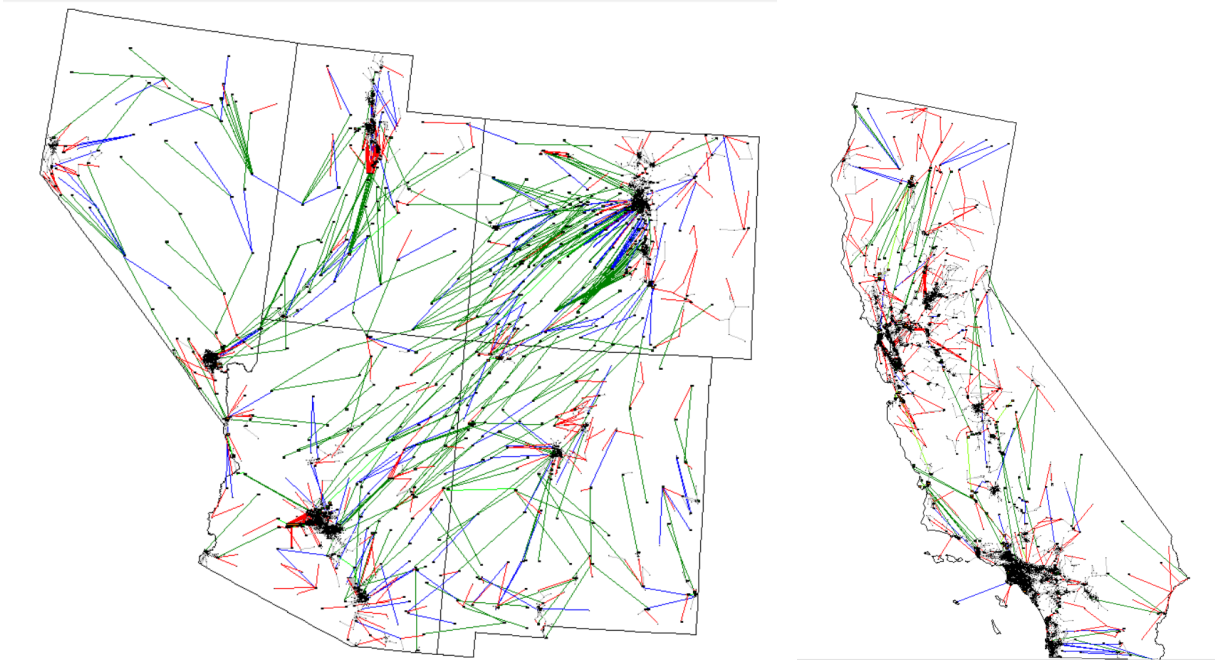
(a) NEISO

(b) NYISO

(c) MISO

(d) NWPP

Figure 6.6: One-Line Diagrams of KCL/KVL-compliant synthetic power system networks



(a) AZNM (b) CAISO

Figure 6.7: One-Line Diagrams of KCL-compliant synthetic power system networks cont.

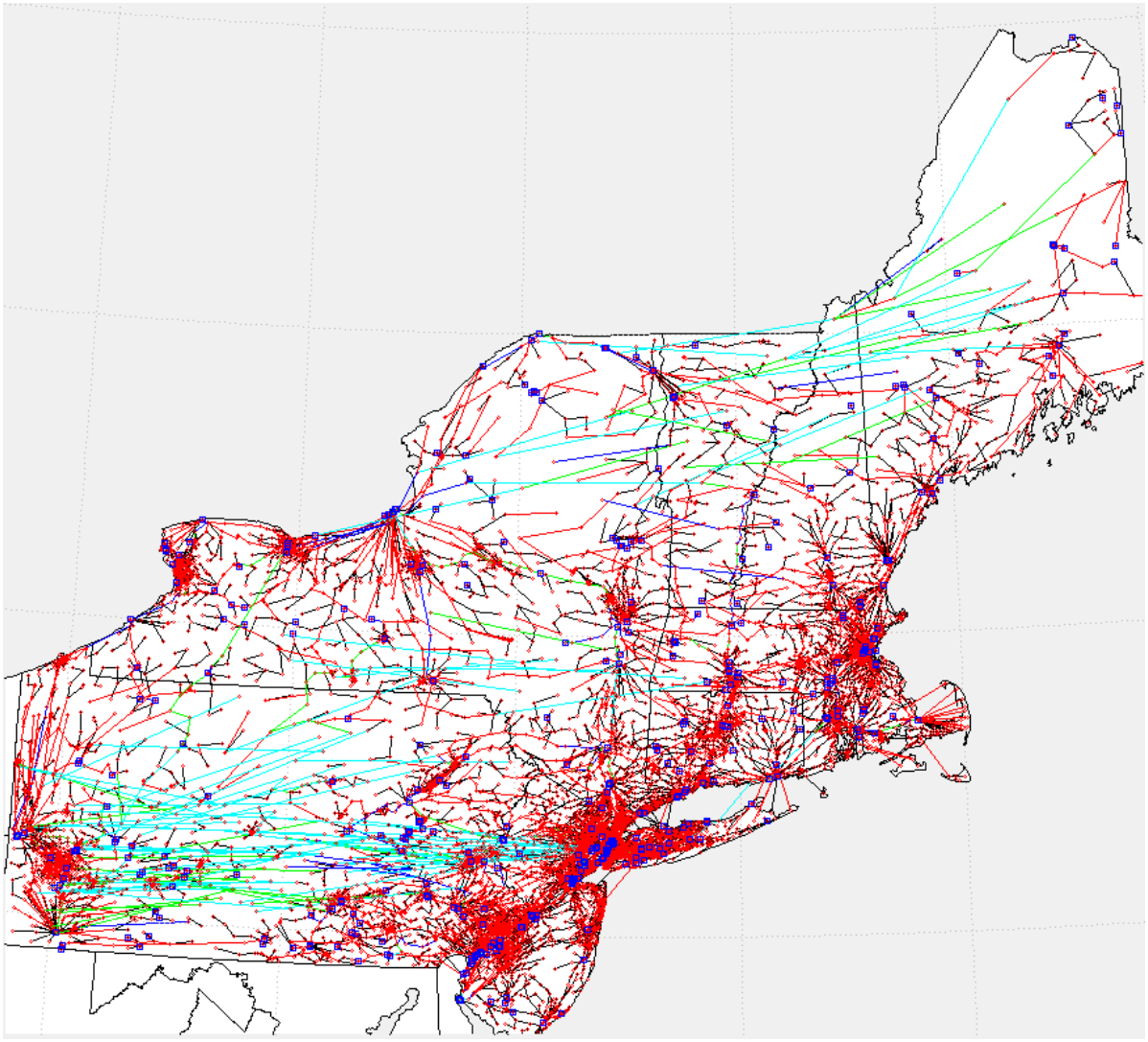


Figure 6.8: One-Line of a KCL/KVL-Compliant synthetic Northeastern USA model

6.2.1 KCL/KVL-Compliant Large Interconnect-Scale Networks

Figures 6.8-6.10 show one-line of the largest synthetic networks created is this dissertation work, ranging from the 1200 substation synthetic PJM model to the 16300 substation synthetic WECC model. These one-line diagrams illustrate that the KCL/KVL-compliant multilevel algorithm can construct large interconnect scale synthetic power system networks.

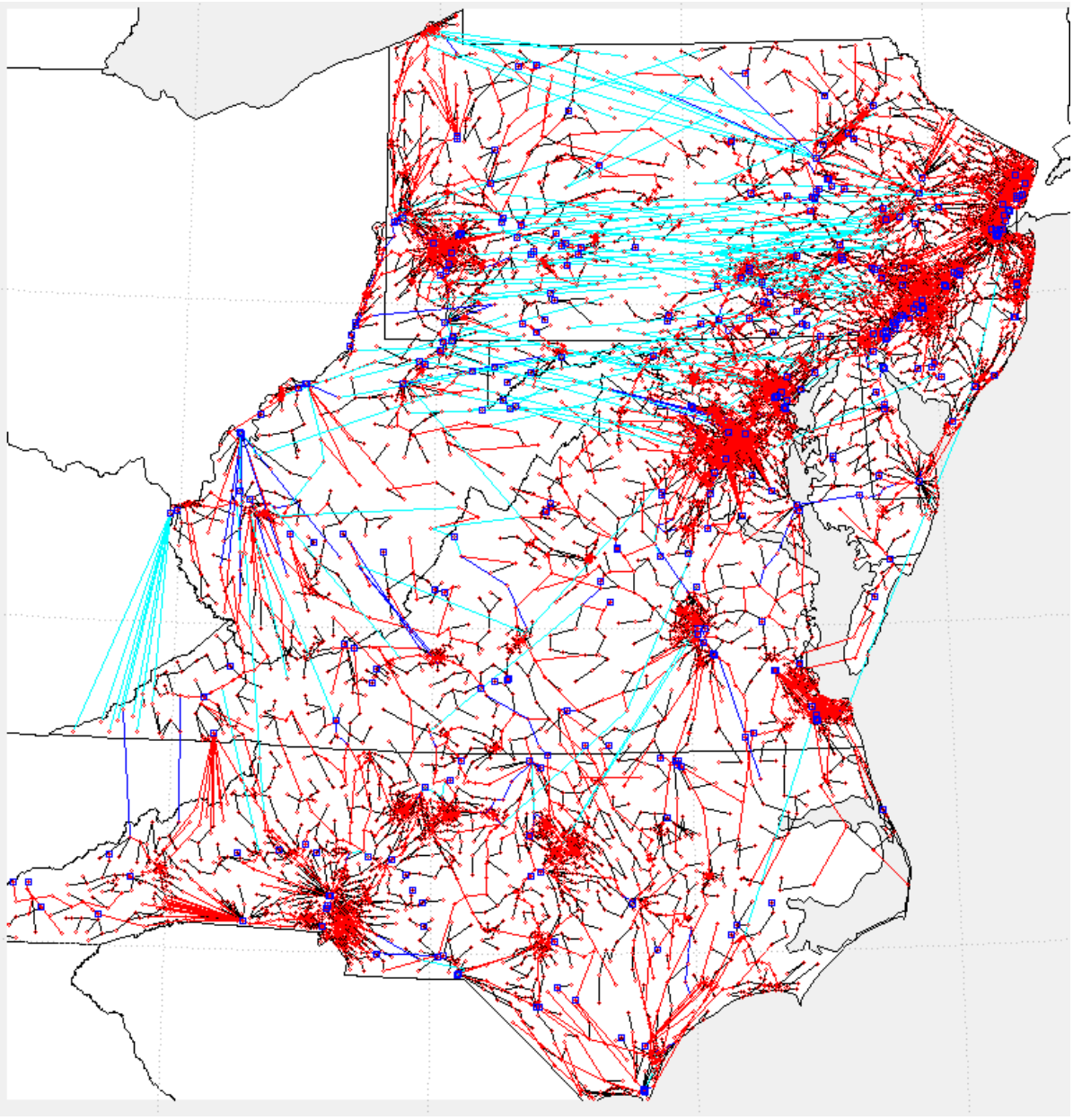


Figure 6.9: One-Line of a KCL/KVL-Compliant synthetic PJM model

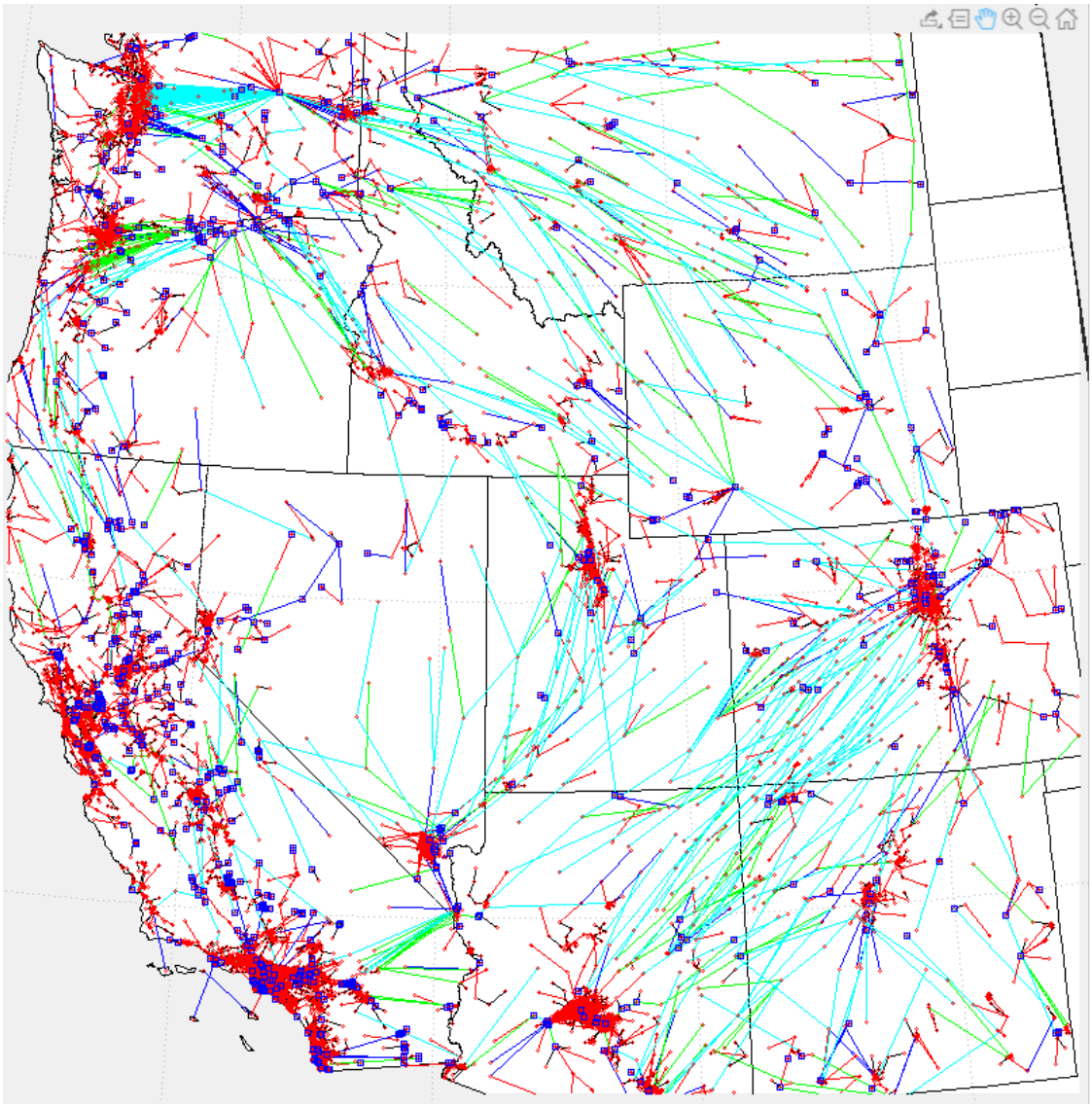


Figure 6.10: One-Line of a KCL/KVL-Compliant synthetic Western Interconnect

6.2.2 Limitations of the KCL/KVL-compliant Algorithm

There are two primary limitations that can be observed in these empirical results of the KCL/KVL algorithm. The first is the high node degree of some generator substations. This characteristic in the synthetic networks will be referred to as the "starburst" issue, because the large number of lines leaving a generator substation sometimes resembles a "starburst". Strong empirical evidence suggests that this issue is caused by a lack of sufficient variety in the input data of the generator dispatch scenarios. Computational experience suggest it can be largely resolved by varying the set of net injection scenarios such that no generators are dispatched for more than three quarters of the total scenarios. This roughly mimics utility practice, since existing transmission systems incorporate generator N-1 contingencies, as well as account for generators being out on schedule maintenance. As detailed in Section 5.2.3, the network creation algorithms presented in this dissertation used a "union" of the generator N-1 contingencies and outage scenarios to avoid having hundreds of the net injection scenarios accounting for individual generator outages and greatly increasing the computational time of the algorithm. However each of the net injection scenarios can be run in parallel, so a compute cluster with a large number of nodes and or compute cores can be utilized to substantially reduce the computational time needed.

The second primary issue is the unrealistically high number of transmission lines routed roughly in parallel in corridors that are geographically close. This issue is visible in the synthetic Pacific Northwest area in Figure 6.6d. Notice that a several of the

buses in Washington and Oregon have a large number of transmission lines leaving a substation and terminating at buses that are in close proximity. Also, the problem is noticeable in the Desert Southwest area in Figure 6.7a, where there are a large number of transmission lines connected to dense urban areas and traveling in adjacent or nearby transmission corridors.

The author hypothesizes that this problem arises from a lack of fixed cost modeling in the objective functions in the inner loops of the multilevel algorithms. The nature of the decision variables for both the KCL-only and KCL/KVL multilevel algorithms are such that it is difficult to directly model the overhead capital cost of procuring a right-of-way for a transmission line, without introducing binary decision variables. Thus, multiple transmission lines can be routed in nearby corridors since the objective functions do not penalize such an assignment. This problem could be addressed using a post-processing step that would identify transmission lines that are routed in nearby corridors and combine them into a few multi-circuited transmission lines. Then short transmission lines to be built at either end of the new corridor to ensure that all the buses are properly connected.

However, for multiple transmission lines routed in very close proximity, this represents multi-circuited transmission lines that follow the same corridor, but diverge at the end and connect to adjacent or nearby buses. This can be observed in Figures 6.6d and 6.7a.

6.3 Comparison of Synthetic and Real Networks

To visual inspection, the sets of one-line diagrams in Figures 6.1-6.7 appear to be fairly realistic power systems. However, one might have concerns that the abnormalities discussed above in Section 6.2.2 could render the KCL/KVL networks less realistic than their KCL-only counterparts. Thus, it is valuable to quantify the degree to which the synthetic power system networks match select features of their real network counterparts on the same geographic footprint. To this end the metrics described in section 2.5 are computed and discussed below.

6.4 Electrical Performance Metrics

The following sections compare the electrical performance of the KCL-only networks to the corresponding networks generated using the KCL/KVL multilevel algorithm. The hypothesis is that because the KCL/KVL algorithm models DC power flow directly in the inner loop, the resulting networks will electrically perform better than networks created using an inner loop that only included the KCl constraints. Thus, three different networks are utilized for the comparison: the networks created after the termination of the KCL-only multilevel algorithm (Section 4.4), the networks modified using a sequential DC power flow (Section 7.1), and the networks created using the KCL/KVL multilevel algorithm (Section 5.3). Impedances were assigned to the first set of networks utilizing the method described in Section 7.1.1 in order to compute a DC power flow.

Two sets of net injection scenarios are utilized for each of the three sets of networks: the injection scenarios utilized to create the KCL-only networks and the net injection scenarios used for the KCL/KVL algorithms. In other words, each network is tested using the net injection scenarios used to create that network and also cross-tested with net injection scenarios utilized to create the "competing" networks (i.e. KCL to KCL/KVL and vice versa).

The three sets of networks utilized are labeled as follows: the networks created after the termination of the KCL-only multilevel algorithm ("KCL-Only"), the KCL-only networks modified using a sequential DC power flow ("KCL-Only DC"), and the networks created using the KCL/KVL multilevel algorithm ("KCL/KVL").

A variety of performance metrics are computed and explained in the respective sections below. Since the DC power flow approximation assumes that transmission lines and transformers are lossless (i.e. $R = 0$), approximate per unit real power "losses" are computed by multiplying the square of the current through each transmission line by its associated resistance, then converted to actual MW values.

For the larger networks comprised of two or more geographic footprints (e.g. WECC), two different generator dispatch algorithms were utilized, and these are respectively termed the "area" dispatch and the "network" dispatch. As described in section 5.2.3, the area dispatch methodology treats each synthetic area in the larger network as a separate balancing area and requires the generators in each area to be dispatched to meet the total load in that area. The "network" dispatch treats the entire super area as

a single area and dispatch as all generators without the requirement for meeting smaller area total load.

Since multiple scenarios are utilized ranging across a variety of loading levels, the total load column represents the average of all of the load scenarios utilized.

6.4.1 DC Power Flow

To quantify the electrical performance of the three sets of networks described above, a DC power flow is run utilizing both the net injection sets used to construct the KCL-only and the KCL/KVL-compliant networks. The following quantities are computed for each of the DC power flow solutions for each of the net injection scenarios: line losses (as described above), number of branches overloaded, and the % of the MVA rating utilized for each branch. The average, or standard deviation, of the quantities for each set of net injection scenarios is computed and presented in Table 6.1.

The "Branch % Loaded" column is the mean of the MVA limit utilization percentage for all lines that are not overloaded. The "% Branches Overloaded" column divides the number of overloaded branches by the total number of branches in the network. The "Mean of the Overload %" column is the average of the percent utilized over the MVA limit for each overloaded branch, and the "Std of the Overload %" standard deviation of the set of data.

There are several trends to be noticed in the data in table 6.1. For the majority of the networks, the KCL/KVL networks perform better electrically across the majority

Region	Algorithm	Generator Dispatch	Total Load (GW)	DC PF Loss %	Branch % Loaded	Over-loaded Branches	% Branches Overloaded	Mean of the Overload %	Std of the Overload %
AZNM	KCL/KVL		27.76	6.13%	26.5%	2	0.03%	9.1%	0.6%
AZNM	KCL-Only		30.63	5.13%	25.0%	159	2.66%	25.5%	26.6%
AZNM	KCL-Only DC		30.63	11.67%	22.9%	168	2.81%	30.6%	33.4%
CAISO	KCL/KVL		29.18	2.55%	23.7%	1	0.01%	2.7%	0.1%
CAISO	KCL-Only		29.18	2.92%	25.0%	331	3.13%	25.6%	25.2%
CAISO	KCL-Only DC		29.18	4.22%	22.7%	163	1.54%	23.7%	23.3%
CUSA	KCL/KVL	Network	104.31	6.78%	22.2%	46	0.19%	15.1%	15.1%
CUSA	KCL/KVL	Area	122.04	4.63%	25.0%	10	0.04%	29.5%	40.0%
CUSA	KCL-Only	Area	155.94	2.09%	19.0%	131	0.53%	26.0%	31.2%
FRCC	KCL/KVL		49.24	3.37%	24.6%	2	0.02%	0.6%	0.2%
FRCC	KCL-Only		51.37	5.19%	28.2%	488	5.05%	31.2%	30.8%
FRCC	KCL-Only DC		51.37	8.86%	28.5%	389	4.02%	45.3%	51.0%
MISO	KCL/KVL		36.02	3.60%	26.2%	2	0.02%	11.3%	2.5%
MISO	KCL-Only		36.02	3.75%	27.5%	249	2.75%	22.1%	21.0%
MISO	KCL-Only DC		36.02	5.69%	21.5%	146	1.61%	22.7%	21.3%
NEISO	KCL/KVL		15.86	3.26%	25.5%	0	0.01%	0.0%	0.0%
NEISO	KCL-Only		15.86	4.15%	25.3%	209	4.98%	40.1%	38.9%
NEISO	KCL-Only DC		15.86	4.04%	19.1%	65	1.56%	35.8%	37.3%
NNEAST	KCL/KVL		68.03	2.90%	26.1%	1	0.01%	12.6%	1.5%
NWPP	KCL/KVL		24.79	4.16%	25.9%	1	0.02%	0.3%	0.1%
NWPP	KCL-Only		30.63	5.13%	25.0%	159	2.66%	25.5%	26.6%
NWPP	KCL-Only DC		30.63	11.67%	22.9%	168	2.81%	30.6%	33.4%
NYISO	KCL/KVL		20.42	2.74%	30.8%	4	0.07%	22.7%	21.4%
NYISO	KCL-Only		20.40	5.03%	29.1%	402	6.54%	32.3%	32.9%
NYISO	KCL-Only DC		20.34	8.93%	29.0%	498	8.10%	45.7%	51.6%
PJM	KCL/KVL	Area	77.41	3.28%	29.5%	1	0.01%	9.4%	6.5%
PJM	KCL/KVL	Network	82.37	5.13%	29.3%	45	0.30%	44.3%	47.0%
RFCE	KCL/KVL		31.31	3.23%	28.0%	4	0.06%	14.9%	11.2%
RFCE	KCL-Only		31.31	4.46%	32.0%	747	10.72%	33.5%	31.1%
RFCE	KCL-Only DC		31.31	5.54%	26.7%	432	6.20%	38.8%	40.2%
RFCW	KCL/KVL		47.76	4.13%	28.2%	6	0.06%	30.6%	33.2%
RFCW	KCL-Only		47.76	5.81%	31.7%	1059	10.99%	47.2%	43.2%
RFCW	KCL-Only DC		47.76	6.83%	21.8%	505	5.24%	62.8%	69.2%
SPP	KCL/KVL		68.14	4.57%	26.0%	6	0.05%	11.0%	6.0%
SPP	KCL-Only		68.14	12.09%	34.9%	1851	16.87%	51.5%	54.2%
SPP	KCL-Only DC		68.14	19.64%	36.0%	1747	15.92%	57.6%	62.1%
SRMV	KCL/KVL		24.67	4.91%	24.6%	3	0.07%	13.36%	4.79%
SRMV	KCL-Only		24.67	6.18%	33.7%	258	6.04%	22.32%	6.52%
SRMV	KCL-Only DC		24.67	9.18%	30.2%	245	5.73%	60.63%	19.98%
SRTV	KCL/KVL		34.68	4.26%	27.6%	22	0.38%	12.41%	62.63%
SRTV	KCL-Only		34.68	9.59%	32.7%	789	13.11%	45.84%	17.30%
SRTV	KCL-Only DC		34.68	9.56%	24.9%	246	4.08%	33.46%	14.63%
SRVC	KCL/KVL		45.09	3.30%	27.3%	0	0.00%	1.8%	0.3%
SRVC	KCL-Only		45.09	3.26%	29.3%	306	3.61%	24.2%	23.1%
SRVC	KCL-Only DC		45.09	2.75%	19.6%	32	0.38%	14.5%	13.7%
WECC	KCL/KVL	Network	82.36	4.53%	23.5%	5	0.02%	23.0%	9.8%
WECC	KCL/KVL	Area	81.74	3.56%	24.0%	10	0.05%	9.3%	3.9%
WECC	KCL-Only	Area	91.55	1.82%	15.9%	47	0.20%	24.7%	26.0%

Table 6.1: DC power flow metrics comparing synthetic networks generated using different algorithms

Quantity	Algorithm	DC PF Loss %	Branch % Loaded	Overloaded Branches	% Branches Overloaded	Mean of the Overload %	Std of the Overload %
Mean	KCL/KVL	4.17%	26.2%	31	0.42%	25.04%	12.04%
	KCL-Only	5.30%	27.7%	484	5.80%	32.70%	10.07%
	KCL-Only DC	8.04%	24.8%	357	4.41%	37.34%	13.97%
Max	KCL/KVL	6.78%	30.8%	271	4.61%	50.94%	42.93%
	KCL-Only	12.09%	34.9%	1851	16.87%	51.47%	22.45%
	KCL-Only DC	19.64%	36.0%	1747	15.92%	63.10%	23.42%
Min	KCL/KVL	2.55%	22.2%	1	0.01%	5.77%	0.00%
	KCL-Only	2.09%	19.0%	56	0.53%	19.74%	3.85%
	KCL-Only DC	2.75%	19.1%	11	0.27%	9.37%	4.65%
Std	KCL/KVL	1.19%	2.1%	57	0.98%	14.21%	10.24%
	KCL-Only	2.30%	4.5%	443	4.38%	10.06%	5.66%
	KCL-Only DC	4.62%	4.7%	433	4.04%	16.40%	6.82%

Table 6.2: Summary of DC power flow metrics comparing synthetic networks generated using different algorithms

of the metrics, i.e. lower percent losses, fewer overloaded branches and a lower mean overload on the lines that are overloaded. Also, it is not surprising that the networks modified using the sequential DC power flow to resize the transmission lines performed better than the networks before modification. Finally, the data in Table 6.1 confirms the hypothesis that modeling the DC power flow directly in the inner loop of the multilevel algorithm results in a network that performs better electrically than modifying networks created using the KCL-only multilevel algorithm.

Finally, it should be noted that although the KCL/KVL networks perform better electrically than the KCL-only networks, both are within the normal percent losses on the real transmission system, which typically range from 3% to 7%.

6.4.2 DC Optimal Power Flow

After the analysis using DC power flow is complete, further analysis is performed using a DC optimal power flow (DCOPF). Since the DC power flow calculates the

Region	Algorithm	Generator Dispatch	Total Load (GW)	DCOPF Loss %	% Cases feasible	% MVA Limit Relaxation	Std % MVA Limit Relaxation	Branch % Loaded OPF
AZNM	KCL/KVL		27.76	4.25%	98.35%	12.36%	2.99%	23.77%
AZNM	KCL-Only		30.63	2.83%	46.51%	45.91%	39.45%	21.60%
AZNM	KCL-Only DC		30.63	5.56%	58.14%	41.44%	30.50%	20.28%
CAISO	KCL/KVL		29.18	2.10%	99.79%	23.49%	1.07%	20.32%
CAISO	KCL-Only		29.18	1.53%	0.00%	23.46%	19.53%	17.59%
CAISO	KCL-Only DC		29.18	2.69%	0.00%	60.71%	41.99%	16.32%
CUSA	KCL/KVL	Network	104.31	4.46%	57.00%	51.52%	91.04%	18.52%
CUSA	KCL/KVL	Area	122.04	4.41%	56.17%	71.35%	63.92%	23.50%
CUSA	KCL-Only	Area	155.94	1.24%	0.00%	32.66%	33.61%	16.38%
FRCC	KCL/KVL		49.24	2.38%	65.23%	6.83%	4.52%	23.34%
FRCC	KCL-Only		51.37	1.99%	0.00%	64.61%	57.78%	22.20%
FRCC	KCL-Only DC		51.37	3.30%	0.00%	62.46%	68.14%	22.76%
MISO	KCL/KVL		36.02	2.62%	80.04%	24.01%	9.86%	23.13%
MISO	KCL-Only		36.02	2.60%	0.00%	36.98%	33.46%	22.73%
MISO	KCL-Only DC		36.02	2.90%	0.00%	40.50%	22.76%	15.79%
NEISO	KCL/KVL		15.86	2.94%	99.59%	34.18%	2.19%	24.66%
NEISO	KCL-Only		15.86	1.53%	0.00%	50.56%	28.58%	20.77%
NEISO	KCL-Only DC		15.86	1.59%	0.00%	32.86%	14.47%	14.33%
NNEAST	KCL/KVL		68.03	1.72%	96.28%	5.69%	1.76%	21.93%
NWPP	KCL/KVL		24.79	4.33%	99.59%	18.18%	1.40%	26.22%
NWPP	KCL-Only		30.63	2.83%	46.51%	45.91%	39.45%	21.60%
NWPP	KCL-Only DC		30.63	5.56%	58.14%	41.44%	30.50%	20.28%
NYISO	KCL/KVL		20.42	2.05%	99.59%	229.12%	19.75%	26.21%
NYISO	KCL-Only		20.40	1.86%	0.00%	24.29%	16.31%	21.67%
NYISO	KCL-Only DC		20.34	3.01%	0.00%	18.51%	9.71%	23.72%
PJM	KCL/KVL	Area	77.41	2.73%	93.18%	8.91%	2.87%	27.02%
PJM	KCL/KVL	Network	82.37	2.25%	86.16%	4.14%	1.73%	25.82%
RFCE	KCL/KVL		31.31	1.91%	96.91%	12.75%	4.58%	22.74%
RFCE	KCL-Only		31.31	1.58%	0.00%	55.11%	45.49%	25.38%
RFCE	KCL-Only DC		31.31	1.74%	0.00%	96.50%	64.78%	19.40%
RFCW	KCL/KVL		47.76	3.00%	97.12%	27.98%	11.19%	24.76%
RFCW	KCL-Only		47.76	2.20%	0.00%	54.24%	39.64%	25.21%
RFCW	KCL-Only DC		47.76	1.84%	0.00%	33.98%	26.95%	14.88%
SPP	KCL/KVL		68.14	3.21%	62.55%	17.79%	10.20%	22.60%
SPP	KCL-Only		68.14	2.37%	0.00%	38.67%	40.88%	23.50%
SPP	KCL-Only DC		68.14	5.81%	0.00%	69.74%	42.02%	26.60%
SRMV	KCL/KVL		24.67	2.52%	99.79%	64.29%	2.92%	21.10%
SRMV	KCL-Only		24.67	1.46%	0.00%	16.97%	10.21%	19.22%
SRMV	KCL-Only DC		24.67	3.15%	0.00%	82.95%	64.11%	20.79%
SRTV	KCL/KVL		34.77	2.76%	96.71%	110.81%	289.61%	24.17%
SRTV	KCL-Only		34.77	3.52%	0.21%	106.73%	81.03%	26.08%
SRTV	KCL-Only DC		34.77	3.41%	0.82%	101.29%	85.76%	17.54%
SRVC	KCL/KVL		45.09	3.07%	95.27%	12.29%	3.64%	26.26%
SRVC	KCL-Only		45.09	1.70%	0.00%	24.30%	16.96%	23.77%
SRVC	KCL-Only DC		45.09	1.54%	0.00%	13.63%	2.55%	15.84%
WECC	KCL/KVL	Network	82.36	2.97%	89.92%	6.41%	5.51%	19.86%
WECC	KCL/KVL	Area	81.74	4.46%	57.00%	61.04%	46.82%	22.59%
WECC	KCL-Only	Area	91.55	1.41%	54.78%	20.75%	23.97%	14.50%

Table 6.3: DC optimal power flow metrics comparing synthetic networks generated using different algorithms

Quantity	Algorithm	DCOPF Loss %	% Cases feasible	% MVA Limit Relaxation	Std % MVA Limit Relaxation	Branch % Loaded OPF
Mean	KCL/KVL	3.07%	81.76%	40.82%	37.26%	23.52%
	KCL-Only	2.04%	44.14%	43.76%	36.54%	21.75%
	KCL-Only DC	3.13%	53.31%	51.78%	42.52%	19.05%
Max	KCL/KVL	4.46%	99.79%	229.12%	205.69%	27.02%
	KCL-Only	3.51%	89.53%	106.07%	81.71%	26.02%
	KCL-Only DC	6.54%	100.00%	101.20%	86.08%	26.60%
Min	KCL/KVL	1.72%	0.41%	4.14%	0.00%	18.52%
	KCL-Only	1.24%	0.00%	16.19%	0.00%	16.38%
	KCL-Only DC	1.54%	0.82%	0.00%	0.00%	14.33%
Std	KCL/KVL	0.88%	24.06%	49.93%	51.53%	2.28%
	KCL-Only	0.62%	28.92%	22.96%	22.47%	2.77%
	KCL-Only DC	1.46%	34.50%	30.44%	24.60%	3.59%

Table 6.4: Summary of DC optimal power flow metrics comparing synthetic networks generated using different algorithms

power flowing on each transmission branch for a fixed set of generator injections and load withdrawals, the resulting metrics from Section 6.4.1 only capture a single "snapshot" of each power system network for each load scenario. Thus it is beneficial to utilize a DC optimal power flow to quantify how well the networks perform when the generators are allowed to deviate from their initial setpoints in the DC power flow.

The formulation for the DCOPF utilized in this dissertation is shown in (6.1). The a, b, c terms in (6.1a) are the coefficients in the generator quadratic cost function, and P_g is the real power output of a generator. The optimal solution must satisfy the DC power flow equations (6.1b) subject to the generator limits (6.1c)-(6.1d) and branch thermal limits 6.1f, where P_G and P_D are the generator real power injections and the real power load withdrawals, respectively. Equation 6.1e is included as an intermediate constraint to relate the branch currents p_j to the bus voltage angles θ.

$$\min_{P_g} \quad \sum a + bP_g + cP_g^2 \tag{6.1a}$$

$$\text{s.t.} \quad B_{\text{Bus}}\theta = P_G - P_D, \tag{6.1b}$$

$$P_g \leq P_{g,\max}, \tag{6.1c}$$

$$P_g \geq P_{g,\min}, \tag{6.1d}$$

$$(\theta_i - \theta_j)b_{i,j} = p_j \forall j, \tag{6.1e}$$

$$|p_j| \leq \text{line thermal limit } \forall j \tag{6.1f}$$

Similar to the DC power flow, both of the injection sets utilized to create the KCL-only and the KCL/KVL networks are utilized for the DCOPF analysis. For each scenario, a DC optimal power flow is run using the load scenarios associated with each base case network. However, sometimes the sum of the load demands is below the sum of the generator lower power limits, thus some generator units need to be decommitted in order to have a feasible DCOPF solution. Since each net injection vector for the KCl KVL compliant multilevel algorithm was created using a heuristic unit commitment algorithm (Section 5.2.3), this unit commitment information was extracted from each net injection vector and utilized in the DC optimal power flow.

Various statistics and metrics were computed from the OPF results and reported in table 6.3. The total number of scenarios with a feasible solution to the base case without modifications to the MVA line limits was counted and reported as a percentage in the "%

cases feasible" column. For cases and net injection scenarios without a feasible solution, the MVA limits of the transmission lines were uniformly scaled up the by the maximum overload percentage from the DC power flow computation for that net injection scenario. Then, the line limits were sequentially reduced by a few %, and the DCOPF was recalculated after each successive tightening of the line limits. At the boundary for which the case was no longer feasible, the % difference between the "relaxed" line limits and the base case MVA line limits was computed and the average of all such relaxations for initially non-feasible cases was recorded in the "% MVA limit relaxation" column. The standard deviation of this set of percent relaxations was computed and recorded in the "Std % MVA Limit Relaxation" column.

The approximate losses were computed by multiplying the square of the per-unit current flowing through each branch by the per unit resistance and converting to actual MVA. The cases with the "tightest" relaxation of the MVA line limits were utilized to compute the approximate losses, or the base case without relaxations if it was feasible. The average of the losses across all cases was computed and recorded in the "DCOPF Loss %" column. Finally, the average of the percent utilization of all branches across all scenarios was calculated and recorded in the branch percent loaded OPF column.

As one can see from the results in Table 6.3, most of the KCL/KVL networks outperform both of the KCL-only networks across most of the metrics, as our hypothesis suggested. The majority of the KCL/KVL cases are feasible for the base case, whereas the majority of KCL-only cases requires substantial relaxation of the MVA limits to

achieve a feasible DCOPF solution. However, there are a couple of exceptions to this rule. The KCL-only AZNM networks perform better than the KCL/KVL cases for most of the metrics. Additionally, the KCL-only cases without modification by the sequential DC power flow have a lower average loss percent for most of the geographic footprints. This may be explained by the fact that these networks tend to have a much higher percent relaxation of the MVA limits, thus allowing the DCOPF to find a solution with lower losses that would otherwise be infeasible with tighter line limits. Additionally, the "KCL-only DC" networks were modified using a sequential DC power flow to correct the majority of overloads, However this modification was done only using the peak load scenario, thus skewing these cases to perform better for the peak load scenario but resulting in poorer performance for other scenarios, as seen by the results in Table 6.3.

6.5 Graph Metrics

Two metrics are used to quantitatively compare the differences in the graph topologies of the one-line diagrams presented above: node degree distribution and cycle basis. The node degree distribution is a measure of how densely edges (transmission lines) connect to nodes (buses), and the cycle basis is an indicator of how "meshed" the grid is. Both of these metrics are discussed in more detail in Section 2.5. Quantitative statistics are computed and associated plots constructed in the two sections below. These metrics assess the similarity of the KCL-Only and KCL/KVL networks to each other as well as the real networks on corresponding geographic footprints (either the Eastern or Western

Interconnect).

6.5.1 Line Mileage Statistics

Table 6.5 shows the mean and standard deviations of the branch and bus metrics from all synthetic networks compared to the real Eastern and Western interconnect models. Statistics for each individual region is included in Appendix C. The KCL-Only and KCL-KVL synthetic networks are compared side-by-side for each geographic area. The branch metrics include the quantiles (10%, 50%, 90%), mean, and standard deviation of the lengths of the transmission lines at each voltage in the network. The bus statistics include the total number and percentage of buses at each voltage level Notice that many of the quantities fairly well aligned between the synthetic and real networks. The primary disagreement between the synthetic and real networks is in the percentage of buses at 69kV and 138kV, however this can be remedied by modifying the bus voltage levels as described in Section 9.1.1.

6.5.2 Graph Node Degree Metrics

Another metric that facilitates comparison of graph structure in the synthetic and real power system networks is based on the node degree distributions in the networks. Figure 6.11 shows the log-linear plots of the normalized node degree distributions for six synthetic power system networks, compared to the real networks on the same geographic footprint.

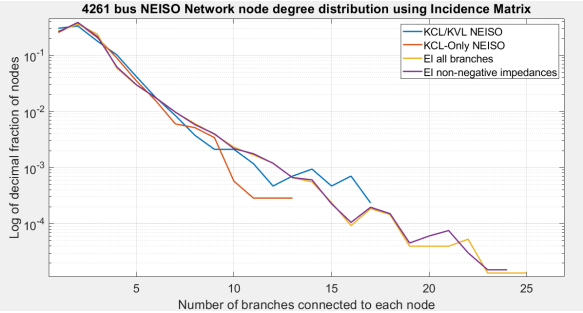
Quantity or Network	Metrics (miles)	KCL-Only Networks (kV)						KCL/KVL Networks (kV)						
		69	**138**	**230**	**345**	**500**	**765**	**12.9**	**69**	**138**	**230**	**345**	**500**	**765**
Synthetic Mean	% of Lines	82.1%	13.8%	2.2%	1.5%	1.3%	0.1%		77.4%	15.5%	2.4%	4.1%	2.3%	0.2%
	Mean	3.8	9.0	12.6	22.6	52.7	65.4		4.3	13.6	30.0	66.0	92.5	172.9
	Std	4.8	13.0	19.8	32.6	62.0	18.9		5.0	19.5	43.9	69.4	69.6	66.3
	10%	0.7	0.9	1.0	1.5	3.7	41.4		1.0	1.0	1.0	1.3	14.5	108.5
	50%	1.9	4.0	5.4	10.7	27.4	72.2		2.4	5.4	4.4	52.2	92.6	148.3
	90%	9.5	23.3	34.2	61.7	149.9	82.8		10.2	45.2	92.5	159.9	160.7	249.5
	% of buses	78.5%	15.9%	2.9%	2.0%	1.8%	0.1%		64.4%	17.5%	3.0%	5.7%	3.5%	0.3%
Synthetic Std	% of Lines	10.6%	8.3%	1.3%	1.3%	2.0%	0.0%	9.2%	6.7%	3.4%	1.1%	2.6%	1.6%	0.1%
	Mean	1.1	3.7	3.4	6.3	30.0	45.4		1.2	5.1	15.1	36.0	61.5	66.2
	Std	1.2	3.9	6.1	11.5	25.2	4.5		1.0	3.9	13.5	11.4	27.3	38.6
	10%	0.2	0.2	0.3	0.5	3.3	37.7		0.1	0.1	0.1	0.6	38.3	106.3
	50%	0.8	1.8	1.6	5.9	25.3	46.7		0.9	8.3	6.2	68.9	90.2	75.8
	90%	3.0	12.5	10.9	30.0	74.9	51.8		3.2	11.5	43.6	45.9	82.5	31.2
	% of buses	10.8%	8.2%	1.5%	1.9%	2.4%	0.0%	2.0%	9.8%	3.0%	1.2%	4.0%	2.2%	0.2%
EI Real Network	% of Lines	42.6%	44.6%	8.9%	2.9%	0.7%	0.2%							
	GVA-Miles	12240	45642	42219	68481	48071	74036							
	Total Len	121605	195218	72335	51422	17146	10313							
	Mean	4.7	7.3	13.5	29.1	37.9	88.9							
	Std	5.3	8.5	17.6	30.8	35.3	55.1							
	10%	0.7	0.7	0.7	2.5	1.0	15.5							
	50%	3.2	4.4	7.4	20.3	28.6	82.8							
	90%	10.4	17.1	33.1	65.1	87.7	160.4							
	Num Buses	35276	24336	4761	1368	347	36							
	% of buses	53.2%	36.7%	7.2%	2.1%	0.5%	0.1%							
WECC Real Network	% of Lines	31.9%	47.4%	17.0%	1.2%	2.4%	0.0%							
	GVA-Miles	2682	8241	26278	11260	46653	0							
	Total Len	29408	56759	52451	11879	18491	0							
	Mean	6.0	7.8	19.9	62.2	49.2	0.0							
	Std	7.2	10.9	44.1	53.2	49.7	0.0							
	10%	0.5	0.5	0.9	4.2	1.0	0.0							
	50%	3.5	3.9	10.6	47.6	36.2	0.0							
	90%	14.3	19.3	48.0	136.6	104.9	0.0							
	Num Buses	8308	5693	1899	184	274	0							
	% of buses	50.8%	34.8%	11.6%	1.1%	1.7%	0.0%							

Table 6.5: Summary of branch and bus statistics from all synthetic networks

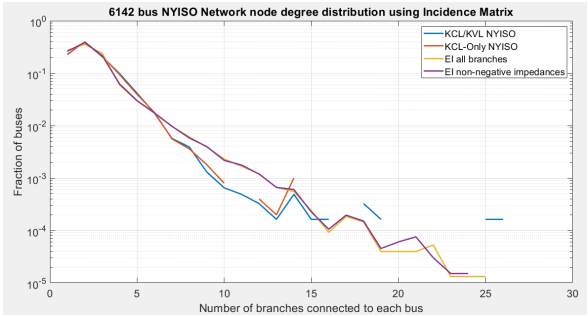
Recall from Section 2.5.2 that two network models for the "real-world" Eastern and Western interconnects were created, one keeping all buses and branches from the largest island, and the other case keeping only branches (and connected buses) with positive reactances in the largest island. These two cases have slightly different node degree distributions, as seen from the plots in Figure 6.11. Thus, some variation between the synthetic and real node degree distributions is to be expected. However notice from figure 2.5.2 that there seems to be excellent agreement between the node degree distributions for most of the six networks and the corresponding real world networks.

To better quantify the differences between the node degree distributions for the synthetic and real cases, affine functions were fit to the node degree distribution plots as shown in Figures 6.12 and 6.13. The percent difference between the synthetic and real networks' areas under the linear curve fits in Figure 6.14 were computed and recorded in Table 6.6 under the "Node Degree % difference" column. The R-Squared values were not included for these linear curve fits Since they were all close to 0.98 to 0.99.

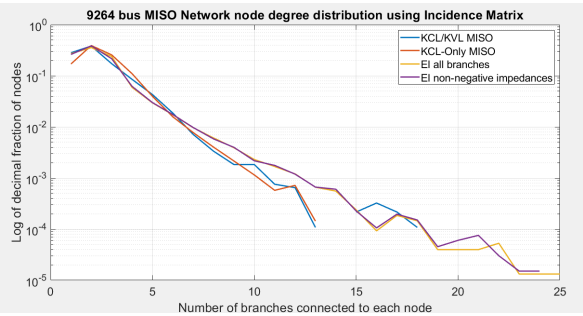
On the average, the KCL/KVL systems have a smaller percent difference compared to the curve fits of the real networks than the percent difference calculated for the KCL-Only networks. This suggests that the KCL/KVL networks have greater similarity in graph structure to the real network, relative to those created using the KCL-Only algorithms.



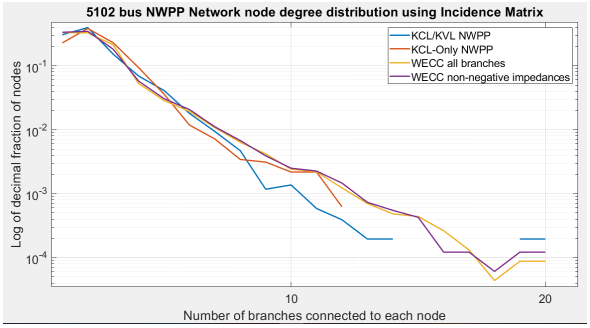
(a) NEISO



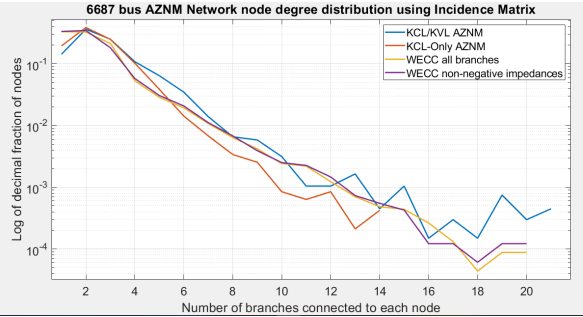
(b) NYISO



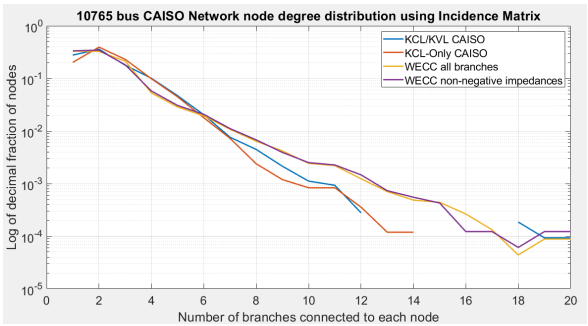
(c) MISO



(d) NWPP

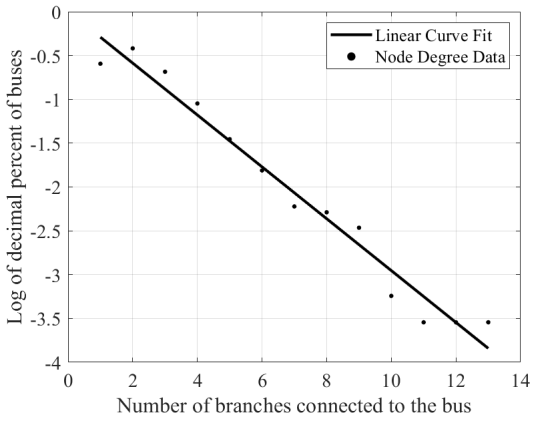


(e) AZNM

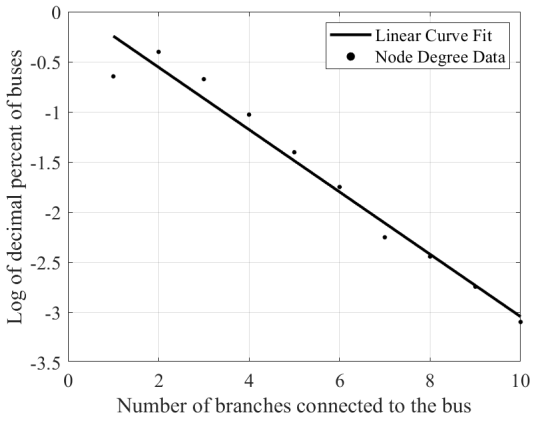


(f) CAISO

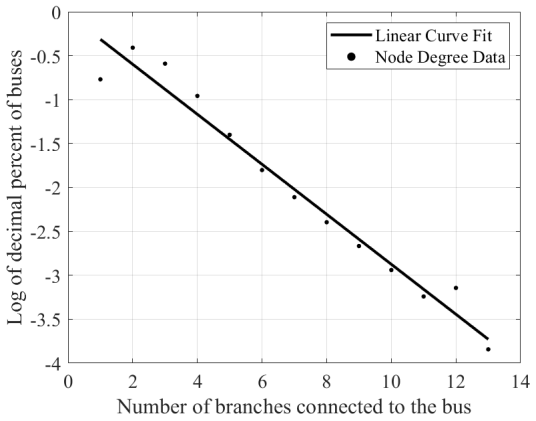
Figure 6.11: Node degree distribution plots for synthetic networks vs real networks



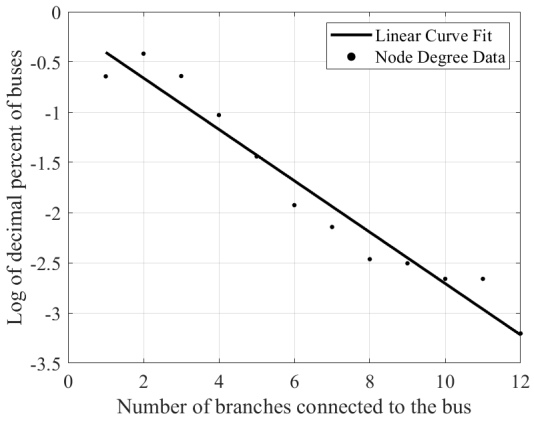
(a) NEISO



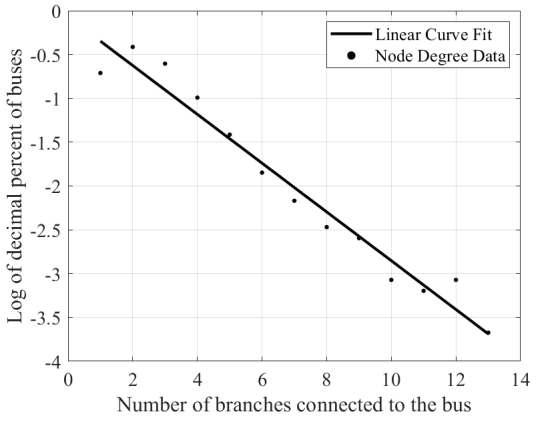
(b) NYISO



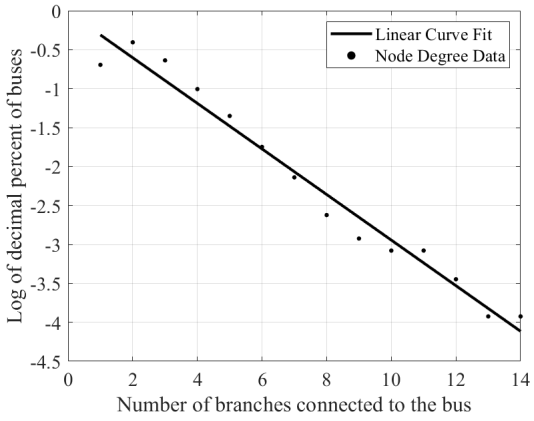
(c) MISO



(d) NWPP

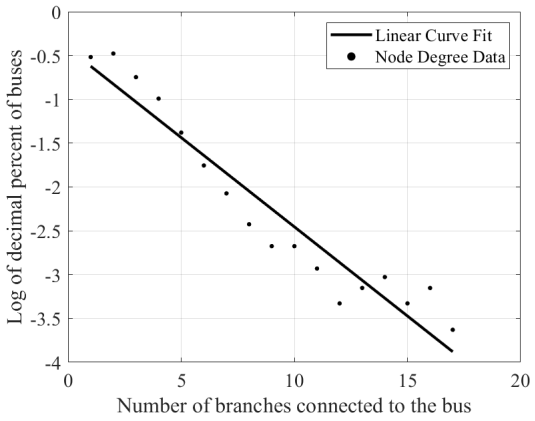


(e) AZNM

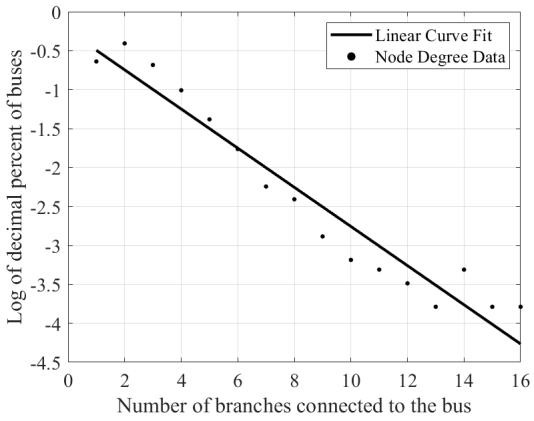


(f) CAISO

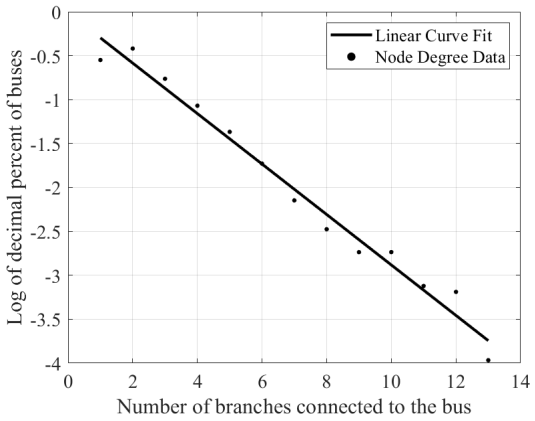
Figure 6.12: Linear fit to the log-linear node degree distribution for the KCL-compliant networks



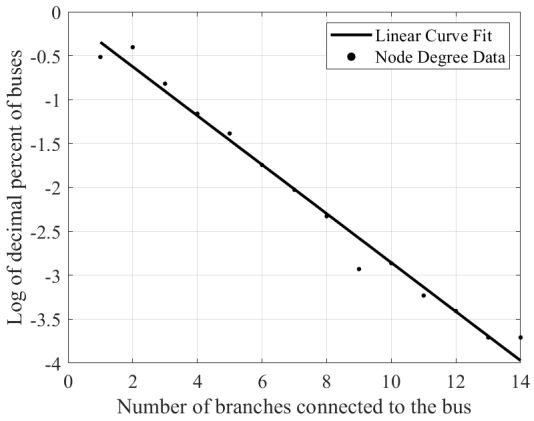
(a) NEISO



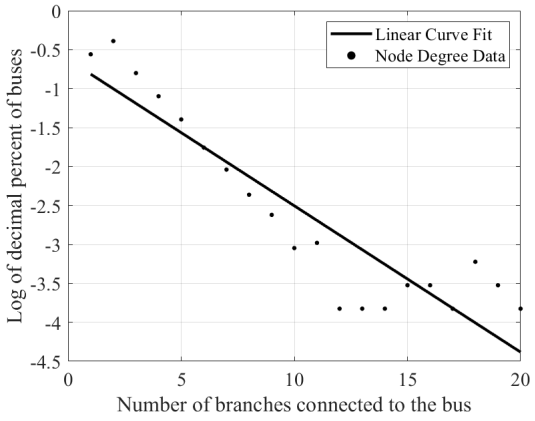
(b) NYISO



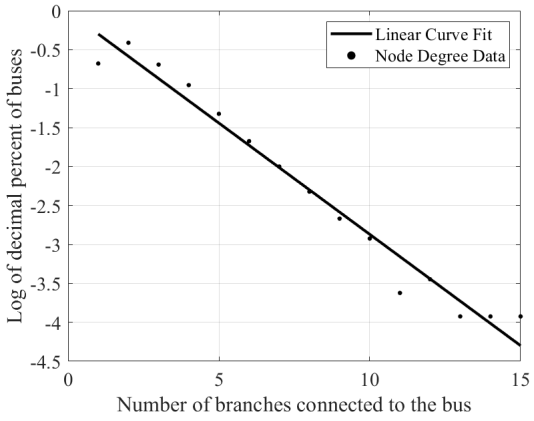
(c) MISO



(d) NWPP

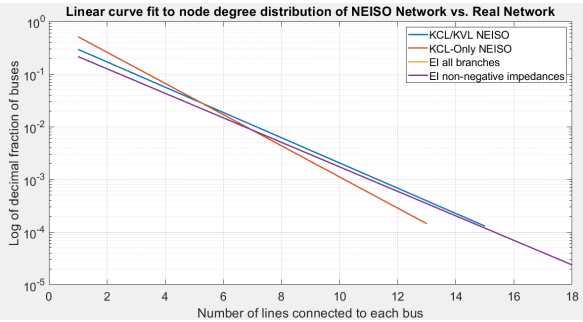


(e) AZNM

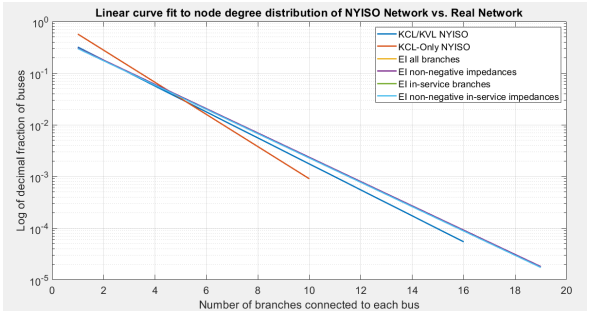


(f) CAISO

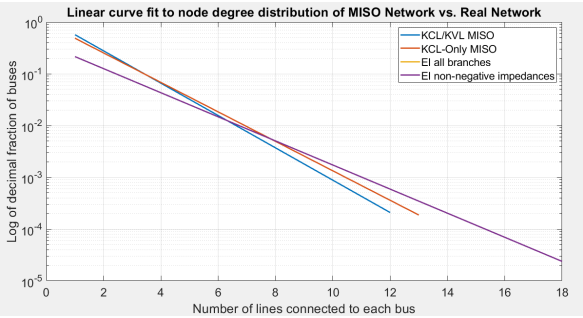
Figure 6.13: Linear fit to the log-linear node degree distribution for the KCL/KVL-compliant networks



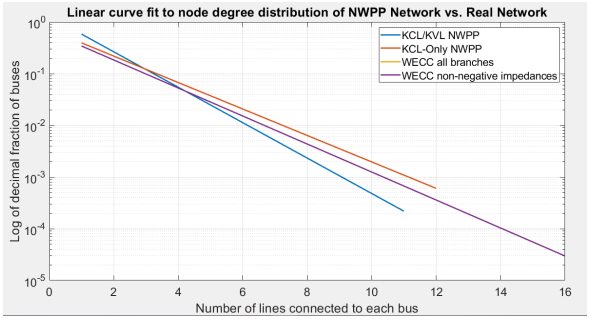
(a) NEISO



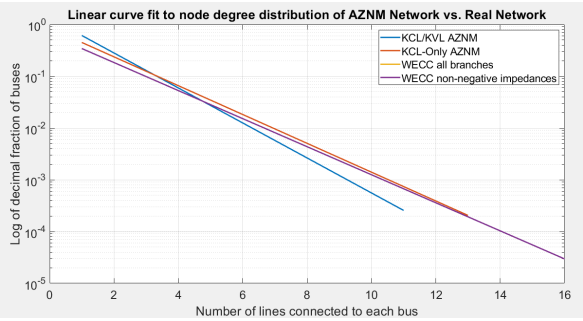
(b) NYISO



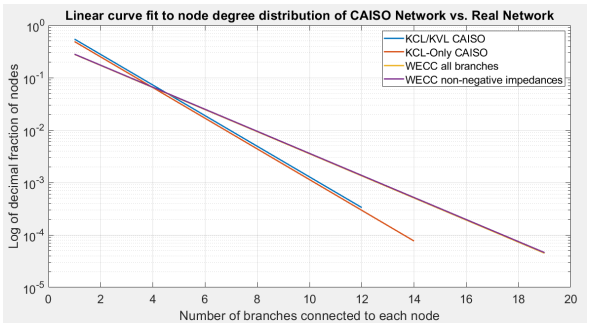
(c) MISO



(d) NWPP



(e) AZNM



(f) CAISO

Figure 6.14: Comparison of the node degree linear fit lines from the synthetic networks with the real networks

Region	Sub-stations	Max Load (GW)	Gen Dispatch	Node Degree Linear Fit			
				% Difference		R-Squared	
				KCL-Only	KCL-KVL	KCL-Only	KCL-KVL
AZNM	4700	43		48.8%	4.3%	0.96	0.84
CAISO	8400	47		39.2%	32.0%	0.97	0.97
MISOSPP	18800	192	Network	9.5%	7.4%	0.97	0.95
MISOSPP	18800	192	Area		19.0%		0.97
FRCC	7500	79		42.7%	25.3%	0.95	0.94
MISO	7000	55		48.8%	48.8%	0.97	0.98
NEISO	3500	25		47.6%	23.9%	0.97	0.92
NEAST	13900	105			23.2%		0.96
NWPP	3200	37		57.9%	40.7%	0.95	0.98
NYISO	5000	32		68.7%	48.5%	0.97	0.97
PJM	11900	118	Network		34.4%	0.00	0.98
PJM	11900	118	Area		1.3%	0.00	0.96
RFCE	5400	48		35.3%	35.0%	0.96	0.96
RFCW	7400	70		25.7%	2.6%	0.92	0.94
SPP	8500	100		49.9%	34.1%	0.95	0.97
SRMV	3300	37		63.5%	41.8%	0.95	0.97
SRTV	4600	55		43.3%	32.3%	0.97	0.99
SRVC	6500	69		50.6%	28.5%	0.96	0.93
WECC	16300	127	Network	19.8%	37.1%	0.98	0.88
WECC	16300	127	Area		27.6%		0.91
WECC	Actual Network					0.981	
EI	Real Network					0.971	
Mean				43.4%	27.4%	0.85	0.95
Max				68.7%	48.8%	0.98	0.99
Min				9.5%	1.3%	0.00	0.84
Std				15.3%	14.0%	0.31	0.04

Table 6.6: Quantitative metrics for network node degree distributions

Chapter 7

Realism Enhancing Heuristics for KCL-Compliant Networks

This chapter and Chapter 8 presents modifications to the KCL-only and KCL/KVL-compliant multi-level algorithms to improve the networks presented in Sections 6.1 and 6.2. First, Section 7.1 presents a method for updating the initial results of the KCL-only multilevel algorithm to be DCOPF feasible, i.e. a network with a solution that satisfies the line limits in a DC optimal power flow. This involves assigning and updating line impedances and associated MVA limits following the method described below. Next, Section 7.2 describes how buses are assigned to each substation and connected by transformers. Finally Section 7.3 describes a methodology for creating large interconnect scale synthetic networks, such as synthetic versions of WECC and the Eastern interconnect.

7.1 Adding or Modifying Branch Parameters

After the initial power system network topology has been established, either using the novel multi-level algorithm described above in Section 4.1, or another existing network creation, parameters need to be assigned to the network branches. Corresponding to item 2 in the general 4-step network construction process (Section 1.2), the per unit impedance values (R, X, B) are assigned and the corresponding line limits are updated, as explained later in this section.

7.1.1 Assigning Initial Transmission Line parameters

The multi-level network creation algorithm presented in Section 4.1 is used to create a synthetic network topology with each transmission line containing length and line limit parameters. After the potential transmission line paths are removed from the linear programming model, the MVA limits of the remaining existing transmission lines are used to inform the assignment of transmission line parameters (R, X, B) consistent with the given limit for each line.

For each transmission line in the synthetic model, the line's MVA capacity was matched to the closest capacity entry in a conductor data table that was created using statistics from the publicly available reports and data. The construction of this table was thoroughly described in Chapter 3, bus is briefly summarized below in Section 7.1.1. To compute the conductor lengths, the node-to-node distances d were multiplied by a factor typically selected equal to 1.4 to account for indirect transmission line path

routing and conductor sag. The phase-to-phase conductor spacing information (GMD) was assigned randomly using the empirically inferred GMD distribution from the FERC Form 1 [111] (see Section 7.1.1). Then, the parameters in the conductor data table were utilized to compute the transmission line parameters (R, X, B) utilizing the equations in [24, 110]. This results in a per-unit system with each transmission having assigned line parameters (R, X, B) with underlying voltage levels, conductor sizes (kcmil), phase-to-phase conductor spacings (GMD), conductor lengths (miles).

If the input to this step of the overall algorithm is a network topology that does not have assigned line limits, a typical (or average) transmission line limit can be assigned to the line, and the initial parameter assignment can be performed as above. While this initial assignment is more crude than matching an existing line limit, the parameters and associated line thermal limits will be updated during the sequential DC power flow described in Section 7.1.2.

Conductor Data Tables

In order to assign or update the conductor parameters including impedance values (R, X, B) and line limits, a "look-up table", termed the conductor data table, was created using statistics from publicly available data from the FERC Form 1 report [111] and parameters taken from conductor manufacturers such as [109]. First, 13 synthetically created regional areas were created that roughly approximated the NERC Reliability Assessment Areas and/or the ISOs, with some aggregation in the western United States. For each of these regional areas, the voltage levels with statistically significant total

Capacity (MVA)	Voltage (kV)	conductor size (kcmil)	conductors per phase	Ra	Xa	Xa'
205	138	715	1	0.128	0.399	0.092
219	138	795	1	0.119	0.403	0.0917
230	138	900	1	0.106	0.399	0.0907
238	138	954	1	0.099	0.395	0.0897

Table 7.1: Excerpt from the 138 kV transmission conductor data table

transmission line mileage from the Form 1 were chosen as candidate voltage levels for the lines in the synthetic networks. For each of these voltage levels, a range of conductor sizes (in kcmil) were assembled using the data from the FERC Form 1 for the utilities in The corresponding R_a (resistance per unit length), X_a, and X'_a values (inductive or capacitive reactance at one-foot spacing in ohms per mile) for each transmission line were used from [110], and the thermal limits were assigned using [109]. For areas were the Form 1 had a limited number of conductors at each voltage level, additional conductor sizes were added between the minimum and maximum size at each voltage level. An excerpt from the full conductor data table is shown in Table 7.1.

Both $X_a = k\log(1/\text{GMR})$ and $X_d = k\log(\text{GMD})$ (where $k = 0.2794$ at 60 Hz) are both required to calculate a transmission line's per unit length inductance, since $X_L = X_d + X_a$ [110]. Thus, inferred GMD values were calculated for the transmission lines in the FERC Form 1, and statistics from these values were used to create potential GMD values for each synthetic region described above.

The FERC Form 1 does not directly contain GMD values, but contains the necessary information (voltage level, transmission structure construction type and material) to

calculate possible GMD values for each transmission line in the report. Thus, typical GMD values for each transmission structure type (tower, single pole, H-frame) and construction material (wood or steel) at voltage level were calculated from the different typical transmission structure information found in [110, 113–115]. The low, average, and high GMD values calculated were chosen as representative GMD values for each transmission structure "type" (i.e. 138kV wood H-Frame, 345 kV Steel Tower, etc). Then, for each transmission line in the FERC Form 1, one of the 3 calculated GMD values for the matching transmission structure type was randomly assigned to the entry in the Form 1. Finally, these inferred GMD values were used to create a statistical distribution of the GMD values in each of the geographic footprints, which are used to calculate the X_d values. Since each utility uses a different percentage of construction materials or designs for their transmission lines, the distribution of GMD values for each of the geographic footprints are different, resulting in synthetic power system networks that better mimic the characteristics and impedances found in the existing transmission networks. For more details regarding the tables of conductor parameters and GMD values see Chapter 3. The final resulting tables of conductor parameters and GMD values are input to the sequential DC power flow (Section 7.1.2) as shown in the overall flowchart in Fig. 4.1, thus allowing the overall KCL-compliant NCA to perform step 4 of the ideal high level NCA mentioned in Section 1.2.

7.1.2 Sequential DC Power Flow

The goal of this step of the overall network construction algorithm (NCA) is to update branch parameters such as impedance (R, X, B) and the branch flow limits (in MVA) so that very few or none of the branches are overloaded for a variety of net injection scenarios. This step can be performed on any synthetic power network that has a topology with branches containing assigned impedances and line limits that are either preexisting or assigned using the methods described above in Section 7.1.1.

Using the given line parameters, multiple DC power flow calculations are performed, and the resulting power flow on each transmission line is compared to the power limit on the transmission line. After each DC power flow calculation, a chosen subset of the lines that are overloaded are selected to be upgraded. In this work, the subset was chosen to be the larger of either 2% of the overloaded lines, or 15 lines. A line is upgraded by changing the underlying conductor assignment to match the next highest MVA rating line in conductor data table 7.1 in Section 7.1.1, and the corresponding parameters (R, X, B) are recalculated using the equations in [24, 110].

As an optional step, lines with a flow of less than a threshold percentage of its limit is deemed to be underutilized (chosen to be 30% in this work), and is marked as a candidate for downgrade. Similar to the overloaded lines, after each DC power flow a small subset are selected to be downgraded. A similar procedure to line upgrading is followed is followed for a line downgrade, except that the next smallest capacity line in table 7.1 is utilized. In this work, the subset is the larger of either 1.5% of the underutilized lines,

or 10 lines.

When lines are upgraded or downsized, typically a random selection the overloaded or underutilized lines was selected to be resized. This method results in different sets of network parameters each time the full NCA is run, which is beneficial for creating differing sets of synthetic power system networks. Alternative methods include upgrading the lines with the highest percentage overload, or largest MVA overload. While this allows the overall NCA to consistently create similar sets of network parameters, computation experience has shown that the resulting networks have unrealistic distributions of their line ratings, with the percentage of lines with "large" MVA ratings far exceeding the percentages in real power system networks.

The conductor data table (Section 7.1.1) contains a discretized range of conductor MVA ratings that span all voltage levels that are present in the FERC Form 1 for the given region. Thus, sometimes when a transmission line's MVA rating is increased (or decreased), the line moves to the next higher (or lower) voltage level. This process of running DC power and modifying lines that are overloaded or underutilized is continued until a suitably small percentage of lines are overloaded or underutilized. For this work, 0.1% and 10% are used as the percentages of lines that can be overloaded or underutilized, respectively. These percentages need to be large enough to allow the sequential DC power flow to terminate, since it is sometimes impossible to correct the majority of overloaded lines without a subset of the lines being underutilized, and vice versa.

This process of sequentially running the DC power flow and updating a subset of line

parameters and limits is performed using multiple net injection scenarios. The scenarios must be provided as inputs to this overall algorithm (as shown in the data input block in Fig 4.1). The scenarios can be calculated using many methods, ranging from production cost modeling to set load and generation scenarios, to the methods described in Section 4.4.1. Running the sequential DC power flow for k injection scenarios results in k different sets of network parameters that must be combined to form a final network. For each network branch, the maximum line rating across all the scenarios is chosen as the final line rating, i.e. the L∞ norm. which results in a network that is more robust to multiple load and generation scenarios. The calculated line rating is then used to calculate the final R, X, B parameters as described in Section 7.1.1. However, Section 5.3.2 will explore methods of considering multiple injection scenarios when modifying the network parameters in each iteration of the sequential DC power flow.

Additionally, overloaded transmission lines that are assigned the highest capacity conductor in their respective voltage level's table are flagged for manual intervention after the algorithm terminates. The normal course of action is to add a double circuit to the transmission corridor, since the algorithm does not create double circuited transmission lines automatically to avoid an unrealistically high number lines with large line limits.

After the first run of the sequential DC power flow, each branch has assigned line limits and parameters (R, X, B). By adding branch parameters and calculating DC power flow, this adds the KVL and network element relations constraint from (1.1e) to

the network model, i.e. modeling constraint (1.3c). Additionally, each node represents a single substation with branches modeled in per unit quantities. Individual buses and transformers connecting those buses will not be modeled until Sections 7.2.1 and 7.2.3 respectively, and as seen in the overall flowchart in the flowchart in Fig 4.1. These assignments are based on the capacity and associated voltage level of each transmission network element connected to the substation.

7.1.3 Network Element N-1 Contingency Modeling and Feasibility

As an optional step, the sequential DC power flow line resizing can be performed considering an N-1 contingency analysis. First, a sequential DC power flow is calculated to correct any overloaded lines, following the method in Section Section 7.1.2. Next, each transmission line is removed one at a time to conduct an N-1 contingency analysis. After removing a line from the network, the sequential DC power flow from Section 7.1.2 is then re-run, but only the overloaded transmission lines are upgraded. Underutilized lines are not modified to avoid creating overloaded lines in previous iterations of the N-1 contingency analysis. sAfter the final N-1 transmission line contingency has been calculated and the overloaded lines upgraded, the lines that were assigned their maximum MVA rating but are still overloaded can be manually double-circuited as explained in Section 7.2.2.

7.2 Additional Topology Modifications for Increased Network Reliability and Realism

After performing the DC power flow-based line resizing algorithm, the power system network has a topology containing branches with parameters (R, X, B) that result in a DC power flow solution that satisfies the majority of the line limits. If the network topology only includes substations without individual buses and transformers at each substation, the algorithms in Section 7.2.1 and later 7.2.3 must be utilized to create a more realistic synthetic network. These modifications correspond to step 3 in the overall 4-step process of an idealized NCA (Section 1.2), and are shown in the overall flowchart in Fig. 4.1. As an additional step, Section 7.2.2 adds transmission lines to each voltage level individually to create a better meshed network. This step is required for network topologies generated using KCL-compliant algorithms such as that in Section 4.1, but optional for network topologies where the sub-networks at each voltage level are mostly connected. Finally, the sequential DC power flow is re-run to resize the new transmission lines added.

7.2.1 Creating Buses with Assigned Voltage Levels

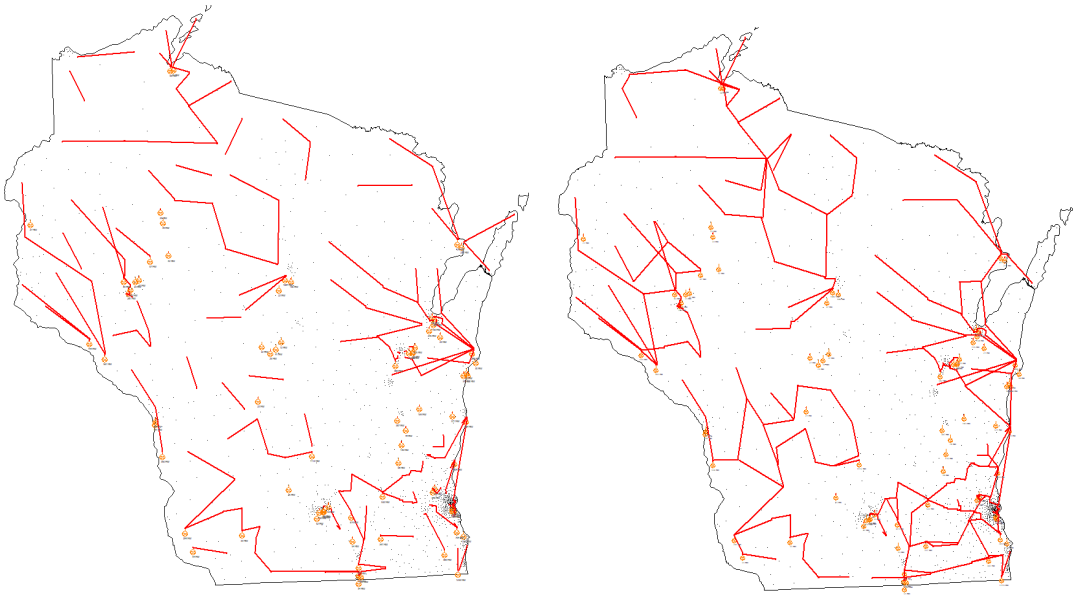
Since KCL-compliant network construction algorithms (NCA) utilize the transportation model, the only network parameters considered are the line limit constraints. This limits KCL-compliant NCAs to represent the network in a per-unit representation without considering different voltage levels at each substation. Since impedances are not

modeled by the multi-level linear program in Section 4.1, the voltage levels are assigned as part of the sequential DC power flow described in Section 7.1.2. However, voltage levels are assigned to the conductors themselves instead of directly to individual buses. Thus, each substation will contain more than one voltage level associated it based on the different voltage levels assigned to the transmission line(s) connected to the substation. Thus, a bus is assigned to a substation for each different voltage level present at the substation. Since discretized data tables of transformers have not yet been created, at this step in the network creation process the buses within a substation are assumed to be connected by lossless, zero-impedance ideal transformers. After the network is more finalized, transformer impedances and ratings are added to the ideal transformers as described in Section 7.2.3.

Since this assignment does not consider network topology, the resulting sub-network at an individual voltage level is often not a connected graph, since voltage levels are assigned to each line individually without consideration of connectivity. Thus, the network should be modified ensure the sub-graphs at each voltage level are connected. An example of this is shown in Fig. 7.1a.

7.2.2 Re-Establishing Connected Networks after Separating by Voltage Levels

To reconnect the voltage levels, each individual voltage is passed back to the multi-level KCL-compliant algorithm described in Section 4.1 to add additional lines and



(a) Fragmented network before reconnection (b) Reconnected sub-networks

Figure 7.1: 138 kV networks of the Wisconsin 1664 bus model

create connected networks at each voltage level. To create connected networks, the effects of the other voltage levels must be considered, since often loads and generators are connected to buses that are at different voltage levels than the one being reconnected. Network reduction techniques such as the Kron reduction [29] are typically utilized to reduce the network to an equivalent, smaller network. When the equivalent network is represented using the network flow model in the KCL-compliant NCA, the resulting line flows are within the line limits of the existing lines and no flow is routed on the potential transmission line paths. This results in no additional lines being added to the network using network equivalencing methods. Thus, a set of net injection vectors must be created that represent the effect of power flows on other lines and net injections at other buses without exactly replicating those quantities. This results in a similar but not exactly equivalent flow pattern on lines, which causes power to be routed on the

potential transmission line paths and resulting in new lines being added to reconnect the network at the given voltage level.

To calculate these new net injection vectors at each substation, the sum of the calculated DC power flows on lines at voltage levels other than the one being reconnected are used to compute random numbers of the same orders of magnitude. These randomly assigned numbers become the new net power injections at each bus of the voltage level being reconnected. For example, to compute the new net injection vector for the 138 kV voltage level show in Fig 7.1a, the sum of the line power flows into and out of each substation at all voltage levels excluding 138 kV. As an example calculation, if a particular substation with 69kV, 138kV and 230kV lines has total flow of 76 MW entering or leaving the substation from the 69kV and 230kV lines, a random net injection ranging from 10 to 100 MW is assigned to the 138 kV bus at that substation.

The multi-level KCL-compliant algorithm is again used to modify the topology by adding lines to the sub-networks at each voltage level. As seen in the bottom portion of the flowchart in Fig. 4.1, the inputs are the new net injection vectors, and the fragmented network at each voltage level with the associated line limits (e.g. the 138kV network in Fig. 7.1a). Since the goal is to increase the meshing of the system and increase the number of transmission corridors utilized, potential transmission line paths that are in parallel with existing branches at other voltage levels are removed. This prevents the outer loop from adding multiple voltage levels along the same corridor, which reduces the reliability of the system if contingencies portraying the loss of a corridor or transmission

towers are considered. The resulting network has additional lines that reconnect the voltage level, thus improving the reliability of the system.

The process in this Section is repeated with each voltage level until all of the voltage levels have been reconnected with additional lines being added to the network. As in previous stages of the overall network construction process, additional net injection scenarios can be utilized to add additional lines to the network as discussed further in Section 4.4. Finally, Fig. 7.1b shows the 138kV networks of the Wisconsin 1664 bus model after the fragmented networks have been reconnected. Note that the 138 kV network has been reconnected into four sub-networks, and this process is explained further in Section 7.3.1.

Sequential DC power flow respecting voltage level

First, the sequential DC power flow from Section 7.1.2 is re-run on the full network with reconnected voltage levels. This allows the new lines added by the linear programming step to be re-sized to account for the modeling differences between the network flow model and the DC power flow model. However, the conductor voltage levels are fixed in this step, so each line is only allowed to be upgraded or downsized within the range of lines in the conductor data table that are at the same voltage level (Section 7.1.1). The termination criteria for the sequential DC power flow is changed so that the algorithm terminates when fewer than 0.5% of the transmission lines are overloaded and fewer than 5% of the transmission lines are underutilized (flow less than 30% of the MVA capacity). These remaining overloaded lines are either corrected manually or are

automatically corrected during the network modifications considering N-1 transmission line contingencies in Section 7.1.3. As in Section 7.1.2, multiple net injection scenarios can be considered, and the maximum line limits across the scenarios is selected as the final line limits.

7.2.3 Adding Transformers

After the second run of the sequential DC power flow, impedances and MVA limits are assigned to the ideal transformers connecting the different buses at each substation. The transformer MVA ratings were determined using the two sums of the ratings of the line(s), generator(s), and/or transformer(s) connected to either side of the transformer. For substations with exactly two voltage levels, the lower of the two sums was used as the MVA rating for the transformer(s) (plus some margin in the case of a large difference). For substations with more than two voltage levels, the parallel combination of transformers was rated at 75% of the smaller sum of the combined MVA ratings of the connected transmission line(s), generator(s) and transformer(s). The transformer MVA capacity ranges, average per unit impedances, and average X/R values for a range of transformer MVA values were computed from the FERC Form 715 data [9], and were used to compute the transformer parameters (R, X) and MVA ratings for the transformers in the synthetic system. The resulting synthetic network contains the parameters (R, X, B) and data (real and reactive net power injections) necessary for a full AC power flow solution. However, for medium and large-scale networks, reactive power support must be added for the AC power flow to successfully converge.

An example system created using this algorithm is the Wisconsin 1664 bus system shown in Fig. 7.2. This system is posted online and can be located at [127].

7.2.4 Further Modifications for AC Feasibility

Once the transformer impedances have been added to the network in Section 7.2.3, the power networks that have been created are DC OPF feasible, meaning there exists an operating point that satisfies all network element limit constraints. However, additional modifications need to be made to the networks for them to be AC power flow or AC OPF feasible. Such modifications include adding reactive power compensation devices such as switched shunt capacitors or SVCs and further modifying the transmission line parameters and limits to make the system fully AC OPF feasible. While the addition and placement of reactive power compensation devices is beyond the scope of this chapter, the work in [128] can be used for this purpose. Additionally, large interconnect-scale models can be constructed using the process described in Section 7.3.

7.3 Creating Interconnection-Scale Networks

To create networks larger than 8,000 buses, the full set of synthetic substations is partitioned into regional geographic footprints of 8,000 substations or fewer for computational efficiency. The multi-level algorithm and first sequential DC power flow is run on each regional area to create buses with associated voltage levels within each substation (Section 7.1.2). Then, the sub-networks at each voltage level are combined into a larger

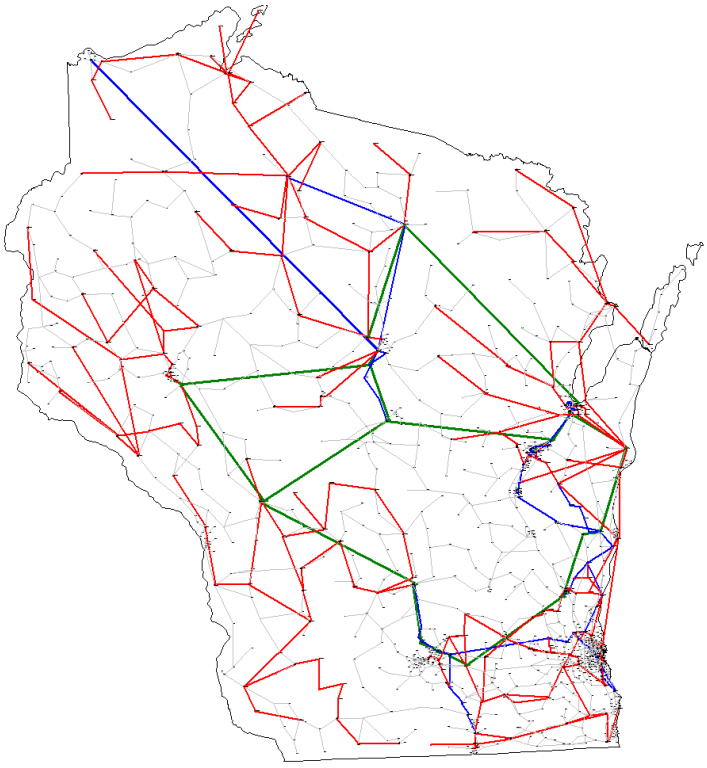


Figure 7.2: One-Line of the full Wisconsin 1664 bus network model

network that are each reconnected using the multi-level optimization framework.

The first step in creating an interconnection-scale network, such as the 20,000 bus synthetic Western United States model discussed in this section, is to create smaller regional networks that are partially built, then a fully interconnected model is constructed using these partially constructed area networks. This mimics historical transmission network design by first creating multiple smaller regional networks, such as the ones that RTOs operate. For the synthetic Western model system, 3 regional areas were used: California, Pacific Northwest, Desert Southwest.

Next, the initial topology for each regional area is established using the multi-level algorithm described in Section 4.1, and buses with voltage levels are established using

the sequential DC power flow-based algorithm in Section 7.1. The results of this step are shown in Fig. 7.3. Notice how the three areas are not yet interconnected, and the sub-networks at each voltage level are fragmented and not fully connected.

The fragmented voltage sub-networks from each area are utilized as inputs to the reconnection algorithm in Section 7.2.2. Since there is sometimes a large geographic distance between sub-graphs of each voltage level, each voltage level's network graph is partitioned using the Fiedler eigenvector-based approach in Section 7.3.1 until each sub-network can be connected using lines that are less than the maximum length transmission line specified for the given voltage level. After each voltage level has been reconnected, the remaining steps in Section 7.2.2 are completed, resulting in a network that is DC N-1 OPF feasible. The final results of this large-scale network creation process is shown in Fig. 7.4. Notice how some voltage levels, such as the 500 kV network depicted in orange, are a single connected network, whereas the 230kV (blue) and 345kV (dark green) networks have been partitioned into smaller networks to allow them to be reconnected with lines that respect the maximum length constraints.

This sub-division and combination approach has the benefit of reducing the computational burden when the network of potential transmission line paths still includes the (slightly modified) complete graph. However, after voltage levels have been assigned, each sub-network has a much smaller set of potential transmission line paths than the original regional networks. The lower voltage networks have large numbers of buses, but the shorter maximum line lengths allow for a significant reduction in the set of realis-

tic potential transmission line paths. The higher voltage network can have much longer paths, but has a significantly lower number of buses, thus achieving a geographic sparsity and also significantly reducing the number of potential transmission line paths in these sub-networks. Note that given enough computational power and time, this initial partitioning into sub-regions is unnecessary. However, sub-dividing into areas is performed for practical implementation of this full KCL-compliant NCA on consumer-grade computer hardware.

7.3.1 Partitioning Larger Networks

Small to medium sized networks, such as the Wisconsin 1664 bus network shown in Fig. 7.2 can have the network associated with each voltage level connected as a single network. However, for larger networks it is often unrealistic to model the entire voltage level using one connected system. Thus, a Fiedler eigenvector-based partitioning method is used to partition the voltage level network into smaller sub-networks before reconnecting them in Section 7.2.2.

First, a symmetrical weighted Laplacian matrix L is created by using a combination of the line capacities and the bus-to-bus distance matrix to create a rank $n-1$ matrix as shown in (7.1)-(7.3). The bus-to-bus distance matrix D is expanded from an upper-triangular matrix to be a full, symmetrical matrix when used to construct the Laplacian matrix. This Laplacian matrix is primarily weighted using the transmission line limits, however the bus-to-bus distance matrix is also used to minimize the size difference between the partitions. The use of L_d also ensures that L is rank $n-1$, where n is the

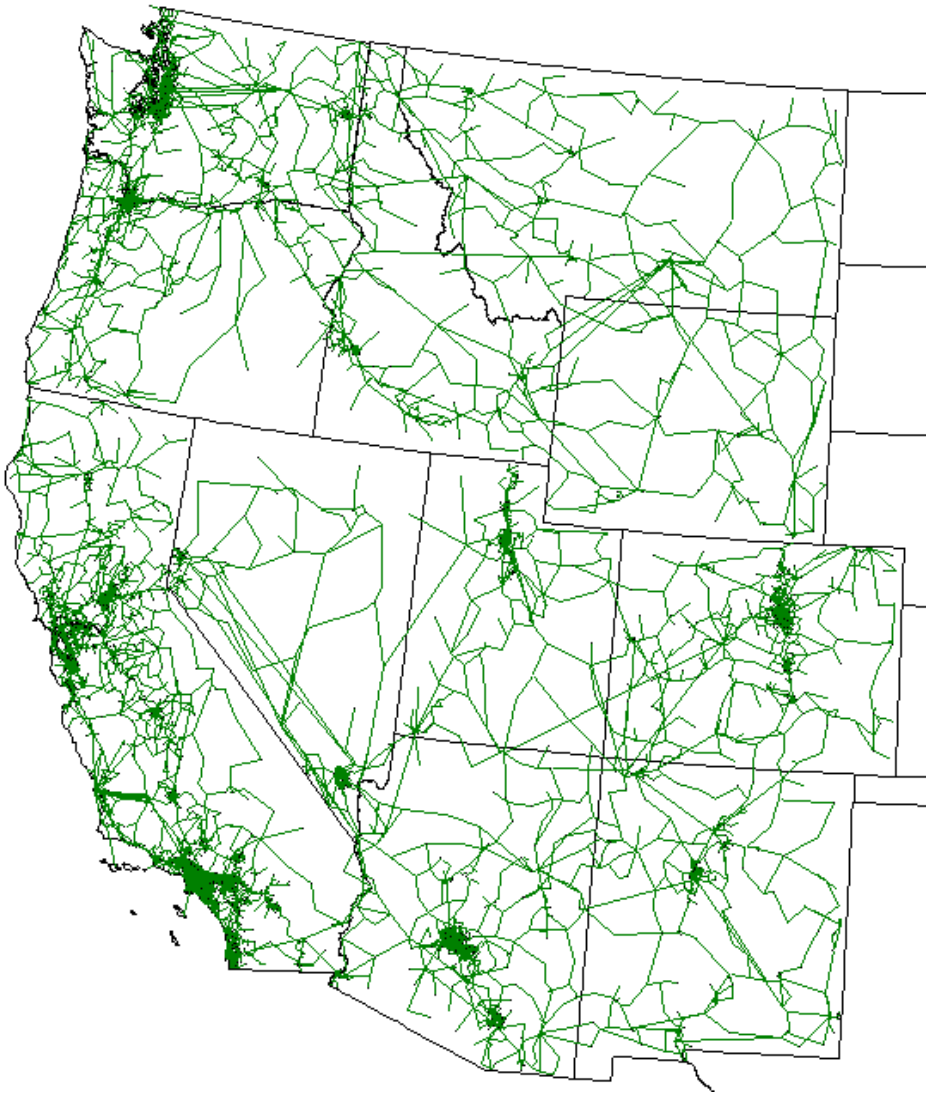


Figure 7.3: Initial topology of 3-area synthetic Western United States power system

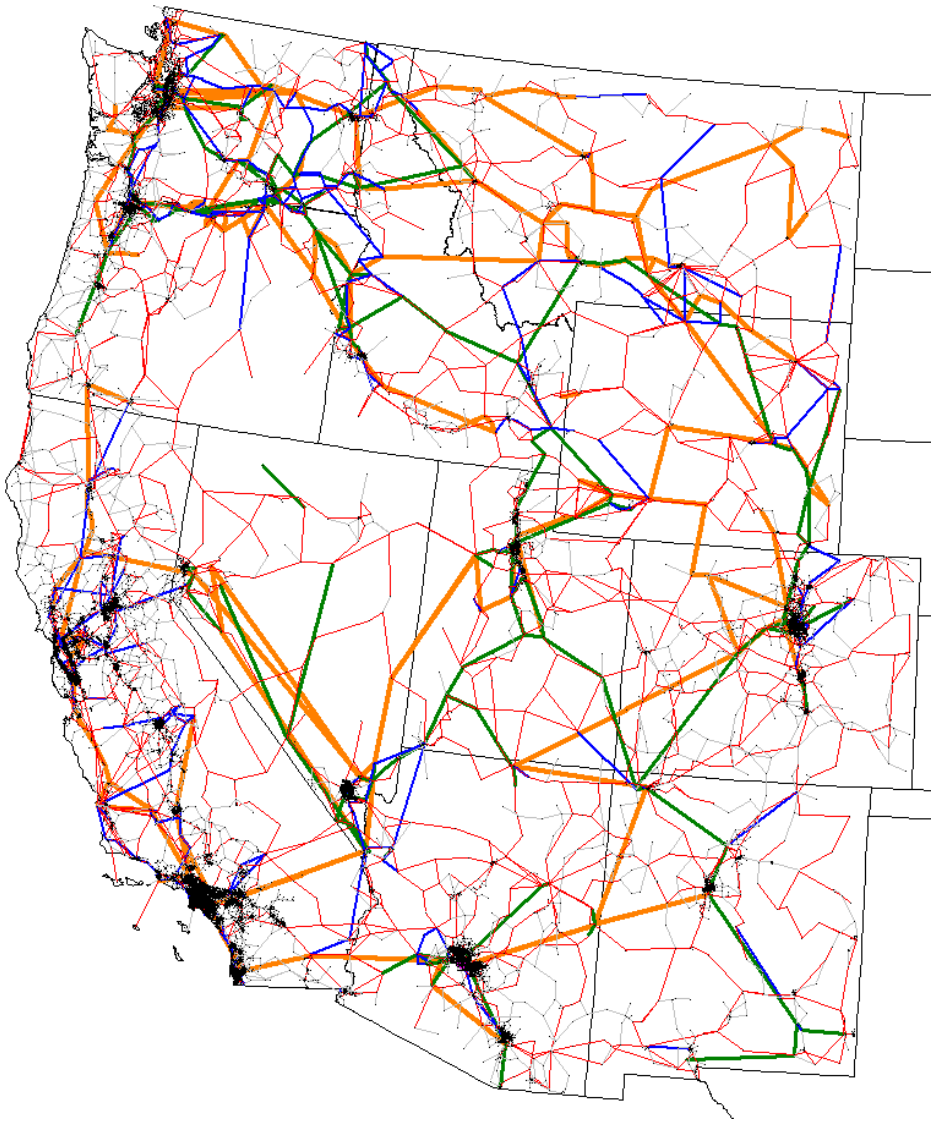


Figure 7.4: One-line of a 20,000 synthetic Western United States system

number of buses, which is not guaranteed if the existing transmission lines do not form a connected graph.

$$L = L_d + L_\ell \tag{7.1}$$

Where,

$$L_{d,i,j} = L_{d,j,i} = \begin{cases} \frac{\alpha}{D_{i,j}} & \text{where } \alpha \text{ is a constant} \\ 0 & \text{when } i = j \end{cases} \tag{7.2}$$

and,

$$L_{\ell,i,j} = \begin{cases} \text{Power MVA limit of element connecting buses } i \text{ to } j \\ 0\text{if no network element connects buses } i \text{ to } j \text{ or if } i = j \end{cases} \tag{7.3}$$

The eigenvectors and eigenvalues of the Laplacian matrix are then calculated, and the second smallest eigenvalue and associated eigenvector, termed the Fiedler eigenvalue, is used for the partitioning. Utilizing the work of Fiedler in [129] which is further clarified in [130], the values in the Fiedler eigenvector that are positive are assigned to one partition, and the negative values are assigned to the other partition. The position of the value in the eigenvector indicates the associated bus number. This partitioning process can be repeated on each sub-partition by recreating a Laplacian matrix using the buses and transmission lines that are contained within each partition.

Experience has shown that setting $\alpha = 10$ results in partitions that minimize the number of transmission tie lines that span partitions, as well as create areas that are more even in size. To keep the tie-lines from being discarded since they are not fully

contained by either partition, they can be assigned to the smaller partition if further sub-partitioning is needed, or kept in a separate list of tie lines. For a result of this Fiedler eigenvalue-based partitioning process, see the 4 areas shown in Fig. 7.1b

The results of the Fiedler eigenvector-based partitioning are show earlier in Fig. 7.1b, which illustrates how the partitioning algorithm created 4 sub-networks of the 138 kV transmission system that were then re-connected in Section 7.2.2 to form the networks shown in the one-line. The 345 kV network (shown in green in 7.2) contained few enough buses to not be partitioned into sub-networks.

7.4 Chapter Summary

The simplest, and most tractable approach to modeling a NCA is the KCL-compliant class of algorithms presented in Section 4.1. This KCL-constrained, weighted L1 norm approach to creating networks has the linear programming advantages of significant scalability and excellent computational tractability. The multi-level algorithm first determines the initial network topology and network branch MVA limits.

Further network element parameters such as transmission tower material and construction and the corresponding element parameters including impedance and power limits are iteratively determined using a sequential DC power flow as shown in Section 7.1. The multi-level algorithm is then re-run on each voltage level of this modified network to reconnect the sub-networks at each voltage level and increase system reliability by adding additional branches (Section 7.2.2. Finally, the sequential DC power flow is

re-run on the resulting network to account for the modeling discrepancies between the minimum MW-mile network flow formulation and the nodal analysis DC power flow formulation (Section 7.2.2). Multiple networks have been created using this method, ranging from a small 250 bus New England synthetic model to the largest 78,000 bus Eastern United States synthetic network. All networks are available for download at [127]

However, there are also disadvantages to the heuristics presented in this chapter. First, accurately creating realistic power system networks requires multiple iterations of the full multi-level linear program (LP) presented in Chapter 4, while the KCL/KVL algorithms in Chapter 5 only require a single iteration of its multi-level algorithms. Also, the sequential and iterative nature of the heuristics remove any optimality guarantees from the algorithm. Each iteration of the KCL-compliant inner loop (Section 4.2) is guaranteed to find an optimal solution solution in finite time if the feasible set is non-empty. However, iterating between sequential DC power flow and the multi-level linear program is not guaranteed to produce an "optimally designed" network. This is because there are substantial differences in the flow calculated using linear programming and a DC power flow, resulting in iterating between the two methods.

Some of these disadvantages can be addressed using a "hybrid" method that models existing lines using DC power flow and potential transmission line paths using the transportation model. This alleviates some of the modeling disadvantages posed by only using the transportation model, but doesn't solve the limitations in modeling multiple voltage levels for the potential network. Thus it is desirable to utilize the new

class of KCL/KVL-compliant network construction algorithms that directly models DC power flow while retaining the computational tractability and scalability of KCL-only NCAs implemented using multi-level linear programs. These algorithms were presented in Chapter 5.

Chapter 8

Additions and Modifications to KCL/KVL Multi-Level Algorithm

The multilevel KCL/KVL algorithm as described in Section 5.3 produces networks that share many characteristics with existing power system networks, as verified in Chapter 6. However, there are still modifications that need to be made to the algorithm in order to add additional modeling detail such that the networks produced by the KCL/KVL multilevel algorithm better resemble real power system networks. Thus, the following sections describe additions to the middle and outer loops that increase realism by adding buses to substations, connecting these buses with transformers, adding reactive power compensation, and ultimately computing and AC power flow solution.

8.1 Outer Loop Additions: Substation to Bus model

The networks created by the multilevel algorithm in Section 5.3 are substation-level models, i.e. each substation has one "bus" and all transmission lines are directly connected to this "bus." This is because the decision variables in the inner loop in Section 5.3.1 are the power flow values on each individual line, thus the multilevel algorithm takes a "line centric" modeling approach as opposed to a "bus centric" modeling approach as is common in power systems. This results in all of the quantities and parameters, such as voltage level, being directly assigned to transmission lines and only indirectly assigned to the substations at either end of the transmission line. Thus, although there are multiple voltage levels represented in the network, there is no direct and straightforward way of ensuring that the subnetworks at each voltage level is connected. In the KCL-only algorithm, Section 7.2.2 explains how additional lines were added to reconnect each subnetworks. In this KCL/KVL algorithm, Section 9.1.2 explains the process of modifying the voltage level of some transmission lines, plus the addition of some lines, to better ensure connected subnetworks at each voltage level.

The first step to creating connected subnetworks is to create buses at each substation connected by transformers as described in the following sections. Next, reactive power support is added by placing shunt capacitors or reactors at a percentage of the buses.

8.1.1 Creating Buses

Since each transmission line has an associated voltage level, the voltage levels present at each substation can be computed by examining the unique set of voltage levels of the lines connected each substation. A bus is created at each voltage level present in the substation, and lines are moved from the original "bus" to the newly created buses. The lowest voltage level at the substation is assigned to this "original bus", and the loads are left directly connected to this bus. Loads over a certain threshold, in this case 50 MW, can be moved to higher voltage buses to avoid overloading the low voltage sub transmission system. Transmission lines that voltage is higher than the lowest voltage level are move to the newly created buses.

Since the EIA Form 860 [131] includes information regarding the point of interconnection (POI) voltage for generators, this information can be utilized to move generators from the original bus to the bus at the closest voltage level to this POI voltage. However, if generation data other than the form 860 is utilized, other methods can be utilized to determine which voltage level generators are assigned to. It is advisable that generators not be left at the lowest voltage level otherwise this creates substantial overloads on the low voltage transmission system.

Optionally, Generators and loads can be moved to their own buses, and these buses be connected via generators step up (GSU) or load step down transformers. An analysis of the FERC Form 715 for the Eastern and Western interconnects revealed that the majority of generators are at their own dedicated generator bus, thus in this dissertation

generators are moved to their own buses connected by GSUs. A similar analysis reveals that only 1/3 to 1/4 of the loads in the Eastern and Western interconnects are at their own dedicated load bus. Since there is no easily identifiable criteria for a load being placed at its own bus instead of directly at the low voltage transmission bus in the Eastern and Western interconnect models, future work will explore placing a subset of the low that their own buses connected by load step down transformers.

8.1.2 Creating Transformers

Transformers are created in order to connect the newly created buses at each substation. Generator step up transformers (GSUs) are sized such that the maximum capacity of the transformer is 20% higher than the maximum output of the generator to allow for potential upgrades to the generator and to keep the transformer from becoming a binding constraint. Load step down transformers are similarly sized such that they are 20% higher than the load at their low voltage load bus.

Since there is no way to determine the power flowing through a transformer until the transformers placed in the network, transformers connecting transmission level buses are all initially size to the largest capacity of 2000 MVA. Next, a DC power flow is run and the resulting flow through each transformer is set as the capacity of the transformer. These newly created transformers are then updated in the middle loop along with the transmission lines as described in Section 8.2. This method results in much more realistic capacity ranges of transformers as compared to the KCL-Only algorithm method Section 7.2.3, which produces a much higher number of parallel transformers than is

realistic.

Once an MVA capacity is assigned to each transformer, this value, along with the primary and secondary voltage levels of the transformers, are used to identify the corresponding %Z and X/R ratio from the transformer lookup table created using statistics from the FERC Form 715, as described in Section 3.2.2. The transformer impedances are then converted from the base MVA of the transformers to the system base of 100 MVA, and the X/R ratios are utilized to calculate the per unit resistance and reactance for each transformer.

8.1.3 Adding reactive support

In order to converted the DC network model to an AC power flow model, reactive power compensation typically needs to be added to a subset of the buses in order to have a converged AC power flow. Following the suggestion in [128], all buses are initially assigned as PV buses with a voltage setpoint of one per unit and an AC power flow is run. The resulting reactive power injection or withdrawal at each bus is calculated, and a percentage of the buses with the lowest reactive power injection or withdrawal are converted to PQ buses. The MVar rating of the capacitor or shunt inductor is set to be a certain percentage higher than the calculated reactive power injection or withdrawal at the connected bus. The AC power flow is then rerun, and newly assigned PQ buses that are over under their voltage limits are changed back to PV buses and marked to have Sean compensation. This process is continued until the number of PV buses in the network is below a target percentage, in this case 20%.

To create a set of buses with reactive power support that is robust across multiple scenarios and contingencies, multiple net injection scenarios can be utilized in the process listed above. The reactive power injection or withdrawal at each PV bus can be computed for each of the scenarios, and either the mean or the minimum across the scenarios can be utilized to decide which buses are change to PQ buses.

8.2 Middle Loop Additions

The outer loop modifications and updates described above change the topology of the network by adding buses, transformers, and reactive power compensation. The next updates to the KCL/KVL multilevel algorithm are middle loop updates that change the parameters of the branches.

8.2.1 Sizing transformers

Once transformers have been created to connect the buses within a substation, the middle loop is modified to resize the transformers in addition to the transmission lines. The same method described in Section 5.3.2 is utilized, with the MVA limits of overloaded transformers being increased, and the MVA limits of underutilized transformers being decreased. The method described above in Section 8.1.2 is utilized to recalculate the R, X values of the transformers. Per analysis of this FERC Form 715, transformer MVA ratings typically range between 20 MVA in 2000 MVA, so if a transformer needs to have a rating higher than 2000 MVA, an additional transformer is added in parallel to the

overloaded transformer. The parallel combination of the transformers is sized such that the combined MVA rating is sufficient such that either transformers overloaded. If there are parallel transformers such that The total utilization is below 30%, one transformer is removed to avoid having an unrealistically large number of parallel transformers.

8.2.2 AC Power Flow

Finally, once the network is AC power flow feasible, the network parameters can be modified using an AC power flow calculation instead of the DC power flow. This is done by replacing the sequential DC power flow in the middle loop with a sequential AC power flow based method. If the network is robust enough, the AC power flow scenarios can be run in parallel. However if the network cannot converge from a flat start, then a set of net injection scenarios with smaller variations between scenarios can be utilized. The solution to the previous scenario can be utilized as the initial condition for the next scenario.

If there are a small number of AC power flow scenarios that are not feasible, these scenarios can be discarded. However if these are important scenarios, or of there are a larger number of infeasible scenarios, the reactive power compensation devices need to be modified as per Section 8.1.3.

Chapter 9

Conclusions and Future Work

The goal of this dissertation is to create electrical power system network models that are similar to existing power systems both electrically and graphically. However, the problem is challenging since the most accurate mathematical decision model for electric power system construction results in a nonlinear, nonconvex set of equations describing the electrical power flow in the network (Chapter 1). If this full AC power flow formulation is utilized to construct power system networks, it results in a mixed-integer, nonlinear, nonconvex optimization problem, which is intractable for constructing medium to large scale networks.

The majority of the network construction algorithms in the existing literature solve this problem by focusing primarily on creating synthetic power system models that match graph metrics gathered from the existing U.S. Eastern or Western interconnect power systems. However, this results in network models that are less realistic than originally

anticipated, and also often do not match the electrical performance of existing power systems as established in Chapter 2.

Thus, this thesis has developed two sets of network construction algorithms (NCAs) are developed that primarily consider the electrical behavior and performance of the power system networks being constructed and utilize graph measures as comparison tests of the similarity of the synthetic networks to existing power systems. Both sets of algorithms utilize a multilevel formulation consisting of three parts: an optimization-based inner loop that computes the flow on each of the network branches, a middle loop that uses this flow information to update the parameters of the branches, and outer loop that either updates the net injection scenarios or makes topology changes to the network.

The first set of algorithms, termed KCL-Only or KCL-compliant network construction algorithms, were developed from the transportation formulation of the network flow problem (Section 1.4.1). These algorithms were able to create large-scale synthetic power systems that matched many of the characteristics of the existing electrical networks (Section 6.1.1.

However, it has been observed that the networks generated using the KCL-compliant algorithms do not perform well electrically when tested with additional load scenarios as seen in Section 6.4. Thus a new set of algorithms, termed KCL/KVL-compliant algorithms, was developed to overcome these shortcomings. The goal of this new set of algorithms is to directly model DC power flow in the inner loop optimization program

to improve the electrical performance of the resulting networks. Since DC power flow results in non-zero power flow on all branches in the network, an L1 regularizer was added to the objective function to enforce a sparse solution to the problem. As shown in Section 6.4, the networks generated using this KCL/KVL-compliant NCA resulted in a substantial improvements in the electrical performance over the networks generated using the KCL-Only algorithms. Although there is no explicit targeting of the graph metrics presented in Sections 6.5.2 and D, the KCL/KVL-compliant algorithms also generated networks that on the average showed a higher degree of similarity to existing power system networks than those generated using the KCL-only algorithms (Section 6.5.

9.1 Future Work

This dissertation has demonstrated that both the KCL-compliant and KCL/KVL-compliant network construction algorithms (NCAs) are successful at creating large-scale synthetic power system networks that bear many similarities with the existing power system networks. Given the primary emphasis on electrical performance in these algorithms, there is nothing fundamentally preventing these algorithms, especially the KCL/KVL-compliant algorithms, from being modified and used for transmission expansion. This is in stark contrast to the existing graph metrics-based NCAs reviewed in Chapter 2, which have inherent fundamental limitations preventing them from being adapted for transmission expansion.

Additionally, the improvements described in the following sections can be made to the KCL/KVL-compliant NCA, in addition to resolving the parallel lines issue in Section 6.2.2.

9.1.1 Creating and Updating Buses Within Substations

Instead of creating the buses at one point in the multilevel algorithm and fixing all voltage levels for the rest of the multilevel algorithm, it would be better to have a number of outer loop iterations where buses are dynamically created and updated.The initial bus assignments to the substations are based on the voltage levels of the transmission lines that are connected to each substation. Since the decision variables in the inner loop are the flows on the paths, the voltage levels are assigned to the transmission lines and only indirectly to the substations. The voltage levels present at a substation is determined by the voltage level of each transmission line connected to that substation.

As an initial assignment, a bus will be created for each voltage level present in a substation and the transmission lines moved as described in Section 7.2.1. Then, transformers will be created that connects all of the buses in an "open ring", i.e. the lowest voltage is connected to the next bus voltage via a transformer and so on until all the buses are connected. Initially, transformer capacity and parameters will be assigned using the methodology in Section 8.1.2. However, in the future the transformers will be treated like the transmission lines in the middle loop, and transformers that are substantially underutilized or overloaded will be downgraded or upgraded incapacity. Then the transformer impedances will be recalculated using the new MVA capacity and

average % impedance values and X/R ratios from the form 715.

Additionally, buses with an unrealistically high node degree can be split into multiple buses you using a bus tie or other configurations such as a ring bus or double bus double breaker.

9.1.2 Voltage Level Assignment Problem

Currently, voltage levels are directly assigned to each branch in the network. For the KCL-compliant algorithm, existing transmission lines are assigned voltage level after the initial topology has been established when the outer loop terminates. However for the KCL-KVL algorithm, all paths including potential transmission line paths in existing transmission lines have voltage levels assigned to them as an underlying feature of their branch impedances.

For both the KCL-compliant and KCL-KVL algorithm, this voltage level assignment is based on the capacity of the line, which is determined by the flow on that line in the inner loop of the respective multilevel algorithms. Currently, this voltage level assignment decision is done for each transmission line individually, independent of connected and neighboring transmission lines. As shown earlier in Section 7.2.2, this creates disconnected subnetworks at each voltage level since no consideration is given to topology when the voltage levels are assigned.

Future work will include modifications to the middle and outer loops of the KCL-KVL multilevel algorithm such that the subnetworks at each voltage level will be connected

by the termination of the outer loop. There are several methods to achieve this goal. One of them is to include an optimization based second "middle loop" that modifies the voltage level of transmission lines to ensure connectivity of each subnetwork, shown below in Section 9.1.2. Other methods include heuristic based methods, including some of the methods described below.

After a specified number of outer loop iterations or another criteria is met, the middle loop would begin deciding voltage levels to better enforce conductivity of the subnetworks at each voltage level. The general idea of this modified algorithm would be to move some lines at lower voltage levels to a higher voltage level if doing so would connect islands in the subnetworks. Additionally, some isolated high voltage transmission lines that were not nearby any other islands at that voltage level could be removed, if the line to be removed was not carrying a large power flow, and removing the line would not cause the network to become disconnected.

This algorithm could be accomplished by using the incidence of adjacency matrix to determine the nearest two or three neighbors to each bus in an island in a given subnetwork. Then the lines along shortest path distance, either electrically or geographically, could be raised to the higher voltage level to interconnect the two islands. If the geographic distance between two islands was short but the path of existing lines to connect those two islands was long, a new transmission line be added between the two islands to directly connect them.

Optimal Transformer Creation Problem

The reason why the number of disconnected islands at each sub voltage level should be minimized is because having multiple disconnected voltage level islands results in a higher number of transformers installed in the system. Heuristics such as the one above can be utilized, or an optimization based method for minimizing the number of transformers in a synthetic network could be developed. The solution must satisfy the target line mileage at each voltage level as specified in the input data, as well as ensure that each voltage level sub-network is connected.

9.1.3 N-1 Contingencies

Additionally, the multilevel algorithms should be modified to account for transmission line N-1 contingencies. Currently, N-1 contingencies are modeled in the KCL-compliant NCA by modifying the transmission line limits and associated impedance parameters as described in Section 7.1.3. However, future work includes modifying the outer loop of the KCL-KVL NCA to include scenarios that model important N-1 branch contingencies, especially transmission lines and transformers that are connected to generator substations.

Another approach would be to utilize the method described in [58] to add additional transmission lines after the termination of the outer loop to ensure the network will remain feasible for a given list of N-1 contingencies.

Appendix A

Supporting information for KCL-Compliant Algorithms

A.1 Review of Garver's method

In [13], Garver proposed an algorithm for constructing a transmission network capable of handling future generation and load scenarios. The algorithm utilized multiple linear programming steps to solve for the power flow on existing network circuits and overload-paths. After each solution of the linear program (LP), a new circuit was assigned to the overload path with the largest calculated flow, duplicating this path in the set of pre-existing lines. This procedure was repeated until no overload paths contain flows. This process is explained further in the sections below.

A.1.1 Required Inputs

Garver's work requires data for existing buses and lines, as well as new buses and overload (candidate) paths. First, it requires that each existing and new bus in the network have specified net power injections P. Existing lines must have power capacity limits and weighting terms for the cost function which are termed the guide numbers c_e. Overload paths do not have capacity limits, but also must be assigned guide numbers c_p. Each overload path should also have an associated new circuit that can be added to the network with a real valued number guide number c_e, and power capacity limit. The guide number is used as a weight in the objective function to represent the costs related with routing power onto the existing circuits or overload paths.

In lieu of existing line and overload path guide numbers (costs), the length of the existing and overload paths can be used instead. Garver's algorithm suggests using $c_e = d_e$ and $c_p = 5d_p$. Recall from the Chapter 4 nomenclature above that d_e are the node-to-node distances of existing transmission lines and d_p are the node-to-node distances of overload paths (potential transmission line paths).

A.1.2 Problem Formulation and Solution

Garver's original 1970 work presented a narrative description of the problem and its accompanying solver. As stated succinctly here in Section 4.2, Garver's method minimizes the sum of the weighted absolute value of the flow on each line, with the guide numbers serving as the line weights. The flow on existing lines is constrained by

the capacity limit, while the flow on the overload paths is unconstrained. The network constraints consistent with Garver's formulation are summarized in (A.1). The quantity of bus potential is calculated and assigned to each bus in the network via the method described in Section A.1.4. Since the bus potential quantity is only used in Garver's solver presented in [13] and is not implemented in the linear programming formulation, further discussion regarding bus potential are moved to Chapter 7.

$$\min \sum_{\forall i,j} c_{i,j}|p_{i,j}| \tag{A.1a}$$

$$\text{subject to} \tag{A.1b}$$

$$\text{KCL: Sum of all power flowing into a bus is 0} \tag{A.1c}$$

$$p_{i,j} = \begin{cases} 0 & \text{if the bus potential difference from } i \text{ to } j \text{ is less than the line guide number} \\ p_{\max} & \text{if bus potential } i \text{ - bus potential } j > \text{guide number} \\ -p_{\max} & \text{if bus potential } j \text{ - bus potential } i > \text{guide number} \\ & \text{Determined using KCL for lines in the chosen spanning tree} \\ & \text{or when bus potential } j \text{ - bus potential } i = \text{guide number} \end{cases} \tag{A.1d}$$

Where $p_{\max}$ is the power limit of the line from bus i to bus j, $p_{i,j}$ is the flowing on the path (existing or potential) connecting bus i to j, and $c_{i,j}$ is the guide number of the path connecting bus i and j.

To compute the flows on the existing and potential transmission line paths, Garver proposed his own linear programming solver in the the appendix of his paper [13]. However, this solver is less effective than classic simplex approaches such as [132], and certainly inferior to modern LP solvers. To realize the benefits these commercially available LP solvers, Garver's problem formulation in A.1 was reformulated into a standard form LP in Section 4.2.

A.1.3 Limitations and shortcomings of Garver's algorithm

First, as one might expect of an optimization formulation that minimizes an L1 norm of flows, this formulation favors solutions for line flows that are sparse while satisfying the KCL constraints. As a result, the networks created are typically highly radial and tree-like, e.g. Fig 4.3a. To resolve this issue and create a more realistic network topology, multiple generator dispatch scenarios are computed and additional iterations of the outer loop are run as explained in Section 4.4.1.

Unfortunately, both the original Garver algorithm in [13] and the reformulation in Section 4.1 generate power system network graphs with a higher percentage of radial "branches" than is observed in existing networks. As further explained in Section A.1.3, this is because the most efficient way to minimize the total MW-miles of flow in the network is by creating a radial, tree-like network. Since only one net injection vector is utilized in Garver's paper and the revised KCL-compliant NCA, multiple net-injection scenarios can be utilized to increase the meshing of the network thus decreasing the average cycle length. Section 4.4 explains how multiple iterations of the KCL-compliant

NCA are run with different net-injection vectors (P).

Next, the sharp division between the cost weights (guide numbers) for existing transmission lines and potential transmission paths creates several buses with abnormally high node degrees, as seen in Fig 4.3a. This is partially caused by Garver's original suggestion of only two weight coefficients, one for "overload paths", hence termed potential transmission line paths, and one for existing lines. This issue is remedied using by assigning an intermediate weight coefficient of $3d$ for potential (overload) paths in parallel with existing lines as explained in Section 4.2.1.

Finally, since the transportation model minimizes the weighted L1 norm of the line flows, power may be routed on any path as long as the KCL constraints are satisfied at every node. The optimal solution of (4.1) results in 0 flow on the majority of the potential transmission line paths. However, when modeled using DC power flow, there are non-zero flows on all existing and potential transmission line paths in the network. Thus, there is a difference, often substantially so, between the flows calculated using the transportation model and the DC power flow model. This discrepancy may be overcome in future work by modifying the networks created in this dissertation using the other non-dashed/dotted portions of the full NCA in Fig. 4.1. Future work also explores the impact of modifying the optimization formulation to directly model DC power flow.

There several categories of difficulties and limitations of Garver's original algorithm that are presented in this section. The first two categories can be solved by the reformulations and improvements suggested in Section 4.1. However, the optimization-based

algorithm presented in Section 4.1 cannot address all of the disadvantages of Garver's algorithm and thus the next three categories of limitations must be addressed by heuristic additions explained in Chapter 7. The addition of these heuristics result in multiple additional iterations of the multi-level linear program as well as other disadvantages listed in later sub-sections, thus motivating the creation of a more robust set of optimization-based network construction algorithms in Chapter 5.

Non-standard problem description and inefficient solver

Although Garver's algorithm is an early work presenting the idea of implementing transmission expansion by minimizing a weighted L1 norm, he does not use standard formulations or notations that were commonly accepted at the time. For example, Garver did not mention that his formulation is utilizing the linear transportation model, although this formulation was known several decades before Garver's work [71]. Also, Garver did not utilize common matrix notation for his formulation summarized in (A.1). This includes his lack of implementation of the incidence matrix [133] utilized for Kirchoff's Current Law [134]. Also, Garver proposed his own linear programming solver instead of utilizing a more efficient solvers that were available at the time. As further explained in Section A.1.4, Garver's own solver utilizes a quantity termed "bus potential" to calculate the flows on each existing and potential network element/path.

While Garver's narrative problem formulation can be easily updated to a standard form linear optimization problem as seen in Section 4.2 (4.1), his proposal to utilize bus potentials in his own solver is both unnecessary and potentially confusing.

Slow algorithmic performance

In addition to the unnecessarily slow algorithmic performance of Garver's proposed solver, his method of adding new elements to the network results in extraneous additional iterations of his transmission expansion algorithm. This process occurs after the linear program finds an optimal solution and assignment to the p vector in what is termed the inner loop in Section 4.2. Then, control is passed to what is termed the outer loop in Section 4.4 where heuristic methods are used to add elements to the existing network or update parameters for existing elements.

After the termination of each instance of the inner loop linear program, Garver suggests adding one line corresponding to the overload path with the highest flow. However, each new line has a capacity of 100 MVA regardless of the value of the flow on the corresponding overload path. This results in multiple unnecessary iterations of the inner loop, since multiple lines are added in parallel to reach the total capacity needed to transport power along a given path. To solve this inefficiency, Section 4.4 proposes adding multiple new transmission lines per iteration of the outer loop with ratings determined by the flows on the potential transmission line paths.

Finally, the sparsity of the p vector requires keeping all potential transmission line paths in the network until the last iteration of the outer loop. This results in a significant computational burden for large networks, since the incidence matrix for the complete graph needs to be used for every iteration of the inner loop until the final outer loop iteration terminates. Thus, it would be advantageous if a new NCA could be developed

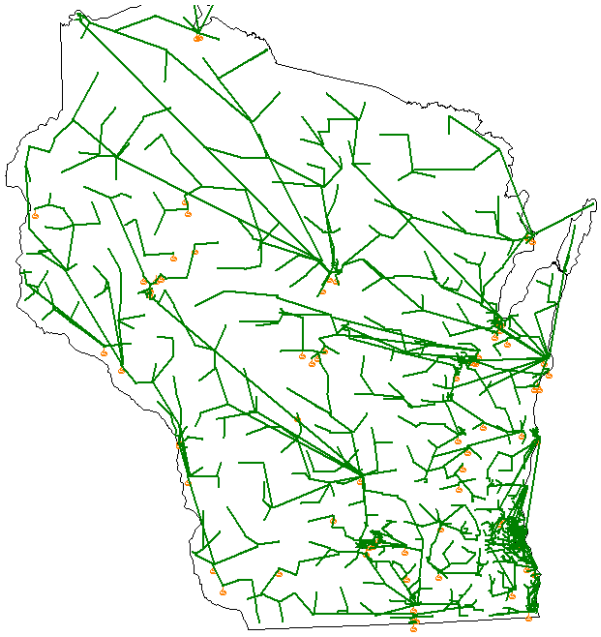


Figure A.1: One-Line of a tree-structured network

that would reduce the number of potential transmission line paths each iteration of the inner or middle loop.

Graph metrics for judging network realism

First, as one might expect of an optimization formulation that minimizes an L1 norm of flows, this formulation favors solutions for line flows that are sparse while satisfying the KCL constraints. This tendency of an objective function minimizing an L1 norm to create sparse solutions is explored in more detail in Section 5.1.2. As a result, the networks created are typically highly radial and tree-like. Since the objective function minimizes total MW-miles in the network, the optimal solution to this problem would be a tree-structured network. Fig. A.1 shows an example network graph (one-line) with this radial tree-like structure. To resolve this issue and create a more realistic network

topology, multiple generator dispatch scenarios are computed and additional iterations of the outer loop are run as explained in Section 4.4.

Next, the sharp division between the cost weights (guide numbers) for existing transmission lines and potential transmission paths creates several buses with abnormally high node degrees. This is partially caused by Garver's original suggestion of only two weight coefficients, one for potential transmission line paths and one for existing lines. When an existing line has power flow p at its capacity limit, any additional flow must be routed on the parallel overload path with $c = 5d$ as explained in Section A.1. Since overload paths are weighted by their length, the least cost route is a direct path between buses i and j. Thus, buses with large loads or generators will often have existing lines that reach their capacity after a given iteration of the inner loop, resulting in additional lines being added to directly connect other buses. However, these additional "direct routes" are sometimes unnecessary, since it is occasionally more realistic to increase the capacity of existing lines that already connect bus i to j instead of building another direct transmission line. This issue is remedied using by assigning an intermediate weight coefficient of $3d$ for potential (overload) paths in parallel with existing lines as explained in Section 4.2.2.

Transportation model to to DC power flow

Unfortunately, the modifications above and further developed in Section 4.1 are unable to address some of the simulation inaccuracies resulting from implementing a modified version of Garver's algorithm. To correct these inaccuracies subsequent algorithmic

additions are developed in Chapter 7, as summarized below.

After the termination of the outer loop, the only graph edge constraints included in the network model are the line limits. Thus, network elements with associated parameters and constitutive equations need to be assigned to each network graph edge, as detailed in Section 7.1.1. Once modeled using DC power flow that includes the constitutive relations of transmission lines, the calculated power flows on the newly assigned network elements are often substantially different from the flows assigned from the transportation model. These differences result in numerous lines being either substantially overloaded or significantly underutilized. Thus, a method is needed to modify the network element parameters so that the DC power flow solution is DC OPF feasible, which is defined as a network that result in a DC power flow solution that satisfies all of the elements' power limits. Due to its computational tractability for large networks, a sequential DC power flow is utilized to modify the element parameters so that the network is DC OPF feasible. The process is explained in further detail in Section 7.1.

Computational experience to date suggests that iterating between running a DC power flow and updating line parameters usually results in a network that mostly satisfies the element power limits. However, this method often results in some elements that are either overloaded or substantially underutilized. Additionally, the parameters assigned are not necessarily reproducible due to the stochastic nature of the sequential DC power flow. These two issues are discussed in more detail in Section 7.1.2.

Also, the stochastic aspects of the sequential DC power flow results in a different set

of lines being modified each time the power flow is computed, resulting in different sets of networks being created for the same input set of buses and net injection(s). While this can be beneficial for creating a family of networks using only one set of input data, it keeps the entire KCL-compliant LP plus heuristics from being truly reproducible.

Modeling of multiple voltage levels

Since KCL-compliant network construction algorithms (NCA) such as Garver's method utilize the transportation model, the only network parameters considered are the L1 norm weighting factors and line limit constraints. This limits KCL-only NCAs to represent the network in a per-unit representation without considering different voltage levels at each substation. Since impedances are not modeled by the multi-level linear program in Section 4.1, the voltage levels are assigned as part of the sequential DC power flow described in Section 7.1. However, voltage levels are assigned to the conductors themselves instead of directly to individual buses in the steps in Section 7.1 after the first run of the multi-level linear program has terminated.

Thus, if the KCL-compliant method is implemented for transmission expansion on an existing network, either two iterations of the multi-level LP must be run, or the potential transmission line paths have to be limited to connecting buses at the same voltage level. Instead of substantially modifying the outer loop in Section 4.4 to address this issue, it is desirable to create a new class of algorithms that directly models network element impedance, and by extension voltage level.

A.1.4 Heuristic Additions and Solver Suggested by Garver

In [13], Garver suggested a solver that utilizes a quantity called bus guide potentials (also called bus potentials, guide potentials, or just potential) to calculate the flow on the existing and overload paths. Each bus has a calculated bus potential, which Garver identifies as having similar properties to voltage in circuit analysis, i.e. the guide potentials are used in the element constitutive relations and Garver's equivalent of KVL. The bus potential is calculated by first selecting any spanning tree from the network of existing circuits and overload paths. The flows on the non-tree overload paths are set to 0, while the flows on non-tree existing lines are set to either 0 or their limit. Then, starting at an end branch in the tree, KCL is utilized to compute the flows on each line in the selected tree. A reference node is selected and assigned a bus potential of 0. Starting from this reference node, each connected node in the tree is assigned a bus potential by subtracting the link guide number (cost) from the sending bus potential. If power is flowing from the receiving bus to the sending bus, the link guide number is added to the sending end bus potential to calculate the receiving end bus potential, otherwise the link guide number is subtracted from the sending end bus potential. After calculating the bus potentials at all buses in the selected tree, [13] evaluates the choice of buses in the tree using the criteria in (A.1) and step 5 of his algorithm in [13].

Garver presents an example 6-bus system where he demonstrates how his formulation (summarized in Section A.1.2) is used for transmission expansion on the 6-bus system. However, in addition to the disadvantages listed in Section A.1.3, there this

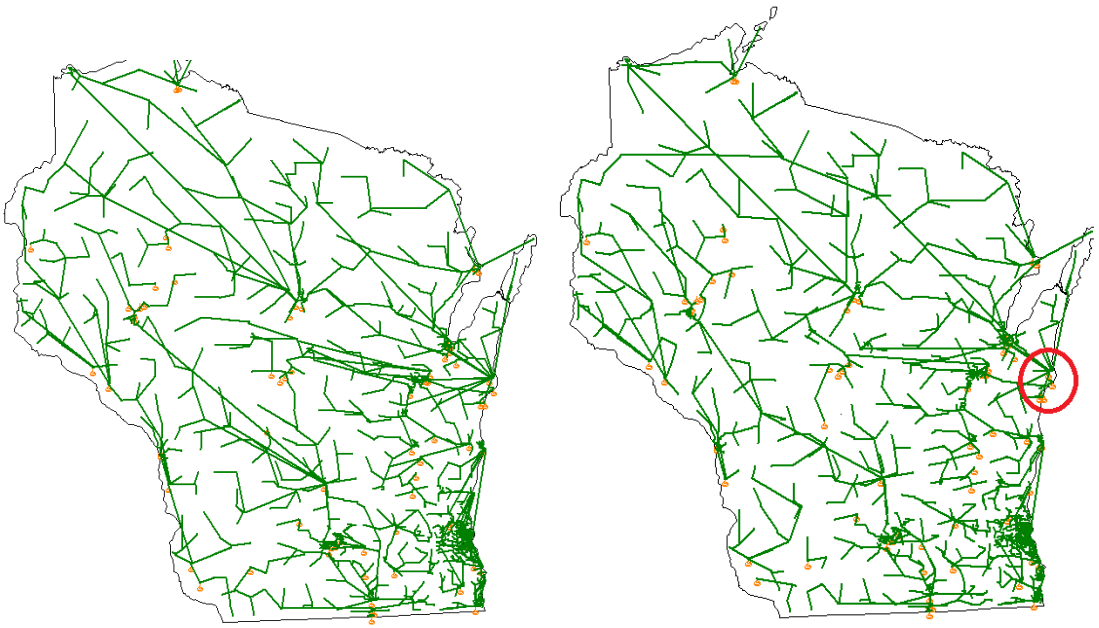
no guarantee that Garver's solver will produce the same results as directly minimizing weighted L1 norm of the line flows as formulated in Section 4.2. Thus, Garver's solver is disregarded, and the multi-level optimization formulation presented in Section 4.1 is utilized instead.

A.2 Refinements to Garver's algorithm: Exploiting its formulation as an L1 minimization subject to linear constraints

The proposed changes suggested in this chapter to Garver's established established linear programming-based algorithm resulted in updated algorithms that were significantly more effective in creating or expanding realistic large-scale electrical power networks. Section 4.2 gives a standard form LP reformulation of Garver's original algorithm that can be solved efficiently using LP solvers. Next, Section 4.4 constructs an outer loop based on Garver's original algorithm that adds or upgrades elements in the existing network. However, to address the limitations of the optimization-based KCL-only formulation listed above, additional heuristic steps and modifications are presented in Chapter 7. Section 7.3) presents additional algorithmic steps that can be used to create large continental-scale synthetic power system models with multiple voltage levels and operational areas (Sections 7.1 - 7.2.2).

A.3 Creating intermediate line cost

As detailed in Section 4.2.2, an intermediate cost of $3d$ was assigned to potential transmission line paths in parallel to existing transmission lines, where d is the length of the line. This value corresponds to the cost associated with replacing circuit conductor(s) or upgrading the voltage on an existing transmission line. Since the process of upgrading an existing transmission line circuit is not as expensive as constructing a new transmission line, the numerical multiplier was chosen to be greater than the weight for existing lines ($1d$) and the cost of constructing new lines ($5d$). After testing values ranging from 2 to $5d$ using 2-3 test cases, $3d$ was empirically determined as the best value to achieve the target node degree distribution. The choice of $3d$ is also supported by [120] which lists the cost of upgrading an existing transmission line as between 35% - 60% of the cost of building a new line. For subtransmission (69kv - 116kV) lines in CAISO, the estimated cost of upgrading an existing transmission line ranges from 60% to 80% of the cost of constructing a new transmission line [121]. The 60% value of $\frac{3}{5}d$ chosen in Section 4.2.2 matches both the upper range of [120] and the lower range of [121]. A one-line diagram illustrating an example network resulting from the use of this intermediate cost weighting factor is shown in figure A.2. The results are somewhat subtle, but can be seen when comparing the bus near Point Beach generating station in eastern Wisconsin near Lake Michigan, which is circled in red in figure A.2b. In figure A.2a, the Point Beach bus has a higher node degree than the same bus in figure A.2b, since adding an intermediate weighting factor causes more lines to be upgraded instead



(a) One-Line of a tree-structured network (b) One-Line with reduced cost of line upgrades

Figure A.2: One-Line Diagrams of Wisconsin 1664 bus system illustrating changes in node degree

of building an excess of new lines.

Appendix B

Additional Data Calculations

B.1 Generator Economic Dispatch Calculations

To compute the vector of net power injections to a bus, P, two categories of input data are needed: total system load and generator cost curves. For the calculations in this dissertation, synthetic load data are utilized with the non-concurrent peak scaled up by 25% to serve as a projection of future load growth. Synthetic generator cost curves were created as described in [108]. Next, an economic dispatch [30] is performed utilizing the sum of the peak planning load and the synthetic generator cost curves. The objective function is to minimize generator cost, and constraints are that the total generation output must equal the total load, and that the generator MW output must be within its upper and lower limits. A normalized plot of the generator outputs for the upper-Midwest region is shown below in Fig. B.1.

The authors also utilized an alternative generator dispatch as assigned in Section B.2. However, any additional generator dispatches that meet the constraints above can be used to increase the number of net power injection scenarios.

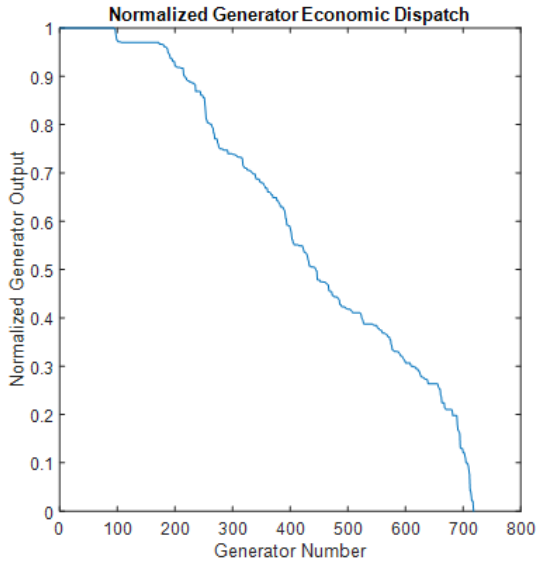


Figure B.1: Normalized generator power outputs from the economic dispatch

=2ws

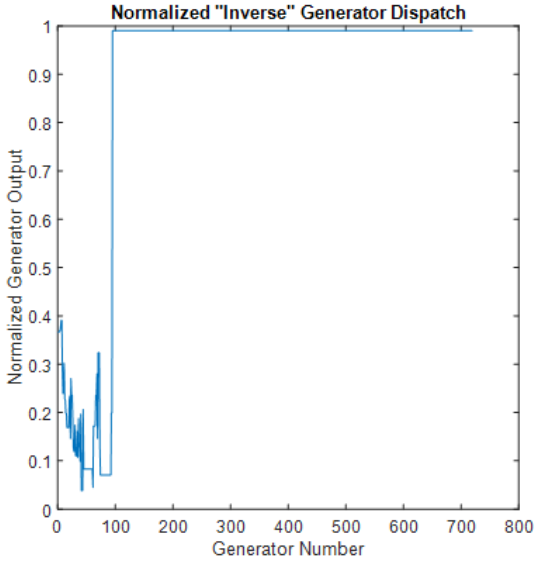


Figure B.2: Normalized generator power outputs from the complementary dispatch

B.2 Complementary Generation Dispatch

After the economic dispatch (Fig. B.1) is performed in Section B.1, the resulting generator output power assignment is used to compute an alternative dispatch, termed the complementary dispatch. Any generator with real power output $P_{G,i}$ less than 85% of the corresponding maximum power output $P_{g,i}^{\max}$ is constrained to be at it's maximum output. Then, the lower bounds of the generators not set to their maximum output are set to 0 to allow the complementary dispatch to be feasible. Finally, the optimization problem is solved, and a dispatch for all of the generators is computed.

However, if the total system load is below the sum of the modified lower bounds for the generators, the total load is proportionally scaled to be equal to the sum of the modified lower bounds. This ensures that the union of the generator power outputs from

the economic dispatch and complementary dispatch results in each generator being at or above 85% of it's maximum power output. This allows for the networks created to have transmission capacity to successfully deliver power from any of the generators in the network.

The result of applying the complementary dispatch is shown in Fig. B.2. First, the normalized result of the economic dispatch is plotted in Fig. B.1. The generators are sorted by their normalized real power output, and assigned a reference number to generate the plot of the complementary dispatch. Notice in Fig. B.1, approximately 100 generators are dispatch at or near their maximum output, with the rest of the generators are dispatched at a lower power output corresponding to their respective costs. Thus, as shown in Fig. B.2, the complementary dispatch sets generators numbered 100 and onward equal to their maximum output. The remaining generators (numbered 1-100) are given a new lower bound of 0, and their output under the complementary dispatch is shown in Fig. B.2.

B.3 Node-to-node distance Calculation

The Haversine formula is used to calculate the distances between each bus [119]. First published by James Andrew in 1805, the Haversine formula calculates the orthodromic distance, or the distance between two points on a sphere as measured along the surface of the sphere. Although the earth is not a perfect sphere, assuming it is a perfect sphere results in at most a 0.5% error for latitude, 0.2% for longitude [135]. The bus-to-bus

distances are represented in matrix D, with $d_{i,j}$ being the distance between bus i and bus j in meters.

Haversine

Nomenclature	**Description**
ϕ_i	**Latitude** of bus i in radians
ψ_i	**Longitude** of bus i in radians
R	**Radius** of the earth in meters (mean radius = $6,371$ km)
$d_{i,j}$	**Distance** from bus i to bus j in meters
h	intermediate variable

$$h = \sin^2(\frac{\phi_i - \phi_j}{2}) + \cos(\phi_i) \cdot \cos(\phi_j) \cdot \sin^2(\frac{\psi_i - \psi_j}{2}) \tag{B.1a}$$

$$d_{i,j} = 2 \cdot R \cdot \text{atan2}(\sqrt{h}, \sqrt{1-h}) \tag{B.1b}$$

Appendix C

Additional Metrics for Synthetic Networks

Tables C.1 - C.5 show the branch and bus metrics for all the networks on the 13 geographic footprints, in addition to the larger networks spanning multiple geographic footprints. The metrics are explained in more detail in Section 6.5.1

Region	Metrics (miles)	KCL-Only Networks (kV)					KCL/KVL Networks (kV)					
		69	**138**	**230**	**345**	**500**	**12.9**	**69**	**138**	**230**	**345**	**500**
AZNM	% of Lines	87.7%	8.9%	1.7%	1.3%	0.3%		71.7%	17.1%	4.1%	6.3%	0.8%
	GVA-Miles	1403	1949	1144	2069	4748		1406	4114	8156	47830	3506
	Total Len	19790	9733	2025	1845	2560		20345	21764	19412	59593	1403
	Mean	3.8	18.3	19.7	23.7	128.0		3.6	15.9	59.6	118.2	23.4
	Std	6.3	23.0	29.9	42.6	116.3		4.9	24.7	59.1	80.1	46.6
	10%	0.6	0.9	1.0	1.8	14.4		1.0	1.0	1.0	1.2	0.8
	50%	1.3	5.6	7.9	7.8	105.6		2.0	2.6	30.2	157.6	4.8
	90%	10.7	53.8	59.8	42.4	321.6		7.4	55.3	130.8	197.7	64.3
	Num Buses	3953	531	115	86	34	630	3784	1247	333	621	72
	% of buses	83.8%	11.3%	2.4%	1.8%	0.7%	9.4%	56.6%	18.6%	5.0%	9.3%	1.1%
CAISO	% of Lines	92.7%	5.9%	0.8%	0.4%	0.1%		80.2%	15.7%	1.7%	1.9%	0.5%
	GVA-Miles	1234	888	433	982	1026		1673	3597	2286	6763	2385
	Total Len	18622	4531	753	774	340		26053	15129	4103	7864	832
	Mean	1.9	7.3	8.5	18.4	24.3		3.0	9.0	23.1	37.6	14.1
	Std	3.3	13.9	15.4	22.7	37.4		3.8	16.6	40.4	64.1	33.2
	10%	0.5	0.6	0.7	1.1	0.6		0.8	0.9	1.0	0.9	0.6
	50%	0.9	1.9	2.6	8.2	7.2		1.9	2.4	3.7	3.4	3.6
	90%	4.0	17.6	26.3	45.3	100.9		6.1	41.1	110.4	157.7	40.7
	Num Buses	7556	658	122	46	22		6537	1448	169	180	70
	% of buses	89.9%	7.8%	1.5%	0.5%	0.3%		77.8%	17.2%	2.0%	2.1%	0.8%
NWPP	% of Lines	82.3%	11.9%	3.3%	2.0%	0.5%		70.4%	18.8%	3.6%	5.1%	2.1%
	GVA-Miles	1212	1840	1328	3099	2720		1309	4224	4185	24639	43426
	Total Len	17343	8691	2364	2855	1012		20169	21216	9727	32725	24932
	Mean	5.3	18.5	17.9	36.6	53.3		4.8	18.8	44.8	106.9	197.9
	Std	7.4	21.2	27.9	52.4	57.5		6.3	24.7	54.3	86.7	71.1
	10%	0.8	1.2	1.5	1.9	4.3		1.0	1.0	1.0	1.0	22.8
	50%	2.0	9.4	6.9	16.4	31.6		2.2	4.5	5.8	153.5	222.2
	90%	15.3	51.8	49.5	110.6	157.2		12.9	57.0	123.2	200.2	238.1
	Num Buses	2520	420	152	86	27	775	2509	1067	219	333	199
	% of buses	78.6%	13.1%	4.7%	2.7%	0.8%	15.2%	49.2%	20.9%	4.3%	6.5%	3.9%
WECC Real Network	% of Lines	31.9%	47.4%	17.0%	1.2%	2.4%						
	GVA-Miles	2682	8241	26278	11260	46653						
	Total Len	29408	56759	52451	11879	18491						
	Mean	6.0	7.8	19.9	62.2	49.2						
	Std	7.2	10.9	44.1	53.2	49.7						
	10%	0.5	0.5	0.9	4.2	1.0						
	50%	3.5	3.9	10.6	47.6	36.2						
	90%	14.3	19.3	48.0	136.6	104.9						
	Num Buses	8308	5693	1899	184	274						
	% of buses	50.8%	34.8%	11.6%	1.1%	1.7%						

Table C.1: Branch and bus statistics for synthetic networks on WECC geographic footprint vs real WECC network

Region	Metrics (miles)	KCL-Only Networks (kV)					KCL/KVL Networks (kV)					
		69	**138**	**230**	**345**	**500**	**12.9**	**69**	**138**	**230**	**345**	**500**
MISOSPP Full Dispatch	% of Lines	60.3%	30.3%	4.6%	4.7%			69.9%	15.9%	4.1%	10.1%	
	GVA-Miles	5614	12015	8861	39245			12409	16383	26656	326504	
	Total Len	70548	54428	16991	28367			76628	71805	58740	399544	
	Mean	4.7	7.2	14.8	24.4			3.4	14.0	44.9	123.4	
	Std	5.4	8.9	19.2	32.7			4.3	22.7	55.3	73.0	
	10%	0.6	0.9	1.6	2.4			1.0	1.0	1.0	2.5	
	50%	2.2	4.3	8.1	12.5			1.9	2.5	2.7	154.4	
	90%	12.0	16.0	36.1	61.8			7.6	50.0	123.1	190.5	
	Num Buses	8284	4910	709	981		2589	14371	4535	1246	4065	
	% of buses	55.7%	33.0%	4.8%	6.6%		9.7%	53.6%	16.9%	4.6%	15.2%	
MISOSPP Area Dispatch	% of Lines	60.3%	30.3%	4.6%	4.7%			72.5%	17.8%	3.4%	6.3%	
	GVA-Miles	5614	12015	8861	39245			13489	18800	17803	183489	
	Total Len	70548	54428	16991	28367			89233	78087	36764	220431	
	Mean	4.7	7.2	14.8	24.4			3.9	13.8	33.9	110.5	
	Std	5.4	8.9	19.2	32.7			5.0	20.7	49.5	77.9	
	10%	0.6	0.9	1.6	2.4			1.0	1.0	1.0	2.3	
	50%	2.2	4.3	8.1	12.5			2.0	3.1	3.1	151.1	
	90%	12.0	16.0	36.1	61.8			9.8	47.4	116.3	190.2	
	Num Buses	8284	4910	709	981		2589	15092	5197	1052	2560	
	% of buses	55.7%	33.0%	4.8%	6.6%		9.8%	57.0%	19.6%	4.0%	9.7%	
PJM Full Dispatch	% of Lines	73.5%	15.5%	3.1%		7.9%		75.3%	15.0%	3.5%		6.2%
	GVA-Miles	2908	4849	3946		31730		7637	5927	4354		31730
	Total Len	44981	17406	4699		11554		48353	17945	4903		11554
	Mean	4.1	7.5	10.3		9.7		3.4	6.3	7.4		9.7
	Std	3.7	12.1	17.2		23.7		3.5	11.2	15.0		23.7
	10%	1.2	1.2	1.3		1.4		1.0	1.0	1.0		1.4
	50%	3.0	3.1	4.4		4.0		2.2	2.5	2.5		4.0
	90%	8.5	16.1	40.3		13.8		7.3	11.8	14.4		13.8
	Num Buses	8145	2171	431		1144	1095	10351	2751	636		1173
	% of buses	68.5%	18.3%	3.6%		9.6%	6.8%	64.7%	17.2%	4.0%		7.3%
PJM Area Dispatch	% of Lines	80.7%	15.8%	1.6%		1.9%		82.2%	14.6%	1.7%		1.5%
	GVA-Miles	4465	7538	2424		40621		8237	8120	2562		40621
	Total Len	63309	28425	3742		23170		66203	28716	3811		23170
	Mean	5.3	12.1	15.6		82.8		4.4	10.9	12.3		82.8
	Std	5.1	16.6	30.4		79.3		4.8	16.1	27.5		79.3
	10%	1.2	1.3	1.2		1.7		1.0	1.0	1.0		1.7
	50%	3.5	4.8	4.9		43.5		2.6	3.9	3.2		43.5
	90%	11.9	43.2	46.0		173.3		10.7	42.4	25.5		173.3
	Num Buses	9400	1935	195		361	1095	11013	2369	297		371
	% of buses	79.1%	16.3%	1.6%		3.0%	7.2%	72.7%	15.6%	2.0%		2.5%
EI Real Network	% of Lines	42.6%	44.6%	8.9%	2.9%	0.7%						
	GVA-Miles	12240	45642	42219	68481	48071						
	Total Len	121605	195218	72335	51422	17146						
	Mean	4.7	7.3	13.5	29.1	37.9						
	Std	5.3	8.5	17.6	30.8	35.3						
	10%	0.7	0.7	0.7	2.5	1.0						
	50%	3.2	4.4	7.4	20.3	28.6						
	90%	10.4	17.1	33.1	65.1	87.7						
	Num Buses	35276	24336	4761	1368	347						
	% of buses	53.2%	36.7%	7.2%	2.1%	0.5%						

Table C.2: Branch and bus statistics for large-scale synthetic networks vs real EI network

Region	Metrics (miles)	KCL-Only Networks (kV)						KCL/KVL Networks (kV)						
		69	**138**	**230**	**345**	**500**	**765**	**12.9**	**69**	**138**	**230**	**345**	**500**	**765**
NNEAST	% of Lines	80.8%	15.2%	1.3%	1.6%	1.1%			81.9%	14.1%	1.6%	1.5%	0.9%	
	GVA-Miles	3871	6483	2178	6113	29611			8168	7197	2404	6199	29611	
	Total Len	57499	23735	3161	6573	16144			60835	24092	3274	6616	16144	
	Mean	4.0	8.9	13.4	23.2	86.3			3.5	7.9	9.4	20.2	86.3	
	Std	4.5	13.6	27.0	49.1	99.4			4.2	13.1	23.0	46.4	99.4	
	10%	0.7	0.9	0.7	1.0	1.6			0.7	1.0	0.8	1.0	1.6	
	50%	2.4	3.6	3.7	4.2	9.9			1.7	3.0	1.5	3.2	9.9	
	90%	9.9	26.9	41.7	146.7	217.5			8.6	18.1	16.5	45.9	217.5	
	Num Buses	10977	2256	213	203	231		1266	12811	2734	346	326	246	
	% of buses	79.1%	16.3%	1.5%	1.5%	1.7%		7.1%	72.3%	15.4%	2.0%	1.8%	1.4%	
NEISO	% of Lines	95.1%	4.5%	0.3%	0.1%				92.6%	6.3%	0.7%	0.5%		
	GVA-Miles	801	353	52	70				1929	1059	713	793		
	Total Len	12417	1394	94	67				33912	5913	1903	1000		
	Mean	3.1	7.3	7.2	22.2				6.9	17.9	54.4	41.7		
	Std	4.1	12.7	6.8	19.1				6.7	24.7	65.9	76.4		
	10%	0.5	0.6	0.8	0.6				1.0	1.0	1.0	0.7		
	50%	1.8	3.4	5.3	29.3				4.7	5.2	3.0	2.7		
	90%	7.0	15.6	18.1	36.7				16.3	58.7	141.4	185.5		
	Num Buses	3255	218	22	6			406	3428	330	54	43		
	% of buses	93.0%	6.2%	0.6%	0.2%			9.5%	80.5%	7.7%	1.3%	1.0%		
NYISO	% of Lines	94.9%	4.2%	0.6%	0.2%		0.0%		85.7%	11.0%	1.5%	1.5%		0.3%
	GVA-Miles	735	283	355	584		1190		802	1958	664	2143		10114
	Total Len	11173	1187	456	465		60		17098	8467	1277	2113		1922
	Mean	1.9	4.6	12.3	31.0		20.1		3.1	11.9	13.3	21.8		106.8
	Std	3.2	9.8	20.6	44.4		14.4		4.6	18.4	31.6	45.0		104.8
	10%	0.2	0.3	0.2	0.5		3.7		0.5	0.6	0.5	0.7		2.2
	50%	0.7	1.6	3.0	5.7		25.5		1.5	3.1	2.7	4.9		72.4
	90%	5.7	9.2	36.6	81.5		31.0		7.8	44.9	27.9	49.9		218.3
	Num Buses	4659	278	46	21		4		4151	646	105	82		24
	% of buses	93.0%	5.6%	0.9%	0.4%		0.1%		82.9%	12.9%	2.1%	1.6%		0.5%
RFCE	% of Lines	92.3%	6.4%	0.8%	0.4%	0.1%			80.2%	16.0%	1.4%	1.5%	0.9%	
	GVA-Miles	1045	457	416	596	1058			1814	3091	970	4059	12051	
	Total Len	15363	1978	641	422	409			25664	12098	1868	4948	6562	
	Mean	2.4	4.5	10.9	16.3	40.9			3.7	8.8	15.2	38.4	86.3	
	Std	2.9	8.0	17.8	25.2	40.8			4.2	14.6	34.3	66.0	99.6	
	10%	0.5	0.6	0.7	1.5	4.9			1.0	1.0	1.0	1.0	1.6	
	50%	1.3	1.9	3.7	6.1	24.4			2.2	3.1	1.9	4.2	9.9	
	90%	5.5	10.2	25.0	40.9	111.2			8.6	41.1	49.9	166.3	217.6	
	Num Buses	4783	466	73	32	17		466	4880	1264	129	160	107	
	% of buses	89.1%	8.7%	1.4%	0.6%	0.3%		6.7%	69.7%	18.0%	1.8%	2.3%	1.5%	
EI Real Network	% of Lines	42.6%	44.6%	8.9%	2.9%	0.7%	0.2%							
	GVA-Miles	12240	45642	42219	68481	48071	74036							
	Total Len	121605	195218	72335	51422	17146	10313							
	Mean	4.7	7.3	13.5	29.1	37.9	88.9							
	Std	5.3	8.5	17.6	30.8	35.3	55.1							
	10%	0.7	0.7	0.7	2.5	1.0	15.5							
	50%	3.2	4.4	7.4	20.3	28.6	82.8							
	90%	10.4	17.1	33.1	65.1	87.7	160.4							
	Num Buses	35276	24336	4761	1368	347	36							
	% of buses	53.2%	36.7%	7.2%	2.1%	0.5%	0.1%							

Table C.3: Branch and bus statistics for synthetic networks on northeast geographic footprint vs real EI network

Region	Metrics (miles)	KCL-Only Networks (kV)					KCL/KVL Networks (kV)					
		69	**138**	**230**	**345**	**500**	**12.9**	**69**	**138**	**230**	**345**	**500**
FRCC	% of Lines	84.7%	11.5%	2.3%	1.2%	0.2%		74.0%	16.4%	1.9%	5.7%	1.9%
	GVA-Miles	1891	1812	991	1917	3579		2774	5267	2138	30157	12421
	Total Len	26498	6893	1672	1541	1386		35513	19903	4153	35283	4156
	Mean	3.2	6.2	7.6	12.8	57.7		3.9	9.8	17.8	49.8	17.7
	Std	3.7	9.2	12.9	21.4	72.3		4.6	16.2	36.5	72.7	37.0
	10%	0.7	1.0	1.0	1.6	2.6		1.0	1.0	1.0	1.0	1.7
	50%	1.8	3.2	3.5	7.0	31.5		2.3	3.0	1.4	5.1	5.1
	90%	7.9	13.1	17.5	30.9	213.0		9.1	41.7	105.8	169.1	43.4
	Num Buses	5938	1057	273	133	39	948	6352	1847	227	816	248
	% of buses	79.8%	14.2%	3.7%	1.8%	0.5%	9.1%	60.9%	17.7%	2.2%	7.8%	2.4%
SRMV	% of Lines	82.3%	11.5%	3.7%	1.7%	0.9%		66.1%	19.7%	3.9%	7.0%	3.3%
	GVA-Miles	1298	1353	872	925	2888		1329	4257	3539	23507	58508
	Total Len	18146	4734	1363	734	1411		18630	20502	7918	29356	33117
	Mean	5.2	9.6	8.7	10.3	38.1		4.8	17.8	35.2	71.6	171.6
	Std	5.9	12.5	13.1	10.7	72.7		6.1	23.2	50.8	79.2	86.7
	10%	0.7	1.0	0.9	1.3	3.1		1.0	1.0	1.0	1.0	4.8
	50%	2.5	5.5	5.4	7.0	10.0		2.3	4.6	2.8	11.3	215.2
	90%	12.9	20.3	17.1	24.4	191.7		12.9	52.7	119.9	175.2	223.5
	Num Buses	2532	444	183	86	46	549	2457	1079	217	410	304
	% of buses	76.9%	13.5%	5.6%	2.6%	1.4%	10.9%	49.0%	21.5%	4.3%	8.2%	6.1%
SRTV	% of Lines	84.9%	10.9%	3.4%		0.8%		87.3%	8.0%	1.3%		3.4%
	GVA-Miles	1970	1441	1901		2890		3455	4873	2399		119213
	Total Len	27412	5798	2527		1315		71991	31051	6041		68145
	Mean	5.4	8.8	12.3		28.6		6.4	30.0	36.4		157.7
	Std	5.0	11.5	17.9		49.1		6.1	23.8	51.9		47.1
	10%	1.0	1.2	1.4		2.0		1.0	1.0	1.0		145.6
	50%	3.9	4.9	6.4		10.2		4.3	41.1	2.7		159.3
	90%	11.5	18.8	32.7		82.9		14.6	56.6	118.2		207.5
	Num Buses	3698	629	239		59	719	3997	1216	277		410
	% of buses	80.0%	13.6%	5.2%		1.3%	10.9%	60.4%	18.4%	4.2%		6.2%
SRVC	% of Lines	89.2%	8.9%	1.6%		0.3%		81.9%	15.4%	1.4%		1.3%
	GVA-Miles	1862	1242	1131		1823		3599	6123	1509		21542
	Total Len	26124	5160	1491		779		49664	27477	2313		12278
	Mean	3.5	6.8	11.0		33.9		5.9	17.2	16.3		90.3
	Std	3.9	9.0	15.3		36.0		6.3	20.5	29.6		83.9
	10%	0.7	0.9	1.3		4.6		1.0	1.0	1.0		1.7
	50%	1.9	4.1	5.5		21.2		3.5	6.4	4.4		145.9
	90%	8.2	15.8	27.5		74.5		14.3	49.0	46.9		186.7
	Num Buses	5551	766	170		33	629	6004	1547	146		191
	% of buses	85.1%	11.7%	2.6%		0.5%	7.4%	70.5%	18.2%	1.7%		2.2%
EI Real Network	% of Lines	42.6%	44.6%	8.9%	2.9%	0.7%						
	GVA-Miles	12240	45642	42219	68481	48071						
	Total Len	121605	195218	72335	51422	17146						
	Mean	4.7	7.3	13.5	29.1	37.9						
	Std	5.3	8.5	17.6	30.8	35.3						
	10%	0.7	0.7	0.7	2.5	1.0						
	50%	3.2	4.4	7.4	20.3	28.6						
	90%	10.4	17.1	33.1	65.1	87.7						
	Num Buses	35276	24336	4761	1368	347						
	% of buses	53.2%	36.7%	7.2%	2.1%	0.5%						

Table C.4: Branch and bus statistics for synthetic networks in southeast geographic footprint vs real EI network

Region	Metrics (miles)	KCL-Only Networks (kV)						KCL/KVL Networks (kV)						
		69	**138**	**230**	**345**	**500**	**765**	**12.9**	**69**	**138**	**230**	**345**	**500**	**765**
SPP	% of Lines	84.8%	12.0%	2.1%	1.1%				69.2%	19.6%	4.1%	7.1%		
	GVA-Miles	3121	3610	2032	3895				3036	8209	10188	103707		
	Total Len	44479	15841	3676	2909				38908	37663	23765	127140		
	Mean	4.8	12.0	16.1	24.7				3.8	13.1	39.3	122.1		
	Std	6.7	16.4	25.7	44.0				5.0	21.2	53.4	74.0		
	10%	0.7	1.1	1.1	1.6				1.0	1.0	1.0	2.8		
	50%	1.7	5.6	6.5	7.7				2.1	2.9	2.7	154.7		
	90%	13.6	35.6	35.7	59.0				8.7	48.2	121.2	192.4		
	Num Buses	6706	1292	281	162			1147	6610	2563	592	1352		
	% of buses	79.4%	15.3%	3.3%	1.9%			9.4%	53.9%	20.9%	4.8%	11.0%		
RFCW	% of Lines	90.1%	8.8%		1.0%		0.1%		81.5%	15.9%		2.6%		0.1%
	GVA-Miles	1968	1729		2304		4434		3861	6343		14647		21141
	Total Len	28078	6460		1802		554		56911	25109		15925		1674
	Mean	3.2	7.6		18.2		110.8		6.1	13.9		53.1		239.1
	Std	3.8	11.8		23.0		23.4		6.5	19.5		55.7		27.7
	10%	0.6	0.9		1.1		79.2		1.0	1.0		1.7		214.8
	50%	1.7	3.6		8.6		118.8		3.6	5.9		16.2		224.1
	90%	7.8	18.6		45.8		134.6		15.2	43.7		127.5		280.7
	Num Buses	6391	887		121		10	721	6715	1718		344		12
	% of buses	86.3%	12.0%		1.6%		0.1%	7.6%	70.6%	18.1%		3.6%		0.1%
MISO	% of Lines	90.6%	7.6%	1.2%	0.6%				78.9%	17.4%	1.7%	2.0%		
	GVA-Miles	2361	1515	595	1915				2726	6077	2939	10551		
	Total Len	33854	6296	1100	1422				39511	24093	5892	11609		
	Mean	4.1	9.1	10.0	28.5				4.6	12.6	31.5	51.8		
	Std	5.5	13.4	17.5	37.9				6.0	19.3	51.9	74.0		
	10%	0.5	1.0	1.2	2.3				1.0	1.0	1.0	1.5		
	50%	1.6	4.5	4.4	15.4				2.0	3.7	4.2	7.4		
	90%	11.3	20.3	27.1	71.4				12.7	51.7	127.4	181.6		
	Num Buses	6046	705	142	75			893	6301	1631	159	280		
	% of buses	86.8%	10.1%	2.0%	1.1%			9.6%	68.0%	17.6%	1.7%	3.0%		
EI Real Network	% of Lines	42.6%	44.6%	8.9%	2.9%	0.7%								
	GVA-Miles	12240	45642	42219	68481	48071								
	Total Len	121605	195218	72335	51422	17146								
	Mean	4.7	7.3	13.5	29.1	37.9								
	Std	5.3	8.5	17.6	30.8	35.3								
	10%	0.7	0.7	0.7	2.5	1.0								
	50%	3.2	4.4	7.4	20.3	28.6								
	90%	10.4	17.1	33.1	65.1	87.7								
	Num Buses	35276	24336	4761	1368	347								
	% of buses	53.2%	36.7%	7.2%	2.1%	0.5%								

Table C.5: Branch and bus statistics for synthetic networks in Midwest and Central geographic footprint vs real EI network

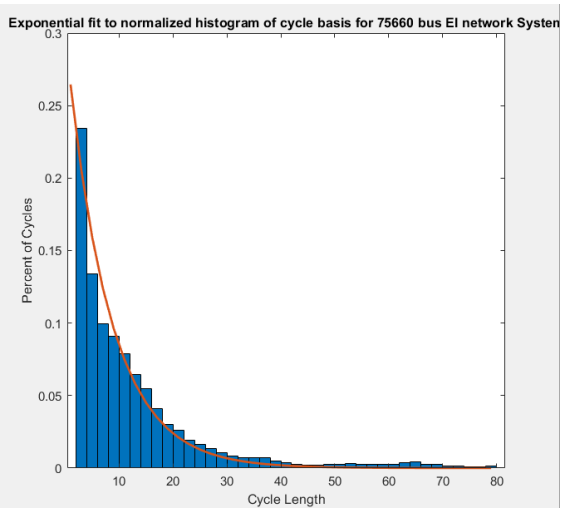
Appendix D

Cycle Basis Analysis

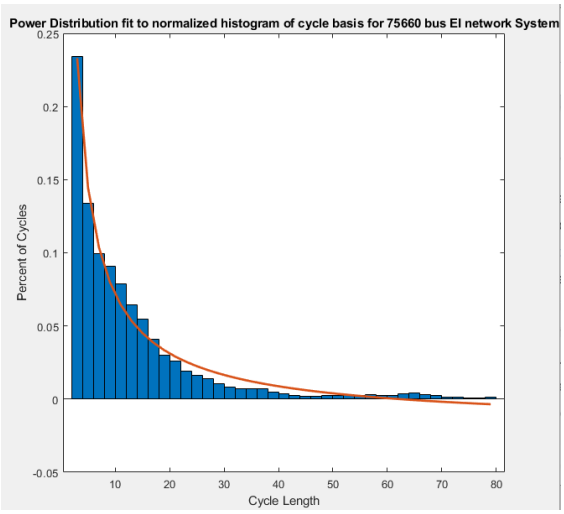
Given the limitations of the cycle basis analysis presented in Section 2.5.3, the preliminary work using this proposed graph analysis metric is presented below in the following sections. Section D.1 presents the initial cycle basis analysis of the existing Eastern and Western interconnect systems, and Section D.2 perform similar analysis on the synthetic networks generated using the algorithms in this dissertation.

D.1 Cycle Basis Analysis for Existing Networks

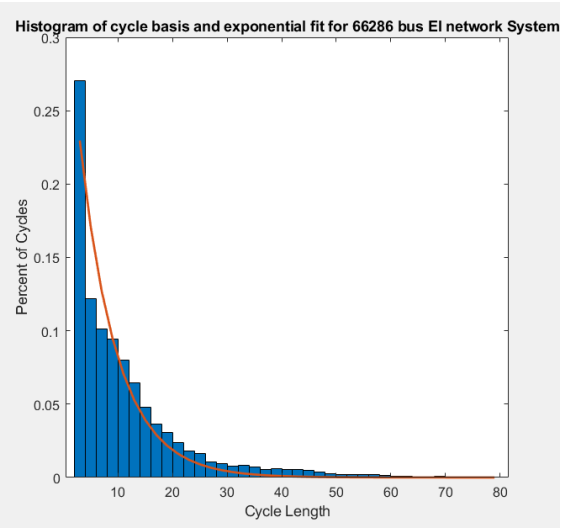
The next graph metric utilized to compare synthetic networks with real power system networks is a cycle basis analysis [136]. A cycle basis is a set of simple cycles that combined form a basis for the cycle space of the graph. The fundamental cycle basis is computed using a minimum spanning tree of the network. Each branch that is not in the minimum spanning tree has an associated fundamental cycle which contains the end



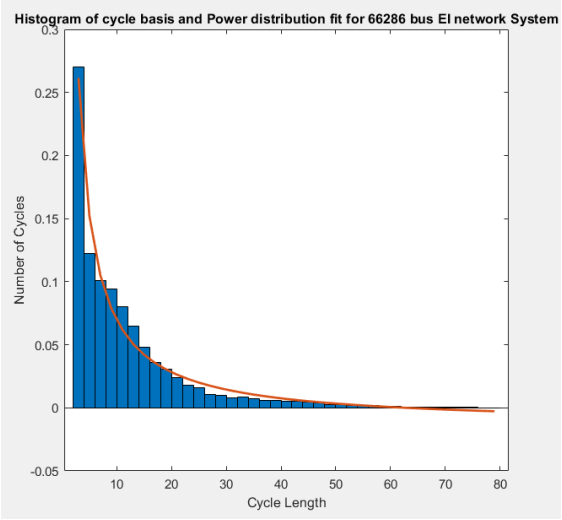
(a) All Branches, Exponential Fit



(b) All Branches, Power Distribution Fit



(c) Only Positive Reactances, Exponential Fit



(d) Only Positive Reactances, Power Distribution Fit

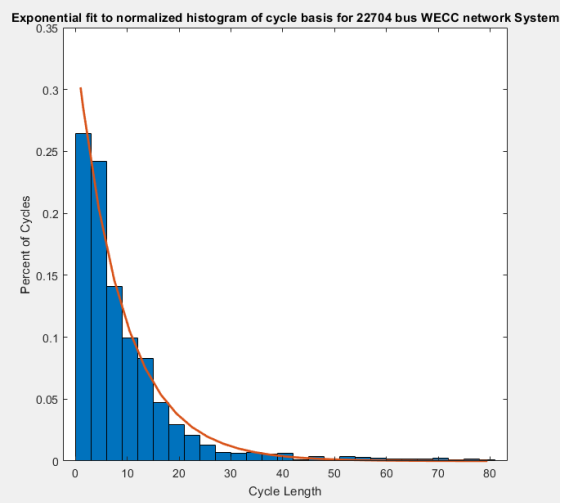
Figure D.1: Eastern Interconnect Curve Fits for Cycle Basis Histograms

nodes of that branch and the path in the minimum spanning tree that connects the two nodes.

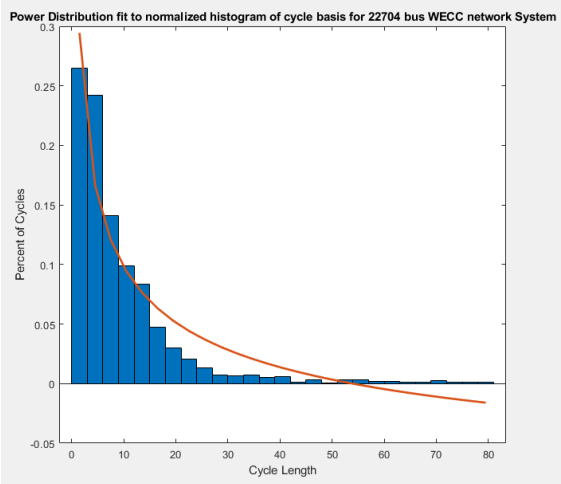
The lengths of the cycles in the cycle basis are utilized here to indicate the degree to which a graph is "meshed". It is hypothesized that graphs with varying randomness, heterogeneity, and modularity will have different shaped histograms of the lengths of the cycles in the cycle basis. While this intuitively makes sense, and is briefly discussed in [137], future work will explore this hypothesis in more detail. However, to compare synthetic with real power system networks, histograms of the lengths of the cycles in the MATLAB-computed cycle basis for each network were created, and were normalized by the total number of cycles to compare networks of different sizes.

A point-to-point comparison of these histograms is difficult since the length of the largest cycle varies across networks, and the discretization of the histogram bins can also vary. Thus, exponential or power distribution curves were fit to this histogram data, and these fit lines were compared across real and synthetic networks.

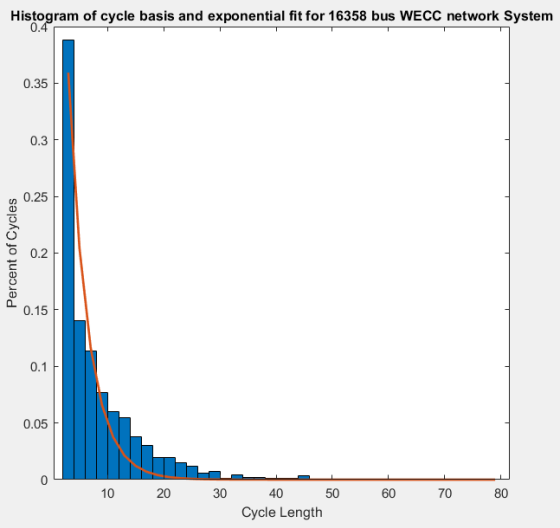
Figures D.1 and D.2 show two different network configurations for the Eastern and Western interconnect, the "all branches" and the "nonnegative impedance" cases discussed previously. For each of these networks, both an exponential and power distribution curve were fit to the histograms of the lengths of the MATLAB-computed cycle basis. While the differences are subtle, the exponential fit is better for the "all branches" cases and the power distribution fit is better for the "only positive reactances" cases. Thus the curve fits in Figures D.1a, D.1d, D.2a, D.2d will be utilized when comparing



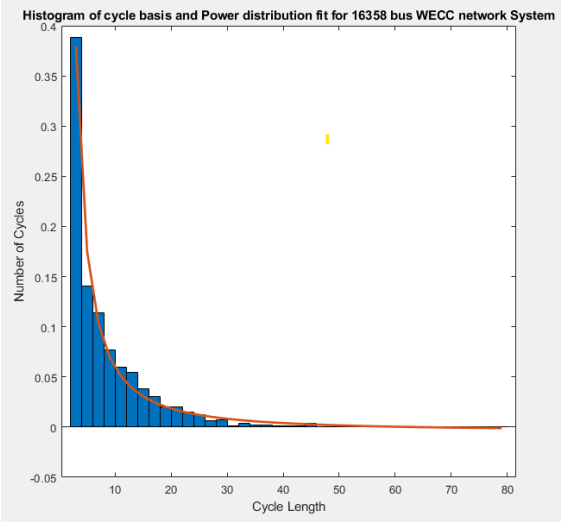
(a) All Branches, Exponential Fit



(b) All Branches, Power Distribution Fit



(c) Only Positive Reactances, Exponential Fit



(d) Only Positive Reactances, Power Distribution Fit

Figure D.2: Western Interconnect Curve Fits for Cycle Basis Histograms

the synthetic networks with the real power system networks.

This difference in the histograms of the MATLAB-computed cycle basis for the two different types of cases is explained by the removal of the fictitious buses used to model three winding transformers. In the all branches case, the three winding transformers have an additional bus in the network that often has negative impedance branches to model the effect of the three winding transformers. Once these negative impedance branches removed, and the fictitious buses are removed, it changes the histograms to have a higher percentage of very short cycles, as seen in figures D.1 and D.2. Additionally, future work will examine metrics for determining how "radial" a system is, and could include statistics such as the ratio of the number of cycles to the number of "radial paths" that are in the minimum spanning tree but not in the MATLAB-computed cycle basis.

In order to quantitatively compare real and synthetic power system networks, the area under the exponential and power distribution curves are computed and used as a "difference metric" in Section D.2.

D.2 Synthetic Network Cycle Basis Analysis

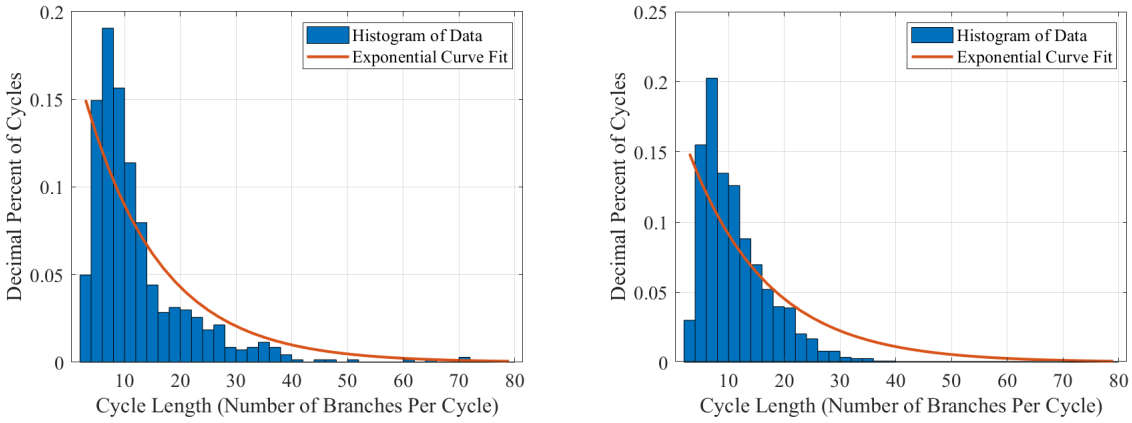
The cycle basis is discussed in more detail in Section D.1, and is utilized in this section to quantify how "meshed" a system is. The cycle basis is computed via the MATLAB routine for each of the six synthetic systems analyze in more detail in this dissertation, and several statistics are calculated. First, the number of branches in each of the cycles in the MATLAB-computed cycle basis is calculated and termed the "length",

and normalized histograms of these values are plotted in Figures D.3 through D.7.

A histogram with larger percentage of short length fundamental cycles may indicate a potential abnormality in the network, most likely caused by a large number of transmission lines traveling in adjacent or nearby corridors. This can be observed in figure D.6e, and verified by the AZNM one-line in figure 6.7a. Additionally, a network with a higher percentage of longer length fundamental cycles indicates that there are a higher percentage of longer loops in the network, as seen in Figures D.6b and D.6d. Although it is harder to see in the synthetic NEISO one-line diagram in Figure 6.6b, the longer length cycles are more visible in the NWPP one-line diagram in Figure 6.6d.

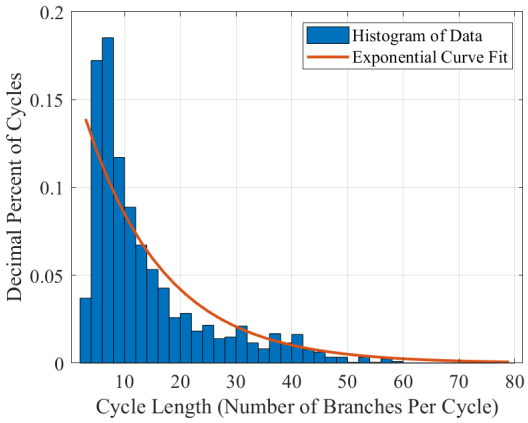
Recall from Section D.1 that the real Eastern and Western interconnect networks have histograms of cycle basis lengths that are reasonably well fit by an exponential or power distribution decay, and thus these distributions are fit to the MATLAB-computed cycle basis histograms for the real networks. The first comparison that can be made between the synthetic and real networks are to compare the plots of these exponential or power distribution fits, as seen in Figures D.8 and D.9.

Several metrics are created and utilized to quantitatively compare the histograms of the lengths of the MATLAB-computed cycle basis across synthetic and real networks, and are presented in Table D.1. First, the average length of the MATLAB-computed cycle basis is computed and listed in the "Mean Cycle Length" column. Then the percent difference between the KCL/KVL synthetic networks and the corresponding Eastern or Western interconnect networks are computed and recorded in the "% Diff" column.

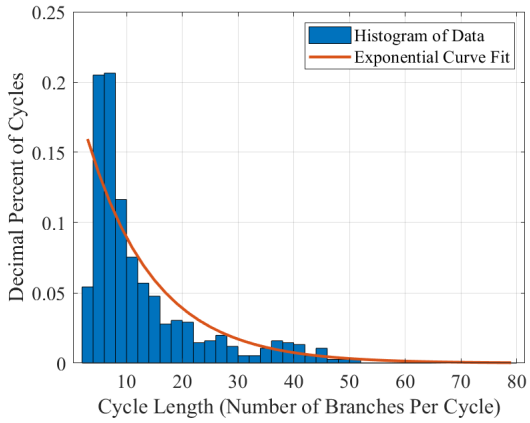


(a) NEISO

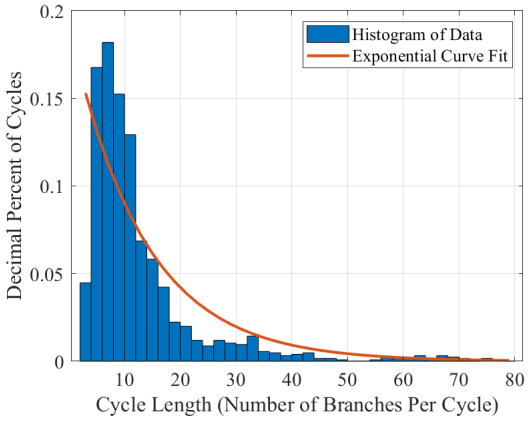
(b) NYISO



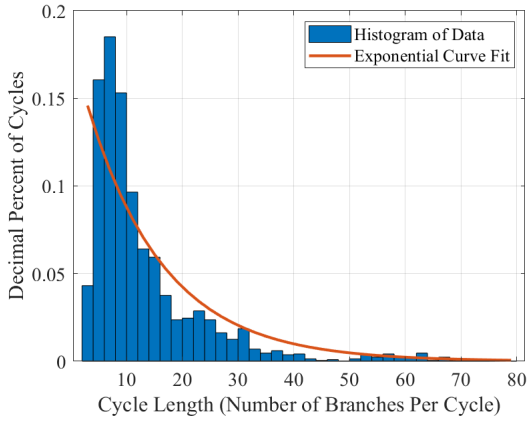
(c) MISO



(d) NWPP

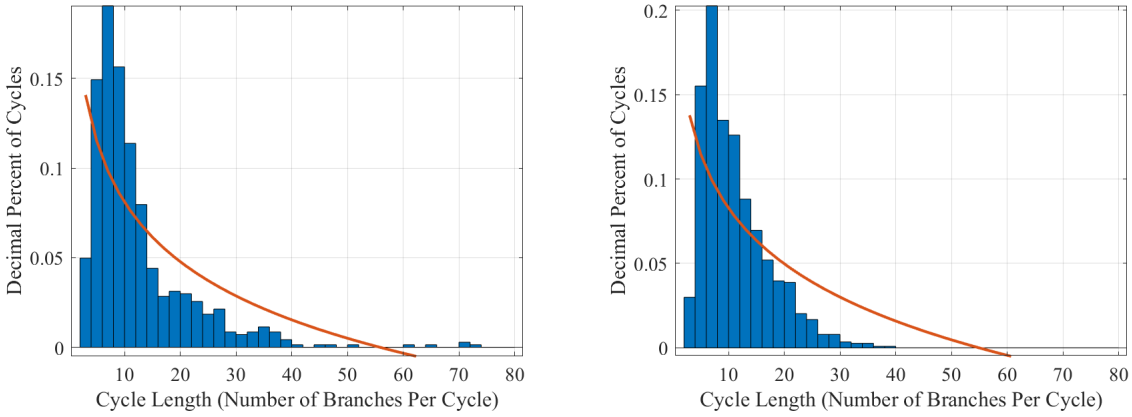


(e) AZNM



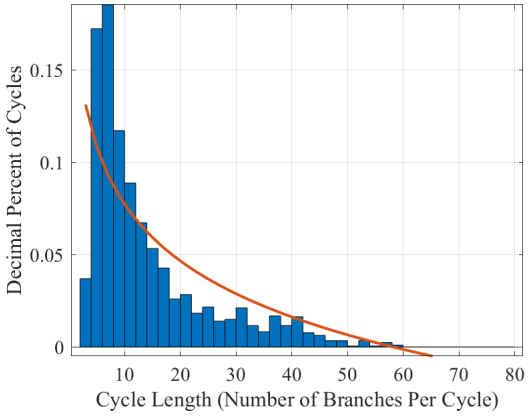
(f) CAISO

Figure D.3: Histograms of the MATLAB-computed cycle basis with exponential curve fits for KCL-compliant synthetic power system networks

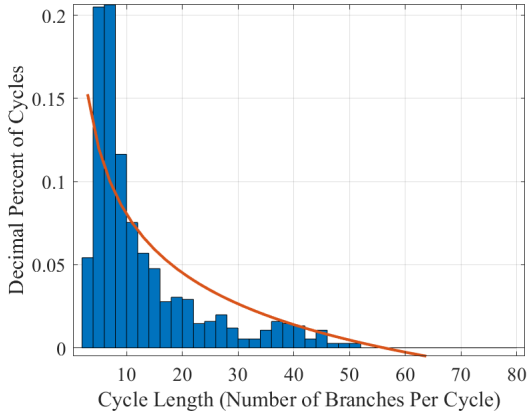


(a) NEISO

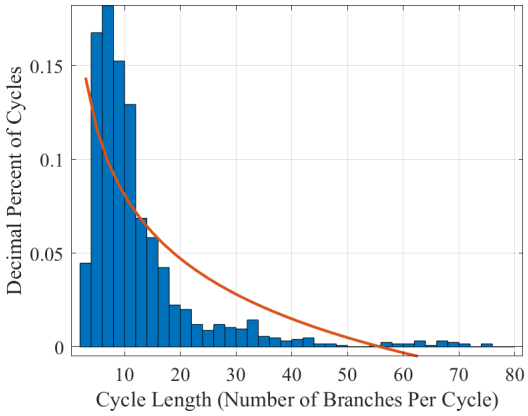
(b) NYISO



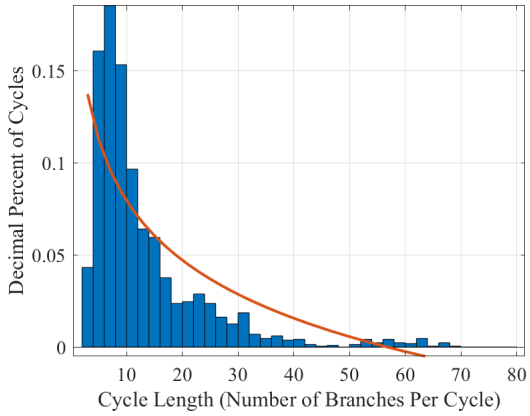
(c) MISO



(d) NWPP

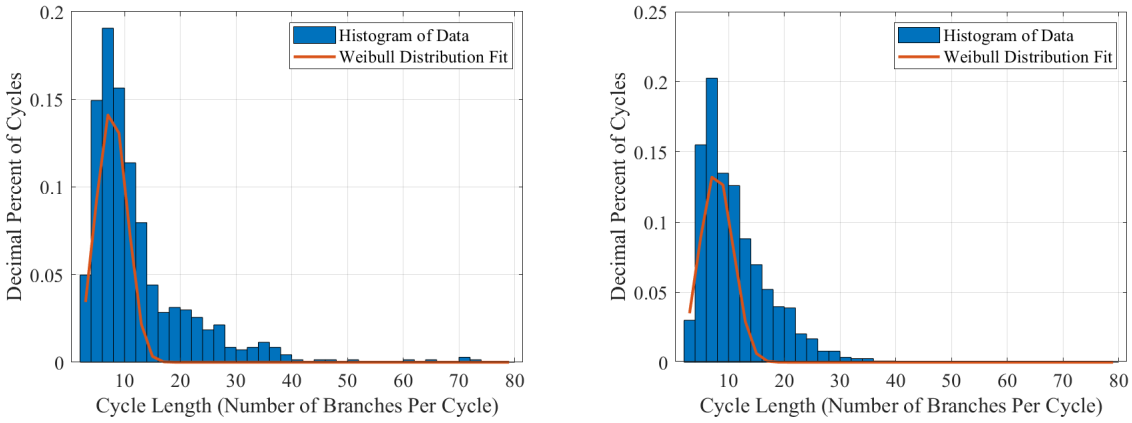


(e) AZNM



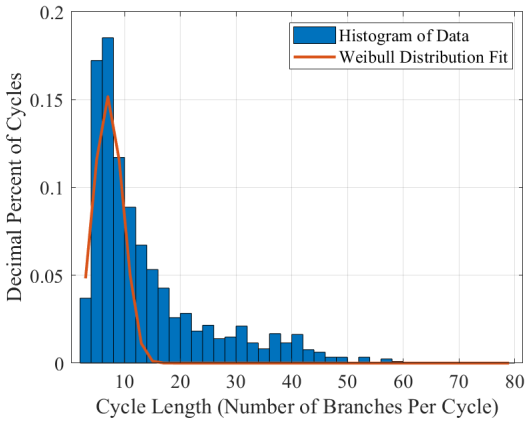
(f) CAISO

Figure D.4: Histograms of the MATLAB-computed cycle basis with power distribution curve fits for KCL-compliant synthetic power system networks

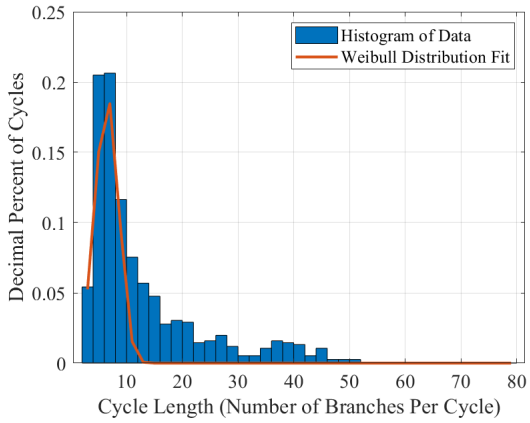


(a) NEISO

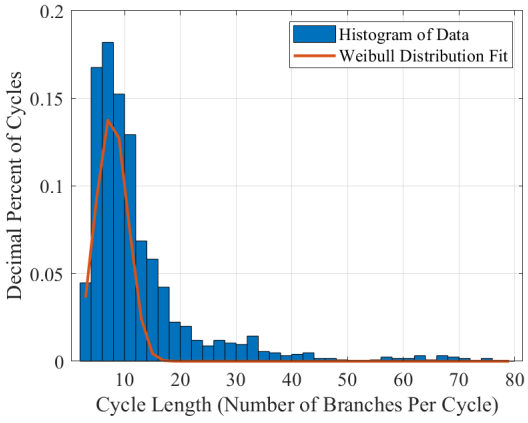
(b) NYISO



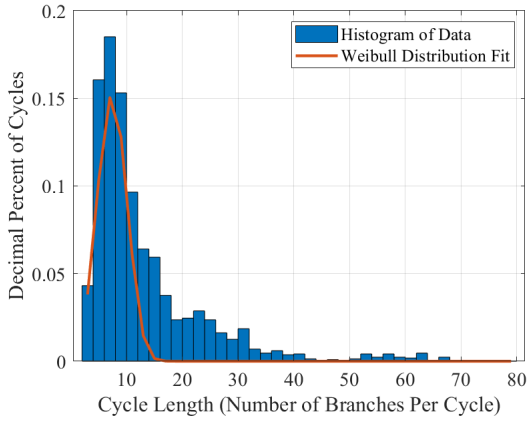
(c) MISO



(d) NWPP

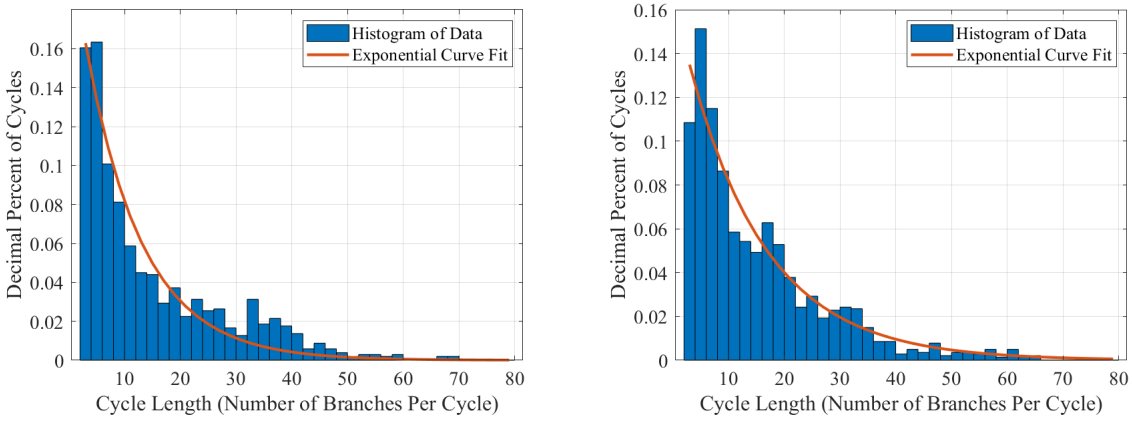


(e) AZNM



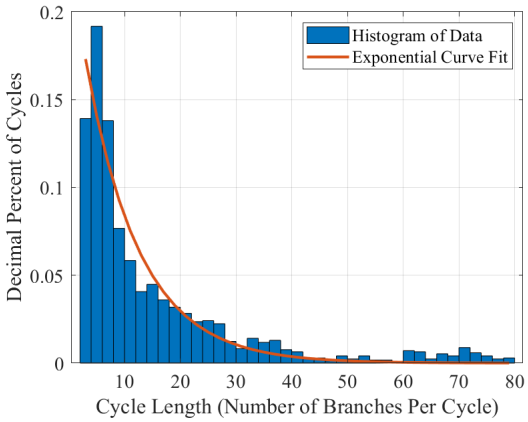
(f) CAISO

Figure D.5: Histograms of the MATLAB-computed cycle basis with Weibull distribution curve fits for KCL-compliant synthetic Weibull system networks

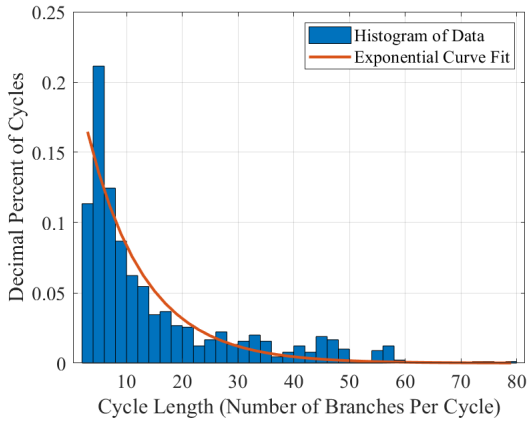


(a) NEISO

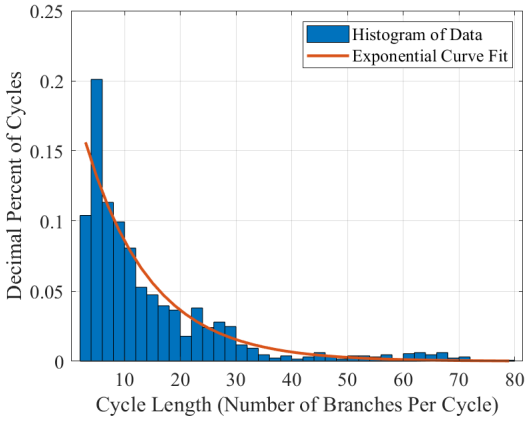
(b) NYISO



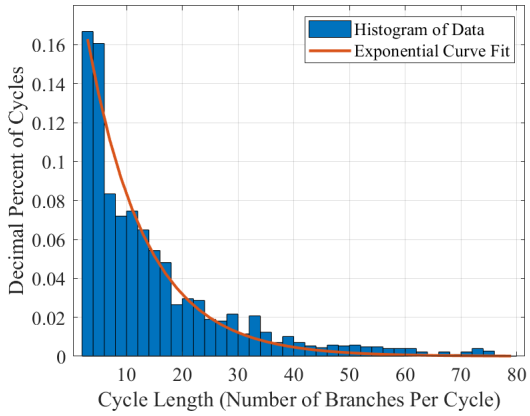
(c) MISO



(d) NWPP

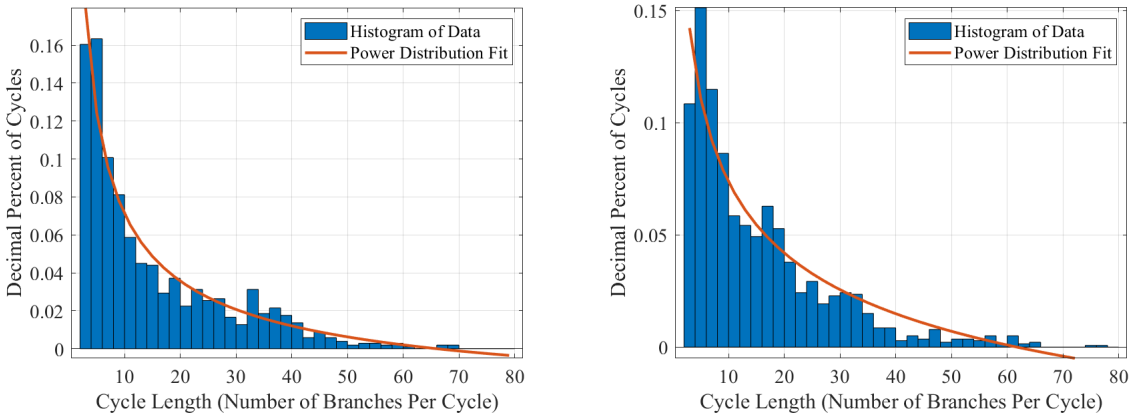


(e) AZNM



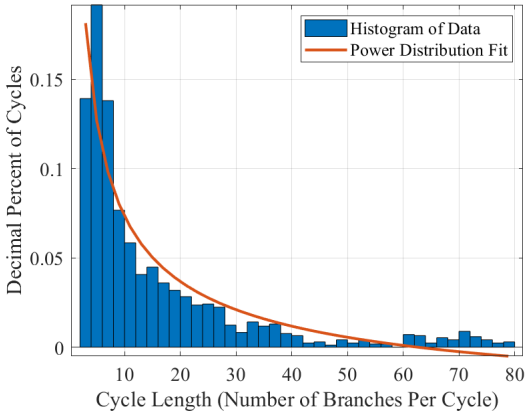
(f) CAISO

Figure D.6: Histograms of the MATLAB-computed cycle basis with exponential curve fits for KCL/KVL-compliant synthetic power system networks

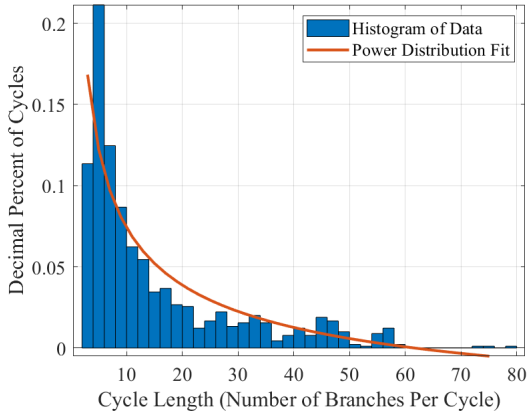


(a) NEISO

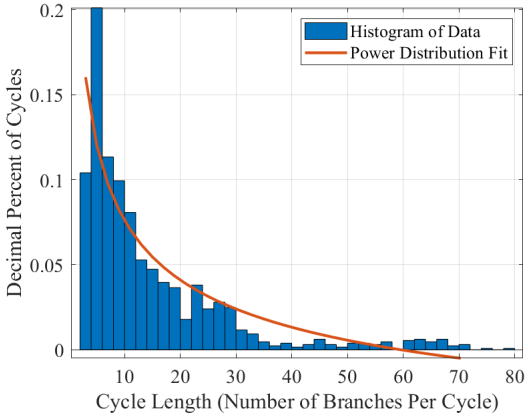
(b) NYISO



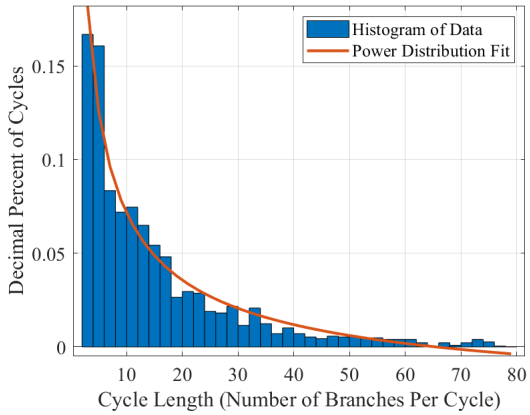
(c) MISO



(d) NWPP

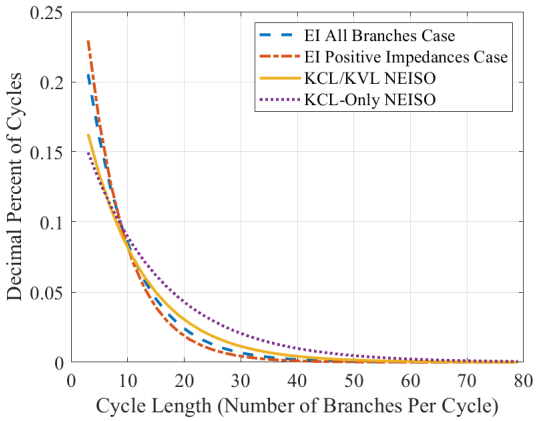


(e) AZNM

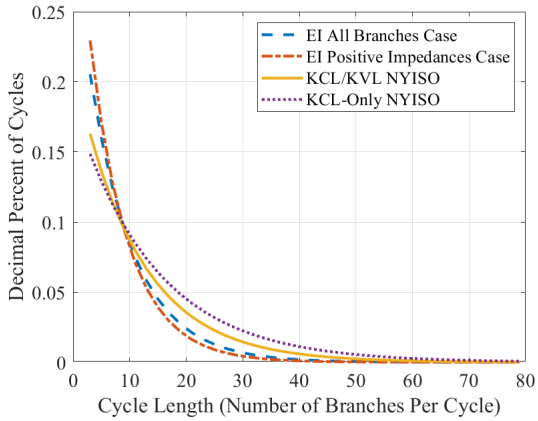


(f) CAISO

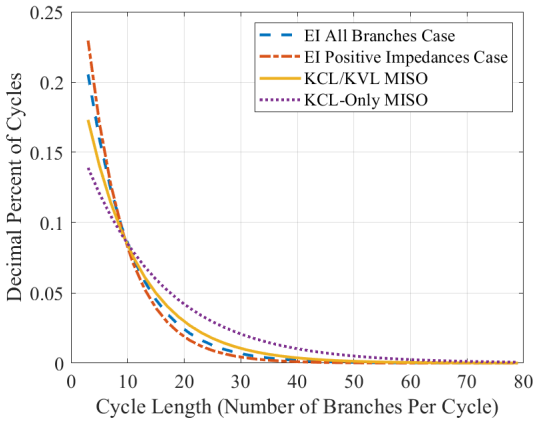
Figure D.7: Histograms of the MATLAB-computed cycle basis with power distribution curve fits for KCL/KVL-compliant synthetic power system networks



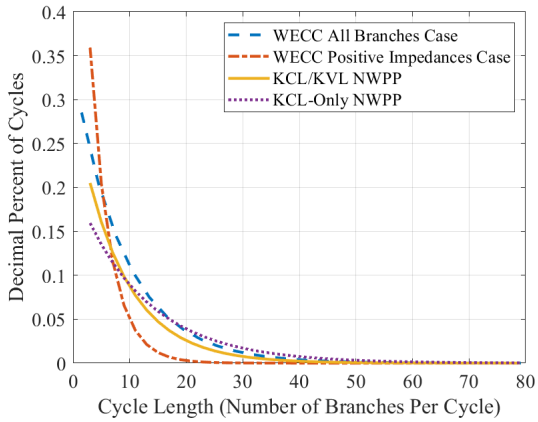
(a) NEISO



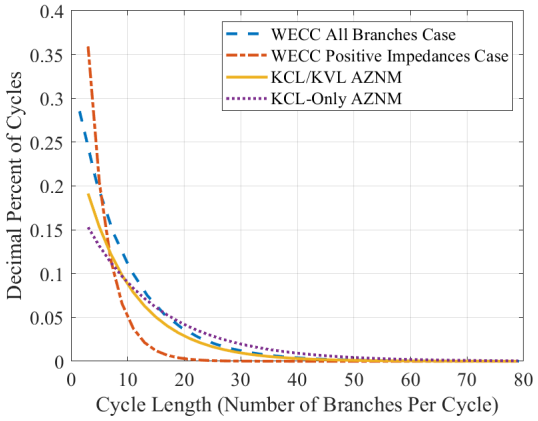
(b) NYISO



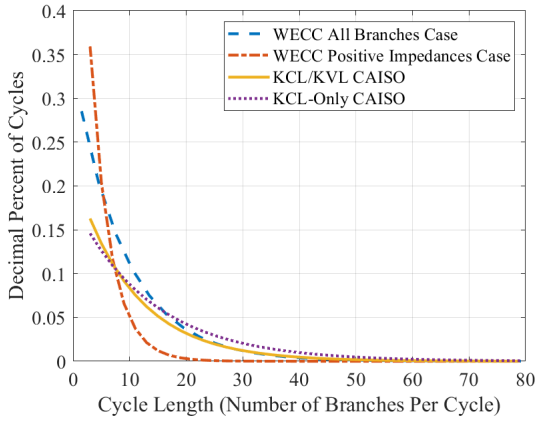
(c) MISO



(d) NWPP

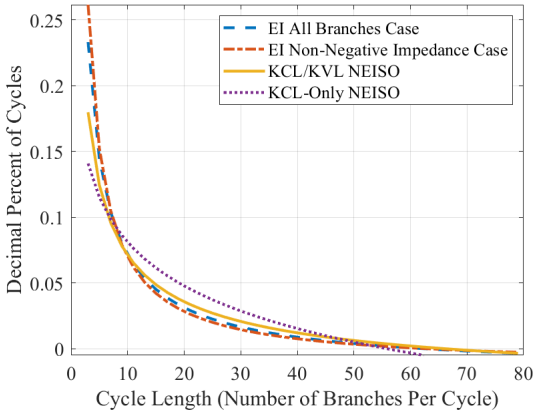


(e) AZNM

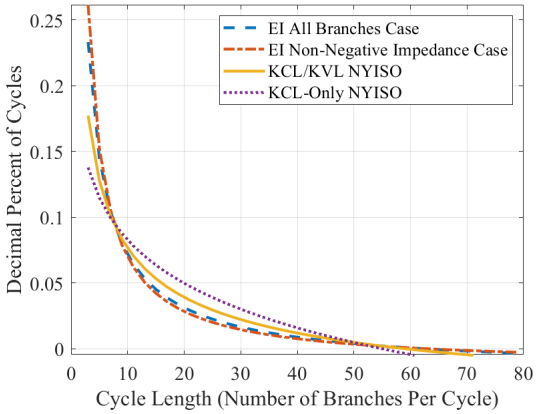


(f) CAISO

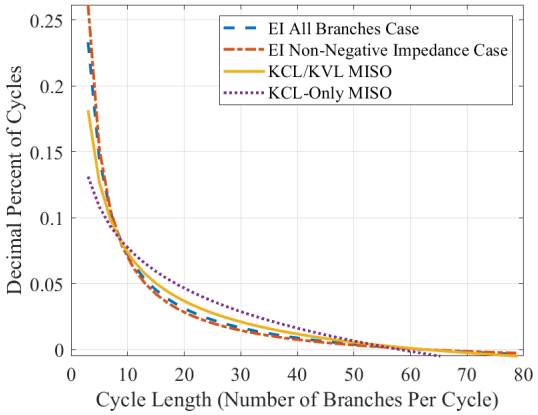
Figure D.8: Comparison of the exponential curve fits from the synthetic systems to real networks



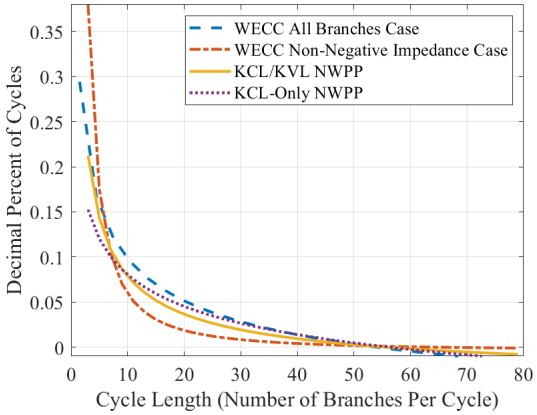
(a) NEISO



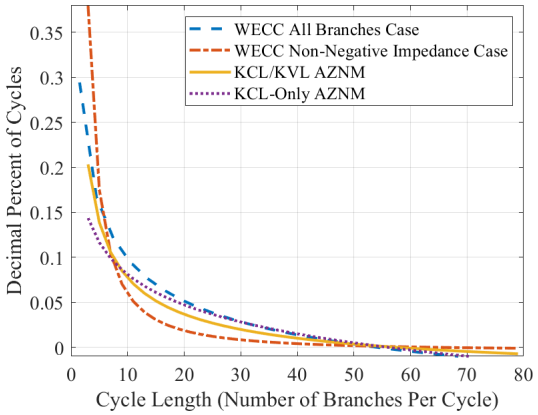
(b) NYISO



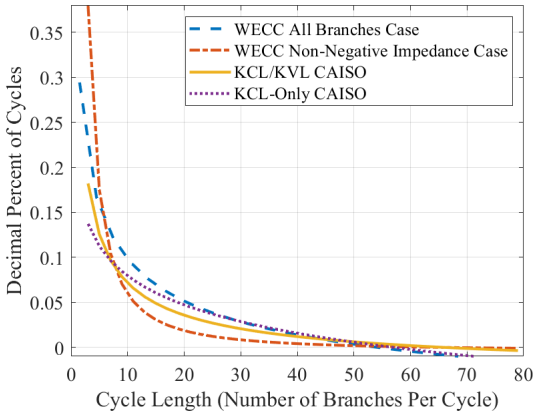
(c) MISO



(d) NWPP



(e) AZNM



(f) CAISO

Figure D.9: Comparison of the power distribution curve fits from the synthetic systems to real networks

Next, the goodness of fit between the exponential or power distributions and the histograms are examined, both qualitatively from the plots in Figures D.8 and D.9, and quantitatively using R-Squared values in Table D.1.

For all six of the KCL/KVL networks presented in this dissertation, either the exponential or power distribution are fairly good fits, as visually evident in Figures D.6 and D.7, and as shown by the R-Squared values in Table D.1. However, the exponential and power distribution fits are not as good for the KCL-Only networks, as indicated by the fact that only three KCL-only networks have an R-Squared value at or above the minimum R-Squared value from the KCL/KVL networks for the exponential fit, and only one network has an R-Squared value above the minimum for the power distribution fit.

Since the exponential and power distributions were not good fits for the KCL-Only network cycle basis length histograms, the Weibull distribution was explored as a distribution curve fit instead. Figure D.5 shows the Weibull distribution curve fits for six KCL-Only networks. Notice that while the Weibull Distribution is not a perfect fit to the histogram data, the general shape of the Weibull fits the shape of the histogram much better than the exponential or power distributions.

The next quantitative comparison metric is the percent difference between the areas under the exponential and power distribution fit curves for the synthetic compared to the real networks. The areas under the power and exponential distribution fit curves from Figures D.8 and D.9 are computed for both the synthetic and real networks, and the

percent difference between the synthetic networks and their corresponding real networks are recorded in the"Exp % diff" and "Power % diff" columns.

Notice that the percent difference is small between the areas under the power distributions for the KCL/KVL networks compared to the real world networks. This suggests that the KCL/KVL networks are similar to the existing real-world networks when comparing lengths of the MATLAB-computed cycle basis. When examining the percent difference in areas under the exponential curve fits to the MATLAB-computed cycle basis histograms for both the KCL-Only and the KCl KVL networks compared to the real networks, recall that the exponential function is not a good representation of the histograms of the KCL-Only networks as seen in Figure D.3. Thus, while the KCL-Only network show a lower percent difference, these numbers are not accurate, as demonstrated by the R-Squared values.

Region	Sub-stations	Max Load (GW)	Gen Dispatch	Mean Cycle Length	Mean Cycle Length % Diff	Area Under Cycle Basis Histogram Fits							
						Exp % diff		R-Squared		Power % diff		R-Squared	
						KCL-Only	KCL-KVL	KCL-Only	KCL-KVL	KCL-Only	KCL-KVL	KCL-Only	KCL-KVL
AZNM	4700	43		12.57	34.7%	23.0%	11.6%	0.71	0.87	36.1%	17.6%	0.63	0.81
CAISO	8400	47		18.03	93.2%	22.5%	3.9%	0.72	0.96	34.7%	2.7%	0.64	0.96
MISOSPP	18800	192	Network	15.15	21.9%	11.5%	9.8%	0.94	0.90	11.7%	16.0%	0.87	0.83
MISOSPP	18800	192	Area	19.19	54.4%		3.5%		0.91		12.9%		0.86
FRCC	7500	79		14.06	13.1%	21.2%	8.7%	0.73	0.91	33.7%	13.6%	0.65	0.85
MISO	7000	55		15.49	24.6%	20.2%	2.3%	0.71	0.91	34.7%	12.2%	0.64	0.86
NEISO	3500	25		13.92	12.0%	24.3%	1.3%	0.73	0.94	34.1%	3.5%	0.65	0.95
NNEAST	13900	105		19.19	54.4%		4.8%		0.94		7.6%		0.91
NWPP	3200	37		11.39	22.1%	17.9%	4.0%	0.71	0.89	35.7%	19.1%	0.63	0.85
NYISO	5000	32		17.58	41.4%		11.4%		0.96	37.2%	4.8%	0.62	0.94
PJM	11900	118	Network	12.33	0.8%		9.6%		0.99		19.9%		0.98
PJM	11900	118	Area	17.28	39.0%		9.1%		0.93		9.3%		0.98
RFCE	5400	48		15.98	28.6%	18.8%	10.2%	0.70	0.93	37.0%	10.2%	0.62	0.89
RFCW	7400	70		15.18	22.1%	21.0%	7.0%	0.70	0.88	36.8%	15.7%	0.62	0.83
SPP	8500	100		15.06	21.2%	18.4%	7.5%	0.75	0.90	31.7%	13.9%	0.67	0.85
SRMV	3300	37		12.72	2.3%	20.7%	7.1%	0.75	0.85	31.2%	19.3%	0.68	0.80
SRTV	4600	55		17.15	38.0%	23.3%	4.2%	0.74	0.93	32.1%	11.5%	0.67	0.88
SRVC	6500	69		15.54	25.0%	15.0%	10.9%	0.72	0.92	35.6%	11.1%	0.63	0.88
WECC	16300	127	Network	16.76	79.6%	16.9%	10.2%	0.90	0.88	17.1%	15.9%	0.82	0.83
WECC2	16300	127	Area	15.62	67.4%		2.8%		0.89		14.0%		0.85
WECC		Actual Network			9.33				0.983				0.912
EI		Actual Network			12.43				0.970				0.985
Mean				15.51	34.8%	19.6%	7.0%	0.75	0.91	32.0%	12.5%	0.67	0.88
Max				19.19	93.2%	24.3%	11.6%	0.94	0.99	37.2%	19.9%	0.87	0.98
Min				11.39	0.8%	11.5%	1.3%	0.70	0.85	11.7%	2.7%	0.62	0.80
Std				2.17	23.9%	3.4%	3.3%	0.07	0.03	7.2%	5.0%	0.07	0.06

Table D.1: Quantitative metrics for cycle basis length histograms

Appendix E

Computational Considerations for Network Construction Algorithms

E.1 Optimization Solvers

The fundamental the goal of the inner loop in both the KCL-only and KCL/KVL inner loop optimization based algorithms is to drive some or many of the flows on the network paths to zero. The direct use of the minimal L1 norm in the transportation model-based KCL-only NCA ensures that the potential transmission line path flows will be zero.

CPLEX [117] was the chosen optimization solver for the KCL-only NCA, and uses a proprietary combination of the simplex method and interior point to solve linear programs. A feature of the simplex method is that it moves between vertices of the feasible

space, and since the optimal solution is at a vertex of the feasible space, the optimal solution calculated by the complex will be a vertex. Since the path flow constraints require the absolute value of the flows to be between zero and the maximum limit (inclusive), the solution of a zero flow is feasible for any of the potential transmission line paths. Thus, the optimal solution calculated by the complex method results in the majority of the flows on potential transmission line paths being numerically and identically zero (within the tolerance of double precision floating point arithmetic).

Gurobi [125] was chosen as the solver for the KCL-KVL network construction algorithm. The feasible space of a linearly constrained quadratic program is not a polygon, and thus the optimal solution does not lie at of vertex. A feature of many quadratic programming (QP) solvers is that of numerical tolerances specified as optional inputs to the solver. Thus, elements of the optimal solution that should be identically zero will instead have some small numerical value, the magnitude of which will vary depending on the chosen solver and tolerance parameters chosen. Solutions below this "cutoff threshold" should be treated as numerically zero even if they are not identically zero. Computational experience has shown that this cutoff threshold is $1e-5$ for CPLEX and Gurobi and $1e-3$ for Mosek. Thus, any path flows beneath these values were treated to be zero for the purposes of the KCL/KVL multilevel algorithm.

Bibliography

[1] Federal Energy Regulatory Commission, "Critical energy/electric infrastructure information (ceii)," tech. rep.

[2] P. Marcos, E. Fernando, C. Mateo Domingo, T. Gómez San Román, B. Palmintier, B.-M. Hodge, V. Krishnan, F. De Cuadra García, and B. Mather, "A review of power distribution test feeders in the united states and the need for synthetic representative networks," *Energies*, vol. 10, no. 11, p. 1896, 2017.

[3] 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. 32, no. 4, pp. 3258–3265, 2017.

[4] L. Tang and M. C. Ferris, "A hierarchical framework for long-term power planning models," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 46–56, 2015.

[5] A. B. Birchfield, T. Xu, K. Shetye, and T. Overbye, "Building synthetic power transmission networks of many voltage levels, spanning multiple areas," in *Hawaii International Conference on System Sciences*, 2018.

[6] 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.

[7] A. H. Escobar, R. A. Gallego, and R. Romero, "Multistage and coordinated planning of the expansion of transmission systems," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 735–744, 2004.

[8] M. Ouyang and K. Yang, "Does topological information matter for power grid vulnerability?," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 24, no. 4, p. 043121, 2014.

[9] FERC, "Form no. 715 - annual transmission planning and evaluation report," tech. rep., 2018.

[10] "Power system test case archive," tech. rep.

[11] ARPA-E, "Grid optimization (go) competition," 2018.

[12] J. D. McCalley and V. Krishnan, "A survey of transmission technologies for planning long distance bulk transmission overlay in us," *International Journal of Electrical Power & Energy Systems*, vol. 54, pp. 559–568, 2014.

[13] L. Garver, "Transmission network estimation using linear programming," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 7, pp. 1688–1697, 1970.

[14] I. Abdou and M. Tkiouat, "Unit commitment problem in electrical power system: a literature review," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 3, p. 1357, 2018.

[15] R. Hemmati, R.-A. Hooshmand, and A. Khodabakhshian, "Comprehensive review of generation and transmission expansion planning," *IET Generation, Transmission & Distribution*, vol. 7, no. 9, pp. 955–964, 2013.

[16] C. O'Malley, L. Roald, D. Kourounis, O. Schenk, and G. Hug, "Security assessment in gas-electric networks," in *2018 Power Systems Computation Conference (PSCC)*, pp. 1–7, IEEE, 2018.

[17] A. Quelhas, E. Gil, J. D. McCalley, and S. M. Ryan, "A multiperiod generalized network flow model of the us integrated energy system: Part i—model description," *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 829–836, 2007.

[18] D. Gan, R. J. Thomas, and R. D. Zimmerman, "Stability-constrained optimal power flow," *IEEE Transactions on Power Systems*, vol. 15, no. 2, pp. 535–540, 2000.

[19] P. Kundur, N. J. Balu, and M. G. Lauby, *Power system stability and control*, vol. 7. McGraw-hill New York, 1994.

[20] G. Latorre, R. D. Cruz, J. M. Areiza, and A. Villegas, "Classification of publications and models on transmission expansion planning," *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 938–946, 2003.

[21] R. Bent, C. Coffrin, R. R. Gumucio, and P. Van Hentenryck, "Transmission network expansion planning: Bridging the gap between ac heuristics and dc approximations," in *2014 Power Systems Computation Conference*, pp. 1–8, IEEE, 2014.

[22] C. Coffrin, H. Hijazi, and P. Van Hentenryck, "Network flow and copper plate relaxations for ac transmission systems," in *2016 Power Systems Computation Conference (PSCC)*, IEEE, 2016.

[23] D. K. Molzahn and I. A. Hiskens, "A survey of relaxations and approximations of the power flow equations," *Foundations and Trends in Electric Energy Systems*, vol. 4, no. 1-2, pp. 1–221, 2019.

[24] J. D. Glover, M. S. Sarma, and T. Overbye, *Power system analysis & design, SI version*. Cengage Learning, 2012.

[25] F. Capitanescu, J. M. Ramos, P. Panciatici, D. Kirschen, A. M. Marcolini, L. Platbrood, and L. Wehenkel, "State-of-the-art, challenges, and future trends in security constrained optimal power flow," *Electric Power Systems Research*, vol. 81, no. 8, pp. 1731–1741, 2011.

[26] B. Stott, O. Alsac, and A. J. Monticelli, "Security analysis and optimization," *Proceedings of the IEEE*, vol. 75, no. 12, pp. 1623–1644, 1987.

[27] A. Monticelli, M. Pereira, and S. Granville, "Security-constrained optimal power flow with post-contingency corrective rescheduling," *IEEE Transactions on Power Systems*, vol. 2, no. 1, pp. 175–180, 1987.

[28] V. Krishnan, J. Ho, B. F. Hobbs, A. L. Liu, J. D. McCalley, M. Shahidehpour, and Q. P. Zheng, "Co-optimization of electricity transmission and generation resources for planning and policy analysis: review of concepts and modeling approaches," *Energy Systems*, vol. 7, no. 2, pp. 297–332, 2016.

[29] A. R. Bergen and V. Vittal, *Power Systems Analysis*. Prentice-Hall, 2000.

[30] J. Lin and F. H. Magnago, *Electricity Markets: Theories and Applications*. John Wiley & Sons, 2017.

[31] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network flows*. Massachusetts Institute of Technology, 1988.

[32] R. Romero, C. Rocha, M. Mantovani, and J. Mantovani, "Analysis of heuristic algorithms for the transportation model in static and multistage planning in network expansion systems," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 150, no. 5, pp. 521–526, 2003.

[33] R. Romero, C. Rocha, J. Mantovani, and I. Sanchez, "Constructive heuristic algorithm for the dc model in network transmission expansion planning," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 152, no. 2, pp. 277–282, 2005.

[34] D. Chéverez-González, *The role of network-related eigenstructure in power system markets and optimization*. Phd thesis, 2009.

[35] E. A. Guillemin, *Theory of linear physical systems: theory of physical systems from the viewpoint of classical dynamics, including Fourier methods*. Courier Corporation, 2013.

[36] J. B. Dennis, *Mathematical programming and electrical networks*. Phd thesis, 1959.

[37] P. Penfield, R. Spence, and S. Duinker, "A generalized form of tellegen's theorem," *IEEE Transactions on Circuit Theory*, vol. 17, no. 3, pp. 302–305, 1970.

[38] G. Strang, "A framework for equilibrium equations," *SIAM Review*, vol. 30, no. 2, pp. 283–297, 1988.

[39] D. P. Bertsekas, "Thevenin decomposition and large-scale optimization," *Journal of Optimization Theory and Applications*, vol. 89, no. 1, pp. 1–15, 1996.

[40] J. C. Baez and B. Fong, "A compositional framework for passive linear networks," *arXiv preprint arXiv:1504.05625*, 2015.

[41] J. A. Momoh, R. Adapa, and M. El-Hawary, "A review of selected optimal power flow literature to 1993. i. nonlinear and quadratic programming approaches," *IEEE Transactions on Power Systems*, vol. 14, no. 1, pp. 96–104, 1999.

[42] M. El-Metwally and Z. Al-Hamouz, "Transmission networks planning using quadratic programming," *Electric machines and power systems*, vol. 18, no. 2, pp. 137–148, 1990.

[43] A. Ghosh and S. Boyd, "Growing well-connected graphs," in *Proceedings of the 45th IEEE Conference on Decision and Control*, pp. 6605–6611, IEEE, 2006.

[44] S. Bornholdt and H. G. Schuster, *Handbook of graphs and networks: from the genome to the internet*. John Wiley & Sons, 2006.

[45] M. Drobyshevskiy and D. Turdakov, "Random graph modeling: A survey of the concepts," *ACM Computing Surveys (CSUR)*, vol. 52, no. 6, pp. 1–36, 2019.

[46] R. V. Solé and S. Valverde, *Information theory of complex networks: on evolution and architectural constraints*, pp. 189–207. Springer, 2004.

[47] R. C. Read, *A survey of graph generation techniques*, pp. 77–89. Springer, 1981.

[48] C. F. Bazlamaçcı and K. S. Hindi, "Minimum-weight spanning tree algorithms a survey and empirical study," *Computers & Operations Research*, vol. 28, no. 8, pp. 767–785, 2001.

[49] P. Pop, "A survey of different integer programming formulations of the generalized minimum spanning tree problem," *Carpathian Journal of Mathematics*, pp. 104–118, 2009.

[50] P. Erdös and A. Rényi, "On random graphs," *Publicationes mathematicae*, vol. 6, no. 26, pp. 290–297, 1959.

[51] D. Volchenkov and P. Blanchard, "An algorithm generating random graphs with power law degree distributions," *Physica A: Statistical Mechanics and its Applications*, vol. 315, no. 3-4, pp. 677–690, 2002.

[52] M. Bayati, J. H. Kim, and A. Saberi, "A sequential algorithm for generating random graphs," *Algorithmica*, vol. 58, no. 4, pp. 860–910, 2010.

[53] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, p. 440, 1998.

[54] Z. Wang, A. Scaglione, and R. J. Thomas, "Generating statistically correct random topologies for testing smart grid communication and control networks," *IEEE transactions on Smart Grid*, vol. 1, no. 1, pp. 28–39, 2010.

[55] E. Cotilla-Sanchez, P. D. 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, 2012.

[56] Z. Wang, S. H. Elyas, and R. J. Thomas, "Generating synthetic electric power system data with accurate electric topology and parameters," in *2016 51st International Universities Power Engineering Conference (UPEC)*, pp. 1–6, IEEE, 2016.

[57] R. Bennon, J. Juves, and A. Meliopoulos, "Use of sensitivity analysis in automated transmission planning," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 1, pp. 53–59, 1982.

[58] A. Birchfield and T. 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.

[59] R. Xu and D. Wunsch, *Clustering*, vol. 10. John Wiley & Sons, 2008.

[60] S. E. Schaeffer, "Graph clustering," *Computer science review*, vol. 1, no. 1, pp. 27–64, 2007.

[61] W. Dong, C. Moses, and K. Li, "Efficient k-nearest neighbor graph construction for generic similarity measures," in *Proceedings of the 20th international conference on World wide web*, pp. 577–586, 2011.

[62] M. T. Thai and P. M. Pardalos, *Handbook of optimization in complex networks: theory and applications*, vol. 57. Springer Science & Business Media, 2011.

[63] R. F. i. Cancho and R. V. Solé, *Optimization in complex networks*, pp. 114–126. Springer, 2003.

[64] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi, "A survey of statistical network models," *Foundations and Trends in Machine Learning*, vol. 2, no. 2, pp. 129–233, 2010.

[65] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.

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

[67] F. P. Preparata and M. I. Shamos, *Computational geometry: an introduction*. Springer Science & Business Media, 2012.

[68] R. Villasana, L. Garver, and S. Salon, "Transmission network planning using linear programming," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-104, no. 2, pp. 349–356, 1985.

[69] C. Coffrin and P. Van Hentenryck, "A linear-programming approximation of ac power flows," *INFORMS Journal on Computing*, vol. 26, no. 4, pp. 718–734, 2014.

[70] J. A. Taylor and F. S. Hover, "Linear relaxations for transmission system planning," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2533–2538, 2011.

[71] A. M. Vershik, "Long history of the monge-kantorovich transportation problem," *The Mathematical Intelligencer*, vol. 35, no. 4, pp. 1–9, 2013.

[72] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[73] R. Romero, A. Monticelli, A. Garcia, and S. Haffner, "Test systems and mathematical models for transmission network expansion planning," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 149, no. 1, pp. 27–36, 2002.

[74] H. Sadeghian, S. H. Elyas, and Z. Wang, "A novel algorithm for statistical assignment of transmission capacities in synthetic grid modeling," in *2018 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, IEEE, 2018.

[75] N. Gupta, R. Shekhar, and P. K. Kalra, "Congestion management based roulette wheel simulation for optimal capacity selection: Probabilistic transmission expansion planning," *International Journal of Electrical Power & Energy Systems*, vol. 43, no. 1, pp. 1259–1266, 2012.

[76] S. Merkli, A. Domahidi, J. Jerez, and R. S. Smith, "Globally optimal ac power system upgrade planning under operational policy constraints," in *2018 European Control Conference (ECC)*, pp. 131–136, IEEE, 2018.

[77] T. Kishore and S. Singal, "Optimal economic planning of power transmission lines: A review," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 949–974, 2014.

[78] S. Haffner, A. Monticelli, A. Garcia, J. Mantovani, and R. Romero, "Branch and bound algorithm for transmission system expansion planning using a transportation model," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 147, no. 3, pp. 149–156, 2000.

[79] M. J. Rider, A. V. Garcia, and R. Romero, "Branch and bound algorithm for transmission network expansion planning using dc model," in *2007 IEEE Lausanne Power Tech*, pp. 1350–1355, IEEE, 2007.

[80] N. Alguacil, A. L. Motto, and A. J. Conejo, "Transmission expansion planning: a mixed-integer lp approach," *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1070–1077, 2003.

[81] X. Ban, H. X. Liu, J. Lu, and M. C. Ferris, "Decomposition scheme for continuous network design problem with asymmetric user equilibria," *Transportation Research Record*, vol. 1964, no. 1, pp. 185–192, 2006.

[82] J. X. Ban, H. X. Liu, M. C. Ferris, and B. Ran, "A general mpcc model and its solution algorithm for continuous network design problem," *Mathematical and Computer Modelling*, vol. 43, no. 5-6, pp. 493–505, 2006.

[83] J. Hu, L. Sankar, and D. J. Mir, "Cluster-and-connect: An algorithmic approach to generating synthetic electric power network graphs," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 223–230, IEEE, 2015.

[84] 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)*, pp. 1–5, IEEE, 2018.

[85] F. Fioretto, T. W. Mak, and P. V. Hentenryck, "Privacy-preserving obfuscation of critical infrastructure networks," in *International Joint Conference on Artificial Intelligence (IJCAI-19)*, 2019.

[86] T. W. K. Mak, F. Fioretto, L. Shi, and P. Van Hentenryck, "Privacy-preserving power system obfuscation: A bilevel optimization approach," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1627–1637, 2020.

[87] C. W. Lee, S. Ng, J. Zhong, and F. Wu, "Transmission expansion planning from past to future," IEEE, 2006.

[88] Eastern Interconnection States' Planning Council, "Transmission planning white paper," tech. rep., 2014.

[89] MidAmerican Energy Company, "Reliability planning criteria for 100 kv and above," tech. rep., 2018.

[90] American Transmission Company, "Transmission system planning criteria," tech. rep., 2015.

[91] American Transmission Company, "Transmission planning assessment practices," tech. rep., 2017.

[92] H. E. Joseph and G. Giulia, "Regional transmission planning: A review of practices following ferc order nos. 890 and 1000," tech. rep., 2017.

[93] Midcontinent Independent System Operator (MISO), "Business practices manuals," tech. rep., 2019.

[94] H. H. Farr, *Transmission Line Design Manual: A Guide for the Investigation, Development, and Design of Power Transmission Lines*. US Department of the Interior, Water and Power Resources Service, 1980.

[95] M. C. Monti, S. Rose, and K. A. Mullins, "Transmission planning and capx2020," tech. rep., University of Minnesota Humphrey School of Public Affairs, 2016.

[96] S. Lumbreras and A. Ramos, "The new challenges to transmission expansion planning. survey of recent practice and literature review," *Electric Power Systems Research*, vol. 134, pp. 19–29, 2016.

[97] E. Schweitzer, A. Scaglione, and R. Thomas, "The validation of synthetic power system cases," in *IREP Symposium*, 2017.

[98] A. Birchfield, E. Schweitzer, M. Athari, T. Xu, T. 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.

[99] D. R. Schwarting, D. K. Molzahn, C. L. DeMarco, and B. C. Lesieutre, "Topological and impedance element ranking (tier) of the bulk-power system," in *2011 44th Hawaii International Conference on System Sciences*, pp. 1–10, IEEE, 2011.

[100] S. Babaeinejadsarookolaee, J. Snodgrass, S. Acharya, S. Greene, B. Lesieutre, and C. L. DeMarco, "Comparison of real and synthetic network models of the western united states with respect to new realism measures," in *2021 IEEE Power and Energy Conference at Illinois (PECI)*, pp. 1–8, IEEE, 2021.

[101] ARPA-E, "Grid data program," 2018.

[102] MATLAB, *version 9.10.0 (R2021a)*. The MathWorks Inc., 2021.

[103] *Matlab cyclebasis function documentation*. The Mathworks Inc., 2021.

[104] J. D. Horton, "A polynomial-time algorithm to find the shortest cycle basis of a graph," *SIAM Journal on Computing*, vol. 16, no. 2, pp. 358–366, 1987.

[105] E. Amaldi, C. Iuliano, and R. Rizzi, "Efficient deterministic algorithms for finding a minimum cycle basis in undirected graphs," in *International Conference on Integer Programming and Combinatorial Optimization*, pp. 397–410, Springer, 2010.

[106] K. Mehlhorn and D. Michail, "Minimum cycle bases: Faster and simpler," *ACM Transactions on Algorithms (TALG)*, vol. 6, no. 1, pp. 1–13, 2009.

[107] S. Bityukov, A. Maksimushkina, and V. Smirnova, "Comparison of histograms in physical research," *Nuclear Energy and technology*, vol. 2, no. 2, pp. 108–113, 2016.

[108] L. Wang, J. McCalley, C. L. DeMarco, F. Akhavizadegan, A. Venkatraman, and J. Snodgrass, "Development of expansion planning methods and tools for handling uncertainty," tech. rep., Power Systems Engineering Research Center, May 2020.

[109] S. Company, "Acsr aluminum conductor. steel reinforced. bare.," tech. rep., 2012.

[110] J. J. LaForest, *Transmission line reference book, 345 kv and above*. Palo Alto, Calif.: General Electric Company. Large Transformer Division and Energy Systems and Technology Divisions Electric Power Research Institute., second edition. ed., 1982.

[111] FERC, "Form no. 1: Annual report of major electric utility," tech. rep., 2010.

[112] US Department of Energy (DOE), "State energy sector risk profile," tech. rep., 2014.

[113] L. Barthold, R. Clayton, I. Grant, V. Longo, J. Stewart, and D. Wilson, *Transmission line reference book: 115–138 kV compact line design*. Palo Alto, CA: Electric Power Research Institute, Inc, 1978.

[114] USDA, *Electric transmission specifications and drawings, 115 kV through 230 kV*. Washington, D.C.: U.S. Dept. of Agriculture, Rural Utilities Service, 1998.

[115] Aluminum Association, *Aluminum electrical conductor handbook*. Washington, D.C. (818 Connecticut Ave., Washington 20006): Aluminum Association, second edition. ed., 1982.

[116] MATLAB, *version 9.0.0 (R2016a)*. The MathWorks Inc., 2016.

[117] *IBM ILOG CPLEX 12.7 User's Manual*. 2017.

[118] R. Tibshirani and L. Wasserman, "A closer look at sparse regression," tech. rep., 2016.

[119] C. C. Robusto, "The cosine-haversine formula," *The American Mathematical Monthly*, vol. 64, no. 1, p. 38, 1957.

[120] J. Hanson, "Upgrading transmission lines," in *Proceedings of the 1991 IEEE Power Engineering Society Transmission and Distribution Conference*, pp. 824–827, IEEE, 1991.

[121] Southern California Edison, "2014 final sce generator interconnection unit cost guide," tech. rep., 2014.

[122] C. A. Desoer and E. S. Kuh, *Basic circuit theory*. New York: McGraw-Hill, 1969.

[123] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[124] T. He, *Lasso and general l1-regularized regression under linear equality and inequality constraints*. Phd thesis, 2011.

[125] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2021.

[126] MATLAB, *version 9.4.0 (R2018a)*. The MathWorks Inc., 2018.

[127] "Epigrids synthetic networks," tech. rep.

[128] A. B. Birchfield, T. Xu, and T. J. Overbye, "Power flow convergence and reactive power planning in the creation of large synthetic grids," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6667–6674, 2018.

[129] M. Fiedler, "A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory," *Czechoslovak Mathematical Journal*, vol. 25, no. 4, pp. 619–633, 1975.

[130] B. Slininger, "Fiedler's theory of spectral graph partitioning," tech. rep., Citeseer, 2013.

[131] US Energy Information Administration (EIA), "Form eia-860," tech. rep., 2016.

[132] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.

[133] G. Kirchhoff, "Ueber die auflösung der gleichungen, auf welche man bei der untersuchung der linearen vertheilung galvanischer ströme geführt wird," *Annalen der Physik*, vol. 148, no. 12, pp. 497–508, 1847.

[134] S. Kirchhoff, "Ueber den durchgang eines elektrischen stromes durch eine ebene, insbesondere durch eine kreisförmige," *Annalen der Physik*, vol. 140, no. 4, pp. 497–514, 1845.

[135] *Admiralty Manual of Navigation: BR 45(1)*. The Stationery Office, 1997.

[136] A. Reich, *Cycle bases of graphs and spanning trees with many leaves-complexity results on planar and regular graphs*. Dsc thesis, 2014.

[137] D. Michail, *Minimum Cycle Basis Algorithms & Applications*. Dr.-ing. thesis, 2006.